



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

«Εξόρυξη Γνώσης για Μελλοντικές Καιρικές Συνθήκες»

ΕΥΑΓΓΕΛΟΣ ΜΠΟΚΟΡΟΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

ΚΟΛΟΜΒΑΤΣΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
Επίκουρος Καθηγητής

Λαμία, 2024



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

«Εξόρυξη Γνώσης για Μελλοντικές Καιρικές Συνθήκες»

ΕΥΑΓΓΕΛΟΣ ΜΠΟΚΟΡΟΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

ΚΟΛΟΜΒΑΤΣΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
Επίκουρος Καθηγητής

Λαμία , 2024



UNIVERSITY OF
THESSALY

SCHOOL OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE & TELECOMMUNICATIONS

“Knowledge Extraction for Future Weather Conditions”

EVANGELOS BOKOROS

FINAL THESIS

ADVISOR

KONSTANTINOS KOLOMVATSOS
Assistant Professor

Lamia , 2024

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ.), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία:/...../20.....

Ο – Η Δηλ.

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία εξετάζει τη χρήση μοντέλων μηχανικής μάθησης για την πρόβλεψη καιρικών συνθηκών σε πόλεις της Ελλάδας. Ειδικότερα, αξιολογούνται η γραμμική παλινδρόμηση, το XGBoost και το AdaBoost όσον αφορά την πρόβλεψη μέσης θερμοκρασίας, ταχύτητας ανέμου, κατακρήμνισης και ατμοσφαιρικής πίεσης. Παράλληλα, παρουσιάζεται και ένα μετα-μοντέλο που συνδυάζει τις προβλέψεις των παραπάνω αλγορίθμων με στόχο τη βελτίωση της ακρίβειας των προβλέψεων. Τα αποτελέσματα δείχνουν ότι το XGBoost και το AdaBoost υπερέχουν της γραμμικής παλινδρόμησης σε πολλές περιπτώσεις, με το μετα-μοντέλο να επιτυγχάνει βελτιωμένες προβλέψεις για την ταχύτητα ανέμου και την κατακρήμνιση. Ωστόσο, η πρόβλεψη της ατμοσφαιρικής πίεσης από το XGBoost παρουσίασε σημαντικά σφάλματα, επηρεάζοντας αρνητικά και το μετα-μοντέλο. Τέλος, προτείνονται βελτιώσεις και μελλοντικές προεκτάσεις της έρευνας, όπως η συλλογή περισσότερων δεδομένων και η εφαρμογή επιπλέον τεχνικών μηχανικής μάθησης.

ABSTRACT

This thesis investigates the use of machine learning models for weather prediction in Greek cities. Specifically, linear regression, XGBoost, and AdaBoost are evaluated for predicting mean temperature, wind speed, precipitation, and atmospheric pressure. Additionally, a meta-model is presented that combines the predictions of these algorithms to improve accuracy. The results show that XGBoost and AdaBoost outperform linear regression in many cases, with the meta-model achieving improved predictions for wind speed and precipitation. However, XGBoost's predictions for atmospheric pressure showed significant errors, negatively affecting the meta-model as well. Finally, improvements and future research extensions are proposed, such as collecting more data and applying additional machine learning techniques.

Table of Contents

ΠΕΡΙΛΗΨΗ	I
ABSTRACT	III
ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ	2
ΣΚΟΠΟΣ ΤΗΣ ΈΡΕΥΝΑΣ	2
ΣΗΜΑΣΙΑ ΤΗΣ ΈΡΕΥΝΑΣ	2
ΔΟΜΗ ΤΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ	3
ΚΕΦΑΛΑΙΟ 2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ	5
ΕΙΣΑΓΩΓΗ	5
ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΣΤΗ ΠΡΟΒΛΕΨΗ ΤΟΥ ΚΑΙΡΟΥ	5
ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ ΚΑΙΡΟΥ	7
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΤΟΥ ΚΑΙΡΟΥ	9
ΚΕΦΑΛΑΙΟ 3 ΣΥΣΤΗΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΚΑΙΡΙΚΩΝ ΣΥΝΘΗΚΩΝ	12
ΠΑΡΑΔΟΣΙΑΚΕΣ ΜΕΘΟΔΟΙ ΠΡΟΒΛΕΨΗΣ	12
ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ ΤΩΝ ΠΑΡΑΔΟΣΙΑΚΩΝ ΜΕΘΟΔΩΝ	12
ΜΕΤΕΩΡΟΛΟΓΙΑ ΣΤΟΝ 19 ^ο – 20 ^ο ΑΙ.	13
ΜΕΤΕΩΡΟΛΟΓΙΑ ΣΤΟΝ 21 ^ο ΑΙ.	14
ΑΡΙΘΜΗΤΙΚΑ ΜΟΝΤΕΛΑ ΚΑΙΡΟΥ	15
ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΑΡΙΘΜΗΤΙΚΩΝ ΜΟΝΤΕΛΩΝ	15
ΜΟΝΤΕΛΑ ΓΕΝΙΚΗΣ ΚΥΚΛΟΦΟΡΙΑΣ (GENERAL CIRCULATION MODELS - GCMs)	17
ΠΕΡΙΦΕΡΕΙΑΚΑ ΜΟΝΤΕΛΑ (REGIONAL CLIMATE MODELS - RCMs)	20
ΣΥΓΚΡΙΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΑΡΙΘΜΗΤΙΚΩΝ ΜΟΝΤΕΛΩΝ	22
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	23
ΔΟΡΥΦΟΡΙΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΑΙΣΘΗΤΗΡΕΣ ΙΟΤ	25
ΔΟΡΥΦΟΡΙΚΗ ΤΕΧΝΟΛΟΓΙΑ	25
ΑΙΣΘΗΤΗΡΕΣ ΙΟΤ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΚΑΙΡΟΥ	30
ΚΕΦΑΛΑΙΟ 4 ΜΟΝΤΕΛΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	33
ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΤΗΝ ΕΞΟΥΣΗ ΔΕΔΟΜΕΝΩΝ	33
ΟΡΙΣΜΟΣ ΚΑΙ ΒΑΣΙΚΕΣ ΈΝΝΟΙΕΣ	33
ΙΣΤΟΡΙΚΗ ΕΞΕΛΙΞΗ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	34
ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΗ ΜΕΤΕΩΡΟΛΟΓΙΑ	35
ΤΕΧΝΙΚΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΚΑΙΡΟΥ	37
ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	37
ΑΛΓΟΡΙΘΜΟΙ ΥΠΟΣΤΗΡΙΞΗΣ ΔΙΑΝΥΣΜΑΤΩΝ (SVMs)	39
ΜΟΝΤΕΛΑ ΕΝΙΣΧΥΣΗΣ ΔΕΝΤΡΩΝ ΑΠΟΦΑΣΕΩΝ (BOOSTING)	41
ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (LINEAR REGRESSION)	47
ΣΥΓΚΡΙΣΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΤΩΝ ΜΟΝΤΕΛΩΝ	48
ΚΡΙΤΗΡΙΑ ΑΞΙΟΛΟΓΗΣΗΣ	48

ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΚΑΙ ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΤΩΝ ΜΟΝΤΕΛΩΝ.....	52
ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	52
SVMs	53
BOOSTING	55
LINEAR REGRESSION	57
 <u>ΚΕΦΑΛΑΙΟ 5 ΠΡΟΤΕΙΝΟΜΕΝΟ ΣΥΣΤΗΜΑ ΚΑΙ ΠΕΙΡΑΜΑΤΙΚΗ ΑΠΟΤΙΜΗΣΗ.....</u>	59
 ΕΙΣΑΓΩΓΗ	59
ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ.....	60
ΣΥΛΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	60
ΕΞΕΡΕΥΝΗΣΗ ΔΕΔΟΜΕΝΩΝ (EDA).....	63
ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ.....	63
ΓΡΑΦΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ.....	64
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΟΚΙΜΗ ΜΟΝΤΕΛΩΝ	71
ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ (LINEAR REGRESSION).....	71
XGBOOST (EXTREME GRADIENT BOOSTING)	76
ADABOOST (ADAPTIVE BOOSTING)	80
ΜΕΤΑ-ΜΟΝΤΕΛΟ	83
ΣΥΓΚΡΙΣΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	89
ΣΥΓΚΡΙΣΗ ΜΕ ΥΦΙΣΤΑΜΕΝΑ ΣΥΣΤΗΜΑΤΑ	90
 <u>ΚΕΦΑΛΑΙΟ 6 ΣΥΜΠΕΡΑΣΜΑΤΑ.....</u>	92
 <u>ΒΙΒΛΙΟΓΡΑΦΙΑ.....</u>	94
 <u>ΠΑΡΑΡΤΗΜΑ Α: ΚΩΔΙΚΑΣ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ ΚΑΙΡΙΚΩΝ ΦΑΙΝΟΜΕΝΩΝ</u>	110

ΚΕΦΑΛΑΙΟ 1 Εισαγωγή

Σκοπός της Έρευνας

Η ακριβής πρόβλεψη καιρού αποτελεί κρίσιμο στοιχείο για την κοινωνία και την οικονομία, επηρεάζοντας πολλούς τομείς όπως η γεωργία, οι μεταφορές και η διαχείριση των φυσικών καταστροφών. Οι ακραίες καιρικές συνθήκες, όπως οι καταιγίδες, οι πλημμύρες και οι ξηρασίες, μπορούν να προκαλέσουν σοβαρές ζημιές στις υποδομές και να θέσουν σε κίνδυνο ανθρώπινες ζωές. Η ανάγκη για βελτιωμένες και ακριβέστερες προβλέψεις καιρού καθίσταται ολοένα και πιο επιτακτική, ιδίως λόγω της κλιματικής αλλαγής που εντείνει τη συχνότητα και τη σοβαρότητα των ακραίων καιρικών φαινομένων (Ritchie & Rosado, 2022).

Ο σκοπός της παρούσας έρευνας είναι η βελτίωση των προβλέψεων καιρού μέσω της εφαρμογής προηγμένων τεχνικών μηχανικής μάθησης. Η έρευνα αυτή στοχεύει στην ανάπτυξη και αξιολόγηση διαφόρων μοντέλων μηχανικής μάθησης, όπως η γραμμική παλινδρόμηση, το XGBoost και το AdaBoost, για την πρόβλεψη μετεωρολογικών παραμέτρων. Η χρήση αυτών των μοντέλων μπορεί να συμβάλει στη βελτίωση της ακρίβειας και της αξιοπιστίας των προβλέψεων, παρέχοντας παράλληλα μια πλατφόρμα για τη σύγκριση των αποδόσεών τους.

Τα κίνητρα για αυτή την έρευνα προκύπτουν από την ανάγκη για ακριβέστερες προβλέψεις καιρού, που μπορούν να βοηθήσουν στην καλύτερη προετοιμασία και ανταπόκριση σε φυσικές καταστροφές. Η ακρίβεια των προβλέψεων καιρού έχει βελτιωθεί σημαντικά τα τελευταία χρόνια, ωστόσο, εξακολουθούν να υπάρχουν σημαντικές προκλήσεις, ιδίως όσον αφορά τις μικροκλιματικές προβλέψεις και τα ακραία καιρικά φαινόμενα. Οι σύγχρονες τεχνολογίες μηχανικής μάθησης προσφέρουν νέες δυνατότητες για την αντιμετώπιση αυτών των προκλήσεων (Mcbean, 1999).

Σημασία της Έρευνας

Η σημασία της έρευνας έγκειται στη δυνατότητα παροχής ακριβέστερων προβλέψεων καιρού, που μπορούν να μειώσουν τις αρνητικές επιπτώσεις των ακραίων καιρικών φαινομένων. Οι ακριβείς προβλέψεις μπορούν να βοηθήσουν στην προστασία των υποδομών και των ζώων, στη βελτιστοποίηση της γεωργικής παραγωγής, και στη μείωση των οικονομικών απωλειών από φυσικές καταστροφές. Επιπλέον, η χρήση καινοτόμων τεχνολογιών, όπως η μηχανική μάθηση, προσφέρει νέες δυνατότητες για την αντιμετώπιση των προκλήσεων που σχετίζονται με την πρόβλεψη του καιρού (Bochenek & Ustrnul, 2022).

Η έρευνα αυτή αναμένεται να συνεισφέρει σημαντικά στην επιστημονική γνώση και να παρέχει πρακτικές λύσεις για την πρόβλεψη καιρού. Τα αποτελέσματα της έρευνας θα μπορούσαν να εφαρμοστούν σε διάφορους τομείς, όπως η γεωργία, οι μεταφορές και η διαχείριση καταστροφών,

βελτιώνοντας τη λήψη αποφάσεων και μειώνοντας τους κινδύνους που συνδέονται με τα ακραία καιρικά φαινόμενα (Ritchie & Rosado, 2022; NOAA, 2024).

Δομή της Διπλωματικής Εργασίας

Η διπλωματική εργασία αποτελείται από πέντε κύρια κεφάλαια, καθένα από τα οποία συμβάλλει στην ολοκληρωμένη ανάλυση και παρουσίαση του θέματος. Η **Εισαγωγή** παρουσιάζει τον σκοπό και τα κίνητρα της έρευνας, καθώς και τη σημασία της για την επιστημονική κοινότητα και την κοινωνία. Αναλύεται η ανάγκη για βελτιωμένες προβλέψεις καιρού λόγω των επιπτώσεων των ακραίων καιρικών φαινομένων και της κλιματικής αλλαγής. Το κεφάλαιο αυτό θεμελιώνει τη βάση για την έρευνα, εξηγώντας γιατί η πρόβλεψη καιρού είναι κρίσιμη και πώς η βελτίωσή της μπορεί να μειώσει τους κινδύνους και τις απώλειες.

Το δεύτερο κεφάλαιο, **Βιβλιογραφική Επισκόπηση**, εξετάζει τις σύγχρονες τεχνολογίες και μεθόδους πρόβλεψης καιρού, περιλαμβάνοντας τα ιστορικά δεδομένα και την εξέλιξη των μεθόδων πρόβλεψης. Αναλύονται οι κύριες προκλήσεις και περιορισμοί των σημερινών συστημάτων πρόβλεψης, όπως η αδυναμία ακριβούς πρόβλεψης μικροκλιματικών φαινομένων και ακραίων καιρικών γεγονότων. Παρουσιάζονται επίσης οι νεότερες τεχνολογίες που αναπτύσσονται για να αντιμετωπίσουν αυτές τις προκλήσεις, όπως τα συστήματα μηχανικής μάθησης και τα μεγάλα δεδομένα. Περιλαμβάνονται αναφορές σε σημαντικές έρευνες και ευρήματα της τελευταίας δεκαετίας, με έμφαση στην πρόοδο που έχει γίνει στον τομέα της πρόβλεψης καιρού.

Στο τρίτο κεφάλαιο, **Μοντέλα Μηχανικής Μάθησης**, περιγράφονται τα διάφορα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται για την πρόβλεψη καιρού. Αναλύονται τα πλεονεκτήματα και τα μειονεκτήματα της γραμμικής παλινδρόμησης, του XGBoost και του AdaBoost, και εξετάζεται ο τρόπος με τον οποίο κάθε μοντέλο χειρίζεται τα δεδομένα και τις σχέσεις μεταξύ των μετεωρολογικών παραμέτρων. Παρέχεται σύγκριση των αποδόσεων των μοντέλων, επιτρέποντας την κατανόηση ως προς το ποιο μοντέλο είναι καταλληλότερο για διαφορετικές συνθήκες πρόβλεψης. Αυτό το κεφάλαιο προσφέρει μια σαφή εικόνα των τεχνολογιών που εξετάστηκαν και των λόγων για τους οποίους επιλέχθηκαν.

Το τέταρτο κεφάλαιο, **Το Προτεινόμενο Σύστημα και Πειραματική Αποτίμηση**, παρουσιάζει το προτεινόμενο σύστημα πρόβλεψης καιρού. Εξηγείται η μεθοδολογία που ακολουθήθηκε για την ανάπτυξη του συστήματος, καθώς και τα δεδομένα που χρησιμοποιήθηκαν. Περιγράφεται η πειραματική διαδικασία και οι παράμετροι και μέθοδοι αξιολόγησης που εφαρμόστηκαν. Παρουσιάζονται τα αποτελέσματα των πειραμάτων και η ανάλυσή τους, καθώς και η σύγκριση των επιδόσεων των μοντέλων. Το κεφάλαιο αυτό αναλύει τη διαδικασία ανάπτυξης του συστήματος και παρέχει λεπτομερή ανάλυση των αποτελεσμάτων που επιτεύχθηκαν.

Το πέμπτο και τελευταίο κεφάλαιο, **Συμπεράσματα και Μελλοντικές Προεκτάσεις**, συνοψίζει τα κύρια ευρήματα της έρευνας και αναλύει τα συμπεράσματα που προκύπτουν από τα πειράματα και την ανάλυση των δεδομένων. Συζητούνται οι περιορισμοί της έρευνας και οι παράγοντες που επηρέασαν τα αποτελέσματα. Τέλος, προτείνονται κατευθύνσεις για μελλοντική έρευνα, εστιάζοντας στη βελτίωση των μεθόδων πρόβλεψης και στην ενσωμάτωση νέων τεχνολογιών και δεδομένων. Αυτό το κεφάλαιο παρέχει μια ολοκληρωμένη εικόνα των επιτευγμάτων της έρευνας και θέτει τις βάσεις για μελλοντική ανάπτυξη και έρευνα στον τομέα της πρόβλεψης καιρού.

ΚΕΦΑΛΑΙΟ 2 Βιβλιογραφική Επισκόπηση

Εισαγωγή

Η ανάγκη για ακριβή πρόβλεψη καιρού έχει βαθιές ρίζες στην ιστορία της ανθρωπότητας, ξεκινώντας από τους αρχαίους πολιτισμούς που βασίζονταν σε πρωτόγονες μεθόδους παρατήρησης του ουρανού και των φυσικών φαινομένων. Σήμερα, η πρόβλεψη του καιρού αποτελεί μια σύνθετη επιστήμη που ενσωματώνει τεχνολογίες αιχμής και προηγμένες μεθόδους ανάλυσης δεδομένων. Η εξέλιξη από τις απλές παρατηρήσεις στις σύγχρονες τεχνικές πρόβλεψης αναδεικνύει την ανάγκη για συνεχή βελτίωση και καινοτομία στον τομέα αυτό.

Η πρόβλεψη του καιρού είναι ζωτικής σημασίας για πολλές πτυχές της ανθρώπινης ζωής, συμπεριλαμβανομένης της γεωργίας, των μεταφορών, της διαχείρισης φυσικών καταστροφών και της καθημερινής δραστηριότητας. Η ακριβής πρόβλεψη μπορεί να βοηθήσει στην προστασία των ζώων και των περιουσιών, καθώς και στη βελτίωση της οικονομικής αποδοτικότητας. Οι φυσικές καταστροφές, όπως τυφώνες, πλημμύρες και ξηρασίες, μπορούν να έχουν καταστροφικές συνέπειες, καθιστώντας την αξιόπιστη πρόβλεψη του καιρού ακόμα πιο κρίσιμη (Ritchie & Rosado, 2022).

Η τεχνολογία και η επιστήμη της πρόβλεψης του καιρού έχουν προχωρήσει δραματικά τις τελευταίες δεκαετίες, με την ανάπτυξη υπολογιστικών μοντέλων που ενσωματώνουν τεράστιες ποσότητες δεδομένων και πολύπλοκους αλγορίθμους. Τα σύγχρονα μοντέλα καιρού βασίζονται σε μεθόδους μηχανικής μάθησης και τεχνητής νοημοσύνης, που επιτρέπουν την ακριβέστερη πρόβλεψη και την ανίχνευση περίπλοκων κλιματικών μοτίβων (Wallace & Hobbs, 2006).

Η πρόβλεψη του καιρού, ωστόσο, παραμένει μια πρόκληση λόγω της φυσικής πολυπλοκότητας και της αστάθειας των κλιματικών συστημάτων. Τα δεδομένα που συλλέγονται από διάφορες πηγές, όπως δορυφόροι, μετεωρολογικοί σταθμοί και αεροσκάφη, πρέπει να επεξεργάζονται και να αναλύονται με ακρίβεια για να παρέχουν χρήσιμες και αξιόπιστες προβλέψεις. Οι ερευνητές συνεχίζουν να αναπτύσσουν και να βελτιώνουν τα μοντέλα πρόβλεψης, αξιοποιώντας τις νέες τεχνολογίες και μεθοδολογίες.

Η σημασία της πρόβλεψης του καιρού ενισχύεται περαιτέρω από τις τρέχουσες περιβαλλοντικές προκλήσεις, όπως η κλιματική αλλαγή. Οι αλλαγές στο κλίμα επιδεινώνουν τα καιρικά φαινόμενα, καθιστώντας την πρόβλεψη ακόμα πιο σημαντική για τη διαχείριση των κινδύνων και την προσαρμογή στις νέες συνθήκες (Cox et al., 2022).

Ιστορική αναδρομή στη Πρόβλεψη του Καιρού

Η πρόβλεψη του καιρού έχει μακρά και ενδιαφέρουσα ιστορία, ξεκινώντας από τις πρώτες προσπάθειες των αρχαίων πολιτισμών μέχρι τις σύγχρονες τεχνολογικές εξελίξεις. Κατανοώντας

αυτήν την εξέλιξη, μπορούμε να αναγνωρίσουμε τη συνεχή αναζήτηση για ακρίβεια και τον ρόλο των τεχνολογικών καινοτομιών στην μεταμόρφωση της πρόβλεψης καιρού.

Οι πρώτες προσπάθειες πρόβλεψης του καιρού βασίζονταν κυρίως σε παρατηρήσεις της φύσης και εμπειρικές γνώσεις. Οι αρχαίοι Βαβυλώνιοι γύρω στο 650 π.Χ. χρησιμοποιούσαν τα πρότυπα των νεφών και άλλα ατμοσφαιρικά φαινόμενα για να προβλέπουν βραχυπρόθεσμες καιρικές αλλαγές. Παρόμοια, οι Κινέζοι αστρονόμοι το 300 π.Χ. ανέπτυξαν ένα ημερολόγιο που συνέδεε συγκεκριμένα καιρικά φαινόμενα με διάφορες περιόδους του έτους, δείχνοντας μια πρώιμη κατανόηση των εποχικών προτύπων (NASA, 2002; NASA Aqua Satellite, 2002).

Οι συμβολές του Αριστοτέλη το 340 π.Χ. με το έργο του "Μετεωρολογικά" ήταν καίριες. Αν και πολλές από τις θεωρίες του ήταν λανθασμένες, το έργο του Αριστοτέλη έθεσε τα θεμέλια για μελλοντικές μετεωρολογικές μελέτες. Οι ιδέες του παρέμειναν αυθεντία για σχεδόν 2000 χρόνια μέχρι που αναδείχθηκαν ακριβέστερες επιστημονικές μέθοδοι κατά την Αναγέννηση (Cahir, 2024).

Η Αναγέννηση σηματοδότησε μια σημαντική στροφή με την εφεύρεση κρίσιμων μετεωρολογικών οργάνων. Ο Γαλιλαίος εφηύρε ένα πρώιμο θερμόμετρο το 1592, και ο Ευαγγελιστής Τορικήλι ανέπτυξε το βαρόμετρο το 1643, το οποίο ήταν απαραίτητο για τη μέτρηση της ατμοσφαιρικής πίεσης. Αυτά τα όργανα επέτρεψαν πιο συστηματικές και ποσοτικές καιρικές παρατηρήσεις, απομακρύνοντας από τις καθαρά ποιοτικές αξιολογήσεις (Morone & Gillis, 2007; Wikipedia, 2023)).

Στα μέσα του 17ου αιώνα, η επιστημονική μελέτη της μετεωρολογίας άρχισε να ανθίζει. Σημαντικές ανακαλύψεις στους νόμους των αερίων από χημικούς και φυσικούς συνέβαλαν περαιτέρω στη μετεωρολογική έρευνα. Η ανάπτυξη των τηλεγραφικών δικτύων στα μέσα του 19ου αιώνα επέτρεψε την ταχεία ανταλλαγή μετεωρολογικών δεδομένων, διευκολύνοντας τη δημιουργία πρώιμων καιρικών χαρτών και την αναγνώριση των προτύπων καταιγίδων (NASA, 2002; NASA Aqua Satellite, 2002).

Στα τέλη του 19ου αιώνα, ιδρύθηκε το Weather Bureau (σημερινή Εθνική Μετεωρολογική Υπηρεσία) στις Ηνωμένες Πολιτείες. Αυτή η περίοδος σηματοδότησε την αρχή της συνόψισης της καιρικής πρόγνωσης, η οποία περιλάμβανε τη σύνταξη και ανάλυση ταυτόχρονων μετεωρολογικών παρατηρήσεων σε ευρεία περιοχή. Αυτή η μέθοδος επέτρεψε την αναγνώριση και παρακολούθηση των καιρικών συστημάτων, βελτιώνοντας σημαντικά την ακρίβεια των προβλέψεων (Morone & Gillis, 2007; Wikipedia, 2023)).

Στον 20ό αιώνα, η μετεωρολογία εξελίχθηκε ραγδαία με την ανάπτυξη των υπολογιστών και των αριθμητικών μοντέλων πρόγνωσης καιρού. Τα πρώτα αριθμητικά μοντέλα πρόγνωσης καιρού αναπτύχθηκαν τη δεκαετία του 1950 και επέτρεψαν την πιο ακριβή πρόβλεψη των μετεωρολογικών φαινομένων. Με την εισαγωγή των δορυφόρων τη δεκαετία του 1960, οι μετεωρολόγοι απέκτησαν

πρόσβαση σε δεδομένα υψηλής ανάλυσης για την ατμόσφαιρα και τις καιρικές συνθήκες σε παγκόσμιο επίπεδο.

Η σύγχρονη πρόβλεψη του καιρού βασίζεται σε προηγμένα μαθηματικά μοντέλα και αλγορίθμους μηχανικής μάθησης, τα οποία επεξεργάζονται τεράστιες ποσότητες δεδομένων από δορυφόρους, ραντάρ και μετεωρολογικούς σταθμούς. Οι σύγχρονες τεχνικές πρόβλεψης επιτρέπουν την ακριβέστερη και πιο έγκαιρη πρόβλεψη ακραίων καιρικών φαινομένων, όπως οι κυκλώνες και οι καταιγίδες, συμβάλλοντας στην προστασία των ζωών και των περιουσιών. Παρόλα αυτά, η πρόβλεψη του καιρού παραμένει μια πρόκληση λόγω της φυσικής πολυπλοκότητας και της αστάθειας των κλιματικών συστημάτων. Οι ερευνητές συνεχίζουν να αναπτύσσουν και να βελτιώνουν τα μοντέλα πρόβλεψης, αξιοποιώντας τις νέες τεχνολογίες και μεθοδολογίες (Bochenek & Ustrnul, 2022; Chen et al., 2023; Pugliese et al., 2021).

Η σημασία της πρόβλεψης του καιρού ενισχύεται περαιτέρω από τις τρέχουσες περιβαλλοντικές προκλήσεις, όπως η κλιματική αλλαγή. Οι αλλαγές στο κλίμα επιδεινώνουν τα καιρικά φαινόμενα, καθιστώντας την πρόβλεψη ακόμα πιο σημαντική για τη διαχείριση των κινδύνων και την προσαρμογή στις νέες συνθήκες. Οι ακριβείς προβλέψεις μπορούν να βοηθήσουν στην προστασία των ζωών και των περιουσιών, καθώς και στη βελτίωση της οικονομικής αποδοτικότητας (Cox et al., 2022).

Η βιβλιογραφική επισκόπηση που ακολουθεί θα εξετάσει την εξέλιξη των τεχνολογιών και μεθόδων πρόβλεψης καιρού, τις εφαρμογές της μηχανικής μάθησης στον τομέα και τις σημαντικές έρευνες και ευρήματα της τελευταίας δεκαετίας. Μέσω αυτής της ανάλυσης, θα αναδειχθούν οι προκλήσεις και οι προοπτικές για τη μελλοντική ανάπτυξη και βελτίωση των προβλέψεων καιρού.

Τεχνολογίες και Μέθοδοι Πρόβλεψης Καιρού

Η πρόβλεψη του καιρού έχει εξελιχθεί σημαντικά τα τελευταία χρόνια, χάρη στις προόδους της τεχνολογίας και της επιστήμης. Σήμερα, η πρόβλεψη του καιρού βασίζεται σε μια ποικιλία τεχνολογιών και μεθόδων που συνδυάζουν δεδομένα από διαφορετικές πηγές, προηγμένα μοντέλα υπολογιστών, και τεχνικές μηχανικής μάθησης για την παροχή ακριβέστερων και πιο αξιόπιστων προβλέψεων.

Η δορυφορική τεχνολογία αποτελεί τη βάση των σύγχρονων συστημάτων πρόβλεψης καιρού. Οι δορυφόροι παρέχουν συνεχή και παγκόσμια κάλυψη, καταγράφοντας δεδομένα όπως θερμοκρασία, υγρασία, και ταχύτητα ανέμου. Οι δορυφόροι, όπως οι δορυφόροι GOES και METEOSAT, επιτρέπουν την παρακολούθηση και την ανάλυση των καιρικών συστημάτων σε πραγματικό χρόνο, συμβάλλοντας σημαντικά στη βελτίωση των προβλέψεων (NASA, 2010). Οι δορυφόροι γεωστατικής τροχιάς παρακολουθούν συνεχώς συγκεκριμένες περιοχές, ενώ οι δορυφόροι πολικής τροχιάς

προσφέρουν παγκόσμια κάλυψη με συχνές επισκέψεις σε κάθε σημείο της γης. Αυτά τα δεδομένα είναι απαραίτητα για την κατανόηση και την πρόγνωση των καιρικών συνθηκών (Cahir, 2024).

Τα μετεωρολογικά ραντάρ χρησιμοποιούνται για την ανίχνευση βροχόπτωσης, χιονόπτωσης, και άλλων καιρικών φαινομένων. Παρέχουν δεδομένα υψηλής ανάλυσης για την ένταση και την κίνηση των βροχοπτώσεων, τα οποία είναι κρίσιμα για τις βραχυπρόθεσμες προβλέψεις καιρού. Η χρήση των ραντάρ σε συνδυασμό με τα δορυφορικά δεδομένα επιτρέπει την ακριβέστερη πρόβλεψη ακραίων καιρικών φαινομένων (Cahir, 2024). Τα ραντάρ Doppler μπορούν να μετρήσουν την ταχύτητα και την κατεύθυνση της κίνησης των σταγονιδίων της βροχής, επιτρέποντας την ανίχνευση ανεμοστροβίλων και ισχυρών ανέμων σε καταιγίδες. Τα δεδομένα ραντάρ συνδυάζονται συχνά με δορυφορικές εικόνες για πιο ολοκληρωμένες και ακριβείς προβλέψεις (NASA, 2002).

Τα αριθμητικά μοντέλα πρόβλεψης καιρού (Numerical Weather Prediction Models) είναι πολύπλοκα προγράμματα που χρησιμοποιούν εξισώσεις της φυσικής για την πρόβλεψη των μετεωρολογικών παραμέτρων. Τα μοντέλα αυτά, όπως το GFS (Global Forecast System) και το ECMWF (European Centre for Medium-Range Weather Forecasts), χρησιμοποιούν δεδομένα από δορυφόρους, ραντάρ, και μετεωρολογικούς σταθμούς για να δημιουργήσουν προβλέψεις καιρού για διαφορετικές χρονικές κλίμακες (NOAA, 2019). Η χρήση υπερυπολογιστών είναι κρίσιμη για την εκτέλεση αυτών των μοντέλων, καθώς απαιτείται τεράστια υπολογιστική ισχύς για την επεξεργασία των δεδομένων και την παροχή ακριβών προβλέψεων (Cahir, 2024).

Η τεχνητή νοημοσύνη (AI) και η μηχανική μάθηση (ML) έχουν αρχίσει να παίζουν σημαντικό ρόλο στην πρόβλεψη καιρού. Τα συστήματα αυτά εκπαιδεύονται να αναγνωρίζουν πρότυπα στα μετεωρολογικά δεδομένα και μπορούν να παρέχουν πιο γρήγορες και ακριβείς προβλέψεις. Τα συστήματα AI χρησιμοποιούνται για τη βελτίωση των αριθμητικών μοντέλων και για την ανάλυση μεγάλων συνόλων δεδομένων που συλλέγονται από διάφορες πηγές (MDPI, 2024). Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ανιχνεύουν ανωμαλίες και να προβλέπουν ακραία καιρικά φαινόμενα με μεγαλύτερη ακρίβεια από τις παραδοσιακές μεθόδους (Bassetti, 2024).

Το Διαδίκτυο των Πραγμάτων (Internet of Things - IoT) περιλαμβάνει τη χρήση δικτυωμένων αισθητήρων για τη συλλογή και μετάδοση μετεωρολογικών δεδομένων σε πραγματικό χρόνο. Οι αισθητήρες αυτοί μπορούν να τοποθετηθούν σε διάφορα σημεία, όπως κτίρια, οχήματα, και γεωργικές εκτάσεις, παρέχοντας λεπτομερείς πληροφορίες για τις τοπικές καιρικές συνθήκες. Η χρήση των IoT συσκευών επιτρέπει την καλύτερη παρακολούθηση των καιρικών αλλαγών και τη βελτίωση των τοπικών προβλέψεων (Oliynyk, 2023).

Τα κλιματικά μοντέλα χρησιμοποιούνται για την πρόβλεψη των μακροπρόθεσμων αλλαγών στο κλίμα. Αυτά τα μοντέλα λαμβάνουν υπόψη τις αλληλεπιδράσεις μεταξύ της ατμόσφαιρας, των ωκεανών, της ξηράς, και της βιόσφαιρας για να προβλέψουν τις μελλοντικές κλιματικές συνθήκες. Οι

προσομοιώσεις αυτές είναι κρίσιμες για την κατανόηση των επιπτώσεων της κλιματικής αλλαγής και τη λήψη μέτρων για την προσαρμογή και τον μετριασμό των επιπτώσεων αυτών (USDA Climate, 2023).

Η συνδυασμένη χρήση αυτών των τεχνολογιών και μεθόδων έχει βελτιώσει σημαντικά την ακρίβεια και την αξιοπιστία των προβλέψεων καιρού, παρέχοντας κρίσιμες πληροφορίες για την προστασία της ζωής και της περιουσίας και την υποστήριξη διαφόρων τομέων της οικονομίας.

Μηχανική Μάθηση στην Πρόβλεψη του Καιρού

Η μηχανική μάθηση έχει φέρει επανάσταση στην πρόβλεψη καιρού, παρέχοντας πιο ακριβείς και έγκαιρες προβλέψεις μέσω της ανάλυσης μεγάλων όγκων δεδομένων και της εφαρμογής σύνθετων αλγορίθμων. Οι τεχνολογίες αυτές επιτρέπουν την κατανόηση και την πρόβλεψη μετεωρολογικών φαινομένων με τρόπους που ήταν αδύνατοι με τις παραδοσιακές μεθόδους.

Οι τεχνικές μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα, οι αλγόριθμοι υποστήριξης διανυσμάτων (Support Vector Machines - SVMs) και τα μοντέλα ενίσχυσης δέντρων αποφάσεων (boosting), χρησιμοποιούνται για να εντοπίσουν μοτίβα σε μετεωρολογικά δεδομένα και να βελτιώσουν τις προβλέψεις. Αυτές οι τεχνικές επιτρέπουν τη χρήση ιστορικών δεδομένων και την εκμάθηση από αυτά, προσφέροντας πιο ακριβείς προβλέψεις (Chattopadhyay et al., 2018).

Η πρόβλεψη καιρού με μηχανική μάθηση περιλαμβάνει τη χρήση δεδομένων από διάφορες πηγές, όπως δορυφόρους, μετεωρολογικούς σταθμούς και αισθητήρες IoT. Τα δεδομένα αυτά εισάγονται σε αλγόριθμους μηχανικής μάθησης που εκτελούν αναλύσεις και προσφέρουν προβλέψεις με υψηλή ακρίβεια. Οι τεχνικές αυτές έχουν αποδειχθεί ιδιαίτερα αποτελεσματικές στην πρόβλεψη βραχυπρόθεσμων καιρικών φαινομένων, όπως βροχοπτώσεις και καταιγίδες (Ghosh & Bhattacharya, 2023).

Επιπλέον, η χρήση μοντέλων μηχανικής μάθησης στην πρόβλεψη καιρού έχει επιτρέψει την ενσωμάτωση πολλαπλών πηγών δεδομένων και την ανάλυση πολύπλοκων σχέσεων μεταξύ τους. Για παράδειγμα, οι αλγόριθμοι μπορούν να συνδυάσουν δεδομένα θερμοκρασίας, υγρασίας, ταχύτητας ανέμου και πίεσης για να παρέχουν πιο ολοκληρωμένες προβλέψεις. Αυτός ο συνδυασμός δεδομένων και τεχνολογιών έχει βελτιώσει σημαντικά την ακρίβεια των προβλέψεων και την κατανόηση των μετεωρολογικών φαινομένων (NASA, 2024).

Ένα παράδειγμα είναι η εφαρμογή νευρωνικών δικτύων για την πρόβλεψη ακραίων καιρικών φαινομένων, όπως οι τυφώνες. Τα μοντέλα αυτά μπορούν να αναλύσουν τεράστιες ποσότητες δεδομένων και να εντοπίσουν πρότυπα που υποδηλώνουν την ανάπτυξη τέτοιων φαινομένων. Οι προβλέψεις αυτές είναι ζωτικής σημασίας για την προετοιμασία και την αντιμετώπιση των καταστροφών, προσφέροντας πολύτιμες πληροφορίες στις αρχές και τις κοινότητες (NASA, 2024).

Οι εφαρμογές της μηχανικής μάθησης στην πρόβλεψη καιρού δεν περιορίζονται μόνο στην ανάλυση των δεδομένων. Χρησιμοποιούνται επίσης για την προσομοίωση μελλοντικών καιρικών συνθηκών και την εκτίμηση των επιπτώσεων της κλιματικής αλλαγής. Τα μοντέλα αυτά επιτρέπουν τη δημιουργία σεναρίων και την ανάλυση των πιθανών επιπτώσεων των αλλαγών στο κλίμα, παρέχοντας πολύτιμες πληροφορίες για τη λήψη αποφάσεων (NOAA, 2024).

Υπάρχει πληθώρα ερευνών που κάνουν χρήση της μηχανικής μάθησης για την μελέτη καιρικών φαινομένων ή για την πρόβλεψη του καιρού. Συγκεκριμένα, στην έρευνα των Catherine και Tennessee Leeuwenburg (2023), παρουσιάζεται μια ανασκόπηση των μεθόδων μηχανικής μάθησης που χρησιμοποιούνται στην πρόβλεψη καιρού και κλίματος. Οι συγγραφείς αναλύουν τις πιο δημοφιλείς τεχνικές, όπως τα Βαθιά Νευρωνικά Δίκτυα, τα Δάση Απόφασης και τα Υποστηρικτικά Διανυσματικά Μηχανήματα, τονίζοντας την ικανότητά τους να βελτιώνουν τις προβλέψεις βραχυπρόθεσμου καιρού και μακροπρόθεσμων κλιματικών αλλαγών.

Επισημαίνεται δε, ότι τα Βαθιά Νευρωνικά Δίκτυα έχουν τη δυνατότητα να ενσωματώσουν πολύπλοκες σχέσεις μεταξύ των μετεωρολογικών παραμέτρων, προσφέροντας υψηλότερη ακρίβεια στις προβλέψεις. Τα Δάση Απόφασης παρέχουν έναν ευέλικτο και ισχυρό τρόπο αντιμετώπισης δεδομένων υψηλής διαστάσεως, ενώ τα Υποστηρικτικά Διανυσματικά Μηχανήματα είναι αποτελεσματικά στην ανάλυση μη γραμμικών σχέσεων στα δεδομένα (Catherine & Tennessee Leeuwenburg, 2023).

Ακόμη, η έρευνα του Collins et al. (2023) εξετάζει τις πρόσφατες και αναδυόμενες εφαρμογές της μηχανικής μάθησης στην κλιματική μεταβλητότητα και τα μετεωρολογικά φαινόμενα. Οι συγγραφείς υπογραμμίζουν τη σημασία της διεπιστημονικής συνεργασίας μεταξύ φυσικών επιστημόνων και επιστημόνων υπολογιστών για την αντιμετώπιση των προκλήσεων στον τομέα αυτό και την περαιτέρω πρόοδο της επιστήμης.

Η έρευνα των Molina et al. (2023) τονίζει τις προκλήσεις που αντιμετωπίζουν τα μοντέλα βαθιάς μάθησης στην πρόβλεψη ακραίων καιρικών φαινομένων. Οι συγγραφείς προτείνουν τη βελτιστοποίηση των υπάρχοντων αρχιτεκτονικών βαθιάς μάθησης για να αντιμετωπιστούν καλύτερα οι ακραίες καιρικές συνθήκες (Molina et al., 2023).

Η μελέτη του ECMWF (2023) παρουσιάζει τις πρόσφατες εξελίξεις στη χρήση μοντέλων μηχανικής μάθησης για την πρόβλεψη καιρού. Τα μοντέλα Pangu-Weather και FourCastNet αποδείχθηκαν ιδιαίτερα αποτελεσματικά, προσφέροντας συγκρίσιμες ή και καλύτερες προβλέψεις σε σχέση με τα παραδοσιακά μοντέλα, ενώ ταυτόχρονα καταναλώνουν λιγότερη ενέργεια και χρόνο (Maskell, 2023).

Η μελέτη των Bochenek et al. (2022) παρέχει μια αναλυτική επισκόπηση της εφαρμογής της μηχανικής μάθησης στην πρόβλεψη καιρού, τονίζοντας την ταχύτητα με την οποία οι τεχνολογίες αυτές εξελίσσονται και τις προκλήσεις που αντιμετωπίζουν (Applied Sciences, 2023).

Συνολικά, η μηχανική μάθηση έχει προσφέρει νέες δυνατότητες και προοπτικές στην πρόβλεψη καιρού, ενισχύοντας την ακρίβεια των προβλέψεων και επιτρέποντας την καλύτερη κατανόηση των μετεωρολογικών φαινομένων. Η συνεχής ανάπτυξη και βελτίωση των αλγορίθμων και των τεχνολογιών αναμένεται να προσφέρει ακόμα πιο αξιόπιστες προβλέψεις στο μέλλον.

ΚΕΦΑΛΑΙΟ 3 Συστήματα Πρόβλεψης Καιρικών Συνθηκών

Παραδοσιακές Μέθοδοι Πρόβλεψης

Ιστορική Αναδρομή των Παραδοσιακών Μεθόδων

Η ιστορία της πρόβλεψης του καιρού ξεκινά από την αρχαιότητα, όταν οι πρώτοι άνθρωποι άρχισαν να παρατηρούν και να καταγράφουν τα καιρικά φαινόμενα με σκοπό να προβλέψουν τις μελλοντικές καιρικές συνθήκες. Αυτή η ανάγκη για πρόβλεψη ήταν ζωτικής σημασίας για την επιβίωση και την ευημερία τους, καθώς οι καιρικές συνθήκες επηρέαζαν άμεσα τη γεωργία, το κυνήγι, τη ναυσιπλοΐα και άλλες δραστηριότητες της καθημερινής ζωής.

Οι πρώτες μέθοδοι πρόβλεψης του καιρού ήταν βασισμένες σε παρατηρήσεις των φυσικών φαινομένων και της συμπεριφοράς των ζώων. Στην Αρχαία Ελλάδα, ο Αριστοτέλης ήταν από τους πρώτους που κατέγραψε συστηματικά τις παρατηρήσεις του για τον καιρό στο έργο του "Μετεωρολογικά" (Tsonis & Zerefos, 2023). Οι Αρχαίοι Έλληνες χρησιμοποιούσαν επίσης τις παρατηρήσεις των ανέμων και των νεφών για να προβλέψουν τις καιρικές συνθήκες. Για παράδειγμα, η εμφάνιση κόκκινου ουρανού κατά την ανατολή θεωρούνταν σημάδι καταιγίδας.

Παράλληλα, στην Αρχαία Κίνα, οι πρώτοι μετεωρολόγοι κατέγραφαν τις παρατηρήσεις τους σε μπαμπού και χρησιμοποιούσαν αυτές τις καταγραφές για να βρουν μοτίβα στις καιρικές συνθήκες (Chen et al., 2020; Di, 2008). Οι παρατηρήσεις των ζώων ήταν επίσης σημαντικές. Η συμπεριφορά των πουλιών, των εντόμων και των θηλαστικών θεωρούνταν συχνά προάγγελος των καιρικών αλλαγών. Για παράδειγμα, οι αρχαίοι Κινέζοι παρατηρούσαν τις κινήσεις των μυρμηγκιών και πίστευαν ότι η αυξημένη δραστηριότητά τους προμήνυε βροχή.

Στη μεσαιωνική Ευρώπη, η πρόβλεψη του καιρού βασιζόταν κυρίως σε εμπειρικές παρατηρήσεις και λαϊκές παραδόσεις. Οι αγρότες χρησιμοποιούσαν ρητά και παροιμίες για να προβλέψουν τις καιρικές συνθήκες, όπως "*Κόκκινος ουρανός το πρωί, οι ναυτικοί πενθούν*" που υποδήλωνε κακές καιρικές συνθήκες (Lawrence-Mathers, 2021). Οι μεσαιωνικοί παρατηρητές πρόσεχαν επίσης τις φάσεις της Σελήνης και προσπαθούσαν να συνδέσουν αυτές τις παρατηρήσεις με τις αλλαγές του καιρού. Οι Αρχαίοι Ρωμαίοι, για παράδειγμα, κατέγραφαν τις φάσεις της Σελήνης και προσπαθούσαν να προβλέψουν τις καιρικές συνθήκες με βάση αυτές τις παρατηρήσεις (McCartney, 1928).

Οι Ινδιάνοι της Βόρειας Αμερικής είχαν επίσης αναπτύξει εμπειρικές μεθόδους πρόβλεψης του καιρού. Βασίζονταν στις παρατηρήσεις της φύσης και στα σημάδια που τους έδινε το περιβάλλον τους. Για παράδειγμα, οι Ινδιάνοι παρατηρούσαν τις μεταναστευτικές κινήσεις των πουλιών και την ανάπτυξη των φυτών για να προβλέψουν τις εποχιακές αλλαγές του καιρού (Peppler, 2010).

Επιπλέον, παρατηρούσαν τις συμπεριφορές των ζώων πριν από την έλευση των καταιγίδων και θεωρούσαν ότι η συμπεριφορά τους μπορούσε να προμηνύει τις καιρικές συνθήκες.

Συνολικά, η πρόβλεψη του καιρού στην αρχαιότητα και το μεσαίωνα ήταν μια διαδικασία που βασιζόταν κυρίως στην παρατήρηση και την εμπειρία. Οι παραδοσιακές μέθοδοι, παρόλο που ήταν απλές και συχνά ατελείς, αποτέλεσαν τη βάση για τις σύγχρονες επιστημονικές μεθόδους πρόβλεψης του καιρού. Με την πάροδο του χρόνου και την ανάπτυξη της τεχνολογίας, οι μέθοδοι αυτές εξελίχθηκαν και βελτιώθηκαν, οδηγώντας στις σύγχρονες τεχνικές που χρησιμοποιούνται σήμερα.

Μετεωρολογία στον 19^ο – 20^ο αι.

Η μετεωρολογία από τον 19ο αιώνα μέχρι τον 20ο αιώνα βίωσε σημαντικές εξελίξεις, με την ανάπτυξη νέων μεθόδων και τεχνολογιών που βελτίωσαν την ακρίβεια και την αξιοπιστία των προβλέψεων. Αυτές οι εξελίξεις προετοίμασαν το έδαφος για τις σύγχρονες προσεγγίσεις που βασίζονται σε μηχανική μάθηση και νέες τεχνολογίες.

Στις αρχές του 19ου αιώνα, η μετεωρολογία άρχισε να εξελίσσεται ως επιστημονικός κλάδος με τη βοήθεια των νέων εργαλείων μέτρησης και των συστηματικών παρατηρήσεων. Οι εφευρέσεις όπως το βαρόμετρο και το θερμόμετρο, που αναπτύχθηκαν ήδη από τον 17ο αιώνα, έγιναν πιο ακριβείς και ευρέως διαθέσιμες, επιτρέποντας τη συγκέντρωση δεδομένων από διαφορετικές περιοχές και τη δημιουργία πιο αξιόπιστων προβλέψεων (Myrtonets, 2023).

Ο 19ος αιώνας είδε επίσης την ανάπτυξη της τηλεγραφίας, που επέτρεψε τη γρήγορη μετάδοση μετεωρολογικών δεδομένων σε μεγάλες αποστάσεις. Αυτό το επίτευγμα ήταν κρίσιμο για την ανάπτυξη των πρώτων μετεωρολογικών δικτύων, επιτρέποντας την ταυτόχρονη καταγραφή και ανάλυση δεδομένων από διαφορετικές περιοχές (Neves et al., 2017). Η δυνατότητα μετάδοσης δεδομένων σε πραγματικό χρόνο ενίσχυσε την ικανότητα πρόβλεψης των καιρικών φαινομένων και της ταχείας ανταπόκρισης σε ακραία καιρικά φαινόμενα.

Ο 20ός αιώνας ήταν μια περίοδος μεγάλης προόδου στη μετεωρολογία. Η εισαγωγή των ηλεκτρονικών υπολογιστών τη δεκαετία του 1950 έφερε επανάσταση στην ανάλυση των μετεωρολογικών δεδομένων. Οι πρώτοι ηλεκτρονικοί υπολογιστές χρησιμοποιήθηκαν για να εκτελούν πολύπλοκους υπολογισμούς και να επεξεργάζονται μεγάλα σύνολα δεδομένων, κάτι που ήταν αδύνατο με τις παραδοσιακές μεθόδους (Lynch, 2006). Αυτό οδήγησε στην ανάπτυξη των πρώτων αριθμητικών μοντέλων πρόγνωσης του καιρού, που βασίζονταν στις εξισώσεις της φυσικής για την προσομοίωση της ατμόσφαιρας.

Η ανάπτυξη των δορυφορικών τεχνολογιών τη δεκαετία του 1960 και 1970 ήταν ένα άλλο σημαντικό βήμα για τη μετεωρολογία. Οι μετεωρολογικοί δορυφόροι επέτρεψαν τη συλλογή δεδομένων από δυσπρόσιτες περιοχές, όπως οι ωκεανοί και οι πολικές περιοχές, παρέχοντας μια ολοκληρωμένη εικόνα των παγκόσμιων καιρικών συνθηκών (Kidder & Vonder, 1995). Οι δορυφόροι

επέτρεψαν την παρακολούθηση των καιρικών συστημάτων σε πραγματικό χρόνο και τη βελτίωση της ακρίβειας των προβλέψεων.

Οι ραδιοβολίσεις, που εισήχθησαν τη δεκαετία του 1920, έδωσαν τη δυνατότητα για τη μέτρηση της θερμοκρασίας, της υγρασίας και της πίεσης σε διάφορα ύψη στην ατμόσφαιρα. Οι μετρήσεις αυτές ήταν κρίσιμες για την κατανόηση των ατμοσφαιρικών διεργασιών και τη βελτίωση των προβλέψεων (Fleming, 2018).

Με τη συνεχή ανάπτυξη των τεχνολογιών και των επιστημονικών γνώσεων, η μετεωρολογία στον 19ο και τον 20ό αιώνα έκανε σημαντικά βήματα προόδου, που έθεσαν τα θεμέλια για τις σύγχρονες μεθόδους πρόβλεψης του καιρού. Η εξέλιξη των εργαλείων μέτρησης, η ανάπτυξη των υπολογιστικών μοντέλων και η χρήση των δορυφορικών δεδομένων βελτίωσαν σημαντικά την ακρίβεια και την αξιοπιστία των προβλέψεων, καθιστώντας τη μετεωρολογία μια από τις πιο δυναμικές και σημαντικές επιστήμες του 21ου αιώνα.

Μετεωρολογία στον 21^ο αι.

Ο 21ος αιώνας έφερε σημαντικές εξελίξεις στη μετεωρολογία, με την εμφάνιση νέων τεχνολογιών και την εφαρμογή της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Οι σύγχρονες μέθοδοι πρόβλεψης του καιρού είναι πλέον πιο ακριβείς και αξιόπιστες από ποτέ, χάρη στην πρόοδο της τεχνολογίας και την αυξανόμενη κατανόηση των ατμοσφαιρικών διεργασιών.

Μία από τις μεγαλύτερες προόδους του 21ου αιώνα είναι η ανάπτυξη και η χρήση των αριθμητικών μοντέλων πρόγνωσης καιρού. Τα μοντέλα αυτά βασίζονται σε μαθηματικές εξισώσεις που περιγράφουν τις φυσικές διεργασίες της ατμόσφαιρας. Τα δεδομένα εισόδου για τα μοντέλα αυτά προέρχονται από δορυφόρους, μετεωρολογικούς σταθμούς και άλλες πηγές, και επεξεργάζονται από ισχυρούς υπολογιστές για την παροχή ακριβών προβλέψεων. Η συνεχής βελτίωση των μοντέλων αυτών έχει οδηγήσει σε σημαντική μείωση των σφαλμάτων στις προβλέψεις καιρού (Bauer et al., 2015).

Η χρήση δορυφόρων στον 21ο αιώνα έχει επίσης εξελιχθεί, παρέχοντας δεδομένα υψηλής ανάλυσης για την παρακολούθηση της ατμόσφαιρας σε παγκόσμιο επίπεδο. Οι δορυφόροι γεωστατικής και πολικής τροχιάς συλλέγουν δεδομένα σχετικά με τη θερμοκρασία, την υγρασία, την ταχύτητα ανέμου και άλλες παραμέτρους, επιτρέποντας τη συνεχή παρακολούθηση των καιρικών συνθηκών (Kidder & Vonder, 1995). Η δυνατότητα να παρακολουθούμε τα καιρικά συστήματα σε πραγματικό χρόνο είναι κρίσιμη για την πρόβλεψη ακραίων καιρικών φαινομένων και την προστασία των πληθυσμών από φυσικές καταστροφές.

Η μηχανική μάθηση και η τεχνητή νοημοσύνη έχουν φέρει επανάσταση στην πρόβλεψη του καιρού. Τεχνικές όπως τα νευρωνικά δίκτυα, οι αλγόριθμοι υποστήριξης διανυσμάτων (SVMs) και τα μοντέλα ενίσχυσης δέντρων αποφάσεων (boosting) έχουν εφαρμοστεί με επιτυχία στην ανάλυση

μεγάλων δεδομένων και στην πρόβλεψη των καιρικών συνθηκών (Andrychowicz et al., 2023). Αυτές οι τεχνικές επιτρέπουν την ανίχνευση μοτίβων στα δεδομένα και την παροχή προβλέψεων με υψηλή ακρίβεια. Ένα παράδειγμα είναι η χρήση των μοντέλων αυτών για την πρόβλεψη ακραίων καιρικών φαινομένων, όπως οι τυφώνες και οι καταιγίδες, προσφέροντας πολύτιμες πληροφορίες στις αρχές για την προετοιμασία και την αντιμετώπιση των καταστροφών.

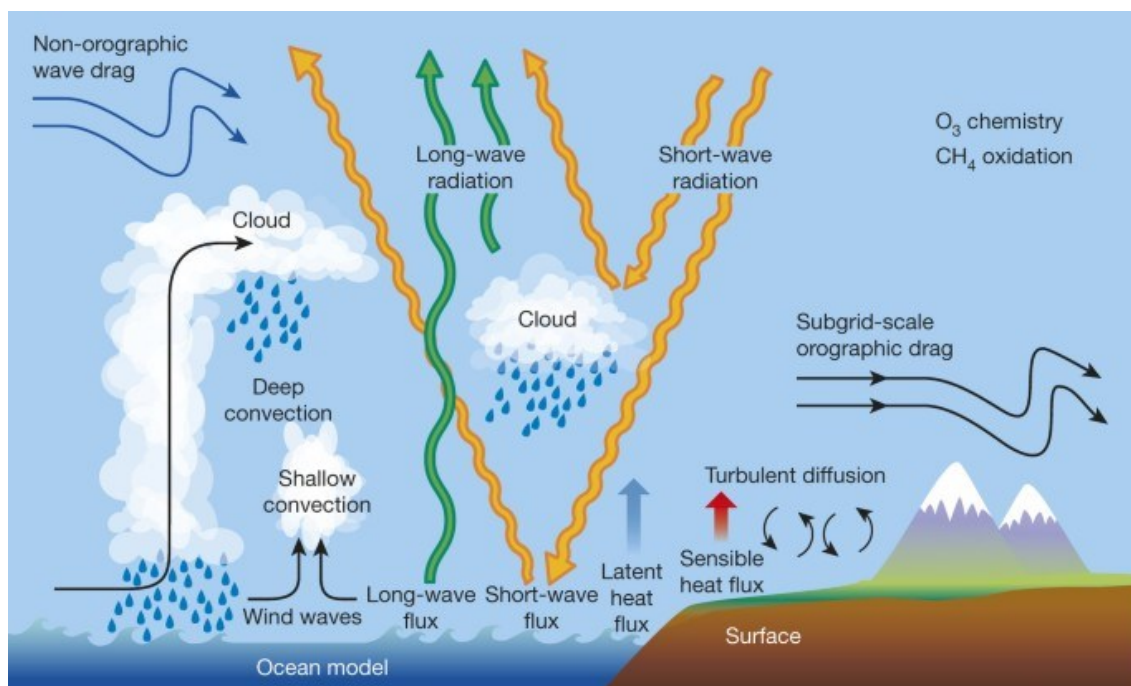
Τα νέα εργαλεία και οι τεχνολογίες έχουν επίσης διευκολύνει την ανάλυση των κλιματικών αλλαγών. Με τη βοήθεια της μηχανικής μάθησης, οι επιστήμονες μπορούν να αναλύσουν τεράστιες ποσότητες δεδομένων και να εντοπίσουν μακροπρόθεσμα μοτίβα στις κλιματικές συνθήκες. Αυτές οι αναλύσεις είναι κρίσιμες για την κατανόηση των επιπτώσεων της κλιματικής αλλαγής και τη λήψη μέτρων για τον μετριασμό των επιπτώσεων αυτών (Rolnick et al., 2019).

Η συνεχής ανάπτυξη των τεχνολογιών και των επιστημονικών γνώσεων στον τομέα της μετεωρολογίας έχει βελτιώσει σημαντικά την ακρίβεια και την αξιοπιστία των προβλέψεων. Η χρήση αριθμητικών μοντέλων, δορυφόρων και τεχνικών μηχανικής μάθησης έχει ανοίξει νέους ορίζοντες για την πρόβλεψη του καιρού και την κατανόηση των κλιματικών φαινομένων. Οι τεχνολογίες αυτές συνεχίζουν να εξελίσσονται, προσφέροντας νέες δυνατότητες και προοπτικές για τη μετεωρολογία στον 21ο αιώνα.

Αριθμητικά Μοντέλα Καιρού

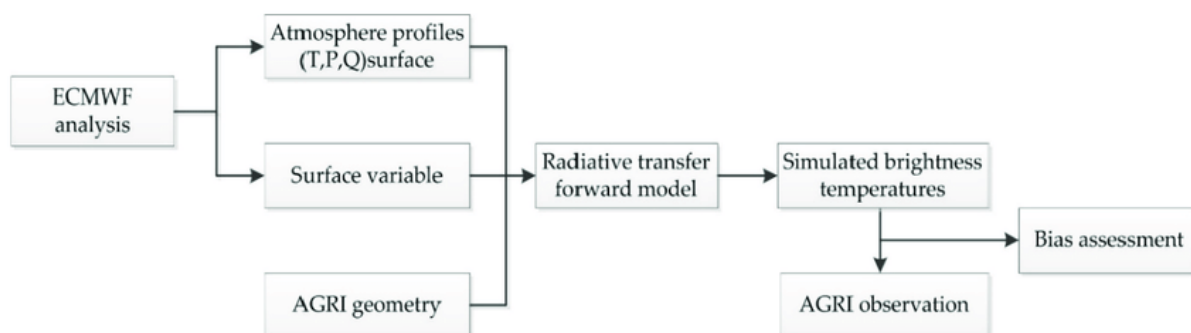
Περιγραφή των Αριθμητικών Μοντέλων

Τα αριθμητικά μοντέλα καιρού (Numerical Weather Prediction - NWP) είναι θεμελιώδη εργαλεία στη σύγχρονη μετεωρολογία, επιτρέποντας την ακριβή πρόβλεψη καιρού μέσω της χρήσης μαθηματικών εξισώσεων που βασίζονται στη φυσική. Αυτά τα μοντέλα περιλαμβάνουν δεδομένα εισόδου από διάφορες πηγές, όπως δορυφόρους, μετεωρολογικούς σταθμούς και αισθητήρες, τα οποία υποβάλλονται σε επεξεργασία από ισχυρούς υπερυπολογιστές. Η διαδικασία αυτή επιτρέπει τη δημιουργία προβλέψεων για τις μελλοντικές καιρικές συνθήκες, οι οποίες είναι ζωτικής σημασίας για την προετοιμασία και την αντιμετώπιση των ακραίων καιρικών φαινομένων (Bauer et al., 2015).



Εικόνα 1: Η εικόνα απεικονίζει διάφορες διεργασίες και μηχανισμούς που λαμβάνουν χώρα στην ατμόσφαιρα και στον ωκεανό, και είναι κρίσιμες για τα αριθμητικά μοντέλα καιρού. Συγκεκριμένα, παρουσιάζει τη μεταφορά θερμότητας μέσω της βαθιάς και ρηχής μεταφοράς (deep and shallow convection), την ακτινοβολία μακρών και βραχέων κυμάτων (long-wave and short-wave radiation), τη διάχυση της θερμότητας μέσω της επιφάνειας και του ωκεανού, την ταραχώδη διάχυση (turbulent diffusion), καθώς και άλλες διεργασίες όπως η ορογραφική έλξη (subgrid-scale orographic drag) και οι χημικές αντιδράσεις στην ατμόσφαιρα. Αυτές οι διαδικασίες συνδυάζονται για να δημιουργήσουν ένα σύνθετο σύστημα που προσομοιώνεται από τα αριθμητικά μοντέλα καιρού για την πρόβλεψη των καιρικών συνθηκών. (Πηγή: Bauer et al., 2015)

Η εξέλιξη των αριθμητικών μοντέλων καιρού ξεκίνησε με τις πρώτες προσπάθειες πρόβλεψης του καιρού στις αρχές του 20ού αιώνα, όταν οι μετεωρολόγοι χρησιμοποίησαν βασικές μαθηματικές εξισώσεις για να κατανοήσουν τις καιρικές διαδικασίες. Με την πρόοδο της τεχνολογίας και την ανάπτυξη των υπολογιστών, τα μοντέλα αυτά έγιναν πιο πολύπλοκα και ακριβή. Σήμερα, τα αριθμητικά μοντέλα καιρού βασίζονται σε ένα σύνθετο σύνολο διαφορικών εξισώσεων που περιγράφουν τη ροή του αέρα, τη μεταφορά θερμότητας, την εξάτμιση και τη συμπύκνωση, καθώς και άλλες διεργασίες της ατμόσφαιρας (Kalnay, 2002).



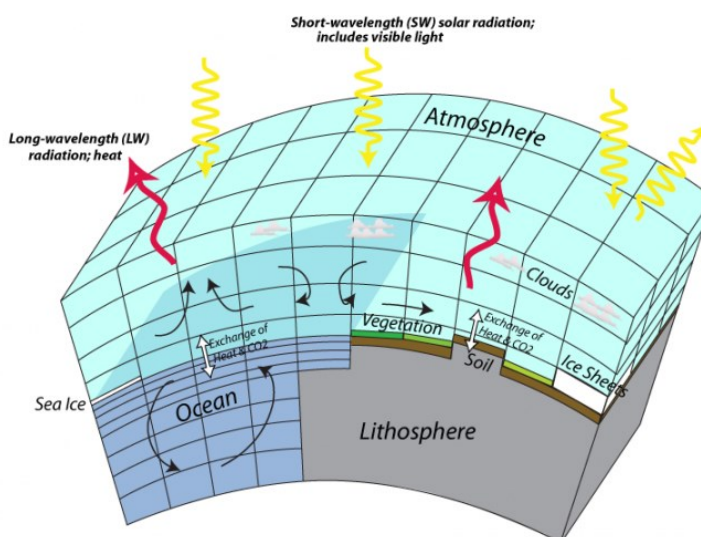
Εικόνα 2: Διαγραμματική αναπαράσταση της διαδικασίας που χρησιμοποιεί το Ευρωπαϊκό Κέντρο Μεσοπρόθεσμων Προγνώσεων Καιρού (ECMWF) για την ανάλυση και πρόγνωση του καιρού. Η διαδικασία περιλαμβάνει την είσοδο ατμοσφαιρικών προφίλ (θερμοκρασία, πίεση, υγρασία) και επιφανειακών μεταβλητών σε ένα μοντέλο προώθησης ακτινοβολίας, το οποίο προσομοιώνει τις φωτεινές θερμοκρασίες. Αυτές συγκρίνονται με τις παρατηρήσεις AGRI για την αξιολόγηση των προκαταλήψεων, βελτιώνοντας έτσι την ακρίβεια των προγνώσεων καιρού. (Πηγή: Zhu et al., 2020)

Τα δεδομένα που χρησιμοποιούνται στα NWP μοντέλα προέρχονται από διάφορες πηγές, όπως δορυφόροι, μετεωρολογικά μπαλόνια και ραντάρ. Αυτά τα δεδομένα συνδυάζονται και επεξεργάζονται για να δημιουργήσουν μια ολοκληρωμένη εικόνα των καιρικών συνθηκών. Τα μοντέλα αυτά ανανεώνονται τακτικά με νέα δεδομένα και οι προβλέψεις ενημερώνονται σε πραγματικό χρόνο. Η συνεχής βελτίωση των αριθμητικών μοντέλων και η αύξηση της υπολογιστικής ισχύος έχουν οδηγήσει σε σημαντικές προόδους στην ακρίβεια και την αξιοπιστία των προβλέψεων, καθιστώντας τα NWP μοντέλα αναπόσπαστο μέρος της σύγχρονης μετεωρολογίας (Blanchonnet, 2022; NOAA, 2019).

Μοντέλα Γενικής Κυκλοφορίας (General Circulation Models - GCMs)

Τα Μοντέλα Γενικής Κυκλοφορίας (General Circulation Models - GCMs) είναι πολύπλοκα μαθηματικά μοντέλα που περιλαμβάνουν μια σειρά εξισώσεων που περιγράφουν τη φυσική των ατμοσφαιρικών και ωκεάνιων ρευμάτων, της ακτινοβολίας, της υγρασίας, και άλλων βασικών παραμέτρων του κλιματικού συστήματος. Τα δεδομένα που χρησιμοποιούνται στα GCMs προέρχονται από διάφορες πηγές, όπως μετεωρολογικούς σταθμούς, δορυφόρους και ωκεανογραφικά πλοία (Bauer et al., 2015).

Τα πρώτα GCMs αναπτύχθηκαν τη δεκαετία του 1950 και του 1960 και ήταν σχετικά απλά σε σύγκριση με τα σύγχρονα μοντέλα. Με την πρόοδο της υπολογιστικής τεχνολογίας και την αύξηση της υπολογιστικής ισχύος, τα GCMs έχουν εξελιχθεί σημαντικά, προσφέροντας υψηλότερη ανάλυση και μεγαλύτερη ακρίβεια στις προβλέψεις. Τα σύγχρονα GCMs μπορούν να προσομοιώνουν με ακρίβεια τη συμπεριφορά της ατμόσφαιρας και των ωκεανών για περιόδους από μερικές ημέρες έως και πολλές δεκαετίες.



Εικόνα 3: Η εικόνα απεικονίζει ένα Μοντέλο Γενικής Κυκλοφορίας (GCM) και τις διάφορες διαδικασίες ανταλλαγής ενέργειας και υλικών μεταξύ των διαφορετικών στοιχείων του κλιματικού συστήματος της Γης. Περιλαμβάνει την ανταλλαγή θερμότητας και διοξειδίου του άνθρακα (CO₂) μεταξύ του ωκεανού, της ατμόσφαιρας, της λιθόσφαιρας και των επιφανειακών στοιχείων, όπως η βλάστηση και οι πάγοι. Επίσης, δείχνει την ακτινοβολία μεγάλου κύματος (LW) και κοντού

κύματος (SW) από και προς την ατμόσφαιρα, καθώς και τις διάφορες μορφές κυκλοφορίας του αέρα και του νερού. (Πηγή: Bralower & Bice, n.d.)

Τα GCMs επιλύουν τις εξισώσεις της ρευστοδυναμικής, της θερμοδυναμικής, και της μεταφοράς ακτινοβολίας σε ένα τρισδιάστατο πλέγμα που καλύπτει την ατμόσφαιρα, τους ωκεανούς και την επιφάνεια της Γης. Κάθε κυψέλη του πλέγματος αντιπροσωπεύει ένα μικρό μέρος της ατμόσφαιρας ή του ωκεανού και οι εξισώσεις επιλύονται για κάθε κυψέλη σε διαδοχικά χρονικά βήματα (Bralower & Bice, n.d.).

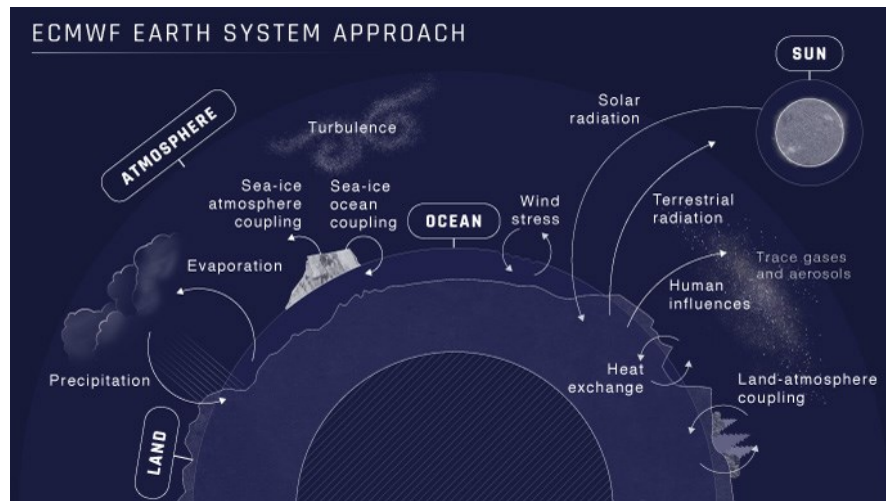
Οι βασικές εξισώσεις που χρησιμοποιούνται στα GCMs περιλαμβάνουν:

Πίνακας 1: Βασικές Εξισώσεις που χρησιμοποιούνται στα GCMs (Πηγή: Bauer et al., 2015; NOAA, 2020; Palmer, 2016)

	Εξίσωση	Περιγραφή
<i>Εξίσωση της Συνέχειας (Continuity Equation)</i>	$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0$	Περιγράφει τη διατήρηση της μάζας, όπου ρ και u είναι η πυκνότητα του αέρα και η ταχύτητα του αέρα.
<i>Εξισώσεις της Κίνησης (Navier-Stokes Equations)</i>	$\frac{\partial u}{\partial t} + (u \cdot \nabla)u = -\frac{1}{\rho} \nabla p + g + F$	Περιγράφουν τη δυναμική του ρευστού, όπου p είναι η πίεση, g είναι η επιτάχυνση λόγω βαρύτητας και F είναι άλλες δυνάμεις.
<i>Εξίσωση της Ενέργειας (Energy Equation)</i>	$\frac{\partial T}{\partial t} + u \cdot \nabla T = \frac{Q}{\rho C_p}$	Περιγράφει τη μεταφορά της θερμότητας, όπου T είναι η θερμοκρασία, Q είναι η θερμότητα που προστίθεται ή αφαιρείται, και C_p είναι η ειδική θερμότητα σε σταθερή πίεση.
<i>Εξίσωση της Μεταφοράς Υγρασίας (Moisture Transport Equation)</i>	$\frac{\partial q}{\partial t} + u \cdot \nabla q = S_q$	Περιγράφει τη μεταφορά της υγρασίας, όπου q είναι η συγκεκριμένη υγρασία και S_q είναι οι πηγές και καταβόθρες υγρασίας.

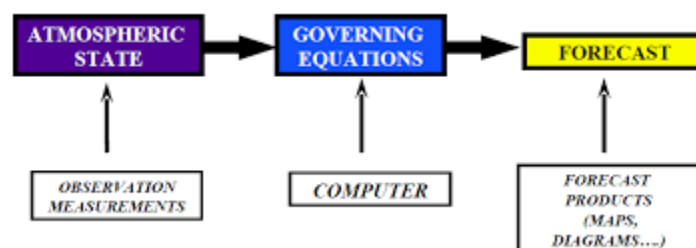
Τα δεδομένα που τροφοδοτούν τις παραπάνω εξισώσεις περιλαμβάνουν μετρήσεις θερμοκρασίας, πίεσης, υγρασίας, ταχύτητας ανέμου, και άλλες μετεωρολογικές παραμέτρους. Αυτά τα δεδομένα συλλέγονται από μια ποικιλία πηγών, όπως δορυφόρους, μετεωρολογικούς σταθμούς, και αισθητήρες IoT. Τα δεδομένα αυτά επεξεργάζονται και ενσωματώνονται στα μοντέλα μέσω τεχνικών αφομοίωσης δεδομένων (Bauer et al., 2015; NOAA, 2020; Palmer, 2016).

Δύο από τα πιο διαδεδομένα μοντέλα είναι το ECMWF (European Centre for Medium-Range Weather Forecasts) και το GFS (Global Forecast System). Η περίπτωση του ECMWF αποτελεί το πιο γνωστό GCM. Φημίζεται για την υψηλή του ακρίβεια στις προβλέψεις καιρού. Χρησιμοποιεί προηγμένες τεχνικές αφομοίωσης δεδομένων και ενσωματώνει δεδομένα από διάφορες πηγές σε πραγματικό χρόνο. Οι προβλέψεις του ECMWF είναι ιδιαίτερα αξιόπιστες για την Ευρώπη και τον υπόλοιπο κόσμο (Woods, 2005).



Εικόνα 4: Η εικόνα παρουσιάζει την προσέγγιση του ECMWF για την προσομοίωση του συστήματος της Γης. Αυτή η προσέγγιση περιλαμβάνει διάφορες διεργασίες και αλληλεπιδράσεις μεταξύ της ατμόσφαιρας, των ωκεανών, της γης και του ήλιου. Στην ατμόσφαιρα, παρατηρούμε την εξάτμιση, την καθίζηση και την ανάμιξη αερίων. Οι ωκεανοί και οι θάλασσες εμπλέκονται μέσω της σύζευξης με την ατμόσφαιρα και των επιδράσεων του ανέμου. Η γη, από την άλλη, εμπλέκεται μέσω της εξάτμισης, της σύζευξης με την ατμόσφαιρα και της ανταλλαγής θερμότητας. Όλα αυτά τα στοιχεία συνδυάζονται για την παραγωγή ακριβών προγνώσεων καιρού. (Πηγή: Jerpesen, 2021)

Ακόμα ένα πολύ γνωστό GCM είναι το GFS (Global Forecast System). Το GFS είναι το κύριο αριθμητικό μοντέλο πρόγνωσης καιρού των ΗΠΑ. Παρέχει προβλέψεις για όλον τον πλανήτη και χρησιμοποιείται ευρέως από τις μετεωρολογικές υπηρεσίες για τις βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις. Το GFS ενημερώνεται τακτικά με νέα δεδομένα και βελτιώσεις στις μεθόδους προσομοίωσης (NOAA, 2020).



Εικόνα 5: Η εικόνα παρουσιάζει τη διαδικασία πρόβλεψης καιρού χρησιμοποιώντας αριθμητικά μοντέλα. Αρχικά, συλλέγονται μετεωρολογικές παρατηρήσεις για την κατάσταση της ατμόσφαιρας. Αυτά τα δεδομένα εισάγονται σε έναν υπολογιστή, ο οποίος εφαρμόζει τις κυρίαρχες εξισώσεις της φυσικής για να υπολογίσει τις μελλοντικές καιρικές συνθήκες. Τα αποτελέσματα της πρόβλεψης περιλαμβάνουν διάφορα προϊόντα, όπως χάρτες και διαγράμματα, που χρησιμοποιούνται για την ενημέρωση των προβλέψεων καιρού. (Πηγή: Durai, 2021)

Συνοψίζοντας, τα GCMs παίζουν κρίσιμο ρόλο στη σύγχρονη μετεωρολογία και κλιματολογία. Παρέχουν προβλέψεις που βοηθούν στην προετοιμασία για ακραία καιρικά φαινόμενα και στην προστασία των πληθυσμών από φυσικές καταστροφές. Επίσης, χρησιμοποιούνται για την εκτίμηση των μελλοντικών κλιματικών συνθηκών και την κατανόηση των επιπτώσεων της κλιματικής αλλαγής. Η συνεχής βελτίωση των GCMs και η αύξηση της υπολογιστικής ισχύος έχουν οδηγήσει σε σημαντικές προόδους στην ακρίβεια των προβλέψεων.

Περιφερειακά Μοντέλα (Regional Climate Models - RCMs)

Τα Περιφερειακά Μοντέλα Κλίματος (RCMs) είναι εξειδικευμένα εργαλεία που επιτρέπουν την πρόβλεψη των κλιματικών συνθηκών σε τοπικό και περιφερειακό επίπεδο με μεγαλύτερη ακρίβεια από τα γενικά μοντέλα κυκλοφορίας (GCMs). Τα RCMs προσφέρουν λεπτομερείς προβλέψεις για περιοχές ενδιαφέροντος, λαμβάνοντας υπόψη τις τοπικές ιδιαιτερότητες που επηρεάζουν το κλίμα. Αυτό είναι ιδιαίτερα σημαντικό για τη μελέτη των κλιματικών φαινομένων που έχουν τοπικό χαρακτήρα και δεν μπορούν να αποτυπωθούν με ακρίβεια από τα GCMs (Giorgi & Gutowski, 2015; Rummukainen, 2016).

Τα RCMs λειτουργούν συνήθως ως "καταβίβαση" των GCMs. Χρησιμοποιούν δεδομένα από τα GCMs ως αρχικές και οριακές συνθήκες και στη συνέχεια εκτελούν προσομοιώσεις για την επιλεγμένη περιοχή σε υψηλότερη ανάλυση. Αυτές οι προσομοιώσεις βασίζονται στις ίδιες βασικές αρχές φυσικής που χρησιμοποιούν τα GCMs, αλλά προσαρμόζονται για να λαμβάνουν υπόψη τις τοπικές γεωγραφικές και κλιματικές συνθήκες (Torma et al., 2015; Laprise, 2008).

Οι βασικές εξισώσεις που χρησιμοποιούνται στα RCMs περιλαμβάνουν:

Πίνακας 2: Βασικές Εξισώσεις που χρησιμοποιούνται στα GCMs (Πηγή: Skamarock et al., 2008)

Εξίσωση	
Εξίσωση της Συνέχειας	$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$
Εξίσωση Navier-Stokes για την κίνηση των ρευστών	$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{g}$
Εξίσωση θερμοδυναμικής ενέργειας	$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \kappa \nabla^2 T + Q$

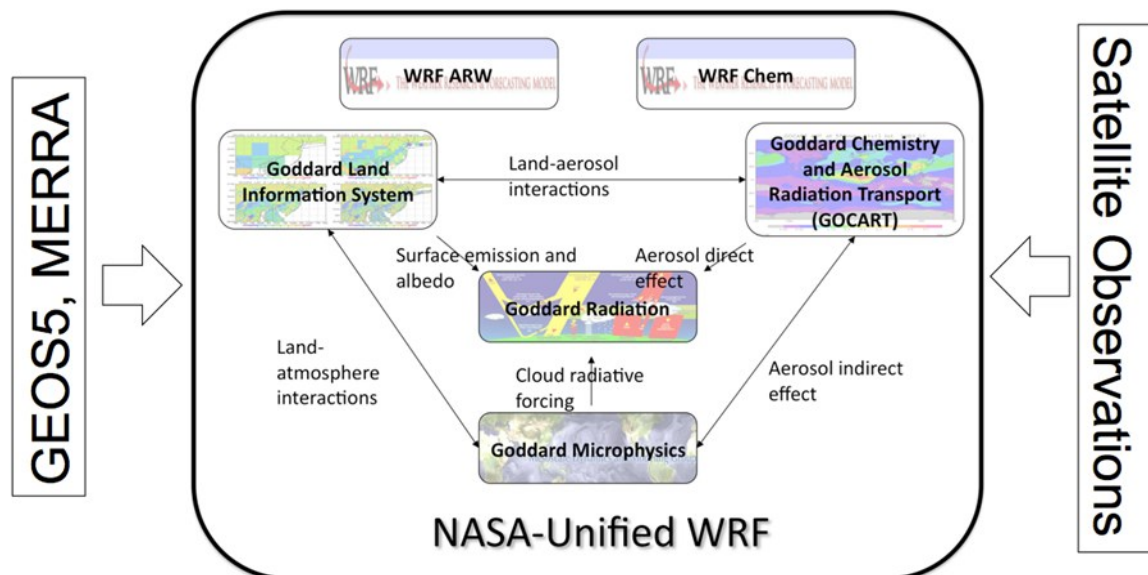
όπου ρ είναι η πυκνότητα, $\mathbf{u}$ είναι το πεδίο ταχύτητας, p είναι η πίεση, ν είναι το κινηματικό ιξώδες, $\mathbf{g}$ είναι η βαρύτητα, T είναι η θερμοκρασία, κ είναι η θερμική αγωγιμότητα και Q είναι η πηγή θερμότητας.

Τα RCMs προσφέρουν σημαντικά πλεονεκτήματα, όπως η δυνατότητα λεπτομερούς ανάλυσης τοπικών κλιματικών φαινομένων, η βελτίωση της ακρίβειας των προβλέψεων και η δυνατότητα προσαρμογής στις ιδιαίτερες συνθήκες μιας περιοχής. Αυτά τα πλεονεκτήματα είναι κρίσιμα για την αντιμετώπιση ακραίων καιρικών φαινομένων, την ανάλυση της κλιματικής αλλαγής και την ανάπτυξη

στρατηγικών προσαρμογής (Laprise, 2008; Torma et al., 2015). Ωστόσο, τα RCMs εξαρτώνται από την ακρίβεια των δεδομένων που παρέχονται από τα GCMs, και η υπολογιστική ισχύς που απαιτείται μπορεί να είναι σημαντική. Επιπλέον, η ανάγκη για υψηλής ποιότητας δεδομένα παρατήρησης και οι περιορισμοί στην ανάλυση μπορεί να οδηγήσουν σε σφάλματα στις προβλέψεις (Giorgi & Gutowski, 2015).

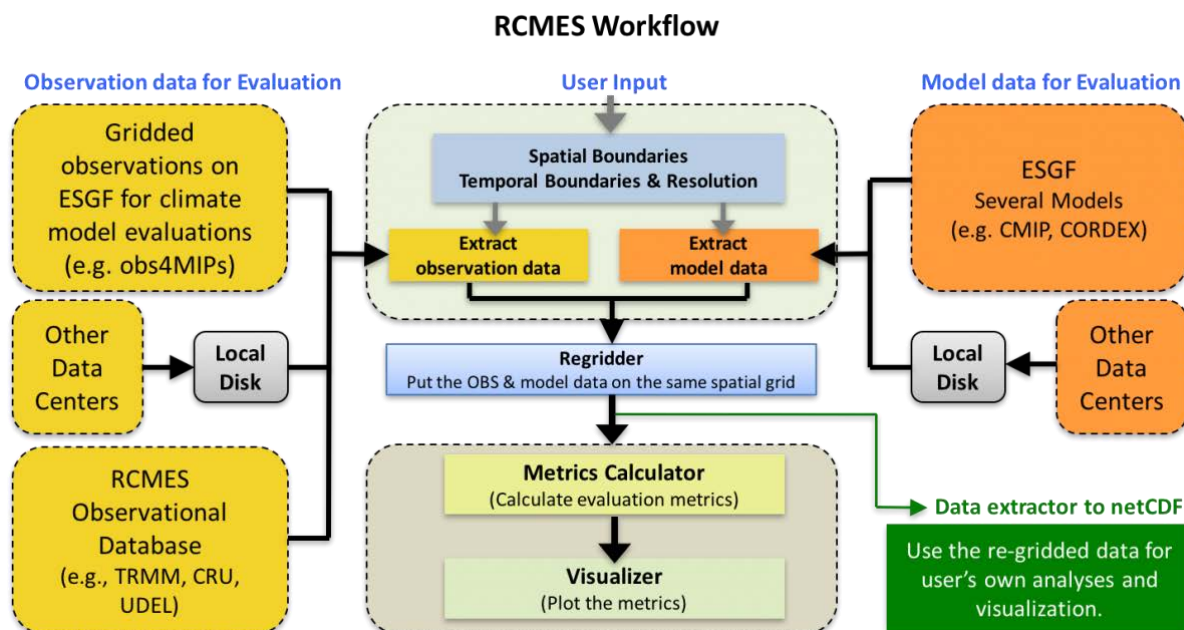
Κάποια παραδείγματα συστημάτων RCMs είναι:

- **Weather Research and Forecasting Model (WRF):** Το WRF είναι ένα ευέλικτο και προσαρμόσιμο μοντέλο που χρησιμοποιείται ευρέως για την προσομοίωση των καιρικών συνθηκών σε τοπικό και περιφερειακό επίπεδο. Αναπτύχθηκε από το Εθνικό Κέντρο Ατμοσφαιρικής Έρευνας (NCAR) και χρησιμοποιείται για την έρευνα και την επιχειρησιακή πρόβλεψη καιρού (Skamarock et al., 2008).



Εικόνα 6: Διάγραμμα που απεικονίζει τη διαδικασία ενσωμάτωσης δεδομένων από τα συστήματα GEOS5 και MERRA, τις δορυφορικές παρατηρήσεις και τα υποσυστήματα Goddard για την παραγωγή και βελτίωση των προβλέψεων μέσω του μοντέλου WRF της NASA. (Πηγή: Mohr, 2023)

- **Regional Climate Model (RegCM):** Το RegCM, αναπτυγμένο από το Διεθνές Κέντρο Θεωρητικής Φυσικής (ICTP), χρησιμοποιείται για την προσομοίωση κλιματικών συνθηκών σε περιοχές όπως η Ευρώπη, η Αφρική και η Ασία. Παρέχει λεπτομερείς προβλέψεις για παραμέτρους όπως η θερμοκρασία και οι βροχοπτώσεις, βοηθώντας στην κατανόηση των τοπικών κλιματικών αλλαγών (Torma et al., 2015).



Εικόνα 7: Διάγραμμα ροής που παρουσιάζει την επεξεργασία δεδομένων παρατήρησης και μοντέλων για την αξιολόγηση κλιματικών μοντέλων. Το διάγραμμα δείχνει τα δεδομένα εισόδου από πηγές παρατήρησης και μοντέλα, τη διαδικασία επανεπεξεργασίας (regridding) και την τελική αξιολόγηση μέσω υπολογισμού μετρικών και απεικόνισης. (Πηγή: NASA, n.d.)

- **Modèle Atmosphérique Régional (MAR):** Το MAR χρησιμοποιείται κυρίως για την προσομοίωση του κλίματος στις πολικές περιοχές και τις περιοχές με υψηλά υψόμετρα. Έχει χρησιμοποιηθεί εκτενώς για τη μελέτη της κλιματικής αλλαγής στην Ανταρκτική και τη Γροιλανδία, προσφέροντας λεπτομερείς προσομοιώσεις των κλιματικών συνθηκών σε αυτές τις περιοχές (Laprise, 2008).

Συγκριτική Αξιολόγηση των Αριθμητικών Μοντέλων

Η συγκριτική αξιολόγηση των αριθμητικών μοντέλων καιρού (Numerical Weather Prediction - NWP) αποτελεί σημαντικό βήμα για την κατανόηση της αποτελεσματικότητας και της ακρίβειάς τους. Τα μοντέλα αυτά βασίζονται σε μαθηματικές εξισώσεις που περιγράφουν τη φυσική της ατμόσφαιρας και απαιτούν μεγάλες υπολογιστικές ικανότητες. Η αξιολόγηση των μοντέλων περιλαμβάνει τη σύγκριση των προβλέψεών τους με τα πραγματικά δεδομένα και τον εντοπισμό των διαφορών.

Τα μοντέλα γενικής κυκλοφορίας (General Circulation Models - GCMs) και τα περιφερειακά μοντέλα (Regional Climate Models - RCMs) είναι δύο βασικοί τύποι αριθμητικών μοντέλων που χρησιμοποιούνται για την πρόβλεψη του καιρού και την κλιματική μοντελοποίηση.

Τα GCMs είναι σχεδιασμένα για να προσομοιώνουν την κυκλοφορία της ατμόσφαιρας και των ωκεανών σε παγκόσμια κλίμακα. Αυτά τα μοντέλα χρησιμοποιούν ένα δίκτυο τρισδιάστατων κυψελών που καλύπτει όλη τη Γη και υπολογίζουν την κίνηση του αέρα και του νερού, καθώς και τις

ανταλλαγές θερμότητας και υγρασίας μεταξύ αυτών των κυψελών. Οι εξισώσεις Navier-Stokes, οι οποίες περιγράφουν τη ροή ρευστών, αποτελούν τη βάση των υπολογισμών αυτών (Randall et al., 2007).

Τα RCMs, από την άλλη πλευρά, επικεντρώνονται σε μικρότερες γεωγραφικές περιοχές και παρέχουν πιο λεπτομερείς προβλέψεις. Αυτά τα μοντέλα λαμβάνουν δεδομένα από τα GCMs και τα χρησιμοποιούν ως όρια για τις προσομοιώσεις τους. Έτσι, τα RCMs μπορούν να προσφέρουν πιο ακριβείς προβλέψεις για συγκεκριμένες περιοχές, όπως πόλεις ή περιοχές με ιδιαίτερα κλιματικά χαρακτηριστικά (Giorgi, 2019).

Πίνακας 3: Παρουσίαση κύριων πλεονεκτημάτων και μειονεκτημάτων μεταξύ των συστημάτων GCM και RCM (Πηγή: Giorgi, 2019)

	GCMs	RCMs
Πλεονεκτήματα	Παρέχουν προβλέψεις σε παγκόσμια κλίμακα.	Υψηλή χωρική ανάλυση για συγκεκριμένες περιοχές.
	Μπορούν να χρησιμοποιηθούν για τη μελέτη των παγκόσμιων κλιματικών αλλαγών.	Βελτιωμένη ακρίβεια στις τοπικές προβλέψεις.
	Κατάλληλα για μακροπρόθεσμες προβλέψεις.	Μπορούν να ενσωματώσουν τοπικά δεδομένα για καλύτερες προβλέψεις.
Μειονεκτήματα	Μειωμένη χωρική ανάλυση.	Εξαρτώνται από τα GCMs για τα δεδομένα ορίων.
	Περιορισμένη ικανότητα πρόβλεψης τοπικών φαινομένων.	Περιορισμένη ικανότητα πρόβλεψης σε παγκόσμια κλίμακα.

Μηχανική Μάθηση και Τεχνητή Νοημοσύνη

Η μηχανική μάθηση και η τεχνητή νοημοσύνη έχουν μετασχηματίσει πολλούς τομείς της σύγχρονης ζωής και επιστήμης, συμπεριλαμβανομένης της μετεωρολογίας. Οι τεχνικές αυτές παρέχουν νέα εργαλεία για την ανάλυση δεδομένων, την πρόβλεψη και τη λήψη αποφάσεων, επιτρέποντας την ανάπτυξη συστημάτων που μπορούν να βελτιώνονται αυτόματα μέσω της εμπειρίας και της χρήσης δεδομένων. Η μηχανική μάθηση είναι ένας υποτομέας της τεχνητής νοημοσύνης που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να μαθαίνουν από δεδομένα, ενώ η τεχνητή νοημοσύνη αναφέρεται σε ένα ευρύτερο σύνολο τεχνολογιών που μιμούνται τις ανθρώπινες γνωστικές λειτουργίες (Goodfellow, Bengio, & Courville, 2016).

Τα συστήματα τεχνητής νοημοσύνης χρησιμοποιούν ποικίλες τεχνικές μηχανικής μάθησης, συμπεριλαμβανομένων των νευρωνικών δικτύων, των δασών απόφασης, των αλγορίθμων ενίσχυσης

και των μοντέλων υποστήριξης διανυσμάτων. Κάθε μία από αυτές τις τεχνικές έχει τα δικά της πλεονεκτήματα και περιορισμούς, και η επιλογή της κατάλληλης τεχνικής εξαρτάται από την εφαρμογή και τον τύπο των δεδομένων που αναλύονται. Για παράδειγμα, τα νευρωνικά δίκτυα είναι ιδιαίτερα αποτελεσματικά στην ανάλυση πολύπλοκων και μη γραμμικών δεδομένων, ενώ τα δάση απόφασης και οι αλγόριθμοι ενίσχυσης είναι γνωστά για την ακρίβειά τους και την ανθεκτικότητά τους σε υπερεκπαίδευση (Hastie, Tibshirani, & Friedman, 2008).

Οι εφαρμογές της μηχανικής μάθησης στην πρόβλεψη καιρού είναι πολυάριθμες και ποικίλες. Τα μοντέλα αυτά μπορούν να χρησιμοποιηθούν για την ανάλυση ιστορικών δεδομένων καιρού, την πρόβλεψη βραχυπρόθεσμων και μακροπρόθεσμων καιρικών φαινομένων, και την ανάλυση των επιπτώσεων της κλιματικής αλλαγής. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ενσωματώσουν δεδομένα από πολλές πηγές, όπως δορυφορικές εικόνες, αισθητήρες και παρατηρήσεις από μετεωρολογικούς σταθμούς, για να παράγουν προβλέψεις με υψηλή ακρίβεια. Η χρήση αυτών των τεχνικών έχει οδηγήσει σε σημαντικές βελτιώσεις στην ακρίβεια και την αξιοπιστία των μετεωρολογικών προβλέψεων (Schneider et al., 2017).

Η μηχανική μάθηση βοηθά στην πρόβλεψη καιρού με την ικανότητα της να διαχειρίζεται μεγάλες ποσότητες δεδομένων και να αναγνωρίζει πρότυπα που δεν είναι εύκολα ανιχνεύσιμα από ανθρώπινους αναλυτές. Οι τεχνικές αυτές επιτρέπουν την ανάλυση δεδομένων από διάφορες πηγές και την ενσωμάτωσή τους σε μοντέλα που μπορούν να προβλέψουν μετεωρολογικά φαινόμενα με υψηλή ακρίβεια. Για παράδειγμα, η χρήση νευρωνικών δικτύων για την επεξεργασία δορυφορικών εικόνων έχει επιτρέψει την βελτίωση της πρόβλεψης βροχοπτώσεων και άλλων ακραίων καιρικών φαινομένων (Shi et al., 2017).

Ένα παράδειγμα αυτής της χρήσης είναι η μελέτη των Shi et al. (2017), η οποία χρησιμοποίησε νευρωνικά δίκτυα για την πρόβλεψη βροχοπτώσεων χρησιμοποιώντας δορυφορικά δεδομένα. Οι συγγραφείς ανέπτυξαν ένα μοντέλο που μπορούσε να προβλέψει τις βροχοπτώσεις με υψηλή ακρίβεια, ενσωματώνοντας δεδομένα από πολλαπλές πηγές και αναλύοντας τις σχέσεις μεταξύ αυτών. Τα αποτελέσματα έδειξαν σημαντική βελτίωση σε σχέση με τις παραδοσιακές μεθόδους πρόβλεψης.

Επιπλέον, στην μελέτη των Chen et al. (2023), η οποία παρέχει μια επισκόπηση των μεθόδων μηχανικής μάθησης που χρησιμοποιούνται σε εφαρμογές καιρού και κλίματος. Οι συγγραφείς αναλύουν διάφορες τεχνικές, συμπεριλαμβανομένων των νευρωνικών δικτύων, των δασών απόφασης και των υποστηρικτικών διανυσματικών μηχανημάτων, και τονίζουν την αποτελεσματικότητά τους στη βελτίωση των προβλέψεων βραχυπρόθεσμου καιρού και μακροπρόθεσμων κλιματικών αλλαγών (Chen et al., 2023).

Η ενσωμάτωση μηχανικής μάθησης και τεχνητής νοημοσύνης στην πρόβλεψη καιρού δεν περιορίζεται μόνο στη βελτίωση των προβλέψεων, αλλά επεκτείνεται και στην κατανόηση των μετεωρολογικών φαινομένων. Για παράδειγμα, η χρήση LSTM νευρωνικών δικτύων για την ανάλυση χρονικών σειρών έχει επιτρέψει την κατανόηση των μακροχρόνιων εξαρτήσεων στα δεδομένα και την βελτίωση της πρόβλεψης φαινομένων όπως οι καταιγίδες και οι τυφώνες (Hochreiter & Schmidhuber, 1997).

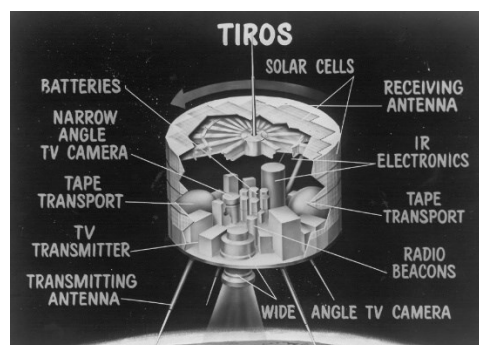
Οι εφαρμογές της μηχανικής μάθησης στην πρόβλεψη καιρού έχουν επίσης επεκταθεί στη χρήση τεχνικών ενίσχυσης (boosting), όπως το XGBoost, για την πρόβλεψη ακραίων καιρικών φαινομένων. Στην έρευνα των Chen & Guestrin (2016), το XGBoost χρησιμοποιήθηκε για την πρόβλεψη καταιγίδων και αποδείχθηκε ότι μπορούσε να παρέχει ακριβείς προβλέψεις, ακόμα και σε περιπτώσεις όπου τα δεδομένα περιείχαν πολύπλοκες και μη γραμμικές σχέσεις.

Συνολικά, η μηχανική μάθηση και η τεχνητή νοημοσύνη έχουν προσφέρει νέες δυνατότητες και προοπτικές στην πρόβλεψη καιρού, ενισχύοντας την ακρίβεια των προβλέψεων και επιτρέποντας την καλύτερη κατανόηση των μετεωρολογικών φαινομένων. Η συνεχής ανάπτυξη και βελτίωση των αλγορίθμων και των τεχνολογιών αναμένεται να προσφέρει ακόμα πιο αξιόπιστες προβλέψεις στο μέλλον (Torma et al., 2011).

Δορυφορικά Συστήματα και Αισθητήρες IoT

Δορυφορική Τεχνολογία

Η ανάπτυξη των δορυφόρων για μετεωρολογικές εφαρμογές έχει επιφέρει μια επανάσταση στον τρόπο που παρακολουθούμε και προβλέπουμε τον καιρό σε παγκόσμιο επίπεδο. Από την εποχή του πρώτου δορυφόρου, του TIROS-1, που εκτοξεύτηκε το 1960, η δυνατότητα λήψης συνεχών δεδομένων για τις καιρικές συνθήκες σε πραγματικό χρόνο έχει βελτιωθεί σημαντικά (NASA, 2020). Οι σύγχρονοι μετεωρολογικοί δορυφόροι, όπως οι σειρές GOES και Meteosat, είναι εξοπλισμένοι με προηγμένους αισθητήρες που μετρούν με ακρίβεια από τη θερμοκρασία επιφανειών έως ατμοσφαιρικά χημικά στοιχεία, παρέχοντας έτσι εξαιρετική κάλυψη και πολυτιμότες πληροφορίες για τον καιρό και το κλίμα (EUMETSAT, 2020).



Εικόνα 8: Η εικόνα απεικονίζει μια σχηματική αναπαράσταση του δορυφόρου TIROS (Television Infrared Observation Satellite), που περιλαμβάνει βασικά εξαρτήματα και λειτουργίες του. Τα κύρια μέρη του δορυφόρου περιλαμβάνουν τις

ηλιακές κυψέλες (solar cells) για την παραγωγή ενέργειας, κεραίες λήψης και μετάδοσης, ηλεκτρονικά συστήματα υπέρυθρων (IR electronics), και διάφορες κάμερες για την καταγραφή τηλεοπτικών εικόνων σε στενή και ευρεία γωνία. Επιπλέον, διαθέτει πομπούς τηλεοπτικών σημάτων, ραδιοφάρους και συστήματα αποθήκευσης ταινίας για την αποθήκευση δεδομένων. Ο TIROS αποτέλεσε τον πρώτο επιτυχημένο μετεωρολογικό δορυφόρο, εκτοξευμένος το 1960, που έδωσε τις πρώτες δορυφορικές εικόνες του καιρού. (Πηγή: DARPA, 2021)

Οι γεωστατικοί δορυφόροι, τοποθετημένοι σε υψηλή Γεωσύγχρονη Τροχιά, παραμένουν ακίνητοι σχετικά με ένα σημείο της γήινης επιφάνειας, παρέχοντας συνεχή εποπτεία μεγάλων περιοχών. Αυτή η ιδιότητα τους καθιστά ιδανικούς για την παρακολούθηση τυφώνων, θερμοκρασιακών ανωμαλιών στους ωκεανούς και άλλων φαινομένων που απαιτούν συνεχή παρακολούθηση σε μεγάλες κλίμακες (NOAA, 2024). Από την άλλη πλευρά, οι πολικοί δορυφόροι ακολουθούν μια χαμηλότερη και περισσότερο ελλειπτική τροχιά, διασχίζοντας τους πόλους της Γης και καλύπτοντας πλήρως κάθε γεωγραφικό πλάτος με την περιστροφή της Γης. Αυτό τους επιτρέπει να συλλέγουν δεδομένα από κάθε περιοχή του πλανήτη, αυξάνοντας την ολοκληρωμένη εικόνα της ατμόσφαιρας και της επιφάνειας της Γης (NOAA, 2018).

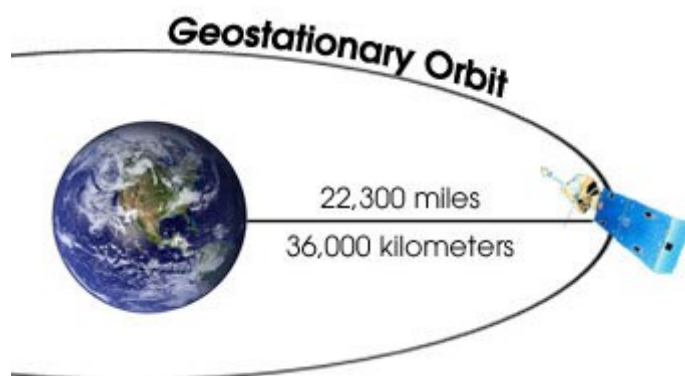


Εικόνα 9: Η εικόνα απεικονίζει έναν γεωστατικό δορυφόρο με τις διάφορες συσκευές και αισθητήρες του. (Πηγή: Wikipedia, 2020)

Οι δορυφόροι αυτοί φέρουν ποικιλία αισθητήρων που προσφέρουν διαφορετικές προοπτικές και δεδομένα απαραίτητα για την μετεωρολογική επιστήμη. Αισθητήρες όπως το Visible Infrared Imaging Radiometer Suite (VIIRS) και το Advanced Baseline Imager (ABI) παρέχουν υψηλής ανάλυσης οπτικές και υπέρυθρες εικόνες, που είναι κρίσιμες για την παρακολούθηση και ανάλυση καιρικών φαινομένων, όπως σύννεφα, καταιγίδες και πυρκαγιές (EUMETSAT, 2020). Αυτές οι τεχνολογίες και τα δεδομένα που παρέχουν, σε συνδυασμό με προηγμένες μεθόδους ανάλυσης και μοντελοποίησης, έχουν μεταμορφώσει την μετεωρολογία σε μια πιο ακριβή και δυναμική επιστήμη.

Γεωστατικοί Δορυφόροι

Οι γεωστατικοί δορυφόροι, επίσης γνωστοί ως δορυφόροι γεωσύγχρονης τροχιάς, αποτελούν κρίσιμο εργαλείο στη μετεωρολογία και τις τηλεπικοινωνίες. Αυτοί οι δορυφόροι τοποθετούνται σε υψηλή γεωσύγχρονη τροχιά περίπου 35,786 χιλιόμετρα πάνω από τον ισημερινό και περιστρέφονται με την ίδια γωνιακή ταχύτητα με τη Γη. Αυτό σημαίνει ότι παραμένουν σταθεροί σε σχέση με ένα συγκεκριμένο σημείο της γήινης επιφάνειας, προσφέροντας συνεχή κάλυψη της ίδιας περιοχής (Kidder & Vonder Haar, 1995).



Εικόνα 10: Η εικόνα παρουσιάζει την καθιερωμένη γεωσύγχρονη τροχιά που ακολουθούν οι γεωστατικοί δορυφόροι (Πηγή: SatMet, 2012)

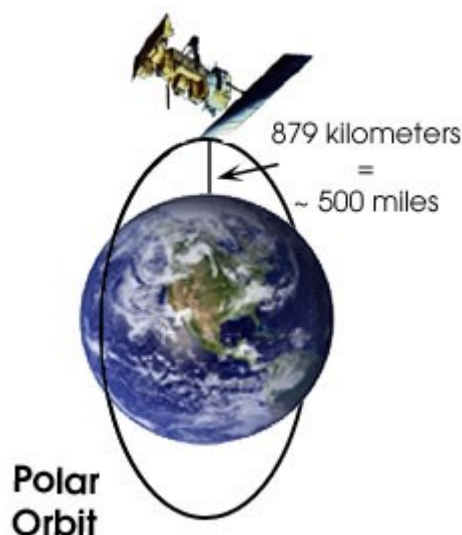
Η ικανότητα των γεωστατικών δορυφόρων να παρακολουθούν διαρκώς την ίδια περιοχή τους καθιστά ιδανικούς για την παρακολούθηση μετεωρολογικών συνθηκών, την εκτέλεση τηλεπικοινωνιακών λειτουργιών και την εκπομπή τηλεοπτικών σημάτων. Αυτοί οι δορυφόροι φέρουν πολύπλοκους αισθητήρες που συλλέγουν δεδομένα για την ατμόσφαιρα, τη θερμοκρασία, την υγρασία, τον άνεμο και άλλα μετεωρολογικά χαρακτηριστικά, παρέχοντας ανεκτίμητες πληροφορίες για την κατανόηση και πρόβλεψη του καιρού (NOAA, 2024).

Επιπλέον, η γεωστατική τροχιά είναι ιδιαίτερα σημαντική για την παρακολούθηση των τροπικών κυκλώνων, των καταιγίδων και άλλων σημαντικών καιρικών φαινομένων σε μεγάλες περιοχές. Αυτό παρέχει τη δυνατότητα στους επιστήμονες και τους μετεωρολόγους να έχουν συνεχή επίβλεψη και να προειδοποιούν το κοινό για επικείμενους κινδύνους, βελτιώνοντας έτσι τη δημόσια ασφάλεια (Wikipedia, 2020).

Η τεχνολογία πίσω από τους γεωστατικούς δορυφόρους συνεχίζει να εξελίσσεται, με την προσθήκη πιο προηγμένων αισθητήρων και την ανάπτυξη νέων τεχνολογικών δυνατοτήτων που επιτρέπουν ακόμη πιο λεπτομερείς και ακριβείς παρατηρήσεις. Αυτή η εξέλιξη αναμένεται να συνεχιστεί και να παρέχει σημαντικές βελτιώσεις στις μετεωρολογικές υπηρεσίες και στην έρευνα του κλίματος (EUMETSAT, 2020).

Πολικοί Δορυφόροι

Οι πολικοί δορυφόροι, αποτελούν μια άλλη κρίσιμη κατηγορία δορυφόρων που χρησιμοποιούνται στην επιστήμη της μετεωρολογίας. Σε αντίθεση με τους γεωστατικούς, αυτοί περιστρέφονται γύρω από τη Γη ακολουθώντας μια σχεδόν πολική τροχιά, καλύπτοντας την επιφάνεια του πλανήτη πλήρως καθώς αυτή περιστρέφεται κάτω από αυτούς. Η τροχιά αυτή τους επιτρέπει να συλλέγουν δεδομένα από κάθε γεωγραφικό πλάτος, προσφέροντας μια ολοκληρωμένη εικόνα των ατμοσφαιρικών διεργασιών σε παγκόσμια κλίμακα (Kidder & Vonder Haar, 1995).



Εικόνα 11: Η εικόνα παρουσιάζει την καθιερωμένη γεωσύγχρονη τροχιά που ακολουθούν οι πολικοί δορυφόροι (Πηγή: SatMet, 2012b)

Η δυνατότητα των πολικών δορυφόρων να συλλέγουν δεδομένα από τους πόλους έχει αναδειχθεί ως ιδιαίτερα σημαντική, καθώς οι πολικές περιοχές είναι κρίσιμες για την κατανόηση των κλιματικών αλλαγών και των παγκόσμιων μετεωρολογικών μοτίβων. Πέρα από τις κλιματικές μελέτες, οι πολικοί δορυφόροι χρησιμοποιούνται για την παρακολούθηση άλλων φαινομένων όπως το λιώσιμο των πάγων, η θαλάσσια ρύπανση και οι αλλαγές στη βιομάζα.

Οι σύγχρονοι πολικοί δορυφόροι είναι εξοπλισμένοι με πολυπλοκότητα αισθητήρων και τεχνολογικών δυνατοτήτων. Για παράδειγμα, οι αισθητήρες που εντοπίζουν τα επίπεδα του θαλάσσιου πάγου, η θερμοκρασία της επιφάνειας των ωκεανών και οι ατμοσφαιρικές διεργασίες είναι απαραίτητοι για την επιστημονική κοινότητα για την κατανόηση και πρόβλεψη του κλίματος (NASA, 2021).

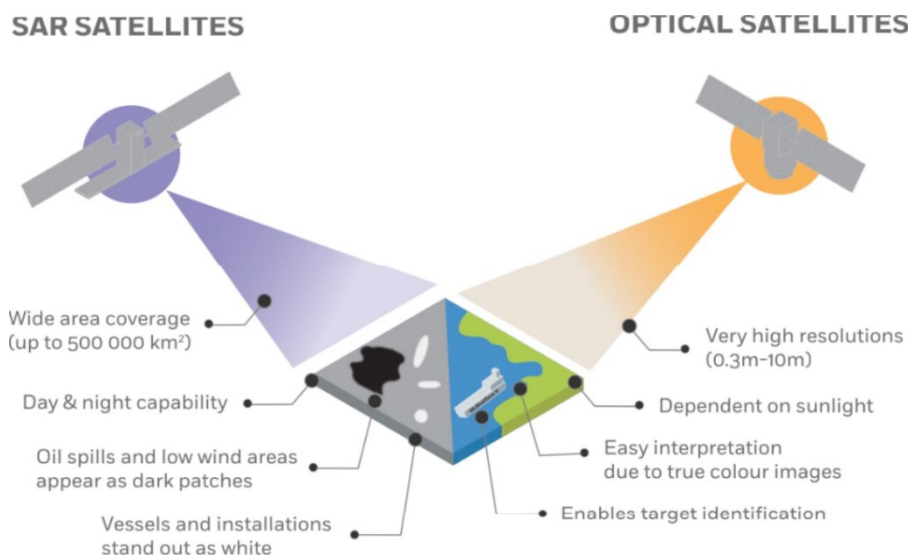
Η εξέλιξη της τεχνολογίας των πολικών δορυφόρων συνεχίζεται, με νέα έργα και αποστολές να σχεδιάζονται για τη βελτίωση της λεπτομέρειας και της συχνότητας της δεδομένων που συλλέγονται. Αυτό θα παρέχει περαιτέρω ευκαιρίες για πιο ακριβείς και έγκαιρες προβλέψεις καιρικών και κλιματικών φαινομένων, ενισχύοντας την παγκόσμια ανταπόκριση σε περιβαλλοντικές αλλαγές και απειλές (EUMETSAT, 2020).

Οι αισθητήρες στους δορυφόρους αποτελούν τα κύρια εργαλεία για τη συλλογή δεδομένων και την παρατήρηση της Γης, παίζοντας θεμελιώδη ρόλο στην ατμοσφαιρική επιστήμη, την ωκεανογραφία, και την κλιματολογία. Με τη χρήση διαφόρων τύπων αισθητήρων, οι δορυφόροι μπορούν να καταγράψουν μια ευρεία γκάμα από φυσικές παραμέτρους που περιλαμβάνουν τη θερμοκρασία, την υγρασία, τα αέρια της ατμόσφαιρας, και τις αλλαγές στην επιφάνεια της Γης (Kidder & Vonder Haar, 1995).



Εικόνα 12: Η εικόνα παρουσιάζει δύο τύπους δορυφορικών αισθητήρων: παθητικούς και ενεργητικούς αισθητήρες. Οι παθητικοί αισθητήρες (αριστερά) ανιχνεύουν και μετρούν τη φυσική ακτινοβολία που εκπέμπεται ή αντανακλάται από αντικείμενα στη Γη, χρησιμοποιώντας πηγές φωτός όπως ο ήλιος. Αντίθετα, οι ενεργητικοί αισθητήρες (δεξιά) εκπέμπουν δικό τους σήμα προς τη Γη και μετρούν την ανακλώμενη ακτινοβολία για να συλλέξουν δεδομένα, επιτρέποντας τη λήψη μετρήσεων ανεξάρτητα από το φως του ήλιου. (Πηγή: NASA, 2023)

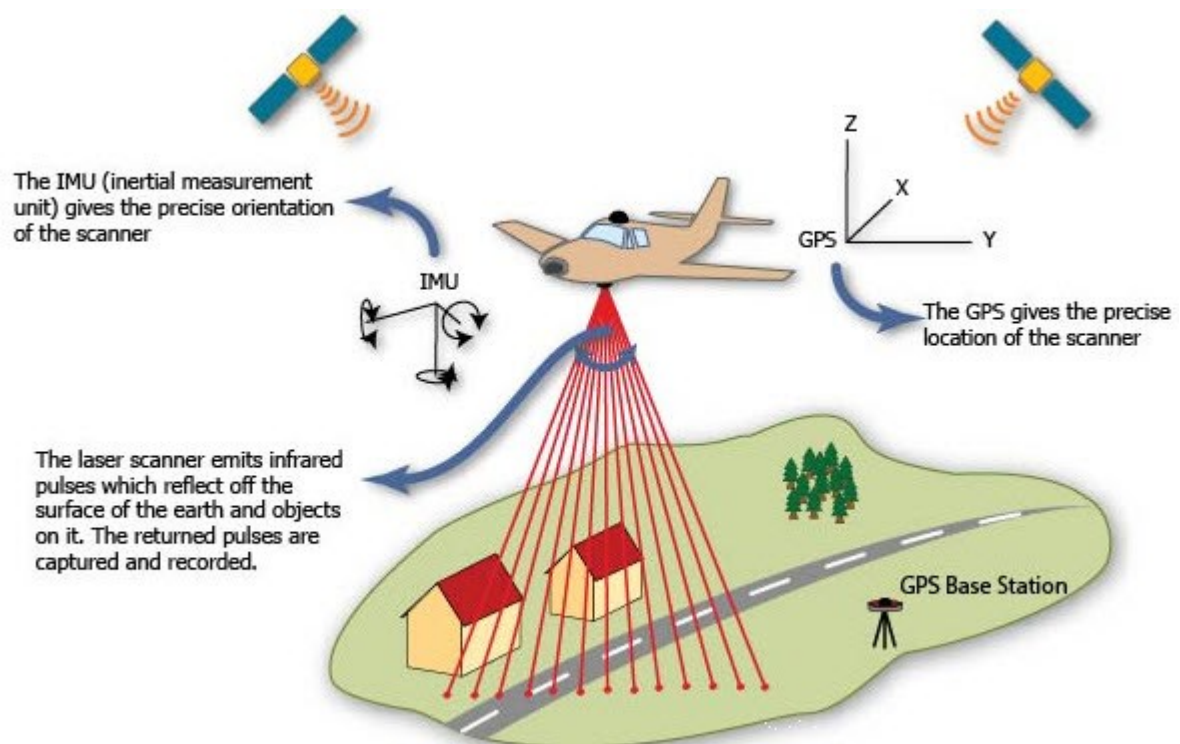
Οι οπτικοί και υπέρυθροι αισθητήρες είναι ίσως οι πιο διαδεδομένοι τύποι αισθητήρων σε μετεωρολογικούς δορυφόρους. Οι οπτικοί αισθητήρες καταγράφουν εικόνες στο ορατό φάσμα, ενώ οι υπέρυθροι αισθητήρες μετρούν τις θερμικές ακτινοβολίες που εκπέμπονται από την Γη, προσφέροντας έτσι σημαντικές πληροφορίες για τη θερμοκρασία και τις καιρικές συνθήκες ακόμη και κατά τη διάρκεια της νύχτας ή σε συνθήκες συννεφιάς (Wikipedia, 2020; Brown et al., 2005).



Εικόνα 13: Η εικόνα παρουσιάζει τη σύγκριση μεταξύ δορυφόρων SAR (Synthetic Aperture Radar) και οπτικών δορυφόρων. Οι δορυφόροι SAR καλύπτουν ευρείες περιοχές (έως 500,000 km²) και μπορούν να λειτουργούν τόσο κατά τη διάρκεια της ημέρας όσο και της νύχτας. Είναι ικανοί να εντοπίζουν πετρελαιοκηλίδες και περιοχές χαμηλής ταχύτητας ανέμου ως σκοτεινές περιοχές, ενώ τα πλοία και οι εγκαταστάσεις εμφανίζονται ως λευκές. Από την άλλη, οι οπτικοί δορυφόροι προσφέρουν πολύ υψηλές αναλύσεις (0,3μ-10μ) και εξαρτώνται από το φως του ήλιου. Οι εικόνες από οπτικούς δορυφόρους είναι εύκολες στην ερμηνεία λόγω των αληθινών χρωμάτων, διευκολύνοντας την αναγνώριση στόχων. (Πηγή: Cervest, 2020)

Μικροκυματικοί αισθητήρες, άλλος ένας τύπος αισθητήρα, παίζουν κρίσιμο ρόλο στην κατανόηση των ατμοσφαιρικών διεργασιών και τη μέτρηση των ωκεάνιων χαρακτηριστικών, όπως η θαλάσσια θερμοκρασία και η υγρασία. Η ικανότητά τους να διαπερνούν τα νέφη τους καθιστά ιδανικούς για την καταγραφή δεδομένων κάτω από σχεδόν κάθε καιρική συνθήκη (Wikipedia, 2020).

Τέλος, τα συστήματα LiDAR (Light Detection and Ranging) χρησιμοποιούνται για την ακριβή μέτρηση των αποστάσεων μεταξύ του δορυφόρου και της επιφάνειας της Γης ή των νεφών, χρησιμοποιώντας την ακτινοβολία λέιζερ. Το LiDAR προσφέρει μεγάλη ακρίβεια στη μέτρηση του υψομέτρου και της δομής των νεφών, καθώς και στην παρακολούθηση των αερολυμάτων και των εκπομπών αερίων στην ατμόσφαιρα (Brown et al., 2005).



Εικόνα 14: Η εικόνα απεικονίζει τη διαδικασία χρήσης συστημάτων LiDAR (Light Detection and Ranging) για χαρτογράφηση της επιφάνειας της Γης. (Πηγή: Elprocus, 2017)

Αισθητήρες IoT στην Πρόβλεψη Καιρού

Το Internet of Things (IoT), ή το Διαδίκτυο των Πραγμάτων, αναφέρεται στη δικτύωση φυσικών αντικειμένων που ενσωματώνουν αισθητήρες, λογισμικό και άλλες τεχνολογίες με σκοπό την ανταλλαγή δεδομένων με άλλες συσκευές και συστήματα μέσω του διαδικτύου. Αυτή η τεχνολογία έχει επαναστατικοποιήσει πολλούς τομείς, από την οικιακή αυτοματοποίηση έως τη βιομηχανική παραγωγή, και ειδικά τη μετεωρολογία, όπου οι αισθητήρες IoT παρέχουν ακριβή και συνεχή παρατήρηση των καιρικών φαινομένων (Borgia, 2014).

Στον τομέα της μετεωρολογίας, οι αισθητήρες IoT χρησιμοποιούνται για την καταγραφή μετρήσεων όπως η θερμοκρασία, η υγρασία, η πίεση της ατμόσφαιρας και άλλες σημαντικές μετεωρολογικές παράμετροι. Αυτοί οι αισθητήρες μπορούν να εγκατασταθούν σε διάφορα μέρη, από αστικές περιοχές μέχρι απομονωμένες τοποθεσίες, και να στείλουν τα δεδομένα σε κεντρικά συστήματα για περαιτέρω ανάλυση, βελτιώνοντας την ακρίβεια των καιρικών προγνώσεων (Perera et al., 2015).

Η ανάπτυξη και η υιοθέτηση των IoT μετεωρολογικών αισθητήρων αντιπροσωπεύει μια μεγάλη βελτίωση στην πρόγνωση και τη διαχείριση καιρικών κινδύνων. Η δυνατότητα αυτών των συστημάτων να παρέχουν συνεχή και αυτόματη καταγραφή δεδομένων ανοίγει νέους δρόμους για πιο έγκαιρη και ακριβή αντίδραση σε καιρικά φαινόμενα, από θυελλώδεις ανέμους μέχρι ακραία κλιματικά φαινόμενα, συμβάλλοντας στην προστασία των ζώων και των περιουσιών (Al-Fuqaha et al., 2015).

Αναφορικά στο τρόπο με τον οποίο ενσωματώνονται τα συστήματα αισθητήρων IoT, παρατηρείται ένα πολυεπίπεδο οικοσύστημα ενσωμάτωσης. Στη βάση του IoT βρίσκονται οι αισθητήρες και οι συσκευές ενδοεπικοινωνίας, οι οποίοι συλλέγουν δεδομένα από το περιβάλλον τους. Αυτά τα δεδομένα μεταδίδονται μέσω δικτύων που υποστηρίζουν την ταχύτητα και την ευρυζωνικότητα που απαιτούνται για την ομαλή λειτουργία του IoT, όπως τα δίκτυα 4G LTE και τώρα το 5G, τα οποία προσφέρουν υψηλότερη ταχύτητα και χαμηλότερη καθυστέρηση στη μετάδοση δεδομένων (Al-Fuqaha et al., 2015).

Η επεξεργασία και η ανάλυση των δεδομένων που συλλέγονται από IoT συσκευές γίνεται συχνά σε cloud-based πλατφόρμες. Αυτές οι πλατφόρμες παρέχουν τους απαραίτητους υπολογιστικούς πόρους και λογισμικά για την αποθήκευση, επεξεργασία και ανάλυση μεγάλων όγκων δεδομένων σε πραγματικό χρόνο. Οι πλατφόρμες αυτές επιτρέπουν τη δυνατότητα ευέλικτης διαχείρισης των συσκευών και των δεδομένων, και υποστηρίζουν την ταχεία ανάπτυξη και ενσωμάτωση νέων εφαρμογών (Perera et al., 2014).

Στο πλαίσιο του IoT, η ασφάλεια των δεδομένων και των δικτύων είναι παράγοντας κρίσιμης σημασίας. Η εγγενής διασυνδεσιμότητα των IoT συσκευών αυξάνει την ευπάθεια σε επιθέσεις και απαιτεί τη χρήση προηγμένων μηχανισμών ασφαλείας. Τεχνολογίες όπως η κρυπτογράφηση δεδομένων, οι τεχνικές αυθεντικοποίησης και η συνεχής παρακολούθηση των δικτύων είναι απαραίτητες για τη διασφάλιση της ακεραιότητας και της ιδιωτικότητας των δεδομένων (Ziegelendorf et al., 2013).

Ένας άλλος σημαντικός τομέας είναι η διαλειτουργικότητα και οι πρότυπες προδιαγραφές που επιτρέπουν στις IoT συσκευές από διαφορετικούς κατασκευαστές να επικοινωνούν μεταξύ τους. Η

ανάπτυξη ανοιχτών προτύπων και πρωτοκόλλων είναι κρίσιμη για τη δημιουργία ενός ενιαίου, αποδοτικού και βιώσιμου IoT περιβάλλοντος. (Bassi et al., 2008).

Τέλος, η εξέλιξη του IoT συνδέεται άμεσα με την πρόοδο στην τεχνητή νοημοσύνη και τη μηχανική μάθηση, καθώς αυτές οι τεχνολογίες επιτρέπουν την πιο αποτελεσματική επεξεργασία και ανάλυση των τεράστιων όγκων δεδομένων που παράγονται από τις IoT συσκευές. Η ικανότητα να μαθαίνει από τα δεδομένα και να προσαρμόζει αυτόματα τις λειτουργίες, χωρίς ανθρώπινη παρέμβαση, οδηγεί σε μεγαλύτερη αποδοτικότητα, ακρίβεια και, τελικά, στην αυτοματοποίηση πολλών διαδικασιών (Atzori et al., 2010).

Σχετικά, με την χρήση τους στην μετεωρολογία η χρήση των IoT μετεωρολογικών αισθητήρων είναι ευρεία και ποικιλόμορφη. Σε αστικές περιοχές, αυτοί οι αισθητήρες μπορούν να εγκατασταθούν σε κτίρια, φωτιστικά στύλους, και άλλες υποδομές, παρέχοντας λεπτομερή δεδομένα για το μικρόκλιμα και βοηθώντας στη διαχείριση της ατμοσφαιρικής ρύπανσης. Σε αγροτικές περιοχές, οι αισθητήρες αυτοί μπορούν να παρακολουθούν τις κλιματικές συνθήκες που επηρεάζουν τις καλλιέργειες, βοηθώντας τους γεωργούς να λαμβάνουν πιο ενημερωμένες αποφάσεις για την άρδευση και τη διαχείριση των καλλιεργειών (Whitmore et al., 2014).

Η δυνατότητα συλλογής και ανάλυσης μεγάλων όγκων δεδομένων από διαφορετικές γεωγραφικές περιοχές επιτρέπει στους επιστήμονες να αναπτύξουν πιο ακριβή μοντέλα καιρού και να παρακολουθούν τις αλλαγές στο κλίμα σε πραγματικό χρόνο. Οι αισθητήρες IoT είναι συχνά εξοπλισμένοι με τεχνολογίες όπως GPS για την ακριβή γεωγραφική τοποθέτηση και επικοινωνούν με κεντρικές πλατφόρμες μέσω δικτύων όπως το LoRaWAN και το NB-IoT, τα οποία είναι σχεδιασμένα για χαμηλή κατανάλωση ενέργειας και μεγάλη εμβέλεια (Borgia, 2014).

Ένα από τα μεγαλύτερα πλεονεκτήματα της χρήσης IoT μετεωρολογικών αισθητήρων είναι η δυνατότητα για απομακρυσμένη παρακολούθηση και συντήρηση. Αυτό σημαίνει ότι οι αισθητήρες μπορούν να λειτουργούν σε απομακρυσμένες ή δύσκολες τοποθεσίες χωρίς την ανάγκη για συχνή ανθρώπινη παρέμβαση. Επίσης, η αυτοματοποίηση των διαδικασιών συλλογής και ανάλυσης δεδομένων μειώνει τα λάθη και αυξάνει την αποτελεσματικότητα (Atzori et al., 2010).

Παρά τις προκλήσεις που αντιμετωπίζουν, όπως η διασφάλιση της ασφάλειας και της ιδιωτικότητας των δεδομένων, η ανάπτυξη των IoT συστημάτων με μετεωρολογικούς αισθητήρες συνεχίζει να εξελίσσεται. Οι τεχνολογίες αυτές προσφέρουν σημαντικά οφέλη, επιτρέποντας την έγκαιρη αντίδραση σε καιρικά φαινόμενα και τη βελτίωση της ποιότητας ζωής των πολιτών, ενώ συμβάλλουν στην κατανόηση και την αντιμετώπιση των κλιματικών αλλαγών (Gubbi et al., 2013).

ΚΕΦΑΛΑΙΟ 4 Μοντέλα Μηχανικής Μάθησης

Εισαγωγή στη Μηχανική Μάθηση και την Εξόρυξη Δεδομένων

Ορισμός και Βασικές Έννοιες

Η μηχανική μάθηση και η εξόρυξη δεδομένων είναι δύο κρίσιμα πεδία της επιστήμης των δεδομένων που έχουν αλλάξει δραστικά τον τρόπο με τον οποίο αντιμετωπίζουμε και αναλύουμε μεγάλους όγκους δεδομένων. Αυτές οι τεχνολογίες παρέχουν ισχυρά εργαλεία για την αναγνώριση προτύπων, την πρόβλεψη μελλοντικών τάσεων και τη λήψη αποφάσεων βασισμένων σε δεδομένα.

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να "μαθαίνουν" από δεδομένα. Αντί να ακολουθούν προγραμματισμένες οδηγίες, τα συστήματα μηχανικής μάθησης βελτιώνουν την απόδοσή τους με την πάροδο του χρόνου καθώς εκτίθενται σε περισσότερα δεδομένα. Τα βασικά στοιχεία της μηχανικής μάθησης περιλαμβάνουν την εκπαίδευση μοντέλων, την αξιολόγηση της απόδοσής τους και τη βελτίωση των αλγορίθμων μέσω της επανατροφοδότησης (Goodfellow, Bengio, & Courville, 2016).

Η εξόρυξη δεδομένων, από την άλλη πλευρά, είναι η διαδικασία ανακάλυψης προτύπων και σχέσεων σε μεγάλα σύνολα δεδομένων μέσω τεχνικών όπως η ανάλυση, η στατιστική, και οι αλγόριθμοι μηχανικής μάθησης. Η εξόρυξη δεδομένων περιλαμβάνει διάφορα στάδια, όπως η συλλογή και η προετοιμασία των δεδομένων, η ανάλυση και η εξαγωγή γνώσεων, και η παρουσίαση των αποτελεσμάτων. Οι τεχνικές εξόρυξης δεδομένων χρησιμοποιούνται σε πολλούς τομείς, από την επιχειρηματική ανάλυση και τη βιοϊατρική έρευνα μέχρι την ανάλυση κοινωνικών δικτύων και την πρόβλεψη χρηματιστηριακών τιμών (Han, Pei, & Kamber, 2011).

Οι εφαρμογές της μηχανικής μάθησης και της εξόρυξης δεδομένων είναι πολυάριθμες και ποικίλες. Για παράδειγμα, στην ιατρική, χρησιμοποιούνται για την ανάλυση ιατρικών εικόνων και την πρόβλεψη της έκβασης των ασθενειών. Στην οικονομία, χρησιμοποιούνται για την πρόβλεψη των χρηματιστηριακών αγορών και την ανίχνευση απάτης. Στην καθημερινή μας ζωή, αλγόριθμοι μηχανικής μάθησης βρίσκονται πίσω από τις προτάσεις προϊόντων στις διαδικτυακές αγορές και τις προτάσεις περιεχομένου στις πλατφόρμες κοινωνικών δικτύων (Murphy, 2012).

Ένα από τα πιο σημαντικά χαρακτηριστικά της μηχανικής μάθησης είναι η ικανότητά της να προσαρμόζεται και να βελτιώνεται με την πάροδο του χρόνου. Μέσω της συνεχούς μάθησης από νέα δεδομένα, τα συστήματα μηχανικής μάθησης μπορούν να αναγνωρίζουν νέα πρότυπα και να βελτιώνουν τις προβλέψεις τους. Αυτή η προσαρμοστικότητα είναι ιδιαίτερα χρήσιμη σε περιβάλλοντα όπου τα δεδομένα αλλάζουν συνεχώς, όπως η ανάλυση αγοράς και η πρόβλεψη καιρού (Bishop, 2006).

Η εξόρυξη δεδομένων και η μηχανική μάθηση συχνά λειτουργούν συνεργατικά. Η εξόρυξη δεδομένων παρέχει τα δεδομένα και τις αρχικές αναλύσεις που χρησιμοποιούνται για την εκπαίδευση των μοντέλων μηχανικής μάθησης. Από την άλλη πλευρά, οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται για την ανάπτυξη πιο σύνθετων και ακριβών μοντέλων που μπορούν να προβλέψουν μελλοντικά γεγονότα και τάσεις με βάση τα δεδομένα που έχουν συλλεχθεί και αναλυθεί (Witten, Frank, Hall, & Pal, 2016).

Συνοψίζοντας, η μηχανική μάθηση και η εξόρυξη δεδομένων είναι αναπόσπαστα στοιχεία της σύγχρονης επιστήμης των δεδομένων, προσφέροντας ισχυρά εργαλεία για την ανάλυση και την αξιοποίηση μεγάλων όγκων δεδομένων. Αυτές οι τεχνολογίες συνεχίζουν να εξελίσσονται και να ενσωματώνονται σε όλο και περισσότερους τομείς της επιστήμης και της καθημερινής ζωής, προσφέροντας νέες δυνατότητες και προοπτικές για την ανάλυση και την κατανόηση του κόσμου γύρω μας.

Ιστορική Εξέλιξη της Μηχανικής Μάθησης

Η ιστορική εξέλιξη της μηχανικής μάθησης είναι μια πορεία γεμάτη καινοτομίες και προόδους που έχουν διαμορφώσει την επιστήμη των υπολογιστών και την τεχνητή νοημοσύνη. Η ιστορία αυτή ξεκινά από τα πρώτα βήματα της πληροφορικής και επεκτείνεται μέχρι τις σύγχρονες εφαρμογές της σε διάφορους τομείς.

Η αρχή της μηχανικής μάθησης τοποθετείται στα τέλη της δεκαετίας του 1950 και τις αρχές της δεκαετίας του 1960. Το 1957, ο Frank Rosenblatt ανέπτυξε το perceptron, το πρώτο νευρωνικό δίκτυο που μπορούσε να μαθαίνει και να αναγνωρίζει μοτίβα. Το perceptron ήταν βασισμένο σε βιολογικούς νευρώνες και αποτελούσε μια απλή μορφή του τι σήμερα ονομάζουμε βαθιά μάθηση (Rosenblatt, 1958).

Στη δεκαετία του 1960, η έρευνα στη μηχανική μάθηση προχώρησε με την ανάπτυξη των αλγορίθμων αναζήτησης και των δενδροειδών δομών. Οι αλγόριθμοι αυτοί χρησιμοποιήθηκαν για την επίλυση προβλημάτων όπως η τακτική στο σκάκι και η βελτιστοποίηση (Nilsson, 1965). Ωστόσο, η πρόοδος ήταν αργή λόγω των περιορισμένων υπολογιστικών πόρων και της έλλειψης δεδομένων.

Η δεκαετία του 1970 έφερε την ανάπτυξη των πρώτων στατιστικών μεθόδων στη μηχανική μάθηση. Οι ερευνητές άρχισαν να ενσωματώνουν στατιστικά μοντέλα και πιθανοθεωρητικές μεθόδους για την ανάλυση των δεδομένων και την κατασκευή μοντέλων πρόβλεψης. Ένα σημαντικό επίτευγμα αυτής της περιόδου ήταν η ανάπτυξη των κρυφών μοντέλων Markov (HMMs), τα οποία χρησιμοποιήθηκαν εκτενώς στην αναγνώριση φωνής και την επεξεργασία φυσικής γλώσσας (Rabiner, 1989).

Στη δεκαετία του 1980, η έρευνα στη μηχανική μάθηση επικεντρώθηκε στην ανάπτυξη των νευρωνικών δικτύων. Η ανακάλυψη του αλγορίθμου της αναδρομικής διάδοσης σφάλματος

(backpropagation) από τους Rumelhart, Hinton και Williams το 1986 αποτέλεσε σταθμό στην ιστορία της μηχανικής μάθησης. Ο αλγόριθμος αυτός επέτρεψε την εκπαίδευση πολυεπίπεδων νευρωνικών δικτύων και βελτίωσε σημαντικά την ικανότητα των μοντέλων να μαθαίνουν από δεδομένα (Rumelhart, Hinton, & Williams, 1986).

Η δεκαετία του 1990 χαρακτηρίστηκε από την ανάπτυξη νέων αλγορίθμων και την εφαρμογή τους σε ποικίλους τομείς. Η ανάπτυξη των αλγορίθμων υποστήριξης διανυσμάτων (SVMs) από τους Vapnik και Cortes αποτέλεσε σημαντική πρόοδο στην ταξινόμηση και την παλινδρόμηση (Cortes & Vapnik, 1995). Επίσης, η εισαγωγή των αλγορίθμων ενίσχυσης (boosting), όπως το AdaBoost, βελτίωσε σημαντικά την απόδοση των μοντέλων πρόβλεψης (Freund & Schapire, 1997).

Στις αρχές της δεκαετίας του 2000, η ανάπτυξη της μηχανικής μάθησης συνέχισε με την εξέλιξη των τεχνικών βαθιάς μάθησης (deep learning). Οι τεχνικές αυτές, βασισμένες σε πολυεπίπεδα νευρωνικά δίκτυα, επέτρεψαν την ανάλυση και την επεξεργασία μεγάλων και σύνθετων συνόλων δεδομένων. Η πρόοδος στην υπολογιστική ισχύ και η διαθεσιμότητα μεγάλων συνόλων δεδομένων συνέβαλαν στην ανάπτυξη των τεχνικών βαθιάς μάθησης (LeCun, Bengio, & Hinton, 2015).

Το 2010 και μετά, η μηχανική μάθηση εισήλθε σε μια νέα εποχή με την ευρεία εφαρμογή της σε διάφορους τομείς. Η ανάπτυξη των αλγορίθμων βαθιάς μάθησης, όπως τα συνελκτικά νευρωνικά δίκτυα (CNNs) και τα επαναληπτικά νευρωνικά δίκτυα (RNNs), επέτρεψαν την πρόοδο στην αναγνώριση εικόνας, την επεξεργασία φυσικής γλώσσας και άλλες εφαρμογές (Krizhevsky, Sutskever, & Hinton, 2012). Επίσης, η εμφάνιση των γεννητικών ανταγωνιστικών δικτύων (GANs) άνοιξε νέους δρόμους στη δημιουργία δεδομένων και την καλλιτεχνική έκφραση (Goodfellow et al., 2014).

Σήμερα, η μηχανική μάθηση είναι πανταχού παρούσα και συνεχώς εξελίσσεται. Οι ερευνητές αναπτύσσουν νέες τεχνικές και αλγόριθμους που μπορούν να αντιμετωπίσουν προκλήσεις όπως η επεξήγηση των μοντέλων και η αντιμετώπιση της μεροληψίας στα δεδομένα. Η ανάπτυξη της κβαντικής υπολογιστικής και οι δυνατότητες της στην επεξεργασία μεγάλων συνόλων δεδομένων υπόσχεται νέες επαναστάσεις στη μηχανική μάθηση (Biamonte et al., 2017).

Συνολικά, η ιστορία της μηχανικής μάθησης είναι γεμάτη καινοτομίες και πρόοδοι που έχουν διαμορφώσει την επιστήμη και την τεχνολογία. Από τις πρώτες προσπάθειες στα τέλη της δεκαετίας του 1950 μέχρι τις σύγχρονες εφαρμογές της βαθιάς μάθησης, η μηχανική μάθηση συνεχίζει να εξελίσσεται και να επηρεάζει την καθημερινή μας ζωή.

Εφαρμογές της Μηχανικής Μάθησης στη Μετεωρολογία

Η μηχανική μάθηση έχει φέρει επανάσταση στη μετεωρολογία, παρέχοντας νέα εργαλεία και τεχνικές για την ανάλυση και την πρόβλεψη καιρικών φαινομένων. Η πρώτη εφαρμογή της μηχανικής μάθησης στη μετεωρολογία ξεκίνησε με τη χρήση απλών αλγορίθμων που μπορούσαν να αναγνωρίζουν μοτίβα σε μεγάλα σύνολα δεδομένων και να βγάζουν συμπεράσματα. Τα νευρωνικά

δίκτυα, τα δάση αποφάσεων και οι μέθοδοι υποστήριξης διανυσμάτων είναι μερικές από τις τεχνικές που χρησιμοποιούνται για την ανάλυση δεδομένων και την πρόβλεψη καιρικών συνθηκών με υψηλή ακρίβεια (Goodfellow et al., 2016; Hastie et al., 2009).

Η ανάλυση μεγάλων δεδομένων είναι ένας από τους κύριους λόγους για την επιτυχία της μηχανικής μάθησης στη μετεωρολογία. Τα σύγχρονα μετεωρολογικά δεδομένα προέρχονται από διάφορες πηγές, όπως δορυφόροι, αισθητήρες IoT, μετεωρολογικοί σταθμοί και ιστορικά αρχεία. Οι αλγόριθμοι μηχανικής μάθησης είναι ικανοί να επεξεργάζονται αυτά τα δεδομένα και να βγάζουν συμπεράσματα που δεν είναι άμεσα εμφανή. Αυτό επιτρέπει στους μετεωρολόγους να κάνουν πιο ακριβείς προβλέψεις για τις μετεωρολογικές συνθήκες (Li et al., 2017).

Ένας άλλος σημαντικός τομέας εφαρμογής της μηχανικής μάθησης είναι η πρόβλεψη ακραίων καιρικών φαινομένων. Τα μοντέλα βαθιάς μάθησης μπορούν να αναλύουν τεράστιες ποσότητες δεδομένων και να εντοπίζουν μοτίβα που προηγούνται ακραίων φαινομένων, όπως τυφώνες, πλημμύρες και καταιγίδες. Αυτές οι προβλέψεις είναι κρίσιμες για την έγκαιρη προειδοποίηση και την προετοιμασία των κοινοτήτων για να αντιμετωπίσουν τις καταστροφές (Wong, 2024).

Η μηχανική μάθηση έχει επίσης σημαντική συμβολή στις μακροπρόθεσμες προβλέψεις καιρού. Οι αλγόριθμοι μπορούν να ενσωματώνουν δεδομένα από διαφορετικές χρονικές περιόδους και να προβλέπουν τις μελλοντικές τάσεις. Αυτό είναι ιδιαίτερα χρήσιμο για την ανάλυση των επιπτώσεων της κλιματικής αλλαγής. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να αναλύσουν τα ιστορικά δεδομένα και να προβλέψουν πώς οι αλλαγές στο κλίμα θα επηρεάσουν τις μελλοντικές καιρικές συνθήκες (Reichstein et al., 2019).

Η εξατομίκευση των προβλέψεων καιρού είναι ένας άλλος τομέας όπου η μηχανική μάθηση έχει κάνει σημαντική πρόοδο. Οι αλγόριθμοι μπορούν να προσαρμόσουν τις προβλέψεις για συγκεκριμένες περιοχές ή κλάδους, όπως η γεωργία, η ναυτιλία και η ενέργεια. Οι εξατομικευμένες προβλέψεις επιτρέπουν στους επαγγελματίες να λαμβάνουν καλύτερες αποφάσεις και να βελτιώνουν την αποδοτικότητά τους (Haupt et al., 2009).

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ενσωματώσουν δεδομένα από πολλές διαφορετικές πηγές, όπως δορυφορικές εικόνες, αισθητήρες IoT, και δεδομένα από μετεωρολογικούς σταθμούς. Αυτό επιτρέπει τη δημιουργία πιο ολοκληρωμένων και ακριβών μοντέλων πρόβλεψης. Οι Li et al. (2017) έδειξαν ότι η ενσωμάτωση πολλαπλών πηγών δεδομένων βελτιώνει σημαντικά την ακρίβεια των προβλέψεων καιρού.

Η χρήση της μηχανικής μάθησης στην πρόβλεψη του καιρού έχει επίσης οδηγήσει σε σημαντικές βελτιώσεις στην απόδοση των υφιστάμενων μοντέλων πρόβλεψης. Οι αλγόριθμοι μπορούν να βελτιώσουν την ακρίβεια των προβλέψεων μέσω της συνεχούς εκπαίδευσης και της προσαρμογής στα νέα δεδομένα. Αυτό είναι ιδιαίτερα σημαντικό για τις μακροπρόθεσμες προβλέψεις, όπου οι

μικρές αλλαγές στα δεδομένα μπορούν να έχουν μεγάλες επιπτώσεις στις προβλέψεις (Rasp et al., 2020).

Τέλος, η μηχανική μάθηση επιτρέπει την ανάλυση πολύπλοκων αλληλεπιδράσεων μεταξύ των διαφορετικών μετεωρολογικών παραμέτρων. Οι αλγόριθμοι μπορούν να αναλύσουν τις σχέσεις μεταξύ της θερμοκρασίας, της υγρασίας, της πίεσης και άλλων παραμέτρων και να βγάλουν συμπεράσματα για το πώς αυτές οι παράμετροι επηρεάζουν η μία την άλλη. Αυτό επιτρέπει στους μετεωρολόγους να κάνουν πιο ακριβείς και λεπτομερείς προβλέψεις (Schneider et al., 2017).

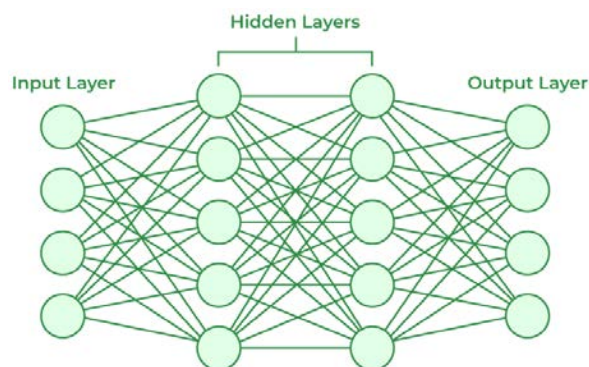
Η μηχανική μάθηση έχει φέρει σημαντικές αλλαγές στη μετεωρολογία, επιτρέποντας την ανάλυση μεγάλων δεδομένων και την πρόβλεψη ακραίων καιρικών φαινομένων με μεγαλύτερη ακρίβεια και αξιοπιστία. Η συνεχής ανάπτυξη και βελτίωση των αλγορίθμων αναμένεται να προσφέρει ακόμα πιο αξιόπιστες προβλέψεις στο μέλλον, βοηθώντας την καλύτερη κατανόηση και προετοιμασία για τις μετεωρολογικές συνθήκες.

Τεχνικές Μηχανικής Μάθησης στην Πρόβλεψη Καιρού

Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αποτελούν μια από τις πιο ισχυρές τεχνολογίες στον τομέα της τεχνητής νοημοσύνης και της μηχανικής μάθησης. Εμπνευσμένα από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου, τα νευρωνικά δίκτυα χρησιμοποιούν τεχνητούς νευρώνες για να επεξεργάζονται δεδομένα και να μαθαίνουν από αυτά. Αυτοί οι τεχνητοί νευρώνες είναι οργανωμένοι σε στρώματα, καθένα από τα οποία εκτελεί συγκεκριμένες λειτουργίες στην επεξεργασία των δεδομένων (Goodfellow, Bengio, & Courville, 2016).

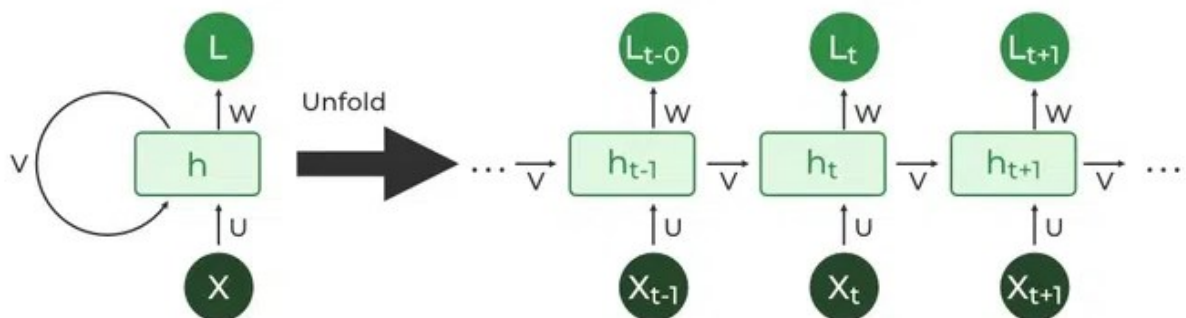
Η βασική αρχιτεκτονική ενός νευρωνικού δικτύου περιλαμβάνει τρία είδη στρωμάτων: το στρώμα εισόδου, τα κρυφά στρώματα και το στρώμα εξόδου. Το στρώμα εισόδου λαμβάνει τα αρχικά δεδομένα και τα εισάγει στο δίκτυο. Τα κρυφά στρώματα, τα οποία μπορεί να είναι πολλά, επεξεργάζονται τις πληροφορίες και εξάγουν τα απαραίτητα χαρακτηριστικά. Το στρώμα εξόδου παράγει την τελική πρόβλεψη ή την απόφαση με βάση τα επεξεργασμένα δεδομένα (LeCun, Bengio, & Hinton, 2015).



Εικόνα 15: Μια τυπική αρχιτεκτονική ενός νευρωνικού δικτύου (Πηγή: Geeksforgeeks, 2024)

Η λειτουργία των νευρωνικών δικτύων βασίζεται στην προώθηση και την αναπροσαρμογή. Κατά την προώθηση, τα δεδομένα περνούν από το στρώμα εισόδου μέσω των κρυφών στρωμάτων μέχρι να φτάσουν στο στρώμα εξόδου. Σε κάθε στρώμα, εφαρμόζονται μαθηματικές συναρτήσεις που μετασχηματίζουν τα δεδομένα. Η αναπροσαρμογή συμβαίνει μετά την παραγωγή της τελικής πρόβλεψης, όταν υπολογίζεται το σφάλμα σε σύγκριση με την πραγματική τιμή. Το σφάλμα αυτό χρησιμοποιείται για την αναπροσαρμογή των βαρών και των παραμέτρων του δικτύου μέσω της διαδικασίας της αναδρομικής διάδοσης (backpropagation) (Rumelhart, Hinton, & Williams, 1986).

Υπάρχουν διάφοροι τύποι νευρωνικών δικτύων, καθένας από τους οποίους είναι κατάλληλος για διαφορετικές εφαρμογές. Τα πολυστρωματικά προωθητικά δίκτυα (Multilayer Perceptrons - MLPs) είναι ο πιο απλός τύπος νευρωνικών δικτύων και χρησιμοποιούνται για βασικές ταξινομήσεις και παλινδρομήσεις. Τα συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs) είναι ιδιαίτερα αποτελεσματικά στην επεξεργασία εικόνας και αναγνώριση προτύπων, ενώ τα επαναληπτικά νευρωνικά δίκτυα (Recurrent Neural Networks - RNNs) είναι κατάλληλα για την επεξεργασία ακολουθιακών δεδομένων, όπως σειρές χρόνου και φυσική γλώσσα (Krizhevsky, Sutskever, & Hinton, 2012; Hochreiter & Schmidhuber, 1997).



Εικόνα 16: Διάγραμμα ενός Αναδρομικού Νευρωνικού Δικτύου (RNN), που δείχνει την αλληλεπίδραση των εσωτερικών καταστάσεων (h) με τις εισόδους (X) και τα βάρη (w , u , v) καθώς και την επέκταση (unfold) του δικτύου σε διαδοχικά χρονικά βήματα (t). (Πηγή: GeeksforGeeks, 2024)

Τα νευρωνικά δίκτυα χρησιμοποιούνται ευρέως στη μετεωρολογία για την πρόβλεψη καιρού. Η ικανότητά τους να επεξεργάζονται μεγάλα σύνολα δεδομένων και να ανιχνεύουν πολύπλοκα μοτίβα τα καθιστά ιδιαίτερα αποτελεσματικά σε αυτήν την εφαρμογή. Ένα παράδειγμα είναι η χρήση των CNNs για την ανάλυση δορυφορικών εικόνων και την πρόβλεψη καιρικών συνθηκών. Τα CNNs μπορούν να ανιχνεύσουν σύννεφα, καταιγίδες και άλλα καιρικά φαινόμενα στις δορυφορικές εικόνες, βελτιώνοντας την ακρίβεια των προβλέψεων (Shi et al., 2015).

Επιπλέον, τα RNNs έχουν χρησιμοποιηθεί για την πρόβλεψη σειρών χρόνου στη μετεωρολογία. Η ικανότητά τους να διατηρούν μνήμη προηγούμενων εισόδων τους επιτρέπει να αναγνωρίζουν μοτίβα και τάσεις στα καιρικά δεδομένα, κάνοντας πιο ακριβείς προβλέψεις για μελλοντικές καιρικές

συνθήκες. Για παράδειγμα, τα Long Short-Term Memory Networks (LSTMs), ένας τύπος RNN, έχουν χρησιμοποιηθεί για την πρόβλεψη της θερμοκρασίας και των βροχοπτώσεων (Hochreiter & Schmidhuber, 1997; Gers, Schmidhuber, & Cummins, 1999).

Η εφαρμογή των νευρωνικών δικτύων στη μετεωρολογία έχει οδηγήσει σε σημαντικές βελτιώσεις στην πρόβλεψη καιρού. Σε πολλές περιπτώσεις, τα νευρωνικά δίκτυα έχουν υπερβεί τις παραδοσιακές μεθόδους πρόβλεψης όσον αφορά την ακρίβεια και την αξιοπιστία. Η ικανότητά τους να ενσωματώνουν δεδομένα από πολλές πηγές και να αναλύουν πολύπλοκες σχέσεις μεταξύ αυτών των δεδομένων τα καθιστά απαραίτητα εργαλεία στη σύγχρονη μετεωρολογία.

Ένα παράδειγμα επιτυχούς εφαρμογής των νευρωνικών δικτύων στη μετεωρολογία είναι η χρήση τους από την Ευρωπαϊκή Υπηρεσία Διαστήματος (ESA) για την πρόβλεψη ακραίων καιρικών φαινομένων. Τα νευρωνικά δίκτυα χρησιμοποιούνται για την ανάλυση δορυφορικών δεδομένων και την ανίχνευση προτύπων που υποδηλώνουν την ανάπτυξη καταιγίδων και τυφώνων. Αυτό επιτρέπει στις αρχές να λαμβάνουν προληπτικά μέτρα και να προστατεύουν τις κοινότητες από καταστροφές (Rasp, Dueben, Scher, Weyn, & Thuerey, 2020).

Τα νευρωνικά δίκτυα έχουν επίσης χρησιμοποιηθεί για την πρόβλεψη των επιπτώσεων της κλιματικής αλλαγής. Με την ικανότητά τους να επεξεργάζονται μεγάλα σύνολα δεδομένων και να αναλύουν πολύπλοκες σχέσεις, τα νευρωνικά δίκτυα μπορούν να βοηθήσουν στην κατανόηση των μακροπρόθεσμων επιπτώσεων της κλιματικής αλλαγής και να υποστηρίξουν τη λήψη αποφάσεων για την προσαρμογή και τον μετριασμό των επιπτώσεων αυτών (Rolnick et al., 2019).

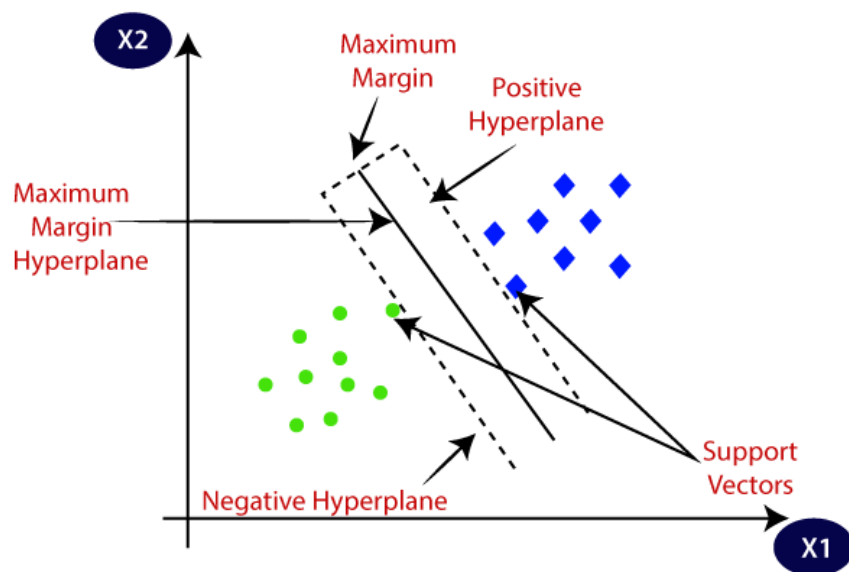
Σε γενικές γραμμές, τα νευρωνικά δίκτυα αποτελούν ένα ισχυρό εργαλείο στη μετεωρολογία, προσφέροντας νέες δυνατότητες για την πρόβλεψη και την ανάλυση καιρικών φαινομένων. Με τη συνεχή ανάπτυξη της τεχνολογίας και την αύξηση της υπολογιστικής ισχύος, αναμένεται ότι τα νευρωνικά δίκτυα θα συνεχίσουν να παίζουν κεντρικό ρόλο στη μετεωρολογία και σε πολλές άλλες επιστημονικές και τεχνολογικές εφαρμογές (Goodfellow et al., 2016; LeCun et al., 2015).

Αλγόριθμοι Υποστήριξης Διανυσμάτων (SVMs)

Οι Αλγόριθμοι Υποστήριξης Διανυσμάτων (Support Vector Machines, SVMs) αποτελούν μία από τις πιο ισχυρές και ευρέως χρησιμοποιούμενες τεχνικές στη μηχανική μάθηση για την επίλυση προβλημάτων ταξινόμησης και παλινδρόμησης. Οι SVMs παρουσιάστηκαν για πρώτη φορά από τον Vladimir Vapnik και το συνεργάτη του Alexey Chervonenkis στις αρχές της δεκαετίας του 1960, αλλά έγιναν δημοφιλείς στα μέσα της δεκαετίας του 1990 (Vapnik, 2000).

Οι SVMs λειτουργούν με την εύρεση του υπερεπίπεδου που διαχωρίζει τα δεδομένα σε δύο κατηγορίες με τον καλύτερο δυνατό τρόπο. Το υπερ-επίπεδο αυτό επιλέγεται έτσι ώστε να έχει τη μέγιστη απόσταση από τα πλησιέστερα σημεία δεδομένων από κάθε κατηγορία, τα οποία ονομάζονται διανύσματα υποστήριξης. Η απόσταση αυτή ονομάζεται περιθώριο και η διαδικασία

αναζήτησης του μέγιστου περιθωρίου αποτελεί το βασικό κριτήριο των SVMs (Cortes & Vapnik, 1995).



Εικόνα 17: Απεικόνιση ενός Υποστηρικτικού Διανύσματος Μηχανής (SVM) (Πηγή: JavaTPoint, n.d.)

Ένα από τα κύρια πλεονεκτήματα των SVMs είναι η δυνατότητά τους να χειρίζονται περιπτώσεις όπου τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα. Για να αντιμετωπίσουν αυτό το πρόβλημα, οι SVMs χρησιμοποιούν μια τεχνική που ονομάζεται πυρηνική μέθοδος (kernel method). Η πυρηνική μέθοδος επιτρέπει τον μετασχηματισμό των δεδομένων σε έναν υψηλότερης διάστασης χώρο, όπου μπορεί να βρεθεί ένα γραμμικό υπερ-επίπεδο διαχωρισμού (Schölkopf & Smola, 2002).

Υπάρχουν διάφοροι τύποι πυρήνων που μπορούν να χρησιμοποιηθούν στις SVMs, όπως οι γραμμικοί, οι πολυωνυμικοί, οι RBF (Radial Basis Function), και οι πυρήνες sigmoidal. Η επιλογή του κατάλληλου πυρήνα εξαρτάται από τη φύση των δεδομένων και το συγκεκριμένο πρόβλημα που πρέπει να λυθεί. Οι πυρήνες RBF είναι ιδιαίτερα δημοφιλείς λόγω της ικανότητάς τους να χειρίζονται δεδομένα με μη γραμμικούς συσχετισμούς (Hastie, Tibshirani, & Friedman, 2009).

Η εκπαίδευση ενός SVM περιλαμβάνει την επίλυση ενός προβλήματος βελτιστοποίησης. Ο στόχος είναι να βρεθούν οι παράμετροι του υπερ-επιπέδου που μεγιστοποιούν το περιθώριο διαχωρισμού, ενώ ταυτόχρονα ελαχιστοποιούν την ταξινόμηση σφάλματος. Αυτό μπορεί να γίνει χρησιμοποιώντας τεχνικές όπως η μέθοδος των παραμέτρων Lagrange ή η μέθοδος του τεμαχισμού (Platt, 1998).

Οι SVMs έχουν αποδειχθεί εξαιρετικά αποτελεσματικοί σε ποικίλες εφαρμογές όπως η αναγνώριση προτύπων, η ανάλυση κειμένου, η αναγνώριση εικόνας, και η βιοπληροφορική. Η ικανότητά τους να χειρίζονται υψηλής διαστάσεως δεδομένα και να αποφεύγουν την υπερ-προσαρμογή τους καθιστά ιδιαίτερα ισχυρούς σε πραγματικές εφαρμογές (Cristianini & Shawe-Taylor, 2001).

Στη μετεωρολογία, οι SVMs χρησιμοποιούνται ευρέως για την πρόβλεψη καιρού, την ανάλυση κλιματικών δεδομένων, και την ανίχνευση ακραίων καιρικών φαινομένων. Για παράδειγμα, οι SVMs έχουν εφαρμοστεί για την πρόβλεψη βροχόπτωσης, θερμοκρασίας, και ταχύτητας ανέμου με υψηλή ακρίβεια. Ένα από τα βασικά πλεονεκτήματα της χρήσης SVMs στη μετεωρολογία είναι η ικανότητά τους να διαχειρίζονται μεγάλα και πολύπλοκα σύνολα δεδομένων, καθώς και να αποδίδουν αξιόπιστες προβλέψεις ακόμα και με περιορισμένα δεδομένα εκπαίδευσης.

Σε σύγκριση με άλλους αλγόριθμους μηχανικής μάθησης, οι SVMs προσφέρουν την ικανότητα για καλή γενίκευση, μειώνοντας τον κίνδυνο υπερ-προσαρμογής στα δεδομένα εκπαίδευσης. Αυτό τους καθιστά ιδιαίτερα χρήσιμους σε περιπτώσεις όπου η ακρίβεια της πρόβλεψης είναι κρίσιμη, όπως στην πρόβλεψη καιρού (Boser et al., 1992).

Μοντέλα Ενίσχυσης Δέντρων Αποφάσεων (Boosting)

Τα μοντέλα ενίσχυσης δέντρων απόφασης (Boosting) είναι ένα ισχυρό σύνολο τεχνικών μηχανικής μάθησης που στοχεύουν στη βελτίωση της απόδοσης των αδύναμων μοντέλων βάσης. Η βασική ιδέα πίσω από το boosting είναι η αλληλουχία εκπαίδευσης πολλαπλών απλών μοντέλων απόφασης (όπως τα δέντρα απόφασης) κατά τέτοιο τρόπο ώστε κάθε νέο μοντέλο να διορθώνει τα σφάλματα των προηγούμενων (Freund & Schapire, 1997).

Τα μοντέλα boosting κατασκευάζονται με την προσθετική προσθήκη δέντρων απόφασης, το καθένα από τα οποία εξειδικεύεται στα παραδείγματα που είχαν τα μεγαλύτερα σφάλματα πρόβλεψης στα προηγούμενα βήματα. Έτσι, δημιουργείται ένα ισχυρό μοντέλο πρόβλεψης που συνδυάζει τις προβλέψεις όλων των ατομικών μοντέλων με τη χρήση των κατάλληλων βαρών (Friedman, 2001).

Ο πιο διαδεδομένος αλγόριθμος ενίσχυσης είναι το AdaBoost (Adaptive Boosting), ο οποίος προσαρμόζει το βάρος των παραδειγμάτων εκπαίδευσης ώστε τα παραδείγματα που έχουν προβλεφθεί εσφαλμένα να αποκτούν μεγαλύτερη βαρύτητα σε επόμενα μοντέλα. Άλλοι σημαντικοί αλγόριθμοι είναι το Gradient Boosting, που βελτιστοποιεί μία διαφορίσιμη απώλεια συνάρτησης, και το XGBoost (Extreme Gradient Boosting), μια εκτεταμένη και βελτιωμένη έκδοση του Gradient Boosting που παρέχει εξαιρετική απόδοση και αποτελεσματικότητα (Chen & Guestrin, 2016).

Η χρήση των μοντέλων ενίσχυσης είναι ευρεία και περιλαμβάνει εφαρμογές όπως η ταξινόμηση και η παλινδρόμηση σε διάφορους τομείς, όπως η χρηματοοικονομική ανάλυση, η ιατρική διάγνωση και η πρόβλεψη καιρού. Τα μοντέλα αυτά είναι ιδιαίτερα αποτελεσματικά στη διαχείριση πολύπλοκων συνόλων δεδομένων και στη μείωση του σφάλματος πρόβλεψης σε σχέση με τα ατομικά δέντρα απόφασης (Freund & Schapire, 1997).

Οι πρόσφατες βελτιώσεις στους αλγόριθμους ενίσχυσης, όπως το LightGBM (Light Gradient Boosting Machine) και το CatBoost, έχουν επιπλέον βελτιώσει την ταχύτητα και την ακρίβεια των

προβλέψεων, καθιστώντας το boosting ένα από τα πιο ισχυρά εργαλεία στη σύγχρονη ανάλυση δεδομένων και τη μηχανική μάθηση (Chen & Guestrin, 2016).

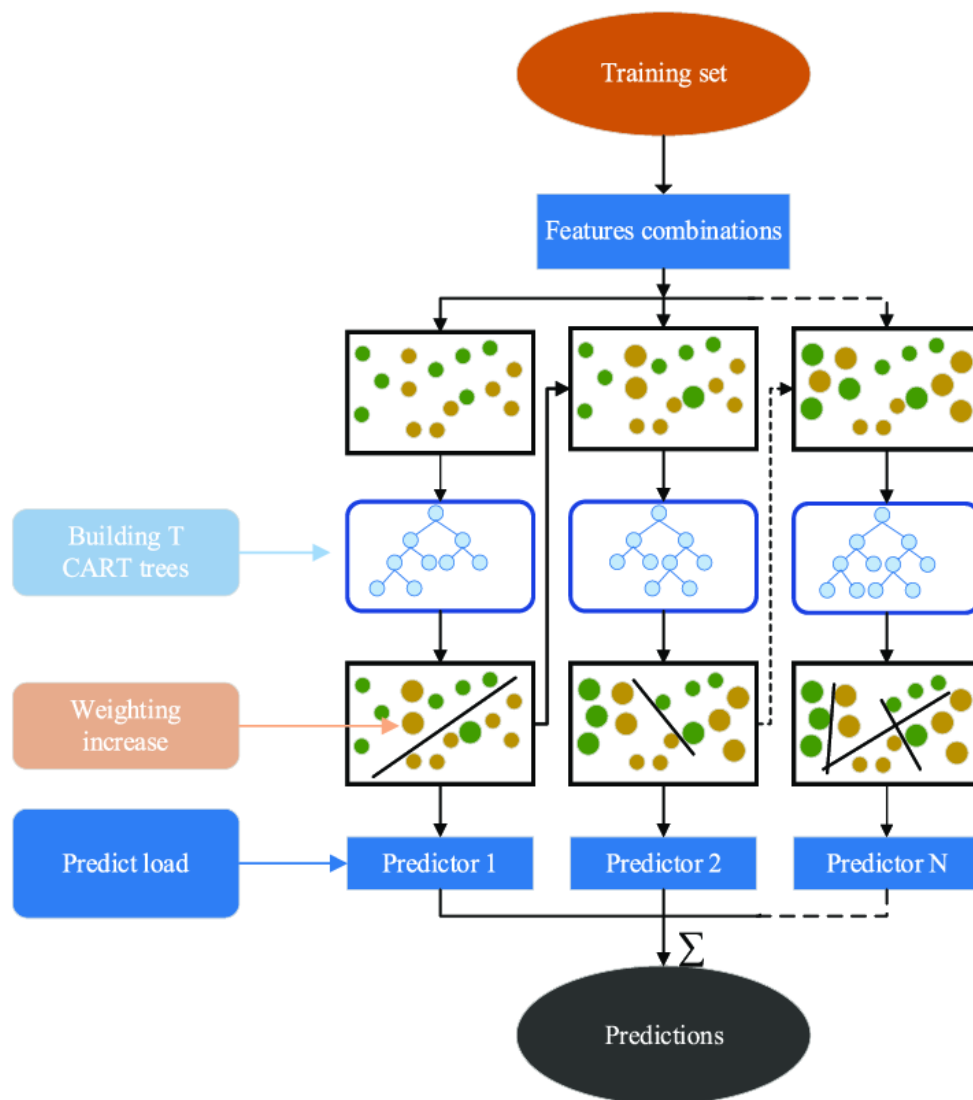
Η πρακτική εφαρμογή των αλγορίθμων boosting έχει αποδείξει την αξία τους σε προβλήματα πραγματικού κόσμου. Για παράδειγμα, το XGBoost χρησιμοποιείται ευρέως σε διαγωνισμούς μηχανικής μάθησης λόγω της υψηλής απόδοσής του και της ικανότητάς του να διαχειρίζεται μεγάλα σύνολα δεδομένων με πολυδιάστατα χαρακτηριστικά (Friedman, 2001).

Επιπλέον, οι αλγόριθμοι boosting είναι εξαιρετικά προσαρμόσιμοι και μπορούν να συνδυαστούν με άλλες τεχνικές μηχανικής μάθησης για τη βελτίωση της απόδοσης. Αυτό καθιστά το boosting μια δημοφιλή επιλογή για ερευνητές και επαγγελματίες που επιδιώκουν να αναπτύξουν ακριβή και αποδοτικά μοντέλα πρόβλεψης (Freund & Schapire, 1997; Chen & Guestrin, 2016).

XGBoost

Ο XGBoost (Extreme Gradient Boosting) είναι ένας από τους πιο δημοφιλείς και ισχυρούς αλγόριθμους μηχανικής μάθησης για εποπτευόμενα μαθησιακά προβλήματα, κυρίως για εφαρμογές παλινδρόμησης και ταξινόμησης. Αναπτύχθηκε από τον Tianqi Chen και παρουσιάστηκε το 2016, αποσπώντας μεγάλη προσοχή και υιοθέτηση από την κοινότητα της μηχανικής μάθησης λόγω της απόδοσής του και της ευελιξίας του (Chen & Guestrin, 2016).

XGBoost βασίζεται στην αρχή του boosting, η οποία περιλαμβάνει την κατασκευή ενός ισχυρού μοντέλου με το συνδυασμό πολλών ασθενών μοντέλων. Συγκεκριμένα, το XGBoost χρησιμοποιεί δέντρα απόφασης ως τα βασικά του μοντέλα και εφαρμόζει την τεχνική του gradient boosting. Σε κάθε βήμα, ένα νέο δέντρο εκπαιδεύεται για να διορθώσει τα σφάλματα του συνδυασμένου μοντέλου από τα προηγούμενα βήματα, και οι προβλέψεις του προστίθενται στις προβλέψεις των προηγούμενων δέντρων για να δημιουργηθεί μια νέα, βελτιωμένη πρόβλεψη (Chen & Guestrin, 2016).



Εικόνα 18: Η εικόνα απεικονίζει τη διαδικασία εκπαίδευσης και λειτουργίας ενός μοντέλου XGBoost. Η διαδικασία ξεκινά με το σύνολο εκπαίδευσης, το οποίο υφίσταται διάφορους συνδυασμούς χαρακτηριστικών. Στη συνέχεια, δημιουργούνται πολλαπλά δέντρα απόφασης (CART trees), το καθένα από τα οποία παράγει έναν προβλεπτικό φορτίο (predict load). Με την ολοκλήρωση της εκπαίδευσης όλων των δέντρων, οι προβλέψεις των επιμέρους μοντέλων συνδυάζονται για να δώσουν την τελική πρόβλεψη. (Πηγή: Yao et al., 2022)

Το XGBoost διακρίνεται για την αποδοτικότητά του και την ικανότητά του να διαχειρίζεται μεγάλα και πολύπλοκα σύνολα δεδομένων. Χρησιμοποιεί διάφορες τεχνικές βελτιστοποίησης, όπως η παράλληλη επεξεργασία και η χρήση του out-of-core computation, που επιτρέπουν την επεξεργασία δεδομένων που δεν χωράνε στη μνήμη του υπολογιστή. Επιπλέον, το XGBoost προσφέρει ρυθμίσεις για την πρόληψη της υπερεκπαίδευσης, όπως το early stopping και η ρύθμιση του learning rate, καθιστώντας το ευέλικτο και ικανό να προσαρμόζεται σε διάφορα προβλήματα (Chen & Guestrin, 2016).

Ένα από τα κύρια πλεονεκτήματα του XGBoost είναι η δυνατότητά του να ενσωματώνει χαρακτηριστικά κανονικοποίησης και να διαχειρίζεται ελλείποντα δεδομένα. Χρησιμοποιεί τη μετρική του κέρδους (gain) για την επιλογή των διασπάσεων κατά τη δημιουργία των δέντρων

απόφασης, η οποία λαμβάνει υπόψη την πληροφορία που παρέχουν τα διαφορετικά χαρακτηριστικά. Αυτό επιτρέπει στο XGBoost να επιλέγει τις πιο σημαντικές διασπάσεις και να αγνοεί τις διασπάσεις που δεν προσφέρουν σημαντική πληροφορία, βελτιώνοντας έτσι την απόδοση και την ερμηνευσιμότητα του μοντέλου (Chen & Guestrin, 2016).

Η ευρεία αποδοχή του XGBoost στην κοινότητα της μηχανικής μάθησης αποδίδεται στις εξαιρετικές του επιδόσεις σε διάφορους διαγωνισμούς μηχανικής μάθησης, όπως ο διαγωνισμός Kaggle. Οι βελτιώσεις και οι επεκτάσεις του XGBoost έχουν επίσης οδηγήσει στην ανάπτυξη νέων αλγορίθμων, όπως το LightGBM και το CatBoost, τα οποία προσφέρουν παρόμοια ή και βελτιωμένη απόδοση σε συγκεκριμένα προβλήματα. Ωστόσο, το XGBoost παραμένει ένα από τα πιο αξιόπιστα και αποδοτικά εργαλεία για την ανάλυση και πρόβλεψη μεγάλων συνόλων δεδομένων (Chen & Guestrin, 2016).

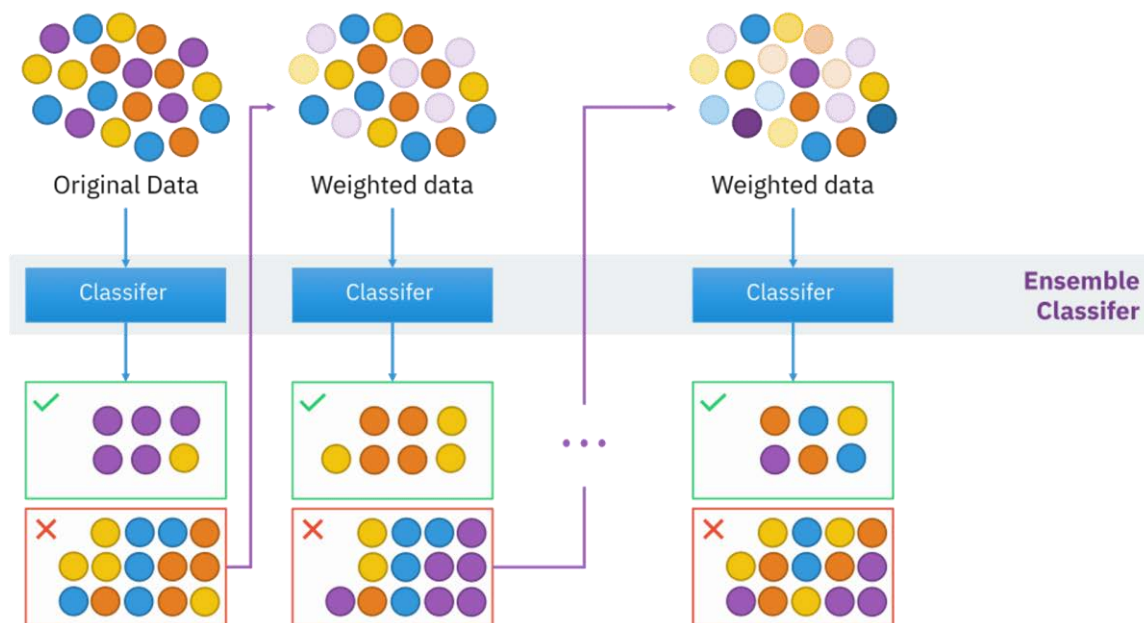
AdaBoost

Το AdaBoost, ή Adaptive Boosting, είναι ένας αλγόριθμος ενίσχυσης που χρησιμοποιείται για την αύξηση της απόδοσης των μοντέλων μηχανικής μάθησης. Ο αλγόριθμος αναπτύχθηκε από τον Yoan Freund και τον Robert Schapire το 1995 και είναι ένας από τους πιο δημοφιλείς αλγόριθμους ενίσχυσης λόγω της αποτελεσματικότητάς του και της απλότητάς του (Freund & Schapire, 1997).

Η βασική ιδέα πίσω από το AdaBoost είναι η δημιουργία ενός ισχυρού ταξινομητή από έναν συνδυασμό ασθενών ταξινομητών. Οι αδύναμοι ταξινομητές είναι απλά μοντέλα που έχουν ελαφρώς καλύτερη απόδοση από την τυχαία επιλογή. Στο AdaBoost, κάθε αδύναμος ταξινομητής εκπαιδεύεται σε μια τροποποιημένη έκδοση του αρχικού συνόλου δεδομένων. Οι παρατηρήσεις που ταξινομούνται εσφαλμένα από το προηγούμενο μοντέλο ενισχύονται, έτσι ώστε ο επόμενος ταξινομητής να επικεντρώνεται περισσότερο σε αυτές τις δύσκολες περιπτώσεις.

Ο αλγόριθμος λειτουργεί ως εξής:

1. Εκπαίδευση ενός αδύναμου ταξινομητή στο αρχικό σύνολο δεδομένων.
2. Αξιολόγηση της απόδοσης του ταξινομητή και υπολογισμός του σφάλματος.
3. Ενίσχυση των βαρών των παρατηρήσεων που ταξινομήθηκαν εσφαλμένα.
4. Εκπαίδευση ενός νέου αδύναμου ταξινομητή στο τροποποιημένο σύνολο δεδομένων.
5. Επανάληψη των παραπάνω βημάτων για έναν προκαθορισμένο αριθμό επαναλήψεων.
6. Συνδυασμός των αδύναμων ταξινομητών για την παραγωγή του τελικού ισχυρού ταξινομητή.



Εικόνα 19: Η εικόνα παρουσιάζει τη διαδικασία του αλγορίθμου AdaBoost, ο οποίος συνδυάζει πολλούς ασθενείς ταξινομητές για να δημιουργήσει έναν ισχυρό ταξινομητή. Τα δεδομένα που ταξινομούνται λάθος από έναν ασθενή ταξινομητή λαμβάνουν μεγαλύτερο βάρος στην επόμενη επανάληψη, επιτρέποντας έτσι στον αλγόριθμο να εστιάσει στα δύσκολα παραδείγματα και να βελτιώσει την ακρίβεια των προβλέψεων του. (Πηγή: Almbetter, n.d.)

Ο τελικός ισχυρός ταξινομητής είναι ένας συνδυασμός όλων των αδύναμων ταξινομητών, όπου κάθε αδύναμος ταξινομητής έχει έναν συντελεστή που εξαρτάται από την απόδοσή του κατά την εκπαίδευση. Οι αδύναμοι ταξινομητές με καλύτερη απόδοση έχουν μεγαλύτερη επιρροή στον τελικό ταξινομητή (Hastie, Tibshirani, & Friedman, 2009).

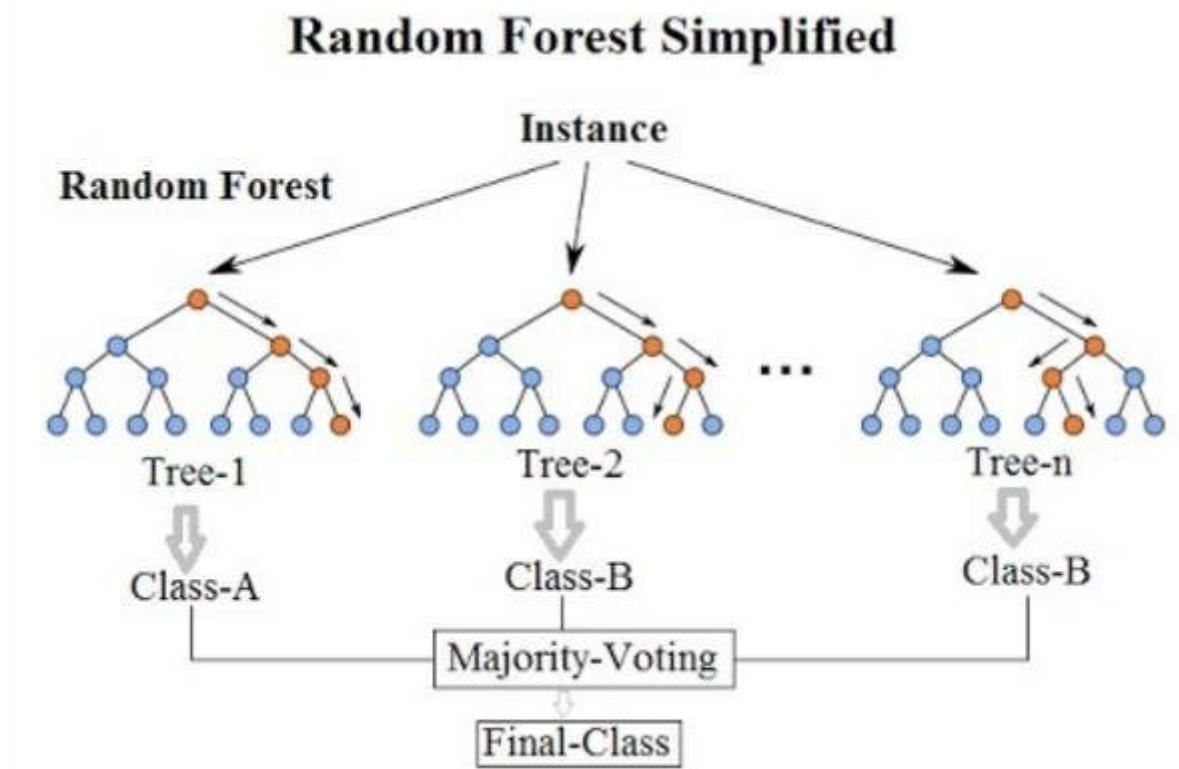
Μία από τις μεγαλύτερες δυνάμεις του AdaBoost είναι η ικανότητά του να προσαρμόζεται στα δεδομένα εκπαίδευσης και να επικεντρώνεται στις πιο δύσκολες περιπτώσεις. Αυτό καθιστά τον αλγόριθμο πολύ αποτελεσματικό σε πολλά προβλήματα ταξινόμησης. Ωστόσο, το AdaBoost μπορεί να είναι ευαίσθητο στον θόρυβο στα δεδομένα εκπαίδευσης, καθώς οι εσφαλμένες παρατηρήσεις μπορούν να ενισχυθούν υπερβολικά, οδηγώντας σε υπερεκπαίδευση (Friedman, Hastie, & Tibshirani, 2001).

Στη μετεωρολογία, το AdaBoost έχει χρησιμοποιηθεί για διάφορες εφαρμογές, όπως η πρόβλεψη βροχοπτώσεων και η ταξινόμηση των τύπων καιρού. Έρευνες έχουν δείξει ότι το AdaBoost μπορεί να βελτιώσει την ακρίβεια των προβλέψεων σε σύγκριση με τους παραδοσιακούς αλγόριθμους, καθιστώντας το ένα ισχυρό εργαλείο για μετεωρολογικές εφαρμογές (Chen et al., 2023).

Random Forest

Ο αλγόριθμος Random Forest αποτελεί έναν από τους πιο δημοφιλείς και ισχυρούς αλγορίθμους μηχανικής μάθησης για ταξινόμηση και παλινδρόμηση. Αναπτύχθηκε από τον Leo Breiman και βασίζεται στη μέθοδο bagging (bootstrap aggregating). Η βασική ιδέα πίσω από τον αλγόριθμο είναι

να συνδυάζει τις προβλέψεις από πολλαπλά δέντρα απόφασης για να βελτιώσει τη γενική απόδοση του μοντέλου (Breiman, 2001). Ο Random Forest λειτουργεί με την εκπαίδευση πολλών δέντρων απόφασης, όπου κάθε δέντρο δημιουργείται από ένα τυχαίο υποσύνολο των δεδομένων εκπαίδευσης και των χαρακτηριστικών τους, μέσω της μεθόδου bootstrap sampling. Η τελική πρόβλεψη προκύπτει μέσω της πλειοψηφικής ψήφου (για ταξινόμηση) ή του μέσου όρου (για παλινδρόμηση) των προβλέψεων όλων των δέντρων (Louppe, 2014).



Εικόνα 20: Απλοποιημένο διάγραμμα του αλγόριθμου Random Forest, που δείχνει τον τρόπο με τον οποίο πολλαπλά δέντρα αποφάσεων συνδυάζονται για να βελτιώσουν την ακρίβεια των προβλέψεων μέσω της διαδικασίας πλειοψηφικής ψήφου. (Πηγή: Roy et al., 2023)

Ένα από τα κύρια πλεονεκτήματα του Random Forest είναι η ικανότητά του να μειώνει το πρόβλημα της υπερεκπαίδευσης (overfitting). Επειδή κάθε δέντρο εκπαιδεύεται σε διαφορετικό υποσύνολο δεδομένων και επειδή τα δέντρα είναι ανεξάρτητα μεταξύ τους, η μέση πρόβλεψη από το σύνολο των δέντρων τείνει να είναι πιο σταθερή και ακριβής από τις προβλέψεις ενός μεμονωμένου δέντρου (Fernández-Delgado et al., 2014; Probst et al., 2019).

Ο Random Forest είναι επίσης εξαιρετικά ανθεκτικός σε δεδομένα με θόρυβο και σε περιπτώσεις όπου υπάρχουν πολλές άσχετες μεταβλητές. Αυτό οφείλεται στο γεγονός ότι ο αλγόριθμος επιλέγει τυχαία ένα υποσύνολο χαρακτηριστικών κατά την κατασκευή κάθε δέντρου, γεγονός που βοηθά στην εξάλειψη των άσχετων χαρακτηριστικών και στη βελτίωση της απόδοσης του μοντέλου (Kuhn & Johnson, 2019). Επιπλέον, ο Random Forest μπορεί να υπολογίσει τη σχετική σημασία των διαφόρων

χαρακτηριστικών, παρέχοντας έτσι πολύτιμες πληροφορίες για την κατανόηση των δεδομένων και τη λήψη αποφάσεων (Wright & Ziegler, 2017).

Συνολικά, ο αλγόριθμος Random Forest είναι ευέλικτος, αποδοτικός και αξιόπιστος, καθιστώντας τον ιδανικό για ποικίλες εφαρμογές μηχανικής μάθησης. Η χρήση του έχει επεκταθεί σε πολλούς τομείς, όπως η βιοϊατρική, η οικονομία, η πληροφορική και η περιβαλλοντική επιστήμη, λόγω της ικανότητάς του να παράγει ακριβείς και σταθερές προβλέψεις (Jiang et al., 2020; Tyralis et al., 2019).

Γραμμική Παλινδρόμηση (Linear Regression)

Η Γραμμική Παλινδρόμηση (Linear Regression) είναι ένας από τους πιο βασικούς και ευρέως χρησιμοποιούμενους αλγορίθμους στην μηχανική μάθηση και τα στατιστικά. Χρησιμοποιείται για την πρόβλεψη μιας εξαρτημένης μεταβλητής (στόχου) με βάση μία ή περισσότερες ανεξάρτητες μεταβλητές (χαρακτηριστικά). Στην ουσία, ο αλγόριθμος προσπαθεί να βρει τη γραμμή που καλύτερα προσαρμόζεται στα δεδομένα μας, ελαχιστοποιώντας το σφάλμα πρόβλεψης (James et al., 2013).

Η βασική ιδέα της γραμμικής παλινδρόμησης είναι η εξής:

Δίνεται ένα σύνολο δεδομένων με χαρακτηριστικά $\bar{X}$ και μια αντίστοιχη εξαρτημένη μεταβλητή Y , και προσπαθούμε να βρούμε την γραμμική σχέση μεταξύ τους.

Αυτή η σχέση περιγράφεται από την εξίσωση της ευθείας γραμμής (Hastie et al., 2008):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Όπου:

- Y είναι η εξαρτημένη μεταβλητή (στόχος)
- X_i είναι οι ανεξάρτητες μεταβλητές (χαρακτηριστικά)
- β_i είναι οι συντελεστές (βάρη) που προσπαθούμε να εκτιμήσουμε
- ϵ είναι το σφάλμα (θόρυβος)

Η πιο συνηθισμένη μέθοδος για την εκτίμηση των συντελεστών είναι η μέθοδος των ελαχίστων τετραγώνων (Ordinary Least Squares, OLS). Η μέθοδος αυτή επιδιώκει να ελαχιστοποιήσει το άθροισμα των τετραγώνων των αποκλίσεων (σφαλμάτων) μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών. Η λύση αυτής της ελαχιστοποίησης μας δίνει τους βέλτιστους συντελεστές β (Goodfellow et al., 2016).

Η γραμμική παλινδρόμηση βασίζεται σε ορισμένες υποθέσεις. Πρώτον, η σχέση μεταξύ των ανεξάρτητων και της εξαρτημένης μεταβλητής πρέπει να είναι γραμμική. Δεύτερον, τα σφάλματα (υπολείμματα) πρέπει να είναι ανεξάρτητα και κατανομημένα κανονικά με μέση τιμή μηδέν και σταθερή διακύμανση (ομοσκεδαστικότητα). Τρίτον, δεν πρέπει να υπάρχει πολυσυγγραμμικότητα, δηλαδή οι ανεξάρτητες μεταβλητές δεν πρέπει να είναι υψηλά συσχετισμένες μεταξύ τους. Αν αυτές

οι υποθέσεις δεν ικανοποιούνται, τα αποτελέσματα της γραμμικής παλινδρόμησης μπορεί να είναι αναξιόπιστα (Hyndman & Athanasopoulos, 2018).

Ένα από τα πλεονεκτήματα της γραμμικής παλινδρόμησης είναι η απλότητά της και η ευκολία ερμηνείας των συντελεστών της. Κάθε συντελεστής β_i αντιπροσωπεύει την αναμενόμενη μεταβολή στην εξαρτημένη μεταβλητή Y για μία μονάδα μεταβολής της αντίστοιχης ανεξάρτητης μεταβλητής X_i , υπό την προϋπόθεση ότι οι άλλες μεταβλητές παραμένουν σταθερές. Επιπλέον, η γραμμική παλινδρόμηση μπορεί να χρησιμοποιηθεί ως βάση για πιο πολύπλοκα μοντέλα παλινδρόμησης, όπως η πολυωνυμική και η λογαριθμική παλινδρόμηση (Géron, 2019).

Παρόλα αυτά, η γραμμική παλινδρόμηση έχει και τα μειονεκτήματά της. Δεν είναι κατάλληλη για μοντέλα που περιλαμβάνουν μη γραμμικές σχέσεις μεταξύ των μεταβλητών. Σε περιπτώσεις όπου οι υποθέσεις της γραμμικής παλινδρόμησης δεν ικανοποιούνται, μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα. Επίσης, η παρουσία εξωτικών τιμών (outliers) μπορεί να επηρεάσει σημαντικά τα αποτελέσματα του μοντέλου.

Η γραμμική παλινδρόμηση έχει πολλές εφαρμογές, από την οικονομία και τα χρηματοοικονομικά μέχρι τη βιολογία και την κοινωνιολογία. Για παράδειγμα, μπορεί να χρησιμοποιηθεί για την πρόβλεψη των πωλήσεων ενός προϊόντος με βάση την τιμή και την ποσότητα της διαφήμισης, ή για την εκτίμηση της επίδρασης μιας ιατρικής θεραπείας με βάση τα δεδομένα ασθενών. Στον τομέα της μετεωρολογίας, η γραμμική παλινδρόμηση μπορεί να χρησιμοποιηθεί για την πρόβλεψη καιρικών συνθηκών όπως η θερμοκρασία και η βροχόπτωση (Schmidhuber, 2015).

Σύγκριση και Αξιολόγηση των Μοντέλων

Κριτήρια Αξιολόγησης

Η σύγκριση και αξιολόγηση των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται στην πρόβλεψη καιρού είναι κρίσιμη για την κατανόηση της αποτελεσματικότητας και της ακρίβειάς τους. Η αξιολόγηση των μοντέλων πραγματοποιείται με βάση συγκεκριμένα κριτήρια που επιτρέπουν την αντικειμενική μέτρηση της απόδοσής τους. Τα κύρια κριτήρια αξιολόγησης περιλαμβάνουν την ακρίβεια (accuracy), το μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE), το μέσο τετραγωνικό σφάλμα (Mean Squared Error - MSE), το ποσοστό διακύμανσης που εξηγείται (R^2 score), και την υπολογιστική αποδοτικότητα.

Ακρίβεια Προβλέψεων (Accuracy)

Η ακρίβεια των προβλέψεων είναι κρίσιμη για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης στη μετεωρολογία. Ένα μοντέλο που παρέχει ακριβείς προβλέψεις μπορεί να βοηθήσει στη λήψη ενημερωμένων αποφάσεων σχετικά με τον καιρό, την κλιματική αλλαγή και την πρόληψη φυσικών καταστροφών.

Η ακρίβεια ορίζεται ως το ποσοστό των σωστών προβλέψεων σε σχέση με το σύνολο των προβλέψεων που έχουν γίνει. Αυτό το μέτρο μπορεί να είναι πολύπλοκο στην εφαρμογή του σε προβλήματα παλινδρόμησης, όπως είναι οι μετεωρολογικές προβλέψεις, όπου οι προβλέψεις είναι συνεχείς τιμές. Σε αυτές τις περιπτώσεις, η ακρίβεια μετριέται συχνά μέσω του μέσου απόλυτου σφάλματος (MAE), του μέσου τετραγωνικού σφάλματος (MSE) και του συντελεστή προσδιορισμού (R^2) (Shalev-Shwartz & Ben-David, 2014).

Το MAE μετρά την μέση απόλυτη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών. Ένα χαμηλό MAE υποδηλώνει υψηλή ακρίβεια. Το MSE, αντίστοιχα, δίνει μεγαλύτερη βαρύτητα στις μεγάλες αποκλίσεις, κάνοντάς το πιο ευαίσθητο στα εξωτερικά στοιχεία (outliers) (James et al., 2013). Ο συντελεστής προσδιορισμού (R^2) μετρά το ποσοστό της διακύμανσης στην εξαρτημένη μεταβλητή που εξηγείται από το μοντέλο. Ένα R^2 κοντά στο 1 υποδηλώνει ότι το μοντέλο εξηγεί καλά τα δεδομένα (Hastie et al., 2008).

Στην εφαρμογή της γραμμικής παλινδρόμησης, τα μοντέλα μηχανικής μάθησης μπορούν να εκπαιδευτούν για να προβλέπουν μετεωρολογικές παραμέτρους όπως η θερμοκρασία, η ταχύτητα ανέμου, και η υγρασία. Η ακρίβεια αυτών των προβλέψεων είναι κρίσιμη για τη βελτίωση των μετεωρολογικών μοντέλων και την ανάπτυξη αποτελεσματικότερων συστημάτων πρόγνωσης καιρού (Friedman et al., 2010).

Η ακρίβεια της πρόβλεψης επηρεάζεται από την ποιότητα και την ποσότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση του μοντέλου. Δεδομένα υψηλής ποιότητας και μεγάλης κλίμακας βοηθούν στην εκπαίδευση μοντέλων που είναι πιο ακριβή και γενικευμένα (Goodfellow et al., 2016). Επιπλέον, η επιλογή των χαρακτηριστικών (feature selection) και η προεπεξεργασία των δεδομένων παίζουν σημαντικό ρόλο στην ακρίβεια των μοντέλων.

Τα σύγχρονα μετεωρολογικά μοντέλα όπως το GFS (Global Forecast System) και το ECMWF (European Centre for Medium-Range Weather Forecasts) χρησιμοποιούν σύνθετα αλγοριθμικά μοντέλα για να παράγουν προβλέψεις υψηλής ακρίβειας. Παράλληλα, η χρήση αλγορίθμων μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα και τα δέντρα αποφάσεων, βελτιώνει περαιτέρω την ακρίβεια αυτών των προβλέψεων (McGovern et al., 2017).

Η χρήση αλγορίθμων μηχανικής μάθησης στη μετεωρολογία είναι σημαντική για την βελτίωση της ακρίβειας των προβλέψεων. Αλγόριθμοι όπως το XGBoost, το AdaBoost, και τα Νευρωνικά Δίκτυα έχουν αποδείξει ότι μπορούν να προσφέρουν ακριβείς προβλέψεις όταν εκπαιδεύονται με κατάλληλα δεδομένα (Chen & Guestrin, 2016; Freund & Schapire, 1997).

Η συνεχής βελτίωση της ακρίβειας των προβλέψεων είναι αναγκαία για την αντιμετώπιση των προκλήσεων που προκύπτουν από την κλιματική αλλαγή και τα ακραία καιρικά φαινόμενα. Η εξέλιξη της τεχνολογίας και η αύξηση της υπολογιστικής ισχύος επιτρέπουν την ανάπτυξη πιο ακριβών και

αξιόπιστων μοντέλων, που είναι ζωτικής σημασίας για την πρόβλεψη και την αντιμετώπιση των επιπτώσεων του καιρού (Shalev-Shwartz & Ben-David, 2014).

Υπολογιστική Αποδοτικότητα

Η υπολογιστική αποδοτικότητα είναι ένας από τους κύριους παράγοντες που επηρεάζουν την απόδοση των αλγορίθμων μηχανικής μάθησης. Αναφέρεται στον χρόνο και τους πόρους που απαιτούνται για την εκπαίδευση και την πρόβλεψη ενός μοντέλου. Η αποδοτικότητα καθορίζεται από παράγοντες όπως η πολυπλοκότητα του αλγορίθμου, η ποσότητα των δεδομένων και οι υπολογιστικοί πόροι που χρησιμοποιούνται (Liu et al., 2014).

Οι σύγχρονοι αλγόριθμοι μηχανικής μάθησης, όπως το XGBoost και το AdaBoost, είναι σχεδιασμένοι για να είναι αποδοτικοί και να εκτελούνται γρήγορα ακόμη και με μεγάλα σύνολα δεδομένων. Το XGBoost, για παράδειγμα, χρησιμοποιεί προηγμένες τεχνικές όπως η παράλληλη επεξεργασία και η αποκοπή δέντρων για να βελτιώσει την ταχύτητα και την απόδοση του μοντέλου (Chen & Guestrin, 2016). Αυτό καθιστά το XGBoost ιδανικό για εφαρμογές σε πραγματικό χρόνο όπου η ταχύτητα είναι κρίσιμη.

Η υπολογιστική αποδοτικότητα είναι επίσης σημαντική για την ανάπτυξη και την υλοποίηση αλγορίθμων σε περιορισμένα περιβάλλοντα πόρων, όπως οι φορητές συσκευές και οι ενσωματωμένοι αισθητήρες. Σε αυτά τα περιβάλλοντα, η βελτιστοποίηση της κατανάλωσης ενέργειας και της μνήμης είναι απαραίτητη για τη διασφάλιση της αποτελεσματικής λειτουργίας του συστήματος (Zoph et al., 2017).

Ένας άλλος σημαντικός παράγοντας που επηρεάζει την υπολογιστική αποδοτικότητα είναι η κλιμάκωση των αλγορίθμων. Καθώς τα δεδομένα αυξάνονται εκθετικά, οι αλγόριθμοι πρέπει να είναι σε θέση να διαχειριστούν την αυξημένη πολυπλοκότητα χωρίς να θυσιάζουν την ταχύτητα ή την απόδοση. Τα κατανεμημένα συστήματα επεξεργασίας δεδομένων, όπως το Hadoop και το Spark, χρησιμοποιούνται συχνά για να επιτύχουν αυτήν την κλιμάκωση (Zaharia et al., 2016).

Η υπολογιστική αποδοτικότητα των αλγορίθμων μηχανικής μάθησης είναι κρίσιμη για την εφαρμογή τους σε πραγματικό χρόνο. Οι βελτιώσεις στους αλγορίθμους, οι οποίες μειώνουν την πολυπλοκότητα και τον χρόνο εκπαίδευσης, μπορούν να κάνουν τη διαφορά σε εφαρμογές όπως η πρόβλεψη καιρού, όπου οι αποφάσεις πρέπει να λαμβάνονται γρήγορα και με ακρίβεια (Raschka & Mirjalili, 2019).

Οι αλγόριθμοι όπως ο Random Forest και το Gradient Boosting είναι επίσης γνωστοί για την υπολογιστική τους αποδοτικότητα. Ο Random Forest, για παράδειγμα, χρησιμοποιεί πολλαπλά δέντρα απόφασης για να βελτιώσει την απόδοση και να μειώσει την υπερβολική προσαρμογή (overfitting), διατηρώντας παράλληλα υψηλή αποδοτικότητα (Breiman, 2001).

Επιπλέον, η ανάπτυξη αλγορίθμων που εκμεταλλεύονται τις δυνατότητες του hardware, όπως οι GPU και οι TPU, έχει βελτιώσει δραματικά την υπολογιστική αποδοτικότητα. Οι GPU επιτρέπουν την παράλληλη επεξεργασία μεγάλων όγκων δεδομένων, βελτιώνοντας την ταχύτητα εκπαίδευσης των μοντέλων (Goodfellow et al., 2016).

Η συνεχής έρευνα και ανάπτυξη νέων τεχνικών βελτιστοποίησης συμβάλλει στη βελτίωση της υπολογιστικής αποδοτικότητας των αλγορίθμων μηχανικής μάθησης. Οι νέες προσεγγίσεις, όπως οι τεχνικές συμπίεσης μοντέλων και η υλοποίηση αλγορίθμων σε εξειδικευμένο hardware, συμβάλλουν στη δημιουργία πιο αποδοτικών και αποτελεσματικών συστημάτων (Tan et al., 2019).

Ευελιξία και Προσαρμοστικότητα

Η ευελιξία και η προσαρμοστικότητα των αλγορίθμων μηχανικής μάθησης είναι καίριας σημασίας για την αποτελεσματική τους εφαρμογή σε διαφορετικά προβλήματα και περιβάλλοντα. Οι αλγόριθμοι που μπορούν να προσαρμόζονται εύκολα στις μεταβαλλόμενες συνθήκες και στα δεδομένα είναι πιο πιθανό να επιτύχουν σε ένα ευρύ φάσμα εφαρμογών (Zhu et al., 2007).

Ένας από τους πιο ευέλικτους αλγόριθμους είναι ο Random Forest, ο οποίος μπορεί να χρησιμοποιηθεί τόσο για προβλήματα ταξινόμησης όσο και για παλινδρόμηση. Ο Random Forest είναι ικανός να διαχειριστεί μεγάλους όγκους δεδομένων και πολλά χαρακτηριστικά, ενώ παράλληλα προσφέρει υψηλή ακρίβεια και σταθερότητα (Biau & Scornet, 2015). Η προσαρμοστικότητά του έγκειται στη δυνατότητα του να δημιουργεί πολλαπλά δέντρα απόφασης και να συνδυάζει τα αποτελέσματά τους για να βελτιώσει την απόδοση.

Οι αλγόριθμοι ενίσχυσης, όπως το XGBoost και το AdaBoost, προσφέρουν επίσης υψηλή ευελιξία. Το XGBoost, για παράδειγμα, επιτρέπει την προσαρμογή πολλών παραμέτρων, όπως ο αριθμός των δέντρων, το βάθος των δέντρων και ο ρυθμός μάθησης, για να βελτιστοποιηθεί η απόδοση του μοντέλου για συγκεκριμένα δεδομένα (Chen & Guestrin, 2016). Αυτή η δυνατότητα προσαρμογής καθιστά το XGBoost ιδανικό για χρήση σε διαφορετικά είδη δεδομένων και προβλήματα.

Η προσαρμοστικότητα των αλγορίθμων μηχανικής μάθησης είναι επίσης σημαντική για την ανάπτυξη σε πραγματικό χρόνο εφαρμογών, όπως η πρόβλεψη καιρού. Για παράδειγμα, οι αλγόριθμοι πρέπει να είναι σε θέση να προσαρμοστούν γρήγορα σε νέες μετεωρολογικές συνθήκες και να ενημερώνουν τα μοντέλα τους με νέα δεδομένα σε τακτά χρονικά διαστήματα (Rasp et al., 2020).

Η ευελιξία ενός αλγόριθμου είναι επίσης κρίσιμη όταν πρόκειται για την επεξεργασία δεδομένων από διαφορετικές πηγές και μορφές. Οι σύγχρονοι αλγόριθμοι μηχανικής μάθησης έχουν σχεδιαστεί για να διαχειρίζονται δεδομένα από διάφορες πηγές, όπως αισθητήρες IoT, δορυφόρους και

δεδομένα κοινωνικών δικτύων, και να ενσωματώνουν αυτές τις πληροφορίες για τη βελτίωση των προβλέψεων τους (Vaquero et al., 2008).

Οι αλγόριθμοι που είναι ευέλικτοι και προσαρμόσιμοι είναι επίσης καλύτερα εξοπλισμένοι για να αντιμετωπίσουν τις προκλήσεις της υπερβολικής προσαρμογής (overfitting). Χρησιμοποιώντας τεχνικές όπως η διασταυρούμενη επαλήθευση και η κανονικοποίηση, οι αλγόριθμοι μπορούν να προσαρμοστούν στα δεδομένα εκπαίδευσης και να διατηρήσουν υψηλή απόδοση στα δεδομένα δοκιμής (Goodfellow et al., 2016).

Η ευελιξία των αλγορίθμων ενισχύεται επίσης από την ικανότητά τους να ενσωματώνουν νέες τεχνολογίες και μεθοδολογίες. Για παράδειγμα, οι αλγόριθμοι μπορούν να επωφεληθούν από τις εξελίξεις στον τομέα της βαθιάς μάθησης και να ενσωματώσουν νευρωνικά δίκτυα για τη βελτίωση της απόδοσής τους (Zhou et al., 2007).

Τέλος, η ευελιξία και η προσαρμοστικότητα των αλγορίθμων μηχανικής μάθησης είναι κρίσιμες για την εφαρμογή τους σε περιβάλλοντα με περιορισμένους πόρους. Οι αλγόριθμοι πρέπει να είναι ικανοί να λειτουργούν αποδοτικά σε φορητές συσκευές και ενσωματωμένα συστήματα, προσαρμόζοντας την κατανάλωση ενέργειας και μνήμης τους για να επιτύχουν υψηλή απόδοση (Han et al., 2015).

Πλεονεκτήματα και Μειονεκτήματα των Μοντέλων

Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα, ως μία από τις πιο δημοφιλείς και ισχυρές τεχνικές μηχανικής μάθησης, προσφέρουν πληθώρα πλεονεκτημάτων και μειονεκτημάτων που επηρεάζουν τη χρήση τους σε διάφορες εφαρμογές. Η πολυπλοκότητά τους και οι δυνατότητές τους τα καθιστούν ιδανικά για πολύπλοκα προβλήματα, αλλά φέρνουν και προκλήσεις που πρέπει να αντιμετωπιστούν προσεκτικά.

Ένα από τα σημαντικότερα πλεονεκτήματα των νευρωνικών δικτύων είναι η ικανότητά τους να μαθαίνουν και να αναγνωρίζουν πολύπλοκα μοτίβα σε μεγάλα σύνολα δεδομένων. Τα νευρωνικά δίκτυα μπορούν να μάθουν από τα δεδομένα χωρίς την ανάγκη για προκαθορισμένους κανόνες, επιτρέποντάς τους να είναι ιδιαίτερα ευέλικτα και να προσαρμόζονται σε διαφορετικές εφαρμογές, όπως η αναγνώριση εικόνων, η ανάλυση φωνής και η επεξεργασία φυσικής γλώσσας (LeCun, Bengio, & Hinton, 2015).

Η ικανότητα των νευρωνικών δικτύων να επεξεργάζονται μεγάλα και σύνθετα σύνολα δεδομένων τα καθιστά εξαιρετικά χρήσιμα σε τομείς όπως η πρόβλεψη καιρού, η ανάλυση οικονομικών δεδομένων και η ιατρική διάγνωση (Goodfellow, Bengio, & Courville, 2016). Επιπλέον, τα δίκτυα αυτά μπορούν να βελτιώνονται συνεχώς με την προσθήκη νέων δεδομένων, γεγονός που τα καθιστά εξαιρετικά δυναμικά και προσαρμόσιμα σε μεταβαλλόμενες συνθήκες.

Αντίθετα, ένα από τα κύρια μειονεκτήματα των νευρωνικών δικτύων είναι η ανάγκη για μεγάλο υπολογιστικό χρόνο και πόρους. Η εκπαίδευση αυτών των μοντέλων μπορεί να είναι εξαιρετικά χρονοβόρα και απαιτεί πρόσβαση σε ισχυρούς επεξεργαστές γραφικών (GPUs) ή ακόμα και υπερυπολογιστές, γεγονός που μπορεί να είναι δαπανηρό (Strubell, Ganesh, & McCallum, 2019).

Επιπλέον, η διαδικασία εκπαίδευσης των νευρωνικών δικτύων απαιτεί μεγάλα σύνολα δεδομένων για να αποδώσουν καλά, και η έλλειψη δεδομένων μπορεί να οδηγήσει σε προβλήματα υπερεκπαίδευσης ή κακής απόδοσης (Goodfellow et al., 2016). Επίσης, η ανάπτυξη και η βελτιστοποίηση των νευρωνικών δικτύων μπορεί να είναι πολύπλοκη, απαιτώντας εξειδικευμένες γνώσεις και εμπειρία (LeCun et al., 2015).

Τα νευρωνικά δίκτυα μπορεί επίσης να είναι ευαίσθητα σε μικρές αλλαγές στα δεδομένα εισόδου, κάτι που μπορεί να επηρεάσει την αξιοπιστία των προβλέψεων τους. Αυτό μπορεί να είναι ιδιαίτερα προβληματικό σε εφαρμογές όπου απαιτείται υψηλή ακρίβεια και σταθερότητα, όπως η πρόβλεψη καιρού και η ιατρική διάγνωση (Schmidhuber, 2015). Τέλος, η χρήση νευρωνικών δικτύων εγείρει ηθικά και κοινωνικά ζητήματα, όπως η ανάγκη για διαφάνεια και δικαιοσύνη στις αποφάσεις που λαμβάνονται από αυτά τα συστήματα (Lipton, 2017).

Πίνακας 4: Συγκεντρωτικός Πίνακας Πλεονεκτημάτων και Μειονεκτημάτων των NN
Τύπος Νευρωνικού Δικτύου

	Πλεονεκτήματα	Μειονεκτήματα
<i>Πλήρως Συνδεδεμένα Δίκτυα (Fully Connected)</i>	Ευελιξία, Ευκολία στην υλοποίηση	Υπολογιστικά απαιτητικά, Τάση υπερεκπαίδευσης
<i>Συνελικτικά Δίκτυα (CNNs)</i>	Ιδανικά για ανάλυση εικόνων, Υψηλή ακρίβεια	Υπολογιστικά απαιτητικά, Πολύπλοκη δομή
<i>Επαναληπτικά Δίκτυα (RNNs)</i>	Κατάλληλα για σειριακά δεδομένα, Ικανότητα μνήμης	Προβλήματα εξαφάνισης/έκρηξης βαθμού προσαρμογής, Αργή εκπαίδευση
<i>Μακροπρόθεσμη Βραχυπρόθεσμη Μνήμη (LSTM)</i>	Αντιμετωπίζει τα προβλήματα των RNNs, Αποτελεσματικά για σειριακά δεδομένα	Υπολογιστικά απαιτητικά, Πολύπλοκη δομή

SVMs

Τα SVMs (Support Vector Machines) είναι μία από τις πιο δημοφιλείς και ευέλικτες τεχνικές μηχανικής μάθησης, προσφέροντας μια σειρά από πλεονεκτήματα και μειονεκτήματα που τα καθιστούν ιδιαίτερα χρήσιμα σε διάφορες εφαρμογές.

Ένα από τα σημαντικότερα πλεονεκτήματα των SVMs είναι η ικανότητά τους να χειρίζονται δεδομένα υψηλής διάστασης. Αυτό τα καθιστά ιδανικά για εφαρμογές όπου τα δεδομένα έχουν πολλές διαστάσεις και οι παραδοσιακές τεχνικές μπορεί να αποτύχουν. Επιπλέον, τα SVMs είναι αποτελεσματικά ακόμα και σε περιπτώσεις όπου ο αριθμός των διαστάσεων υπερβαίνει τον αριθμό των δειγμάτων, προσφέροντας αξιόπιστες προβλέψεις και αναλύσεις (Vapnik, 2000).

Τα SVMs χρησιμοποιούν την έννοια των διανυσμάτων υποστήριξης και των υπερ-επιπέδων για να διαχωρίσουν τις κατηγορίες δεδομένων με τον καλύτερο δυνατό τρόπο. Αυτή η προσέγγιση επιτρέπει στα SVMs να βρουν το βέλτιστο όριο διαχωρισμού που μεγιστοποιεί την απόσταση μεταξύ των διαφορετικών κατηγοριών, βελτιώνοντας την γενίκευση και την ακρίβεια των προβλέψεων (Cristianini & Shawe-Taylor, 2000).

Ένα άλλο σημαντικό πλεονέκτημα των SVMs είναι η ευελιξία τους μέσω της χρήσης των πυρηνικών συναρτήσεων (kernels). Οι πυρηνικές συναρτήσεις επιτρέπουν στα SVMs να διαχειρίζονται μη γραμμικά διαχωρίσιμα δεδομένα μετατρέποντάς τα σε έναν υψηλότερης διάστασης χώρο όπου μπορούν να διαχωριστούν γραμμικά. Αυτό καθιστά τα SVMs πολύ ισχυρά και ευέλικτα, καθώς μπορούν να προσαρμοστούν σε διάφορα είδη προβλημάτων με τη χρήση κατάλληλων πυρηνικών συναρτήσεων (Schölkopf & Smola, 2002).

Η ακρίβεια των SVMs σε προβλήματα ταξινόμησης και παλινδρόμησης είναι εξαιρετικά υψηλή, καθιστώντας τα κατάλληλα για εφαρμογές όπως η αναγνώριση εικόνων, η ανάλυση κειμένου και η βιοπληροφορική. Τα SVMs έχουν αποδειχθεί εξαιρετικά αποδοτικά στην ανίχνευση ανεπιθύμητης αλληλογραφίας (spam detection), στην πρόβλεψη ασθενειών και στην αναγνώριση μοτίβων σε βιολογικά δεδομένα (Cortes & Vapnik, 1995).

Τέλος, τα SVMs είναι γνωστά για την ανθεκτικότητά τους σε προβλήματα υπερπροσαρμογής (overfitting), ειδικά όταν χρησιμοποιούνται σε συνδυασμό με κατάλληλες τεχνικές κανονικοποίησης και επιλογής χαρακτηριστικών. Η καλή γενίκευση και η σταθερότητα των SVMs τα καθιστούν αξιόπιστα εργαλεία για πολλές εφαρμογές μηχανικής μάθησης, διασφαλίζοντας ότι οι προβλέψεις τους παραμένουν αξιόπιστες και ακριβείς ακόμα και σε διαφορετικά σύνολα δεδομένων (Burges, 1998).

Ωστόσο, τα SVMs έχουν και ορισμένα μειονεκτήματα. Ένα από τα βασικά μειονεκτήματα είναι η δυσκολία στην ερμηνεία των αποτελεσμάτων τους. Οι υπερ-επιφάνειες που χρησιμοποιούν για τον διαχωρισμό των δεδομένων μπορεί να είναι πολύπλοκες και δύσκολο να ερμηνευθούν, καθιστώντας δύσκολη την εξαγωγή σαφών συμπερασμάτων από τα μοντέλα (Guyon, Weston, Barnhill, & Vapnik, 2002).

Επιπλέον, η εκπαίδευση των SVMs μπορεί να είναι υπολογιστικά δαπανηρή, ειδικά για μεγάλα σύνολα δεδομένων. Η πολυπλοκότητα της εκπαίδευσης αυξάνεται σημαντικά με τον αριθμό των

δειγμάτων και των διαστάσεων των δεδομένων, κάτι που μπορεί να οδηγήσει σε μεγάλους χρόνους εκπαίδευσης και υψηλή κατανάλωση πόρων (Bottou & Lin, 2007).

Τα SVMs μπορεί επίσης να είναι ευαίσθητα στην επιλογή των παραμέτρων και των πυρηνικών συναρτήσεων. Η απόδοση των SVMs μπορεί να ποικίλλει σημαντικά ανάλογα με τις παραμέτρους που επιλέγονται, και η επιλογή της κατάλληλης πυρηνικής συνάρτησης μπορεί να απαιτεί πειραματισμό και εξειδικευμένες γνώσεις (Hsu, Chang, & Lin, 2003).

Ένα άλλο μειονέκτημα των SVMs είναι η έλλειψη ευελιξίας τους σε μη σταθμισμένα δεδομένα. Τα SVMs τείνουν να επικεντρώνονται στην εύρεση του βέλτιστου ορίου διαχωρισμού και μπορεί να μην αποδίδουν καλά όταν τα δεδομένα είναι έντονα μη ισορροπημένα, απαιτώντας επιπλέον τεχνικές για την αντιμετώπιση αυτών των περιπτώσεων (Akbari, Kwek, & Jarokowicz, 2004).

Τέλος, τα SVMs μπορεί να έχουν περιορισμούς σε προβλήματα που απαιτούν πιθανές προβλέψεις ή εκτιμήσεις. Ενώ είναι εξαιρετικά αποδοτικά σε προβλήματα ταξινόμησης και παλινδρόμησης, η επέκτασή τους σε εφαρμογές όπου απαιτούνται πιθανές προβλέψεις μπορεί να είναι πιο δύσκολη και να απαιτεί πρόσθετες προσαρμογές (Platt, 1999).

Πίνακας 5: Συγκεντρωτικός Πίνακας Πλεονεκτημάτων και Μειονεκτημάτων για τα SVMs

Πλεονεκτήματα των SVMs	Μειονεκτήματα των SVMs
Ικανότητα διαχείρισης δεδομένων υψηλής διάστασης	Δυσκολία στην ερμηνεία των αποτελεσμάτων
Εύρεση βέλτιστου ορίου διαχωρισμού	Υπολογιστική δαπάνη στην εκπαίδευση
Ευελιξία μέσω πυρηνικών συναρτήσεων	Ευαισθησία στην επιλογή παραμέτρων και πυρηνικών συναρτήσεων
Υψηλή ακρίβεια σε ταξινόμηση και παλινδρόμηση	Περιορισμένη απόδοση σε μη σταθμισμένα δεδομένα
Ανθεκτικότητα σε υπερπροσαρμογή	Περιορισμοί σε προβλήματα πιθανών προβλέψεων

Boosting

Τα μοντέλα ενίσχυσης δέντρων απόφασης (boosting) είναι μια ισχυρή τεχνική μηχανικής μάθησης που έχει αποδειχθεί εξαιρετικά αποτελεσματική σε ποικίλες εφαρμογές. Το boosting συνδυάζει αδύναμα μοντέλα για να δημιουργήσει ένα ισχυρό μοντέλο με υψηλή απόδοση. Υπάρχουν διάφορες μέθοδοι boosting, όπως το AdaBoost και το XGBoost, που προσφέρουν σημαντικά πλεονεκτήματα αλλά και κάποια μειονεκτήματα.

Ένα από τα κύρια πλεονεκτήματα του boosting είναι η ικανότητά του να βελτιώνει συνεχώς την απόδοση του μοντέλου. Το boosting λειτουργεί εκπαιδεύοντας διαδοχικά μια σειρά από αδύναμα

μοντέλα, κάθε ένα από τα οποία επικεντρώνεται στα σφάλματα των προηγούμενων μοντέλων. Αυτή η διαδικασία επιτρέπει στο σύστημα να μαθαίνει και να διορθώνει συνεχώς τα λάθη του, οδηγώντας σε ένα ισχυρότερο τελικό μοντέλο (Chen & Guestrin, 2016).

Τα μοντέλα boosting είναι επίσης γνωστά για την υψηλή τους ακρίβεια. Χρησιμοποιούνται ευρέως σε διαγωνισμούς μηχανικής μάθησης και σε εφαρμογές όπου η ακρίβεια είναι κρίσιμη, όπως στην ανάλυση χρηματοοικονομικών δεδομένων, στην ανίχνευση απάτης και στην πρόβλεψη ασθενειών. Η ικανότητα των μοντέλων boosting να ενσωματώνουν πολλά αδύναμα μοντέλα σε ένα ισχυρό μοντέλο καθιστά δυνατή την επίτευξη υψηλών επιπέδων ακρίβειας (Natekin & Knoll, 2013).

Ένα άλλο σημαντικό πλεονέκτημα των μοντέλων boosting είναι η ευελιξία τους. Τα μοντέλα boosting μπορούν να χρησιμοποιηθούν για ποικίλα προβλήματα μηχανικής μάθησης, όπως η ταξινόμηση, η παλινδρόμηση και η ανάλυση χρονικών σειρών. Επιπλέον, μπορούν να ενσωματώσουν διαφορετικούς τύπους δεδομένων και να προσαρμόσουν τη λειτουργία τους ανάλογα με τις ανάγκες της εφαρμογής (Friedman, 2001).

Ωστόσο, τα μοντέλα boosting έχουν και μειονεκτήματα. Ένα από τα βασικά μειονεκτήματα είναι η αυξημένη υπολογιστική πολυπλοκότητα. Η διαδικασία εκπαίδευσης των μοντέλων boosting μπορεί να είναι χρονοβόρα και απαιτεί σημαντικούς πόρους, ειδικά για μεγάλα σύνολα δεδομένων. Αυτό μπορεί να αποτελέσει πρόβλημα σε εφαρμογές όπου οι πόροι είναι περιορισμένοι ή απαιτούνται γρήγορες προβλέψεις (Bühlmann & Hothorn, 2007).

Τα μοντέλα boosting μπορεί επίσης να είναι ευαίσθητα σε υπερπροσαρμογή (overfitting), ειδικά όταν δεν εφαρμόζονται κατάλληλες τεχνικές κανονικοποίησης. Η διαδικασία ενίσχυσης μπορεί να επικεντρωθεί υπερβολικά στα σφάλματα των εκπαιδευτικών δεδομένων, οδηγώντας σε ένα μοντέλο που αποδίδει καλά στα δεδομένα εκπαίδευσης αλλά όχι στα νέα, άγνωστα δεδομένα (Raschka & Mirjalili, 2019).

Επιπλέον, η ανάπτυξη και η βελτιστοποίηση των μοντέλων boosting μπορεί να είναι πολύπλοκη. Απαιτούνται προσεκτικές ρυθμίσεις των υπερπαραμέτρων και επιλογή των κατάλληλων αδύναμων μοντέλων για να επιτευχθεί η βέλτιστη απόδοση. Αυτό μπορεί να απαιτεί εξειδικευμένες γνώσεις και εμπειρία, καθιστώντας τη διαδικασία δύσκολη για αρχάριους χρήστες (Zhang & Haghani, 2015).

Ένα άλλο μειονέκτημα των μοντέλων boosting είναι η δυσκολία στην ερμηνεία των αποτελεσμάτων. Τα μοντέλα αυτά μπορούν να γίνουν πολύπλοκα λόγω του μεγάλου αριθμού αδύναμων μοντέλων που συνδυάζονται, καθιστώντας δύσκολη την κατανόηση του τρόπου με τον οποίο το μοντέλο λαμβάνει αποφάσεις (Hastie, Tibshirani, & Friedman, 2009).

Συνολικά, τα μοντέλα boosting προσφέρουν σημαντικά πλεονεκτήματα σε όρους ακρίβειας και ευελιξίας, αλλά φέρουν και προκλήσεις όσον αφορά την υπολογιστική πολυπλοκότητα, την ευαισθησία σε υπερπροσαρμογή και τη δυσκολία στην ερμηνεία. Η σωστή εφαρμογή και

βελτιστοποίηση των μοντέλων αυτών απαιτεί εξειδικευμένες γνώσεις και προσεκτική διαχείριση των υπερπαραμέτρων.

Πίνακας 6: Συγκεντρωτικός Πίνακας Πλεονεκτημάτων και Μειονεκτημάτων για το Boosting

Πλεονεκτήματα των Boosting Models	Μειονεκτήματα των Boosting Models
Συνεχής βελτίωση απόδοσης	Αυξημένη υπολογιστική πολυπλοκότητα
Υψηλή ακρίβεια	Ευαισθησία σε υπερπροσαρμογή
Ευελιξία σε διάφορα προβλήματα	Πολυπλοκότητα ανάπτυξης και βελτιστοποίησης
Ανθεκτικότητα σε προβλήματα	Δυσκολία στην ερμηνεία των αποτελεσμάτων

Linear Regression

Η γραμμική παλινδρόμηση (Linear Regression) είναι μία από τις πιο βασικές και ευρέως χρησιμοποιούμενες μεθόδους στην ανάλυση δεδομένων και στη μηχανική μάθηση. Χρησιμοποιείται για να μοντελοποιήσει τη σχέση μεταξύ μίας εξαρτημένης μεταβλητής και μίας ή περισσότερων ανεξάρτητων μεταβλητών. Η απλότητα και η ευκολία εφαρμογής της γραμμικής παλινδρόμησης την καθιστούν ιδανική για πολλές πρακτικές εφαρμογές.

Ένα από τα μεγαλύτερα πλεονεκτήματα της γραμμικής παλινδρόμησης είναι η απλότητα και η ευκολία ερμηνείας της. Οι παράμετροι του μοντέλου είναι εύκολο να εξηγηθούν και να κατανοηθούν, καθιστώντας την ιδανική για καταστάσεις όπου η ερμηνευσιμότητα του μοντέλου είναι σημαντική (Hastie, Tibshirani, & Friedman, 2009). Επιπλέον, η γραμμική παλινδρόμηση έχει χαμηλές απαιτήσεις σε υπολογιστικούς πόρους, καθιστώντας την κατάλληλη για μεγάλες ποσότητες δεδομένων και γρήγορες προβλέψεις.

Ένα άλλο πλεονέκτημα είναι η ανθεκτικότητα του μοντέλου σε περιπτώσεις όπου οι εξαρτημένες και ανεξάρτητες μεταβλητές έχουν γραμμική σχέση. Σε τέτοιες περιπτώσεις, η γραμμική παλινδρόμηση μπορεί να προσφέρει ακριβείς και αξιόπιστες προβλέψεις (James et al., 2013). Επιπλέον, υπάρχουν πολλές διαθέσιμες παραλλαγές και επεκτάσεις της γραμμικής παλινδρόμησης, όπως η πολυωνυμική παλινδρόμηση και η παλινδρόμηση Ridge και Lasso, που επιτρέπουν την προσαρμογή σε πιο περίπλοκες σχέσεις μεταξύ των μεταβλητών.

Ωστόσο, η γραμμική παλινδρόμηση έχει και ορισμένα μειονεκτήματα. Ένα βασικό μειονέκτημα είναι ότι υποθέτει μια γραμμική σχέση μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών. Αν η πραγματική σχέση είναι μη γραμμική, το μοντέλο μπορεί να αποδώσει κακώς, παράγοντας ανακριβείς προβλέψεις (Montgomery, Peck, & Vining, 2012). Επιπλέον, η γραμμική παλινδρόμηση μπορεί να είναι ευαίσθητη σε outliers, τα οποία μπορούν να επηρεάσουν σημαντικά τις παραμέτρους του μοντέλου και την ακρίβεια των προβλέψεων.

Ένα άλλο μειονέκτημα είναι η υπόθεση της ομοσκεδαστικότητας (σταθερής διακύμανσης των υπολειμμάτων), η οποία μπορεί να μην ισχύει πάντα στα πραγματικά δεδομένα. Αν αυτή η υπόθεση

παραβιάζεται, το μοντέλο μπορεί να αποδώσει κακώς, οδηγώντας σε ανακριβείς εκτιμήσεις των σφαλμάτων και των παραμέτρων (Seber & Lee, 2003). Επιπλέον, η γραμμική παλινδρόμηση υποθέτει ότι οι ανεξάρτητες μεταβλητές είναι γραμμικά ανεξάρτητες μεταξύ τους, κάτι που μπορεί να μην ισχύει σε πραγματικές εφαρμογές.

Τέλος, η γραμμική παλινδρόμηση μπορεί να είναι περιορισμένη στην αντιμετώπιση πολύπλοκων προβλημάτων με πολυδιάστατες αλληλεπιδράσεις και μη γραμμικές σχέσεις. Σε τέτοιες περιπτώσεις, πιο προηγμένες μέθοδοι μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα ή τα μοντέλα ενίσχυσης, μπορεί να αποδώσουν καλύτερα (Friedman, Hastie, & Tibshirani, 2001).

Πίνακας 7: Συγκεντρωτικός Πίνακας Πλεονεκτημάτων και Μειονεκτημάτων Γραμμικής Παλινδρόμησης

Πλεονεκτήματα	Μειονεκτήματα
<i>Απλότητα και Ερμηνευσιμότητα: Η γραμμική παλινδρόμηση είναι εύκολη στην κατανόηση και την εφαρμογή, και οι συντελεστές της μπορούν να ερμηνευθούν άμεσα.</i>	Υπόθεση Γραμμικότητας: Υποθέτει ότι υπάρχει μια γραμμική σχέση μεταξύ των μεταβλητών, κάτι που δεν ισχύει πάντα στην πραγματικότητα.
<i>Υπολογιστική Αποδοτικότητα: Είναι υπολογιστικά αποδοτική και δεν απαιτεί μεγάλους υπολογιστικούς πόρους.</i>	Ευαισθησία σε Ακραία Σημεία (Outliers): Τα ακραία σημεία μπορούν να επηρεάσουν σημαντικά τις προβλέψεις, οδηγώντας σε λανθασμένα αποτελέσματα.
<i>Ευελιξία: Μπορεί να χρησιμοποιηθεί τόσο για προβλήματα παλινδρόμησης όσο και για προβλήματα ταξινόμησης με κατάλληλες τροποποιήσεις.</i>	Πρόβλημα Υπερπροσαρμογής (Overfitting): Μπορεί να υπερπροσαρμόσει τα δεδομένα εκπαίδευσης όταν υπάρχουν πολλοί συντελεστές, οδηγώντας σε κακή γενίκευση.
<i>Καλή Εκκίνηση: Συχνά χρησιμοποιείται ως σημείο εκκίνησης πριν την εφαρμογή πιο πολύπλοκων μοντέλων.</i>	Ανάγκη Μεγάλων Συνόλων Δεδομένων: Απαιτεί μεγάλα σύνολα δεδομένων για να αποδώσει καλά και η έλλειψη δεδομένων μπορεί να οδηγήσει σε προβλήματα απόδοσης.
<i>Εφαρμογή σε Πολλά Πεδία: Χρησιμοποιείται ευρέως σε οικονομία, βιοϊατρική, κοινωνικές επιστήμες, και άλλους τομείς.</i>	Ευαισθησία σε Μη Γραμμικές Σχέσεις: Αν η σχέση μεταξύ των μεταβλητών είναι μη γραμμική, η γραμμική παλινδρόμηση μπορεί να αποτύχει να παράγει ακριβείς προβλέψεις.

ΚΕΦΑΛΑΙΟ 5 Προτεινόμενο Σύστημα και Πειραματική Αποτίμηση

Εισαγωγή

Η ανάπτυξη ενός συστήματος πρόβλεψης καιρού αποτελεί μια πολυσύνθετη διαδικασία που ενσωματώνει τη συλλογή, επεξεργασία, ανάλυση και πρόβλεψη μετεωρολογικών δεδομένων. Το σύστημα αυτό βασίζεται σε προηγμένες τεχνικές μηχανικής μάθησης και τεχνητής νοημοσύνης για την παροχή ακριβών και έγκαιρων προβλέψεων.

Η συλλογή δεδομένων είναι το πρώτο και θεμελιώδες βήμα. Δεδομένα συλλέγονται από μετεωρολογικούς σταθμούς, δορυφόρους και αισθητήρες IoT. Οι μετεωρολογικοί σταθμοί καταγράφουν πληροφορίες όπως η θερμοκρασία, η ταχύτητα και η κατεύθυνση του ανέμου, η υγρασία, η ατμοσφαιρική πίεση και η κατακρήμνιση. Δορυφόροι όπως της NOAA και του EUMETSAT παρέχουν δεδομένα για τη νεφοκάλυψη και τη θερμοκρασία της επιφάνειας της θάλασσας. Αισθητήρες IoT καταγράφουν δεδομένα σε πραγματικό χρόνο από διάφορα σημεία, προσφέροντας λεπτομερείς πληροφορίες για τις τοπικές καιρικές συνθήκες (Bauer, Thorpe, & Brunet, 2015).

Η επεξεργασία των συλλεγόμενων δεδομένων είναι κρίσιμη για την εξαγωγή χρήσιμων πληροφοριών. Τα δεδομένα καθαρίζονται από θόρυβο και σφάλματα, κανονικοποιούνται και αποθηκεύονται σε μορφή που είναι εύκολα προσβάσιμη από τα μοντέλα πρόβλεψης. Αυτό περιλαμβάνει την αφαίρεση ανωμαλιών, τη συμπλήρωση ελλειπόντων δεδομένων και τη μετατροπή τους σε κατάλληλη μορφή για επεξεργασία από αλγόριθμους μηχανικής μάθησης (Kalnay, 2002).

Η ανάλυση των δεδομένων πραγματοποιείται με τη χρήση στατιστικών και μαθηματικών μεθόδων για την εξαγωγή σημαντικών πληροφοριών. Οι τεχνικές μηχανικής μάθησης επιτρέπουν την ανάλυση μεγάλων όγκων δεδομένων και την αναγνώριση σύνθετων σχέσεων. Αυτές οι τεχνικές περιλαμβάνουν την αναγνώριση μοτίβων, την ανίχνευση τάσεων και την ανάλυση σχέσεων μεταξύ διαφορετικών μετεωρολογικών παραμέτρων (Goodfellow, Bengio, & Courville, 2016).

Η πρόβλεψη του καιρού γίνεται με τη χρήση εκπαιδευμένων μοντέλων μηχανικής μάθησης. Τα μοντέλα αυτά περιλαμβάνουν γραμμική παλινδρόμηση, XGBoost, AdaBoost και άλλα. Κάθε μοντέλο έχει πλεονεκτήματα και περιορισμούς. Η γραμμική παλινδρόμηση είναι απλή και γρήγορη αλλά περιορίζεται στην ανάλυση γραμμικών σχέσεων. Το XGBoost και το AdaBoost είναι πιο σύνθετα και μπορούν να αναγνωρίσουν μη γραμμικές σχέσεις στα δεδομένα, προσφέροντας πιο ακριβείς προβλέψεις (Schneider, Kaul, & Pressel, 2019).

Η αξιολόγηση των προβλέψεων είναι απαραίτητη για την κατανόηση της ακρίβειας και της αξιοπιστίας των μοντέλων. Διάφοροι δείκτες απόδοσης, όπως το μέσο απόλυτο σφάλμα (MAE) και το μέσο τετραγωνικό σφάλμα (MSE), χρησιμοποιούνται για την αξιολόγηση των προβλέψεων.

Βασιζόμενοι στα αποτελέσματα, τα μοντέλα βελτιστοποιούνται συνεχώς για να βελτιωθεί η απόδοσή τους, προσαρμόζοντας τις υπερπαραμέτρους, δοκιμάζοντας νέους αλγόριθμους και ενσωματώνοντας περισσότερα δεδομένα (Hastie, Tibshirani, & Friedman, 2009).

Περιγραφή του Συστήματος

Το σύστημα που έχει αναπτυχθεί για την πρόβλεψη καιρού χρησιμοποιεί προηγμένες τεχνικές μηχανικής μάθησης και δεδομένα από διάφορες πηγές, όπως μετεωρολογικούς σταθμούς, δορυφόρους και αισθητήρες IoT. Αυτά τα δεδομένα συλλέγονται και επεξεργάζονται μέσω API, όπως το OpenWeatherMap, επιτρέποντας την ενσωμάτωση πληροφοριών σχετικά με τη θερμοκρασία, την υγρασία, την ταχύτητα ανέμου και την πίεση. Τα δεδομένα αυτά επεξεργάζονται σε πραγματικό χρόνο και αποθηκεύονται σε κατάλληλες μορφές για περαιτέρω ανάλυση.

Κανονικά η τυπική διαδικασία θα όριζε το σύστημα να προχωρά στη φάση της προεπεξεργασίας, όπου τα δεδομένα καθαρίζονται και κανονικοποιούνται για να αφαιρεθούν οι ανωμαλίες και να αντιμετωπιστούν οι ελλείπουσες τιμές. Καθώς όμως γίνεται χρήση API τα δεδομένα δεν χρειάζονται κάποια επεξεργασία διότι είναι ήδη επεξεργασμένα και οι τιμές που λαμβάνονται είναι «καθαρές». Στη συνέχεια, τα δεδομένα εισάγονται σε μοντέλα μηχανικής μάθησης, όπως η γραμμική παλινδρόμηση, το XGBoost και το AdaBoost. Αυτά τα μοντέλα εκπαιδεύονται για να αναγνωρίζουν μοτίβα στα μετεωρολογικά δεδομένα και να κάνουν ακριβείς προβλέψεις για τις μελλοντικές συνθήκες καιρού. Οι τεχνικές μηχανικής μάθησης επιτρέπουν την ανάλυση ιστορικών δεδομένων και τη δημιουργία προβλέψεων που είναι ζωτικής σημασίας για την προετοιμασία και την αντίδραση σε καιρικά φαινόμενα.

Το σύστημα προσφέρει μια ολοκληρωμένη προσέγγιση στην πρόβλεψη καιρού, συνδυάζοντας τη συλλογή δεδομένων, την προεπεξεργασία και την ανάλυση μέσω μηχανικής μάθησης. Αυτό επιτρέπει την παροχή προβλέψεων υψηλής ακρίβειας, βελτιώνοντας την αξιοπιστία των μετεωρολογικών προβλέψεων. Η συνεχής βελτίωση των μοντέλων και η ενσωμάτωση περισσότερων πηγών δεδομένων μπορούν να ενισχύσουν περαιτέρω την ακρίβεια και την αποτελεσματικότητα του συστήματος, συμβάλλοντας στην καλύτερη κατανόηση και διαχείριση των καιρικών φαινομένων.

Συλλογή και Επεξεργασία Δεδομένων

Το Meteostat είναι μια ισχυρή βιβλιοθήκη της Python που παρέχει πρόσβαση σε ιστορικά και πραγματικού χρόνου μετεωρολογικά δεδομένα. Αυτή η βιβλιοθήκη συλλέγει και ενοποιεί δεδομένα από ποικίλες πηγές, συμπεριλαμβανομένων των εθνικών μετεωρολογικών υπηρεσιών, δορυφόρων, και ανεξάρτητων μετεωρολογικών σταθμών. Ο κύριος στόχος του Meteostat είναι να προσφέρει στους χρήστες εύκολη και γρήγορη πρόσβαση σε υψηλής ποιότητας μετεωρολογικά δεδομένα μέσω ενός απλού και ευέλικτου API (Meteostat, 2022).

Η συλλογή των δεδομένων γίνεται από διάφορους οργανισμούς και υπηρεσίες μετεωρολογίας παγκοσμίως, όπως η National Oceanic and Atmospheric Administration (NOAA), η Deutscher Wetterdienst (DWD), και η Environment Canada. Τα δεδομένα περιλαμβάνουν παρατηρήσεις από μετεωρολογικούς σταθμούς, δορυφορικές εικόνες, και άλλα συστήματα παρακολούθησης του καιρού. Αυτά τα δεδομένα συλλέγονται και επεξεργάζονται σε πραγματικό χρόνο για να προσφέρουν ακριβείς προβλέψεις και αναλύσεις (Meteostat, 2022).

Τα δεδομένα που παρέχει το Meteostat περιλαμβάνουν θερμοκρασία, υγρασία, ατμοσφαιρική πίεση, ταχύτητα και κατεύθυνση ανέμου, καθώς και κατακρημνίσεις. Αυτές οι πληροφορίες είναι διαθέσιμες σε διάφορες χρονικές κλίμακες, από ωριαίες έως ημερήσιες και μηνιαίες καταγραφές. Η ποιότητα και η ακρίβεια των δεδομένων είναι εξαιρετικά σημαντικές για την πρόβλεψη και την ανάλυση καιρικών συνθηκών, και το Meteostat εξασφαλίζει τη συνέπεια και την αξιοπιστία τους μέσω προσεκτικής επεξεργασίας και επικύρωσης.

Η σύνδεση με το Meteostat είναι εξαιρετικά απλή και γίνεται μέσω της Python. Ο χρήστης μπορεί να εγκαταστήσει τη βιβλιοθήκη μέσω του pip και να τη χρησιμοποιήσει για να αντλήσει δεδομένα από το API. Η βιβλιοθήκη παρέχει μια απλή και κατανοητή διεπαφή για την επιλογή της περιοχής ενδιαφέροντος και της χρονικής περιόδου, επιτρέποντας στους χρήστες να αντλούν δεδομένα με ελάχιστο κώδικα. Η τεκμηρίωση του Meteostat παρέχει αναλυτικές οδηγίες και παραδείγματα για τη χρήση του API, διευκολύνοντας την ενσωμάτωση των μετεωρολογικών δεδομένων σε διάφορες εφαρμογές και αναλύσεις.

Ο κώδικας της εφαρμογής που παρουσιάζεται αντλεί δεδομένα από τη βιβλιοθήκη Meteostat. Τα δεδομένα που αντλούνται, περιλαμβάνουν τη θερμοκρασία, την ταχύτητα και κατεύθυνση του ανέμου, την ατμοσφαιρική πίεση, και τις κατακρημνίσεις. Αυτά τα δεδομένα είναι κρίσιμα για την πρόβλεψη καιρού, καθώς παρέχουν μια λεπτομερή εικόνα των καιρικών συνθηκών σε πραγματικό χρόνο ή για συγκεκριμένες χρονικές περιόδους. Η συλλογή αυτών των δεδομένων γίνεται μέσω των συντεταγμένων και της χρονικής περιόδου που ορίζουμε στον κώδικα.

Η διαδικασία συλλογής δεδομένων ξεκινάει με την εισαγωγή των απαραίτητων βιβλιοθηκών, όπως η meteostat, και τον ορισμό της χρονικής περιόδου που μας ενδιαφέρει, η οποία είναι από 1 Ιανουαρίου του 2023 έως και 31 Δεκεμβρίου 2023.

```
1. #import libraries
2. from datetime import datetime
3. import warnings
4.
5. import matplotlib.pyplot as plt
6. from meteostat import Point, Daily
7.
8. warnings.filterwarnings("ignore")
```

```

9. # Ορισμός χρονικής περιόδου
10. start = datetime(2023, 1, 1) # Έναρξη χρονικής περιόδου
11. end = datetime(2023, 12, 31) # Λήξη χρονικής περιόδου

```

Πηγαίος Κώδικας 1: Ο παρακάτω κώδικας εισάγει τις απαραίτητες βιβλιοθήκες και ορίζει τη χρονική περίοδο για την οποία θα αντληθούν τα μετεωρολογικά δεδομένα χρησιμοποιώντας τη βιβλιοθήκη Meteostat. Συγκεκριμένα, καθορίζει την έναρξη και τη λήξη της περιόδου ως την 1η Ιανουαρίου 2023 και την 31η Δεκεμβρίου 2023 αντίστοιχα, και καταστέλλει τυχόν προειδοποιήσεις που μπορεί να εμφανιστούν κατά την εκτέλεση του κώδικα.

Στη συνέχεια, ο χρήστης επιλέγει την πόλη για την οποία θέλει να συλλέξει δεδομένα, και ο κώδικας χρησιμοποιεί τις συντεταγμένες αυτής της πόλης για να δημιουργήσει ένα σημείο ενδιαφέροντος (Point). Για λόγους ευκολίας χρησιμοποιήθηκαν 4 χαρακτηριστικές πόλεις, Αθήνα, Θεσσαλονίκη, Πάτρα και Βόλος. Ανάλογα με την επιλογή του χρήστη, η βιβλιοθήκη Meteostat αντλεί τα ημερήσια δεδομένα για την επιλεγμένη χρονική περίοδο χρησιμοποιώντας τη συνάρτηση Daily.

```

1. # Εκτύπωση επιλογών πόλεων
2. print("Please Press 1 For Athens:") # Επιλογή για Αθήνα
3. print("Please Press 2 For Thessaloniki:") # Επιλογή για Θεσσαλονίκη
4. print("Please Press 3 For Patras:") # Επιλογή για Πάτρα
5. print("Please Press 4 For Volos:") # Επιλογή για Βόλο
6.
7. # Λήψη επιλογής χρήστη
8. x = int(input("Please Give your choice:")) # Εισαγωγή επιλογής από τον χρήστη
9.
10. # Έλεγχος επιλογής και δημιουργία σημείου για την επιλεγμένη πόλη
11. if x == 1: # Αν η επιλογή είναι 1
12.     # Δημιουργία σημείου για Αθήνα
13.     location = Point(37.983810, 23.727539, 20) # Συντεταγμένες και ύψος για Αθήνα
14.     # Λήψη ημερήσιων δεδομένων
15.     data = Daily(location, start, end) # Λήψη δεδομένων για την Αθήνα
16.
17. elif x == 2: # Αν η επιλογή είναι 2
18.     # Δημιουργία σημείου για Θεσσαλονίκη
19.     location = Point(40.629269, 22.947412, 36) # Συντεταγμένες και ύψος για Θεσσαλονίκη
20.     # Λήψη ημερήσιων δεδομένων
21.     data = Daily(location, start, end) # Λήψη δεδομένων για τη Θεσσαλονίκη
22.
23. elif x == 3: # Αν η επιλογή είναι 3
24.     # Δημιουργία σημείου για Πάτρα
25.     location = Point(38.246639, 21.734573, 4) # Συντεταγμένες και ύψος για Πάτρα
26.     # Λήψη ημερήσιων δεδομένων
27.     data = Daily(location, start, end) # Λήψη δεδομένων για την Πάτρα
28.
29. elif x == 4: # Αν η επιλογή είναι 4
30.     # Δημιουργία σημείου για Βόλο
31.     location = Point(39.366669, 22.933332, 15) # Συντεταγμένες και ύψος για Βόλο

```

```
32. # Λήψη ημερήσιων δεδομένων
33. data = Daily(location, start, end) # Λήψη δεδομένων για το Βόλο
34. # Λήψη των δεδομένων καιρού από τη βιβλιοθήκη meteostat
35. data = data.fetch() # Μετατροπή των δεδομένων σε μορφή DataFrame
36.
```

Πηγαίος Κώδικας 2: Ο παραπάνω κώδικας εκτυπώνει επιλογές για τέσσερις ελληνικές πόλεις και λαμβάνει την επιλογή του χρήστη μέσω πληκτρολογίου. Με βάση την επιλογή, δημιουργείται ένα σημείο για την επιλεγμένη πόλη χρησιμοποιώντας τις συντεταγμένες της και αντλούνται ημερήσια μετεωρολογικά δεδομένα για την καθορισμένη χρονική περίοδο από τη βιβλιοθήκη Meteostat. Τα δεδομένα αυτά μετατρέπονται σε μορφή DataFrame για περαιτέρω ανάλυση και επεξεργασία.

Η βιβλιοθήκη Meteostat προσφέρει τα δεδομένα σε μορφή DataFrame, διευκολύνοντας την επεξεργασία και ανάλυση τους. Τα δεδομένα αυτά περιλαμβάνουν ακριβείς καταγραφές των μετεωρολογικών παραμέτρων για κάθε ημέρα της χρονικής περιόδου που έχει οριστεί. Οι παράμετροι αυτές είναι απαραίτητες για την ανάπτυξη μοντέλων πρόβλεψης καιρού, την εκτέλεση αναλύσεων και την εξαγωγή συμπερασμάτων σχετικά με τις καιρικές συνθήκες. Η απλότητα και η ευελιξία της βιβλιοθήκης Meteostat καθιστούν τη διαδικασία συλλογής δεδομένων ιδιαίτερα αποτελεσματική και εύκολη για τους χρήστες.

Εξερεύνηση Δεδομένων

Η Εξερεύνηση Δεδομένων (Exploratory Data Analysis - EDA) αποτελεί ένα κρίσιμο στάδιο στην ανάλυση και επεξεργασία δεδομένων. Κατά τη διάρκεια αυτού του σταδίου, οι ερευνητές συγκεντρώνουν, καθαρίζουν και αναλύουν τα δεδομένα, προκειμένου να αποκαλύψουν σημαντικά πρότυπα και τάσεις που κρύβονται μέσα τους. Η EDA περιλαμβάνει την εφαρμογή στατιστικών τεχνικών και τη δημιουργία γραφημάτων που βοηθούν στην κατανόηση των δεδομένων και στην εξαγωγή αρχικών συμπερασμάτων. Στην παρούσα μελέτη, η EDA χρησιμοποιείται για την ανάλυση των μετεωρολογικών δεδομένων από την πλατφόρμα Meteostat, με στόχο την εξαγωγή πολύτιμων πληροφοριών σχετικά με τις καιρικές συνθήκες και την προετοιμασία των δεδομένων για τα μοντέλα μηχανικής μάθησης που θα ακολουθήσουν.

Στατιστική Ανάλυση

Κατά την ανάλυση των μετεωρολογικών δεδομένων για την Αθήνα, προέκυψε ο παρακάτω πίνακας στατιστικής που παρέχει σημαντικές πληροφορίες για διάφορες μετεωρολογικές παραμέτρους. Ο πίνακας αυτός περιλαμβάνει τις εξής παραμέτρους: τη μέση θερμοκρασία (tavg), την ελάχιστη θερμοκρασία (tmin), τη μέγιστη θερμοκρασία (tmax), την κατακρήμνιση (prcp), το χιόνι (snow), την κατεύθυνση του ανέμου (wdir), την ταχύτητα του ανέμου (wspd), την υψηλότερη ταχύτητα του ανέμου (wpgt), την ατμοσφαιρική πίεση (pres) και τη διάρκεια ηλιοφάνειας (tsun). Αξιοσημείωτο είναι ότι ο πίνακας περιλαμβάνει πληροφορίες για 365 ημέρες, που αντιστοιχούν σε ένα πλήρες έτος.

	tavg	tmin	tmax	prcp	snow	wdir
count	365.0	365.0	365.0	365.0	2.0	365.0
mean	19.89698630136986	16.280547945205477	24.449863013698632	1.0950684931506847	10.0	174.32602739726028
std	6.864086079432264	6.4460593402180635	7.464979132102223	5.081536045376499	0.0	133.4039371498726
min	3.0	0.0	4.9	0.0	10.0	0.0
25%	14.3	11.0	18.7	0.0	10.0	34.0
50%	19.4	16.0	23.7	0.0	10.0	157.0
75%	24.8	20.9	30.0	0.0	10.0	315.0
max	34.4	30.5	41.6	83.1	10.0	359.0

Εικόνα 21: Απόσπασμα του πίνακα περιγραφικής στατιστικής για τα δεδομένα του συνόλου δεδομένων data

Για τη μέση θερμοκρασία (tavg), ο μέσος όρος ήταν 19,9°C με τυπική απόκλιση 6,86°C, που δείχνει σημαντική ποικιλία στις θερμοκρασίες καθ' όλη τη διάρκεια του έτους. Οι θερμοκρασίες κυμαίνονταν από 3°C έως 34,4°C, με τις τιμές του 25^{ου} (25% των παρατηρήσεων), 50^{ου} (50% των παρατηρήσεων) και 75^{ου} (75% των παρατηρήσεων) εκατοστημορίου να είναι 14,3°C, 19,4°C και 24,8°C αντίστοιχα. Αυτά τα στοιχεία υποδεικνύουν ένα σχετικά ήπιο κλίμα με κάποιες περιόδους υψηλών θερμοκρασιών.

Η κατακρήμνιση (prcp) είχε μέσο όρο 1,1 mm με τυπική απόκλιση 5,08 mm. Η μέγιστη τιμή κατακρήμνισης ήταν 83,1 mm, ενώ το 25ο, 50ο και 75ο εκατοστημόριο ήταν 0 mm, που υποδεικνύει ότι τις περισσότερες ημέρες δεν υπήρχαν καθόλου βροχοπτώσεις. Αυτό δείχνει έναν ξηρό κλιματικό χαρακτήρα με κάποιες σποραδικές έντονες βροχοπτώσεις.

Η ατμοσφαιρική πίεση (pres) είχε μέσο όρο 1014,6 hPa με τυπική απόκλιση 5,6 hPa. Οι τιμές της πίεσης κυμαίνονταν από 993,7 hPa έως 1033,3 hPa, με τις ενδιάμεσες τιμές (25ο, 50ο και 75ο εκατοστημόριο) να είναι 1011,1 hPa, 1014,0 hPa και 1016,9 hPa αντίστοιχα. Αυτές οι πληροφορίες είναι σημαντικές για την κατανόηση των καιρικών συνθηκών και την πρόβλεψη των καιρικών φαινομένων.

Ο κώδικας για την εξαγωγή των στατιστικών στοιχείων αποτελείται από την εντολή describe().

```
1. # Παροχή περιγραφικής στατιστικής των δεδομένων
2. data.describe() # Εμφάνιση στατιστικών μέτρων όπως μέσος όρος, τυπική απόκλιση, ελάχιστο και μέγιστο για κάθε στήλη δεδομένων
```

Πηγαίος Κώδικας 3: Κώδικας για την εξαγωγή στατιστικών στοιχείων

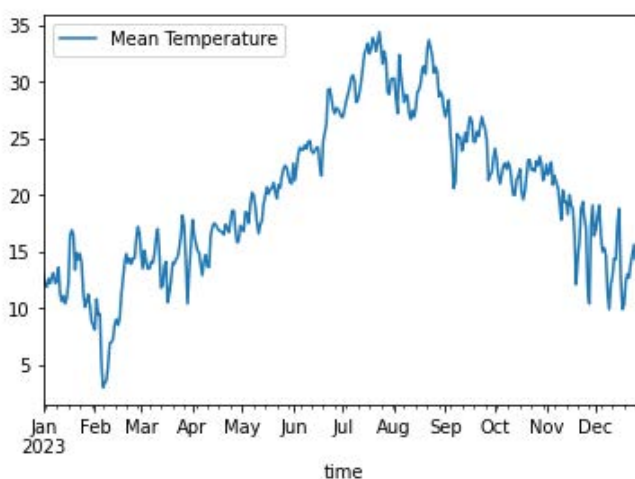
Γραφική Αναπαράσταση

Στη συνέχεια έγινε η γραφική απεικόνιση των διάφορων τιμών. Μέσω της γραφικής παράστασης, μπορούμε να κατανοήσουμε καλύτερα τις τάσεις, τις σχέσεις και τις διακυμάνσεις των μετεωρολογικών παραμέτρων με την πάροδο του χρόνου. Η οπτικοποίηση των δεδομένων διευκολύνει την ανίχνευση μοτίβων και ανωμαλιών, που μπορεί να μην είναι

προφανή μέσω απλών στατιστικών πινάκων. Επιπλέον, η χρήση γραφημάτων μας επιτρέπει να παρουσιάσουμε τα ευρήματά μας με τρόπο που είναι εύκολο να κατανοηθεί από ένα ευρύτερο κοινό, συμπεριλαμβανομένων και των μη ειδικών.

Στην ενότητα αυτή, θα παρουσιάσουμε και θα αναλύσουμε γραφήματα που απεικονίζουν διάφορες μετεωρολογικές παραμέτρους, όπως η θερμοκρασία, οι βροχοπτώσεις και η ατμοσφαιρική πίεση, για το έτος 2023. Θα χρησιμοποιήσουμε πραγματικά δεδομένα από τη Meteostat, τα οποία έχουν συλλεχθεί από μετεωρολογικούς σταθμούς στην Αθήνα, και θα τα απεικονίσουμε μέσω γραφημάτων για να κατανοήσουμε τις εποχιακές μεταβολές και τις κλιματικές τάσεις.

Το ακόλουθο γράφημα απεικονίζει τη μέση θερμοκρασία (avg) στην Αθήνα για το έτος 2023. Οι τιμές της μέσης θερμοκρασίας παρουσιάζονται στον κατακόρυφο άξονα (y), ενώ οι ημερομηνίες βρίσκονται στον οριζόντιο άξονα (x). Παρατηρούμε ότι η θερμοκρασία αυξάνεται σταδιακά από τον Ιανουάριο έως τον Αύγουστο, φτάνοντας σε κορύφωση κατά τους καλοκαιρινούς μήνες, και στη συνέχεια μειώνεται ξανά προς το τέλος του έτους. Αυτή η κυκλική τάση αντικατοπτρίζει τις εποχιακές αλλαγές που είναι χαρακτηριστικές του μεσογειακού κλίματος της Αθήνας.



Εικόνα 22: Γράφημα της μέσης θερμοκρασίας στην Αθήνα για το έτος 2023

Ο κώδικας που χρησιμοποιήθηκε για την εξαγωγή του παραπάνω γραφήματος είναι:

1. `# Σχεδίαση γραφήματος της μέσης θερμοκρασίας`
2. `data.plot(y=['avg'])` `# Σχεδίαση της μέσης θερμοκρασίας (avg) με την ημερομηνία στον άξονα x`
- 3.
4. `# Προσθήκη υπόμνημα (legend) στο γράφημα`

```

5. plt.legend(['Mean Temperature']) # Ορισμός του κειμένου του υπομνήματος ως 'Mean
   Temperature'
6.
7. # Εμφάνιση του γραφήματος
8. plt.show() # Εμφάνιση του γραφήματος στην οθόνη

```

Πηγαίος Κώδικας 4: Κώδικας Python για τη δημιουργία γραφήματος που απεικονίζει τη μέση θερμοκρασία στην Αθήνα για το έτος 2023

Στην περίπτωση της μέσης ταχύτητας ανέμου, ο κώδικας που χρησιμοποιούμε λαμβάνει τα ημερήσια δεδομένα για την ταχύτητα του ανέμου από τη βιβλιοθήκη Meteostat, και τα απεικονίζει σε ένα διάγραμμα χρονικής σειράς (χρονοσειρά). Το διάγραμμα παρουσιάζει τη μέση ταχύτητα του ανέμου (wspd) σε km/h για την χρονική περίοδο του 2023.

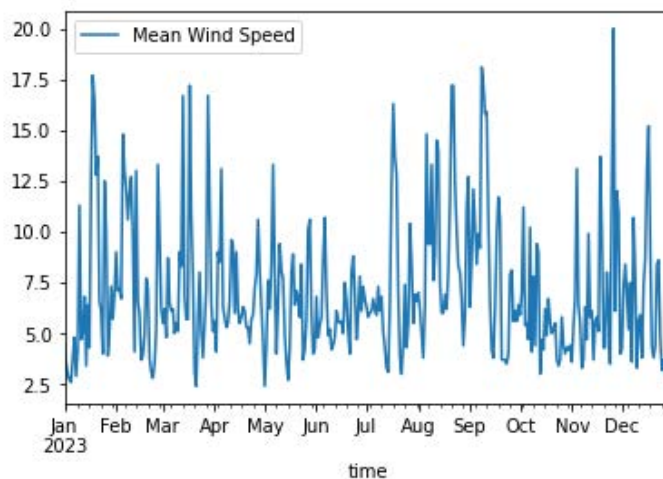
```

1. # Σχεδίαση γραφήματος της μέσης ταχύτητας ανέμου
2. data.plot(y=['wspd']) # Σχεδίαση της μέσης ταχύτητας ανέμου (wspd) με την ημερομηνία στον
   άξονα x
3.
4. # Προσθήκη υπόμνημα (legend) στο γράφημα
5. plt.legend(['Mean Wind Speed']) # Ορισμός του κειμένου του υπομνήματος ως 'Mean Wind
   Speed'
6.
7. # Εμφάνιση του γραφήματος
8. plt.show() # Εμφάνιση του γραφήματος στην οθόνη
9.

```

Πηγαίος Κώδικας 5: Κώδικας για την απεικόνιση της μέσης ταχύτητας ανέμου

Η ακόλουθη απεικόνιση αυτή δείχνει τις μεταβολές στην ταχύτητα του ανέμου κατά τη διάρκεια του έτους, επιτρέποντας την οπτική αναγνώριση των διακυμάνσεων και των τάσεων στον άνεμο. Το γράφημα παρουσιάζει αιχμές και χαμηλά σημεία, τα οποία υποδηλώνουν περιοχές με υψηλές ή χαμηλές ταχύτητες ανέμου.



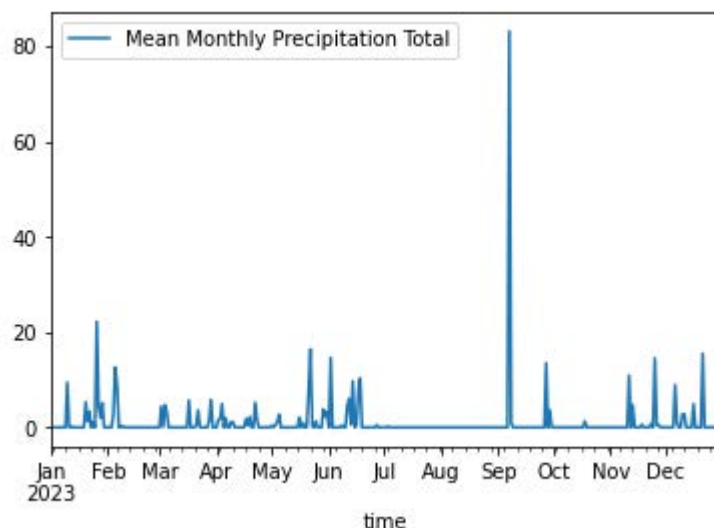
Εικόνα 23: Απεικόνιση της μέσης ταχύτητας ανέμου για την χρονική περίοδο του 2023

Συνεχίζοντας, δημιουργείται το διάγραμμα που απεικονίζει τη συνολική μηνιαία κατακρήμνιση για την πόλη της Αθήνας. Αυτό επιτυγχάνεται με την εντολή `data.plot(y=['prcp'])`, η οποία σχεδιάζει την κατακρήμνιση στον άξονα y σε συνάρτηση με τον χρόνο. Το υπόμνημα προστίθεται στο γράφημα για την καλύτερη κατανόηση των δεδομένων, και τελικά το γράφημα εμφανίζεται στην οθόνη με την εντολή `plt.show()`.

```
1. # Σχεδίαση γραφήματος της μέσης ατμοσφαιρικής πίεσης στη στάθμη της θάλασσας
2. data.plot(y=['pres']) # Σχεδίαση της μέσης ατμοσφαιρικής πίεσης (pres) με την ημερομηνία
   στον άξονα x
3.
4. # Προσθήκη υπόμνημα (legend) στο γράφημα
5. plt.legend(['Mean Sea Level Pressure']) # Ορισμός του κειμένου του υπομνήματος ως 'Mean
   Sea Level Pressure'
6.
7. # Εμφάνιση του γραφήματος
8. plt.show() # Εμφάνιση του γραφήματος στην οθόνη
9.
```

Εικόνα 24: Το τμήμα κώδικα δημιουργεί ένα γράφημα που απεικονίζει τη συνολική μηνιαία κατακρήμνιση (υγρασία) για την πόλη της Αθήνας κατά το έτος 2023

Το γράφημα της συνολικής μηνιαίας κατακρήμνισης παρουσιάζει την κατανομή της κατακρήμνισης κατά τη διάρκεια του έτους 2023. Παρατηρούμε ότι οι μεγαλύτερες τιμές κατακρήμνισης εμφανίζονται μεταξύ του Σεπτεμβρίου και Οκτωβρίου, ενώ κατά τους θερινούς μήνες η κατακρήμνιση είναι σχεδόν μηδενική, όπως είναι αναμενόμενο για το κλίμα της Αθήνας. Αυτό το γράφημα είναι ιδιαίτερα χρήσιμο για την κατανόηση των καιρικών συνθηκών και των μεταβολών στην κατακρήμνιση καθ' όλη τη διάρκεια του έτους.



Εικόνα 25: Γράφημα που δείχνει τη συνολική μηνιαία κατακρήμνιση στην Αθήνα για το 2023

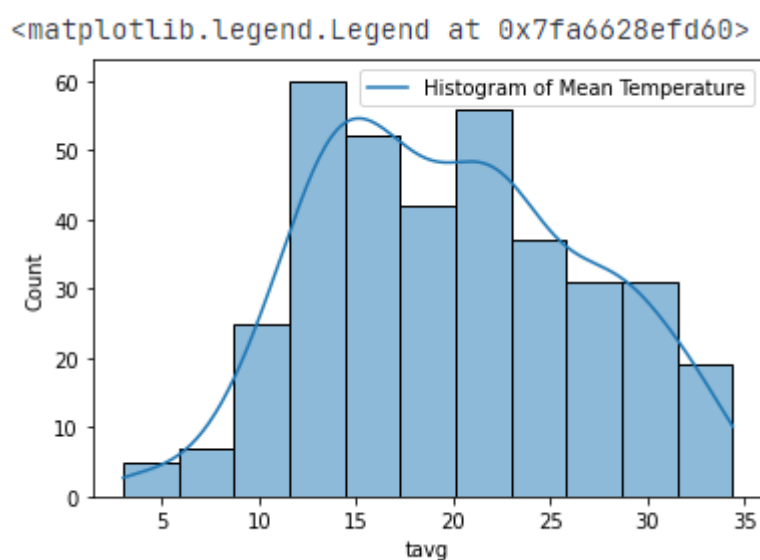
Ο ακόλουθος κώδικας εισάγει τη βιβλιοθήκη Seaborn για την οπτικοποίηση δεδομένων. Η Seaborn είναι γνωστή για τις ανώτερες γραφικές δυνατότητες που προσφέρει στην οπτικοποίηση

στατιστικών μοντέλων. Στη συνέχεια, δημιουργείται ένα ιστόγραμμα για την κατανομή της μέσης θερμοκρασίας (tavg) χρησιμοποιώντας την εντολή `sns.histplot`. Το ιστόγραμμα συνοδεύεται από μία καμπύλη πυκνότητας πυρήνα εκτίμησης (Kernel Density Estimator - KDE) για καλύτερη απεικόνιση της κατανομής των δεδομένων. Τέλος, προστίθεται ένα υπόμνημα στο γράφημα με την εντολή `plt.legend`, το οποίο ορίζει το κείμενο του υπομνήματος ως 'Histogram of Mean Temperature'.

```
1. # Εισαγωγή της βιβλιοθήκης Seaborn για την οπτικοποίηση δεδομένων
2. import seaborn as sns # Η βιβλιοθήκη Seaborn παρέχει ανώτερες γραφικές δυνατότητες για την
   οπτικοποίηση στατιστικών μοντέλων
3.
4. # Δημιουργία ιστογράμματος για την κατανομή της μέσης θερμοκρασίας
5. sns.histplot(data = data['tavg'], kde = True) # Δημιουργία ιστογράμματος για τη μέση
   θερμοκρασία (tavg) με πυκνότητα πυρήνα εκτίμησης (KDE)
6.
7. # Προσθήκη υπόμνημα (legend) στο γράφημα
8. plt.legend(['Histogram of Mean Temperature']) # Ορισμός του κειμένου του υπομνήματος ως
   'Histogram of Mean Temperature'
9.
```

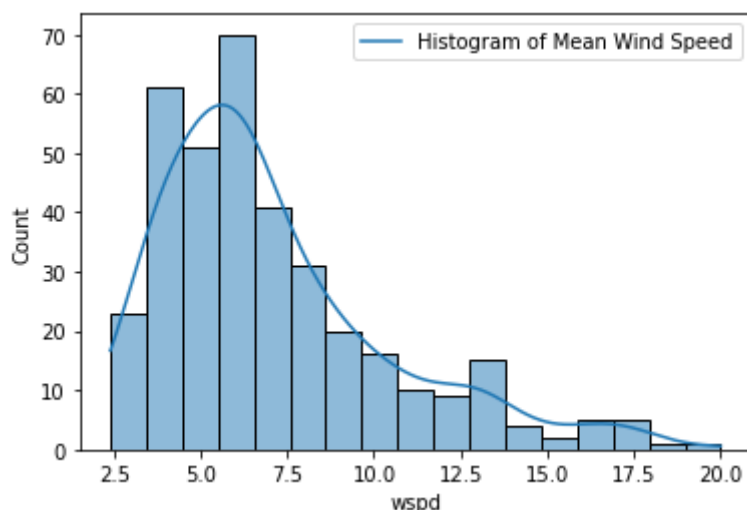
Πηγαίος Κώδικας 6: Κώδικας για τη δημιουργία ιστογράμματος της κατανομής της μέσης θερμοκρασίας στην Αθήνα

Το γράφημα απεικονίζει την κατανομή της μέσης θερμοκρασίας στην Αθήνα για το έτος 2023. Το ιστόγραμμα δείχνει την συχνότητα εμφάνισης διαφόρων τιμών μέσης θερμοκρασίας κατά τη διάρκεια του έτους. Η καμπύλη πυκνότητας πυρήνα εκτίμησης (KDE) προστίθεται για να δώσει μια πιο ομαλή απεικόνιση της κατανομής των θερμοκρασιών. Από το γράφημα μπορούμε να παρατηρήσουμε ότι η μέση θερμοκρασία στην Αθήνα κατά το έτος 2023 κυμαινόταν κυρίως μεταξύ 10°C και 30°C, με τις πιο συχνές τιμές να βρίσκονται γύρω στους 15-20°C.

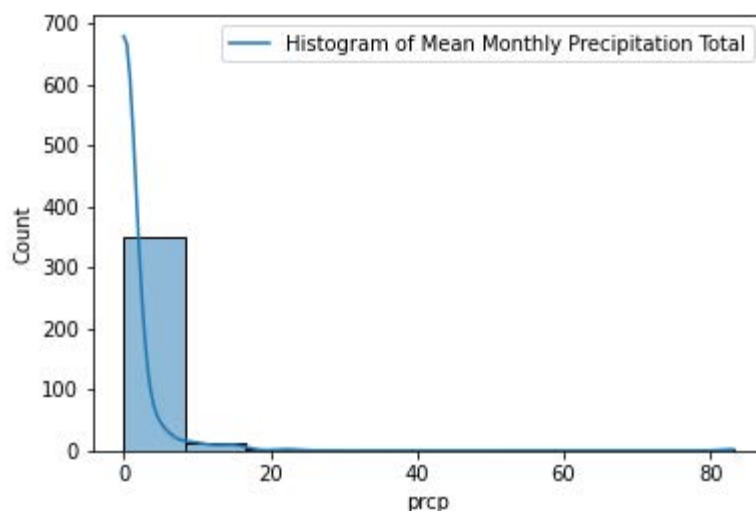


Εικόνα 26: Ιστόγραμμα της κατανομής της μέσης θερμοκρασίας στην Αθήνα για το έτος 2023 με καμπύλη πυκνότητας πυρήνα εκτίμησης (KDE)

Το ακόλουθο διάγραμμα παρουσιάζει το ιστόγραμμα της μέσης ταχύτητας ανέμου στην Αθήνα για το έτος 2023. Τα δεδομένα δείχνουν την κατανομή της ταχύτητας ανέμου με τιμές που κυμαίνονται από 2.5 έως 20 m/s. Η πλειονότητα των παρατηρήσεων συγκεντρώνεται σε ταχύτητες μεταξύ 5 και 7.5 m/s, με την κορύφωση να σημειώνεται κοντά στα 6 m/s. Η κατανομή εμφανίζει μια ασυμμετρία προς τα δεξιά, υποδεικνύοντας ότι οι υψηλότερες ταχύτητες ανέμου είναι λιγότερο συχνές.



Εικόνα 27: Ιστόγραμμα της Μέσης Ταχύτητας Ανέμου στην Αθήνα για το έτος 2023 (Ο κώδικας υπάρχει στο [Παράρτημα Α](#))



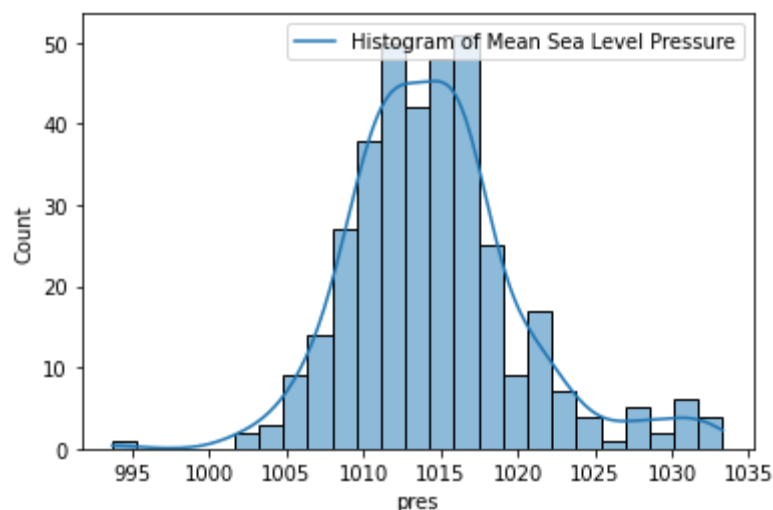
Εικόνα 28: Ιστόγραμμα των συνολικών μηνιαίων κατακρημνίσεων στην Αθήνα για το 2023 (Ο κώδικας υπάρχει στο [Παράρτημα Α](#))

Το παραπάνω ιστόγραμμα παρουσιάζει την κατανομή των συνολικών μηνιαίων κατακρημνίσεων στην Αθήνα για το έτος 2023. Παρατηρούμε ότι οι περισσότερες ημέρες έχουν πολύ χαμηλές κατακρημνίσεις, κοντά στο μηδέν, κάτι που υποδηλώνει ότι η Αθήνα βιώνει συχνά ξηρές περιόδους.

Ωστόσο, υπάρχουν κάποιες ημέρες με σημαντικές κατακρημνίσεις, όπως φαίνεται από τις λίγες υψηλές τιμές, οι οποίες όμως είναι σπάνιες.

Το παρακάτω γράφημα παρουσιάζει την κατανομή της μέσης ατμοσφαιρικής πίεσης για την Αθήνα κατά το έτος 2023. Το γράφημα είναι ένα ιστόγραμμα που απεικονίζει τη συχνότητα εμφάνισης διαφορετικών τιμών ατμοσφαιρικής πίεσης, με την πυκνότητα πυρήνα εκτίμησης (Kernel Density Estimator, KDE) να περιγράφει την κατανομή της πίεσης.

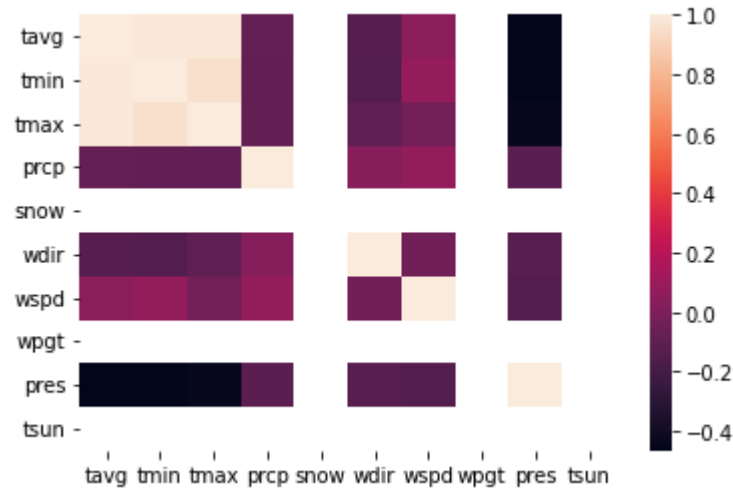
Συνολικά, η ατμοσφαιρική πίεση κυμαίνεται από περίπου 995 hPa έως 1035 hPa, με τις περισσότερες τιμές να συγκεντρώνονται γύρω από το μέσο όρο των 1015 hPa. Το ιστόγραμμα δείχνει μια κανονική κατανομή, με τις περισσότερες τιμές να βρίσκονται κοντά στο μέσο όρο και λιγότερες να απομακρύνονται σημαντικά από αυτόν. Η κανονική κατανομή της ατμοσφαιρικής πίεσης υποδηλώνει σταθερές και προβλέψιμες καιρικές συνθήκες για την περιοχή της Αθήνας.



Εικόνα 29: Ιστόγραμμα της Μέσης Ατμοσφαιρικής Πίεσης - Το γράφημα δείχνει την κατανομή της μέσης ατμοσφαιρικής πίεσης στην Αθήνα για το έτος 2023, παρουσιάζοντας μια κανονική κατανομή γύρω από το μέσο όρο των 1015 hPa (Ο κώδικας υπάρχει στο [Παράρτημα Α](#))

Στη συνέχεια ακολουθεί ο θερμικός χάρτης των δεδομένων. Ένας θερμικός χάρτης (heatmap) είναι μια γραφική αναπαράσταση δεδομένων όπου οι ατομικές τιμές ενός πίνακα απεικονίζονται ως χρώματα. Χρησιμοποιείται ευρέως για την οπτικοποίηση της μήτρας συσχέτισης μεταξύ διαφορετικών μεταβλητών, επιτρέποντας την εύκολη αναγνώριση προτύπων, σχέσεων και εξαρτήσεων στα δεδομένα.

Στη μετεωρολογία και την ανάλυση καιρού, οι θερμικοί χάρτες είναι ιδιαίτερα χρήσιμοι για την κατανόηση του πώς οι διάφορες κλιματικές παράμετροι, όπως η θερμοκρασία, η υγρασία, η ταχύτητα ανέμου και η ατμοσφαιρική πίεση, συνδέονται μεταξύ τους.



Εικόνα 30: Θερμικός χάρτης της μήτρας συσχέτισης των κλιματικών παραμέτρων
(Ο κώδικας υπάρχει στο [Παράρτημα Α](#))

Ο συγκεκριμένος θερμικός χάρτης παρουσιάζει τη μήτρα συσχέτισης των κλιματικών δεδομένων που έχουν συλλεχθεί. Οι τιμές συσχέτισης κυμαίνονται από -1 έως 1, όπου το 1 υποδεικνύει ισχυρή θετική συσχέτιση, το -1 ισχυρή αρνητική συσχέτιση και το 0 καμία συσχέτιση. Στον χάρτη αυτό, οι φωτεινότερες περιοχές υποδεικνύουν ισχυρές θετικές συσχετίσεις, ενώ οι σκοτεινότερες περιοχές υποδεικνύουν αρνητικές συσχετίσεις.

Για παράδειγμα, παρατηρούμε ότι η μέση θερμοκρασία (tavg) έχει ισχυρή θετική συσχέτιση με τη μέγιστη θερμοκρασία (tmax) και την ελάχιστη θερμοκρασία (tmin), όπως αναμενόταν. Η συνολική κατακρήμνιση (prcp) εμφανίζει χαμηλή συσχέτιση με άλλες μεταβλητές, εκτός από την αρνητική συσχέτιση με τη μέση θερμοκρασία, που μπορεί να αντανακλά τα καιρικά μοτίβα της περιοχής.

Εκπαίδευση και Δοκιμή Μοντέλων

Η ενότητα αυτή επικεντρώνεται στην εκπαίδευση και δοκιμή των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται για την πρόβλεψη καιρού. Οι αλγόριθμοι μηχανικής μάθησης έχουν αποδειχθεί ιδιαίτερα αποτελεσματικοί στην ανάλυση μεγάλων όγκων δεδομένων και στην αναγνώριση μοτίβων που μπορούν να βοηθήσουν στην ακριβή πρόβλεψη καιρικών συνθηκών.

Η διαδικασία εκπαίδευσης περιλαμβάνει την εφαρμογή αυτών των αλγορίθμων σε ιστορικά δεδομένα καιρού, ενώ η δοκιμή επικεντρώνεται στην αξιολόγηση της ακρίβειας και της αποτελεσματικότητας των προβλέψεων των μοντέλων αυτών σε δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση. Στην ενότητα αυτή θα παρουσιαστούν τα βήματα εκπαίδευσης και δοκιμής για διάφορους αλγόριθμους, αναλύοντας τον κώδικα, την προσέγγιση και τα αποτελέσματα για κάθε μοντέλο.

Γραμμική Παλινδρόμηση (Linear Regression)

Η γραμμική παλινδρόμηση είναι ένας από τους βασικότερους αλγόριθμους μηχανικής μάθησης που χρησιμοποιείται για την πρόβλεψη συνεχών τιμών. Η βασική ιδέα πίσω από τη γραμμική παλινδρόμηση είναι η εξεύρεση της καλύτερης γραμμής που προσαρμόζεται στα δεδομένα μας, χρησιμοποιώντας μια γραμμική εξίσωση της μορφής:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Όπου:

- y είναι η εξαρτημένη μεταβλητή, δηλαδή η μεταβλητή που θέλουμε να προβλέψουμε,
- x_1, x_2, x_n είναι οι ανεξάρτητες μεταβλητές, και οι
- $\beta_0, \beta_1, \beta_n$ είναι συντελεστές της παλινδρόμησης

Η εκπαίδευση του μοντέλου περιλαμβάνει την εύρεση των τιμών των συντελεστών αυτών που ελαχιστοποιούν το σφάλμα πρόβλεψης (Goodfellow, Bengio, & Courville, 2016).

Η εφαρμογή της γραμμικής παλινδρόμησης στον κώδικα περιλαμβάνει την πρόβλεψη τεσσάρων διαφορετικών μετεωρολογικών παραμέτρων: τη μέση θερμοκρασία, την ταχύτητα ανέμου, την κατακρήμνιση και την ατμοσφαιρική πίεση. Η διαδικασία ξεκινά με τον διαχωρισμό των δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ, χρησιμοποιώντας τη συνάρτηση `train_test_split` από τη βιβλιοθήκη `scikit-learn`. Ο διαχωρισμός αυτός γίνεται για να εξασφαλίσουμε ότι το μοντέλο θα εκπαιδευτεί σε ένα υποσύνολο των δεδομένων και θα αξιολογηθεί σε ένα διαφορετικό υποσύνολο, επιτρέποντας μια αντικειμενική εκτίμηση της απόδοσής του.

Στη συνέχεια, το μοντέλο γραμμικής παλινδρόμησης εκπαιδεύεται χρησιμοποιώντας τη συνάρτηση `fit` της κλάσης `LinearRegression`. Κατά την εκπαίδευση, το μοντέλο μαθαίνει τις σχέσεις μεταξύ των ανεξάρτητων μεταβλητών (όπως η ταχύτητα ανέμου, η κατακρήμνιση και η πίεση) και της εξαρτημένης μεταβλητής (όπως η μέση θερμοκρασία). Μετά την εκπαίδευση, το μοντέλο χρησιμοποιείται για να κάνει προβλέψεις στο δοκιμαστικό σετ, και η ακρίβεια αυτών των προβλέψεων αξιολογείται με τη χρήση του μέσου απόλυτου σφάλματος (Mean Absolute Error - MAE).

Για την πρόβλεψη της μέσης θερμοκρασίας, οι ανεξάρτητες μεταβλητές είναι η ταχύτητα ανέμου (`wspd`), η κατακρήμνιση (`prcp`) και η πίεση (`pres`), ενώ η εξαρτημένη μεταβλητή είναι η μέση θερμοκρασία (`tavg`). Με παρόμοιο τρόπο, για την πρόβλεψη της ταχύτητας ανέμου, οι ανεξάρτητες μεταβλητές είναι η μέση θερμοκρασία (`tavg`), η κατακρήμνιση (`prcp`) και η πίεση (`pres`), και η εξαρτημένη μεταβλητή είναι η ταχύτητα ανέμου (`wspd`). Η ίδια διαδικασία εφαρμόζεται και για την πρόβλεψη της κατακρήμνισης και της ατμοσφαιρικής πίεσης.

Η αξιολόγηση της απόδοσης του μοντέλου γραμμικής παλινδρόμησης παρέχει σημαντικά συμπεράσματα για την ικανότητα του μοντέλου να προβλέπει μετεωρολογικές παραμέτρους. Συγκεκριμένα, το μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE) των προβλέψεων δίνει μια

μέτρηση της απόκλισης των προβλεπόμενων τιμών από τις πραγματικές τιμές. Για την πρόβλεψη της μέσης θερμοκρασίας, το MAE είναι 5.398, για την ταχύτητα ανέμου είναι 2.995, για την κατακρήμνιση είναι 1.691, και για την ατμοσφαιρική πίεση είναι 3.918. Αυτά τα αποτελέσματα υποδεικνύουν την ακρίβεια του μοντέλου για τις διαφορετικές μετεωρολογικές παραμέτρους.

Η σημασία αυτών των αποτελεσμάτων είναι πολλαπλή. Πρώτον, η χαμηλή τιμή του MAE για την κατακρήμνιση (1.691) δείχνει ότι το μοντέλο μπορεί να προβλέψει την ποσότητα κατακρήμνισης με αρκετή ακρίβεια, κάτι που είναι κρίσιμο για την πρόβλεψη ακραίων καιρικών φαινομένων όπως οι καταιγίδες και οι πλημμύρες. Δεύτερον, η τιμή του MAE για την ταχύτητα ανέμου (2.995) υποδεικνύει ότι το μοντέλο είναι ικανό να προβλέψει με σχετική ακρίβεια τις μεταβολές της ταχύτητας του ανέμου, κάτι που είναι σημαντικό για τις αεροπορικές μεταφορές και τις ανανεώσιμες πηγές ενέργειας, όπως οι ανεμογεννήτριες.

Αν και το MAE για την μέση θερμοκρασία (5.398) και την ατμοσφαιρική πίεση (3.918) είναι υψηλότερο, αυτά τα αποτελέσματα εξακολουθούν να παρέχουν χρήσιμες προβλέψεις που μπορούν να ενσωματωθούν σε μεγαλύτερα συστήματα πρόβλεψης καιρού. Η ακρίβεια των προβλέψεων αυτών δείχνει ότι η γραμμική παλινδρόμηση μπορεί να χρησιμοποιηθεί ως ένα πρώτο βήμα για την ανάλυση και την πρόβλεψη των μετεωρολογικών συνθηκών, αν και μπορεί να απαιτούνται πιο προηγμένα μοντέλα για μεγαλύτερη ακρίβεια.

```
1. # Εισαγωγή απαραίτητων βιβλιοθηκών από το scikit-learn
2. from sklearn.model_selection import train_test_split # Εισαγωγή της συνάρτησης για τη
διαίρεση των δεδομένων σε εκπαιδευτικά και δοκιμαστικά σετ
3. from sklearn.metrics import mean_absolute_error # Εισαγωγή της συνάρτησης για τον
υπολογισμό του μέσου απόλυτου σφάλματος
4. from sklearn.linear_model import LinearRegression # Εισαγωγή της κλάσης για την εκτέλεση
γραμμικής παλινδρόμησης
5.
6. # Πρόβλεψη της μέσης θερμοκρασίας βασισμένη στην πίεση, την ταχύτητα ανέμου και την
κατακρήμνιση
7. X_temp = data[['wspd', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
8. y_temp = data[['tavg']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
9.
10. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
11. X_train_temp, X_test_temp, y_train_temp, y_test_temp = train_test_split(X_temp, y_temp,
test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
12.
13. # Εφαρμογή γραμμικής παλινδρόμησης
14. reg_temp = LinearRegression().fit(X_train_temp, y_train_temp) # Εκπαίδευση του μοντέλου
γραμμικής παλινδρόμησης
15.
16. # Προβλέψεις
```

```

17. y_pred_temp = reg_temp.predict(X_test_temp) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
18.
19. # Υπολογισμός σφάλματος
20. error_temp = mean_absolute_error(y_test_temp, y_pred_temp) # Υπολογισμός του μέσου
απόλυτου σφάλματος
21.
22. # Εκτύπωση του σφάλματος
23. print("Mean absolute error in predictions of mean temperature - Linear Regression:" +
str(error_temp)) # Εκτύπωση του μέσου απόλυτου σφάλματος
24.
25. # Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμνιση
26. X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
27. y_wind = data[['wspd']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
28.
29. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
30. X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_wind,
test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
31.
32. # Εφαρμογή γραμμικής παλινδρόμησης
33. reg_wind = LinearRegression().fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου
γραμμικής παλινδρόμησης
34.
35. # Προβλέψεις
36. y_pred_wind = reg_wind.predict(X_test_wind) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
37.
38. # Υπολογισμός σφάλματος
39. error_wind = mean_absolute_error(y_test_wind, y_pred_wind) # Υπολογισμός του μέσου
απόλυτου σφάλματος
40.
41. # Εκτύπωση του σφάλματος
42. print("Mean absolute error in predictions of wind speed - Linear Regression:" +
str(error_wind)) # Εκτύπωση του μέσου απόλυτου σφάλματος
43.
44. # Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την πίεση
45. X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
46. y_precip = data[['prcp']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
47.
48. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
49. X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_split(X_precip,
y_precip, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
50.
51. # Εφαρμογή γραμμικής παλινδρόμησης

```

```

52. reg_precip = LinearRegression().fit(X_train_precip, y_train_precip) # Εκπαίδευση του
μοντέλου γραμμικής παλινδρόμησης
53.
54. # Προβλέψεις
55. y_pred_precip = reg_precip.predict(X_test_precip) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
56.
57. # Υπολογισμός σφάλματος
58. error_precip = mean_absolute_error(y_test_precip, y_pred_precip) # Υπολογισμός του μέσου
απόλυτου σφάλματος
59.
60. # Εκτύπωση του σφάλματος
61. print("Mean absolute error in predictions of Monthly Precipitation total - Linear
Regression:" + str(error_precip)) # Εκτύπωση του μέσου απόλυτου σφάλματος
62.
63. # Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την
κατακρήμνιση
64. X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
65. y_pressure = data[['pres']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
66.
67. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
68. X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure =
train_test_split(X_pressure, y_pressure, test_size=0.33, random_state=42) # Διαίρεση
δεδομένων με το 33% για δοκιμαστικό σετ
69.
70. # Εφαρμογή γραμμικής παλινδρόμησης
71. reg_pressure = LinearRegression().fit(X_train_pressure, y_train_pressure) # Εκπαίδευση
του μοντέλου γραμμικής παλινδρόμησης
72.
73. # Προβλέψεις
74. y_pred_pressure = reg_pressure.predict(X_test_pressure) # Χρήση του εκπαιδευμένου
μοντέλου για προβλέψεις στο δοκιμαστικό σετ
75.
76. # Υπολογισμός σφάλματος
77. error_pressure = mean_absolute_error(y_test_pressure, y_pred_pressure) # Υπολογισμός του
μέσου απόλυτου σφάλματος
78.
79. # Εκτύπωση του σφάλματος
80. print("Mean absolute error in predictions of pressure - Linear Regression:" +
str(error_pressure)) # Εκτύπωση του μέσου απόλυτου σφάλματος
81.

```

```

Mean absolute error in predictions of mean temperature - Linear Regression:5.398268648600534
Mean absolute error in predictions of wind speed - Linear Regression:2.995384113513937
Mean absolute error in predictions of Monthly Precipitation total - Linear
Regression:1.6906130646703028

```

XGBoost (Extreme Gradient Boosting)

Ο αλγόριθμος XGBoost (Extreme Gradient Boosting) είναι ένας από τους πιο δημοφιλείς και αποδοτικούς αλγορίθμους ενίσχυσης (boosting) που χρησιμοποιείται ευρέως στη μηχανική μάθηση. Ο XGBoost βελτιώνει τα παραδοσιακά μοντέλα ενίσχυσης εφαρμόζοντας διάφορες τεχνικές βελτιστοποίησης, όπως η χρήση παραλληλισμού και βελτιωμένων αλγορίθμων για τη διαίρεση των δεδομένων. Η κύρια ιδέα πίσω από τον XGBoost είναι να ενισχύσει τα απλά μοντέλα (όπως τα δέντρα απόφασης) για να δημιουργήσει ένα πιο ισχυρό και ακριβές σύνολο μοντέλων που μπορεί να διαχειριστεί πολύπλοκα δεδομένα και να βελτιώσει την ακρίβεια των προβλέψεων.

Ο XGBoost χρησιμοποιεί την τεχνική gradient boosting, όπου τα επόμενα μοντέλα διορθώνουν τα λάθη των προηγούμενων μοντέλων. Η διαδικασία αυτή συνεχίζεται έως ότου τα σφάλματα μειωθούν στο ελάχιστο ή έως ότου φτάσουμε στον καθορισμένο αριθμό μοντέλων. Οι δυνατότητες βελτιστοποίησης και ο παραλληλισμός που προσφέρει ο XGBoost τον καθιστούν ιδανικό για εφαρμογές σε μεγάλα σύνολα δεδομένων και σε προβλήματα που απαιτούν υψηλή ακρίβεια, όπως είναι η πρόβλεψη καιρού.

Ο κώδικας για την εφαρμογή του αλγορίθμου XGBoost ξεκινά με την εισαγωγή των απαραίτητων βιβλιοθηκών. Η βιβλιοθήκη `xgboost` χρησιμοποιείται για την εκτέλεση της ενισχυμένης παλινδρόμησης, ενώ η `train_test_split` από το `scikit-learn` χρησιμοποιείται για τη διαίρεση των δεδομένων σε εκπαιδευτικά και δοκιμαστικά σύνολα, και η `mean_absolute_error` για τον υπολογισμό του μέσου απόλυτου σφάλματος. Αυτός ο αρχικός διαχωρισμός των δεδομένων είναι κρίσιμος, καθώς διασφαλίζει ότι το μοντέλο εκπαιδεύεται σε ένα υποσύνολο δεδομένων και δοκιμάζεται σε διαφορετικό υποσύνολο, προκειμένου να εκτιμηθεί η απόδοσή του.

Η διαδικασία ξεκινά με την πρόβλεψη της μέσης θερμοκρασίας (`tavg`) χρησιμοποιώντας ως χαρακτηριστικά την ταχύτητα ανέμου (`wspd`), την κατακρήμνιση (`prcp`) και την πίεση (`pres`). Τα δεδομένα χωρίζονται σε εκπαιδευτικά και δοκιμαστικά σετ με αναλογία 67% προς 33% αντίστοιχα. Το μοντέλο `XGBRegressor` εκπαιδεύεται στο εκπαιδευτικό σετ και χρησιμοποιείται για να προβλέψει τις τιμές στο δοκιμαστικό σετ. Το μέσο απόλυτο σφάλμα (MAE) υπολογίζεται για να εκτιμηθεί η ακρίβεια των προβλέψεων του μοντέλου. Η ίδια διαδικασία επαναλαμβάνεται για την πρόβλεψη της ταχύτητας ανέμου, της κατακρήμνισης και της ατμοσφαιρικής πίεσης, χρησιμοποιώντας τα αντίστοιχα χαρακτηριστικά και στόχους για κάθε πρόβλεψη.

Ο κώδικας περιλαμβάνει επίσης την εκπαίδευση του μοντέλου XGBoost για την πρόβλεψη της ταχύτητας ανέμου (`wspd`) με βάση τη μέση θερμοκρασία (`tavg`), την κατακρήμνιση (`prcp`) και την

πίεση (pres). Όπως και με την πρόβλεψη της θερμοκρασίας, τα δεδομένα χωρίζονται σε εκπαιδευτικά και δοκιμαστικά σετ, και το μοντέλο εκπαιδεύεται στο εκπαιδευτικό σετ. Οι προβλέψεις γίνονται στο δοκιμαστικό σετ και υπολογίζεται το μέσο απόλυτο σφάλμα για να εκτιμηθεί η ακρίβεια των προβλέψεων. Αυτή η διαδικασία εφαρμόζεται επίσης για την πρόβλεψη της κατακρήμνισης (prcp) και της ατμοσφαιρικής πίεσης (pres), χρησιμοποιώντας τα αντίστοιχα χαρακτηριστικά για κάθε περίπτωση.

Τέλος, τα αποτελέσματα των προβλέψεων παρουσιάζονται με τη μορφή μέσου απόλυτου σφάλματος για κάθε μετεωρολογική παράμετρο. Τα αποτελέσματα δείχνουν ότι το μέσο απόλυτο σφάλμα για τις προβλέψεις της μέσης θερμοκρασίας είναι 4.968, για την ταχύτητα ανέμου είναι 2.903, για την κατακρήμνιση είναι 1.358 και για την ατμοσφαιρική πίεση είναι 29.184. Αυτά τα αποτελέσματα υπογραμμίζουν την ικανότητα του αλγορίθμου XGBoost να παράγει σχετικά ακριβείς προβλέψεις για τις μετεωρολογικές παραμέτρους, αν και παρατηρείται μεγαλύτερο σφάλμα στην πρόβλεψη της ατμοσφαιρικής πίεσης, το οποίο μπορεί να υποδηλώνει την ανάγκη για βελτιστοποίηση του μοντέλου ή την επιλογή διαφορετικών χαρακτηριστικών για την πρόβλεψη αυτής της παραμέτρου.

```
1. # Εισαγωγή της βιβλιοθήκης XGBoost
2. import xgboost as xg # Εισαγωγή της βιβλιοθήκης XGBoost για την εκτέλεση ενισχυμένης
    παλινδρόμησης
3.
4. # Πρόβλεψη της μέσης θερμοκρασίας βασισμένη στην πίεση, την ταχύτητα ανέμου και την
    κατακρήμνιση
5. X_temp = data[['wspd', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
    μεταβλητές)
6. y_temp = data[['tavg']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
7.
8. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
9. X_train_temp, X_test_temp, y_train_temp, y_test_temp = train_test_split(X_temp, y_temp,
    test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
10.
11. # Εφαρμογή XGBoost
12. reg_temp = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10,
    seed=123).fit(X_train_temp, y_train_temp) # Εκπαίδευση του μοντέλου XGBoost με 10 δέντρα
13.
14. # Προβλέψεις
15. y_pred_temp = reg_temp.predict(X_test_temp) # Χρήση του εκπαιδευμένου μοντέλου για
    προβλέψεις στο δοκιμαστικό σετ
16.
17. # Υπολογισμός σφάλματος
18. error_temp = mean_absolute_error(y_test_temp, y_pred_temp) # Υπολογισμός του μέσου
    απόλυτου σφάλματος
19.
```

```

20. # Εκτύπωση του σφάλματος
21. print("Mean absolute error in predictions of mean temperature - XGBoost:" +
str(error_temp)) # Εκτύπωση του μέσου απόλυτου σφάλματος
22.
23. # Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμνιση
24. X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
25. y_wind = data[['wspd']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
26.
27. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
28. X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_wind,
test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
29.
30. # Εφαρμογή XGBoost
31. reg_wind = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10,
seed=123).fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου XGBoost με 10 δέντρα
32.
33. # Προβλέψεις
34. y_pred_wind = reg_wind.predict(X_test_wind) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
35.
36. # Υπολογισμός σφάλματος
37. error_wind = mean_absolute_error(y_test_wind, y_pred_wind) # Υπολογισμός του μέσου
απόλυτου σφάλματος
38.
39. # Εκτύπωση του σφάλματος
40. print("Mean absolute error in predictions of wind speed - XGBoost:" + str(error_wind)) #
Εκτύπωση του μέσου απόλυτου σφάλματος
41.
42. # Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την πίεση
43. X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
44. y_precip = data[['prcp']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
45.
46. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
47. X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_split(X_precip,
y_precip, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
48.
49. # Εφαρμογή XGBoost
50. reg_precip = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10,
seed=123).fit(X_train_precip, y_train_precip) # Εκπαίδευση του μοντέλου XGBoost με 10 δέντρα
51.
52. # Προβλέψεις
53. y_pred_precip = reg_precip.predict(X_test_precip) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
54.

```

```

55. # Υπολογισμός σφάλματος
56. error_precip = mean_absolute_error(y_test_precip, y_pred_precip) # Υπολογισμός του μέσου
    απόλυτου σφάλματος
57.
58. # Εκτύπωση του σφάλματος
59. print("Mean absolute error in predictions of Monthly Precipitation total - XGBoost:" +
    str(error_precip)) # Εκτύπωση του μέσου απόλυτου σφάλματος
60.
61. # Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την
    κατακρήμνιση
62. X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
    μεταβλητές)
63. y_pressure = data[['pres']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
64.
65. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
66. X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure =
    train_test_split(X_pressure, y_pressure, test_size=0.33, random_state=42) # Διαίρεση
    δεδομένων με το 33% για δοκιμαστικό σετ
67.
68. # Εφαρμογή XGBoost
69. reg_pressure = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10,
    seed=123).fit(X_train_pressure, y_train_pressure) # Εκπαίδευση του μοντέλου XGBoost με 10
    δέντρα
70.
71. # Προβλέψεις
72. y_pred_pressure = reg_pressure.predict(X_test_pressure) # Χρήση του εκπαιδευμένου
    μοντέλου για προβλέψεις στο δοκιμαστικό σετ
73.
74. # Υπολογισμός σφάλματος
75. error_pressure = mean_absolute_error(y_test_pressure, y_pred_pressure) # Υπολογισμός του
    μέσου απόλυτου σφάλματος
76.
77. # Εκτύπωση του σφάλματος
78. print("Mean absolute error in predictions of pressure - XGBoost:" +
    str(error_pressure)) # Εκτύπωση του μέσου απόλυτου σφάλματος

```

Mean absolute error in predictions of mean temperature -

XGBoost:4.968115572495893

Mean absolute error in predictions of wind speed - XGBoost:2.903117922317883

Mean absolute error in predictions of Monthly Precipitation total -

XGBoost:1.3577709051603375

Mean absolute error in predictions of pressure - XGBoost:29.184360371739412

AdaBoost (Adaptive Boosting)

Ο αλγόριθμος AdaBoost (Adaptive Boosting) είναι μια τεχνική ενισχυτικής μάθησης που συνδυάζει πολλαπλά ασθενή μοντέλα για να δημιουργήσει ένα ισχυρό μοντέλο. Αναπτύχθηκε από τον Yoan Freund και τον Robert Schapire το 1996 και αποτελεί έναν από τους πιο σημαντικούς αλγόριθμους ενισχυτικής μάθησης. Ο στόχος του AdaBoost είναι να βελτιώσει την απόδοση των ασθενών μοντέλων, τα οποία από μόνα τους δεν είναι ιδιαίτερα ακριβή, αλλά όταν συνδυάζονται μπορούν να παράγουν προβλέψεις υψηλής ακρίβειας.

Η διαδικασία του AdaBoost περιλαμβάνει την εκπαίδευση μιας σειράς ασθενών μοντέλων, συνήθως δέντρων απόφασης με μικρό βάθος, σε διαφορετικά υποσύνολα των δεδομένων εκπαίδευσης. Κάθε επόμενο μοντέλο επικεντρώνεται στα παραδείγματα που δεν έχουν προβλεφθεί σωστά από τα προηγούμενα μοντέλα. Αυτό επιτυγχάνεται με την προσαρμογή των βαρών των δεδομένων, δίνοντας μεγαλύτερη βαρύτητα στα λάθη. Τελικά, τα ασθενή μοντέλα συνδυάζονται με τρόπο που δίνει μεγαλύτερη έμφαση στα πιο ακριβή μοντέλα, δημιουργώντας ένα ενισχυμένο μοντέλο με καλύτερη συνολική απόδοση (Freund & Schapire, 1996).

Ο AdaBoost έχει εφαρμοστεί με επιτυχία σε διάφορα πεδία, όπως η αναγνώριση προτύπων, η ανάλυση εικόνας και η πρόβλεψη χρόνου. Ένα από τα πλεονεκτήματα του AdaBoost είναι η ικανότητά του να βελτιώνει την απόδοση των αδύναμων μοντέλων χωρίς να απαιτεί λεπτομερή ρύθμιση των υπερπαραμέτρων. Επίσης, μπορεί να συνδυαστεί με διαφορετικούς τύπους ασθενών μοντέλων, καθιστώντας τον έναν ευέλικτο και ισχυρό αλγόριθμο για την ενισχυτική μάθηση (Hastie, Tibshirani, & Friedman, 2009).

Ο αλγόριθμος AdaBoost είναι μια τεχνική ενισχυτικής μάθησης που χρησιμοποιείται για τη βελτίωση της απόδοσης των μοντέλων παλινδρόμησης. Στον κώδικα που δόθηκε, η εφαρμογή του AdaBoost ξεκινά με την εισαγωγή της βιβλιοθήκης AdaBoostRegressor από το scikit-learn. Αυτή η βιβλιοθήκη παρέχει εργαλεία για την εκτέλεση της ενισχυμένης παλινδρόμησης χρησιμοποιώντας τον αλγόριθμο AdaBoost.

Ο κώδικας συνεχίζεται με την πρόβλεψη της μέσης θερμοκρασίας (tavg) χρησιμοποιώντας χαρακτηριστικά όπως η ταχύτητα ανέμου (wsprd), η κατακρήμνιση (prcp) και η πίεση (pres). Τα δεδομένα χωρίζονται σε εκπαιδευτικά και δοκιμαστικά σετ με αναλογία 67% προς 33%. Στη συνέχεια, το μοντέλο AdaBoostRegressor εκπαιδεύεται στο εκπαιδευτικό σετ και χρησιμοποιείται για την πρόβλεψη των τιμών στο δοκιμαστικό σετ. Το μέσο απόλυτο σφάλμα (MAE) υπολογίζεται για την εκτίμηση της ακρίβειας των προβλέψεων του μοντέλου.

Η ίδια διαδικασία επαναλαμβάνεται για την πρόβλεψη της ταχύτητας ανέμου (wsprd), της κατακρήμνισης (prcp) και της ατμοσφαιρικής πίεσης (pres), χρησιμοποιώντας τα αντίστοιχα χαρακτηριστικά για κάθε περίπτωση. Για την πρόβλεψη της ταχύτητας ανέμου, τα δεδομένα

χωρίζονται και πάλι σε εκπαιδευτικά και δοκιμαστικά σετ, και το μοντέλο εκπαιδεύεται στο εκπαιδευτικό σετ. Οι προβλέψεις γίνονται στο δοκιμαστικό σετ και υπολογίζεται το μέσο απόλυτο σφάλμα για την εκτίμηση της ακρίβειας των προβλέψεων.

Για την πρόβλεψη της κατακρήμνισης, χρησιμοποιούνται χαρακτηριστικά όπως η ταχύτητα ανέμου (wsprd), η μέση θερμοκρασία (tavg) και η πίεση (pres). Τα δεδομένα χωρίζονται σε εκπαιδευτικά και δοκιμαστικά σετ, και το μοντέλο εκπαιδεύεται στο εκπαιδευτικό σετ. Οι προβλέψεις γίνονται στο δοκιμαστικό σετ και υπολογίζεται το μέσο απόλυτο σφάλμα. Η ίδια διαδικασία εφαρμόζεται και για την πρόβλεψη της ατμοσφαιρικής πίεσης.

Τέλος, τα αποτελέσματα των προβλέψεων παρουσιάζονται με τη μορφή μέσου απόλυτου σφάλματος για κάθε μετεωρολογική παράμετρο. Τα αποτελέσματα δείχνουν ότι το μέσο απόλυτο σφάλμα για τις προβλέψεις της μέσης θερμοκρασίας είναι 5.086, για την ταχύτητα ανέμου είναι 2.879, για την κατακρήμνιση είναι 1.682 και για την ατμοσφαιρική πίεση είναι 3.145. Αυτά τα αποτελέσματα υπογραμμίζουν την ικανότητα του αλγορίθμου AdaBoost να παράγει σχετικά ακριβείς προβλέψεις για τις μετεωρολογικές παραμέτρους, δείχνοντας τη δύναμη της ενισχυτικής μάθησης στη βελτίωση της ακρίβειας των μοντέλων παλινδρόμησης.

```
1. # Εισαγωγή της βιβλιοθήκης για τον ενισχυτή παλινδρόμησης AdaBoost
2. from sklearn.ensemble import AdaBoostRegressor # Εισαγωγή της κλάσης AdaBoostRegressor
για την εκτέλεση ενισχυμένης παλινδρόμησης
3.
4. # Πρόβλεψη της μέσης θερμοκρασίας βασισμένη στην πίεση, την ταχύτητα ανέμου και την
κατακρήμνιση
5. X_temp = data[['wsprd', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
6. y_temp = data['tavg'] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
7.
8. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
9. X_train_temp, X_test_temp, y_train_temp, y_test_temp = train_test_split(X_temp, y_temp,
test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
10.
11. # Εφαρμογή AdaBoost
12. reg_temp = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_temp,
y_train_temp) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα
13.
14. # Προβλέψεις
15. y_pred_temp = reg_temp.predict(X_test_temp) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
16.
17. # Υπολογισμός σφάλματος
18. error_temp = mean_absolute_error(y_test_temp, y_pred_temp) # Υπολογισμός του μέσου
απόλυτου σφάλματος
19.
```

```

20. # Εκτύπωση του σφάλματος
21. print("Mean absolute error in predictions of mean temperature - AdaBoost:" +
str(error_temp)) # Εκτύπωση του μέσου απόλυτου σφάλματος
22.
23. # Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμνιση
24. X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
25. y_wind = data[['wspd']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
26.
27. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
28. X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_wind,
test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
29.
30. # Εφαρμογή AdaBoost
31. reg_wind = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_wind,
y_train_wind) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα
32.
33. # Προβλέψεις
34. y_pred_wind = reg_wind.predict(X_test_wind) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
35.
36. # Υπολογισμός σφάλματος
37. error_wind = mean_absolute_error(y_test_wind, y_pred_wind) # Υπολογισμός του μέσου
απόλυτου σφάλματος
38.
39. # Εκτύπωση του σφάλματος
40. print("Mean absolute error in predictions of wind speed - AdaBoost:" + str(error_wind)) #
Εκτύπωση του μέσου απόλυτου σφάλματος
41.
42. # Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την πίεση
43. X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
44. y_precip = data[['prcp']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
45.
46. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
47. X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_split(X_precip,
y_precip, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
48.
49. # Εφαρμογή AdaBoost
50. reg_precip = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_precip,
y_train_precip) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα
51.
52. # Προβλέψεις
53. y_pred_precip = reg_precip.predict(X_test_precip) # Χρήση του εκπαιδευμένου μοντέλου για
προβλέψεις στο δοκιμαστικό σετ
54.

```

```

55. # Υπολογισμός σφάλματος
56. error_precip = mean_absolute_error(y_test_precip, y_pred_precip) # Υπολογισμός του μέσου
    απόλυτου σφάλματος
57.
58. # Εκτύπωση του σφάλματος
59. print("Mean absolute error in predictions of Monthly Precipitation total - AdaBoost:" +
    str(error_precip)) # Εκτύπωση του μέσου απόλυτου σφάλματος
60.
61. # Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την
    κατακρήμνιση
62. X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
    μεταβλητές)
63. y_pressure = data[['pres']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
64.
65. # Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
66. X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure =
    train_test_split(X_pressure, y_pressure, test_size=0.33, random_state=42) # Διαίρεση
    δεδομένων με το 33% για δοκιμαστικό σετ
67.
68. # Εφαρμογή AdaBoost
69. reg_pressure = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_pressure,
    y_train_pressure) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα
70.
71. # Προβλέψεις
72. y_pred_pressure = reg_pressure.predict(X_test_pressure) # Χρήση του εκπαιδευμένου
    μοντέλου για προβλέψεις στο δοκιμαστικό σετ
73.
74. # Υπολογισμός σφάλματος
75. error_pressure = mean_absolute_error(y_test_pressure, y_pred_pressure) # Υπολογισμός του
    μέσου απόλυτου σφάλματος
76.
77. # Εκτύπωση του σφάλματος
78. print("Mean absolute error in predictions of pressure - AdaBoost:" +
    str(error_pressure)) # Εκτύπωση του μέσου απόλυτου σφάλματος
79.

```

```

Mean absolute error in predictions of mean temperature -
AdaBoost:5.0858560549248235
Mean absolute error in predictions of wind speed -
AdaBoost:2.879222720818808
Mean absolute error in predictions of Monthly Precipitation total -
AdaBoost:1.6823025461273449
Mean absolute error in predictions of pressure - AdaBoost:3.1451020838958046

```

Μετα-μοντέλο

Ένα μεταμοντέλο, ή αλλιώς μετα-μάθηση, αναφέρεται σε μια τεχνική μηχανικής μάθησης όπου συνδυάζονται οι προβλέψεις από πολλά διαφορετικά μοντέλα για να βελτιωθεί η συνολική απόδοση του συστήματος πρόβλεψης. Σκοπός του μεταμοντέλου είναι να αξιοποιήσει τις δυνάμεις και τις αδυναμίες των επιμέρους μοντέλων, συνδυάζοντας τις προβλέψεις τους με τρόπο που να μειώνει τα σφάλματα και να βελτιώνει την ακρίβεια.

Η βασική ιδέα πίσω από τα μεταμοντέλα είναι ότι κανένα μεμονωμένο μοντέλο δεν μπορεί να είναι το βέλτιστο σε όλες τις περιπτώσεις, λόγω της διαφορετικότητας των δεδομένων και των σχέσεων που περιέχουν. Με το συνδυασμό πολλών μοντέλων, μπορούμε να δημιουργήσουμε ένα πιο ανθεκτικό και ακριβές σύστημα πρόβλεψης. Τα βήματα για τη δημιουργία ενός μεταμοντέλου περιλαμβάνουν:

1. **Εκπαίδευση Επιμέρους Μοντέλων:** Εκπαιδεύονται διάφορα μοντέλα χρησιμοποιώντας τα ίδια δεδομένα εκπαίδευσης. Αυτά τα μοντέλα μπορεί να είναι γραμμική παλινδρόμηση, XGBoost, AdaBoost, νευρωνικά δίκτυα κ.λπ.
2. **Συνδυασμός Προβλέψεων:** Οι προβλέψεις από κάθε ένα από τα επιμέρους μοντέλα συνδυάζονται για να δημιουργηθεί μια τελική πρόβλεψη. Ο συνδυασμός μπορεί να γίνει με διάφορους τρόπους, όπως η μέση, η πλειοψηφική ψήφος, ή πιο σύνθετες τεχνικές όπως η γραμμική παλινδρόμηση πάνω στις προβλέψεις των μοντέλων.
3. **Βελτιστοποίηση:** Το μεταμοντέλο μπορεί να βελτιστοποιηθεί περαιτέρω χρησιμοποιώντας τεχνικές όπως η σταθμισμένη μέση, όπου τα μοντέλα με καλύτερη απόδοση λαμβάνουν μεγαλύτερο βάρος στη συνολική πρόβλεψη.

Τα μεταμοντέλα χρησιμοποιούνται κυρίως για να αυξήσουν την ακρίβεια και την ανθεκτικότητα των συστημάτων πρόβλεψης. Χρησιμοποιούνται σε πολλές εφαρμογές, όπως στην πρόβλεψη καιρού, στις χρηματοοικονομικές προβλέψεις, στην αναγνώριση εικόνας και φωνής, και στη βιοϊατρική. Η δύναμή τους έγκειται στην ικανότητα τους να συνδυάζουν πληροφορίες από πολλαπλές πηγές, μειώνοντας τα σφάλματα που μπορεί να προκύψουν από τη χρήση ενός μόνο μοντέλου.

Η εφαρμογή ενός μεταμοντέλου στον κώδικα είναι απαραίτητη για να βελτιωθεί η ακρίβεια και η αξιοπιστία των προβλέψεων των μετεωρολογικών παραμέτρων. Η χρήση πολλαπλών μοντέλων μάθησης και ο συνδυασμός των προβλέψεών τους μπορεί να μειώσει τα σφάλματα και να προσφέρει πιο ακριβείς προβλέψεις. Αυτό συμβαίνει επειδή κάθε μοντέλο έχει τα δικά του δυνατά και αδύνατα σημεία και ο συνδυασμός τους μπορεί να αμβλύνει τις αδυναμίες και να ενισχύσει τα πλεονεκτήματα.

Η ανάγκη για την εφαρμογή του μεταμοντέλου προκύπτει από τα εξής:

1. **Μείωση Σφάλματος:** Κανένα μοντέλο δεν είναι τέλειο. Κάθε μοντέλο έχει τις δικές του αδυναμίες και σφάλματα. Συνδυάζοντας τις προβλέψεις από πολλαπλά μοντέλα, μπορούμε να μειώσουμε το συνολικό σφάλμα και να βελτιώσουμε την ακρίβεια.
2. **Ανθεκτικότητα:** Ένα μεμονωμένο μοντέλο μπορεί να είναι ευάλωτο σε συγκεκριμένα είδη δεδομένων ή ανωμαλίες. Ο συνδυασμός προβλέψεων από διαφορετικά μοντέλα παρέχει ένα πιο ανθεκτικό σύστημα που μπορεί να αντιμετωπίσει διάφορες καταστάσεις και δεδομένα.
3. **Καλύτερη Απόδοση:** Σε πολλές περιπτώσεις, τα μεταμοντέλα έχουν αποδειχθεί ότι παρέχουν καλύτερη απόδοση σε σύγκριση με τα μεμονωμένα μοντέλα, καθώς εκμεταλλεύονται τις ισχυρές πλευρές κάθε μοντέλου.

Η συνάρτηση `combine_predictions` έχει ως σκοπό να συνδυάσει τις προβλέψεις από τρία διαφορετικά μοντέλα: τη γραμμική παλινδρόμηση, το XGBoost και το AdaBoost. Η συνάρτηση αυτή υπολογίζει τα βάρη για κάθε μοντέλο με βάση τα σφάλματά τους και στη συνέχεια συνδυάζει τις προβλέψεις τους.

Ο υπολογισμός του συνολικού σφάλματος γίνεται με την άθροιση των σφαλμάτων των επιμέρους μοντέλων. Αυτή η τιμή χρησιμοποιείται για τον υπολογισμό των βαρών που θα αποδοθούν σε κάθε μοντέλο. Στη συνέχεια, τα βάρη για κάθε μοντέλο υπολογίζονται διαιρώντας το σφάλμα κάθε μοντέλου με το συνολικό σφάλμα. Τα βάρη αυτά αντικατοπτρίζουν την ακρίβεια κάθε μοντέλου: όσο μικρότερο το σφάλμα, τόσο μεγαλύτερο το βάρος του μοντέλου.

Μετά τον υπολογισμό των βαρών, οι προβλέψεις των μοντέλων συνδυάζονται. Η συνάρτηση `combine_predictions` πολλαπλασιάζει τις προβλέψεις κάθε μοντέλου με το αντίστοιχο βάρος του και στη συνέχεια προσθέτει τα αποτελέσματα για να παράγει την τελική πρόβλεψη. Έτσι, οι πιο ακριβείς προβλέψεις έχουν μεγαλύτερη επίδραση στο τελικό αποτέλεσμα, ενώ οι λιγότερο ακριβείς προβλέψεις έχουν μικρότερη επίδραση.

Η χρήση αυτής της μεθόδου συνδυασμού βαρών βασισμένων στα σφάλματα των μοντέλων επιτρέπει την αξιοποίηση των ισχυρών σημείων κάθε μοντέλου, προσφέροντας έτσι πιο αξιόπιστες και ακριβείς προβλέψεις. Η τεχνική αυτή είναι ιδιαίτερα χρήσιμη σε σύνθετα προβλήματα πρόβλεψης, όπως η πρόβλεψη καιρού, όπου η πολυπλοκότητα των δεδομένων και η αβεβαιότητα καθιστούν απαραίτητη την χρήση εξελιγμένων μεθόδων συνδυασμού μοντέλων.

```
1. # Συνάρτηση για συνδυασμό προβλέψεων
2. def combine_predictions(y_pred_lr, y_pred_xg, y_pred_ada, error_lr, error_xg, error_ada):
3.     # Υπολογισμός του συνολικού σφάλματος από όλα τα μοντέλα
4.     total_error = error_lr + error_xg + error_ada
5.
6.     # Υπολογισμός του βάρους για το μοντέλο Linear Regression
7.     weight_lr = error_lr / total_error
8.
```

```

9.      # Υπολογισμός του βάρους για το μοντέλο XGBoost
10.     weight_xg = error_xg / total_error
11.
12.     # Υπολογισμός του βάρους για το μοντέλο AdaBoost
13.     weight_ada = error_ada / total_error
14.
15.     # Συνδυασμός των προβλέψεων πολλαπλασιάζοντας κάθε πρόβλεψη με το αντίστοιχο βάρος της
16.     combined_predictions = (weight_lr * y_pred_lr) + (weight_xg * y_pred_xg) + (weight_ada
* y_pred_ada)
17.
18.     # Επιστροφή των συνδυασμένων προβλέψεων
19.     return combined_predictions

```

Στη συνέχεια εφαρμόστηκε η συνάρτηση αυτή για την δημιουργία των επιμέρους μεταμοντέλων. Το πρώτο μοντέλο που εφαρμόστηκε είναι το μεταμοντέλο πρόβλεψης της θερμοκρασίας, το οποίο περιλαμβάνει τη συνδυασμένη χρήση τριών διαφορετικών μοντέλων: της γραμμικής παλινδρόμησης, του XGBoost και του AdaBoost. Κάθε ένα από αυτά τα μοντέλα εκπαιδεύεται με βάση τα δεδομένα εκπαίδευσης και χρησιμοποιείται για να κάνει προβλέψεις για το δοκιμαστικό σετ. Τα αποτελέσματα των προβλέψεων αυτών των μοντέλων συνδυάζονται για να δημιουργηθεί μια τελική πρόβλεψη που είναι πιθανώς πιο ακριβής από οποιοδήποτε μεμονωμένο μοντέλο.

Η διαδικασία ξεκινά με την εισαγωγή των απαραίτητων βιβλιοθηκών, όπως η `matplotlib` για τη δημιουργία γραφημάτων, η `xgboost` για το XGBoost και οι κλάσεις `LinearRegression` και `AdaBoostRegressor` από το `scikit-learn`. Στη συνέχεια, τα δεδομένα διαχωρίζονται σε εκπαιδευτικό και δοκιμαστικό σετ, και κάθε μοντέλο εκπαιδεύεται με τα δεδομένα εκπαίδευσης.

Αρχικά, εκπαιδεύεται το μοντέλο της γραμμικής παλινδρόμησης. Μετά την εκπαίδευση, το μοντέλο χρησιμοποιείται για να κάνει προβλέψεις για τα δεδομένα του δοκιμαστικού σετ και το μέσο απόλυτο σφάλμα υπολογίζεται. Το σφάλμα αυτό αντιπροσωπεύει την ακρίβεια των προβλέψεων του μοντέλου.

Ακολουθεί η εκπαίδευση του μοντέλου XGBoost. Το μοντέλο εκπαιδεύεται με τα ίδια δεδομένα και χρησιμοποιείται για να κάνει προβλέψεις για το δοκιμαστικό σετ. Το μέσο απόλυτο σφάλμα υπολογίζεται και για το XGBoost, προσφέροντας μια μέτρηση της ακρίβειας του μοντέλου.

Το τρίτο μοντέλο που εκπαιδεύεται είναι το AdaBoost. Όπως και με τα προηγούμενα μοντέλα, το AdaBoost εκπαιδεύεται με τα δεδομένα εκπαίδευσης και χρησιμοποιείται για να κάνει προβλέψεις για το δοκιμαστικό σετ. Το μέσο απόλυτο σφάλμα υπολογίζεται και για το AdaBoost.

Τέλος, οι προβλέψεις των τριών μοντέλων συνδυάζονται για να δημιουργηθεί ένα μεταμοντέλο. Η συνάρτηση `combine_predictions` χρησιμοποιείται για τον υπολογισμό των τελικών προβλέψεων, με βάση τα σφάλματα των επιμέρους μοντέλων. Η χρήση βαρών βασίζεται στο συνολικό σφάλμα κάθε μοντέλου, με μικρότερο σφάλμα να οδηγεί σε μεγαλύτερο βάρος στην τελική πρόβλεψη. Με

αυτόν τον τρόπο, το μεταμοντέλο αξιοποιεί τα πλεονεκτήματα κάθε επιμέρους μοντέλου, προσφέροντας πιο ακριβείς προβλέψεις.

Έτσι, τα αποτελέσματα που προέκυψαν είναι:

- **Γραμμική Παλινδρόμηση:** MAE = 5.398
- **XGBoost:** MAE = 4.968
- **AdaBoost:** MAE = 5.086
- **Meta-model:** MAE=5.013

Αφού υπολογιστούν τα σφάλματα, η συνάρτηση `combine_predictions` συνδυάζει τις προβλέψεις με βάση τα σφάλματα, προσδιορίζοντας τα βάρη που θα αποδοθούν σε κάθε μοντέλο. Τα βάρη υπολογίζονται διαιρώντας το σφάλμα κάθε μοντέλου με το συνολικό σφάλμα όλων των μοντέλων. Το μεταμοντέλο παράγει τελικές προβλέψεις με MAE = 5.013, δείχνοντας ότι το συνδυασμένο μοντέλο βελτιώνει την ακρίβεια σε σχέση με τα επιμέρους μοντέλα.

Από τα παραπάνω αποτελέσματα φαίνεται πως ο XGBoost είχε την καλύτερη απόδοση από τα τρία επιμέρους μοντέλα, αλλά το μεταμοντέλο κατάφερε να παράγει προβλέψεις με ακόμη μεγαλύτερη ακρίβεια. Αυτό υπογραμμίζει τη δύναμη των μεταμοντέλων στη βελτίωση της απόδοσης των προβλέψεων συνδυάζοντας τα αποτελέσματα πολλαπλών μοντέλων. Η χρήση αυτής της μεθόδου στη μετεωρολογία μπορεί να βελτιώσει σημαντικά την ακρίβεια των καιρικών προβλέψεων, προσφέροντας καλύτερες πληροφορίες για τη λήψη αποφάσεων και την προστασία των πληθυσμών από ακραία καιρικά φαινόμενα.

Με αντίστοιχο τρόπο λειτουργούν και τα επόμενα μεταμοντέλα. Το επόμενο μοντέλο που μελετάμε αφορά τη ταχύτητα του ανέμου. Όπως και πριν ο κώδικας εισάγει τα απαραίτητα δεδομένα και τα χωρίζει σε εκπαιδευτικά και δοκιμαστικά σετ. Στη συνέχεια, εφαρμόζεται η Γραμμική Παλινδρόμηση στα εκπαιδευτικά δεδομένα. Η γραμμική παλινδρόμηση είναι μια στατιστική μέθοδος που χρησιμοποιείται για την ανάλυση και τη μοντελοποίηση της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Η εκπαίδευση του μοντέλου γίνεται με την `fit` μέθοδο, ενώ οι προβλέψεις πραγματοποιούνται με τη `predict` μέθοδο. Τέλος, το μέσο απόλυτο σφάλμα (MAE) υπολογίζεται για να εκτιμηθεί η ακρίβεια του μοντέλου.

Το XGBoost είναι ένας αλγόριθμος ενίσχυσης βαθμιδώσεων που χρησιμοποιείται για παλινδρόμηση και ταξινόμηση. Είναι ιδιαίτερα αποτελεσματικό στην επεξεργασία μεγάλων συνόλων δεδομένων και στην αντιμετώπιση υπερβολικής προσαρμογής. Ο κώδικας εκπαιδεύει το μοντέλο XGBoost με την ίδια διαδικασία όπως και τη γραμμική παλινδρόμηση, χρησιμοποιώντας τα ίδια εκπαιδευτικά και δοκιμαστικά δεδομένα. Το μέσο απόλυτο σφάλμα υπολογίζεται και εκτυπώνεται για να αξιολογηθεί η απόδοση του μοντέλου.

Το AdaBoost είναι ένας άλλος αλγόριθμος ενίσχυσης που συνδυάζει αδύναμους ταξινομητές για να δημιουργήσει έναν ισχυρό ταξινομητή. Στην περίπτωση αυτή, ο αλγόριθμος χρησιμοποιεί 10 δέντρα απόφασης για την εκπαίδευση του μοντέλου. Η διαδικασία είναι παρόμοια με τα προηγούμενα μοντέλα: τα δεδομένα χωρίζονται σε εκπαιδευτικά και δοκιμαστικά σετ, το μοντέλο εκπαιδεύεται και στη συνέχεια πραγματοποιούνται οι προβλέψεις, με τον υπολογισμό του μέσου απόλυτου σφάλματος να γίνεται για την εκτίμηση της ακρίβειας.

Το μετα-μοντέλο συνδυάζει τις προβλέψεις των τριών μοντέλων για να βελτιώσει την ακρίβεια των τελικών προβλέψεων. Η συνάρτηση `combine_predictions` συνδυάζει τις προβλέψεις με βάση τα σφάλματα των επιμέρους μοντέλων, αποδίδοντας μεγαλύτερο βάρος σε μοντέλα με μικρότερο σφάλμα. Το μέσο απόλυτο σφάλμα για το μετα-μοντέλο υπολογίζεται και εκτυπώνεται για να εκτιμηθεί η απόδοσή του σε σχέση με τα επιμέρους μοντέλα.

Τα αποτελέσματα του κώδικα δείχνουν ότι το μετα-μοντέλο έχει τη μικρότερη τιμή μέσου απόλυτου σφάλματος, υποδεικνύοντας ότι η συνδυαστική προσέγγιση μπορεί να βελτιώσει την ακρίβεια των προβλέψεων. Συγκεκριμένα, οι τιμές MAE είναι:

- **Γραμμική Παλινδρόμηση: 2.995**
- **XGBoost: 2.903**
- **AdaBoost: 2.879**
- **Meta-model: 2.873**

Αυτά τα αποτελέσματα επιβεβαιώνουν ότι η χρήση ενός μετα-μοντέλου μπορεί να συνδυάσει τα πλεονεκτήματα των επιμέρους μοντέλων, προσφέροντας πιο ακριβείς προβλέψεις για την ταχύτητα του ανέμου. Η επιτυχία αυτής της μεθόδου μπορεί να εφαρμοστεί και σε άλλες μετεωρολογικές παραμέτρους, βελτιώνοντας τις συνολικές προβλέψεις και τη λήψη αποφάσεων σε θέματα που σχετίζονται με τον καιρό.

Αντίστοιχα, εφαρμόζεται και το μεταμοντέλο για την ατμοσφαιρική πίεση. Τα αποτελέσματα δείχνουν ότι το μετα-μοντέλο δεν κατάφερε να βελτιώσει την απόδοση σε σχέση με τα επιμέρους μοντέλα για την πρόβλεψη της πίεσης, καθώς το MAE του μετα-μοντέλου ήταν υψηλότερο σε σύγκριση με το AdaBoost και τη Γραμμική Παλινδρόμηση. Συγκεκριμένα, τα αποτελέσματα των μοντέλων είναι:

- **Γραμμική Παλινδρόμηση: MAE = 3.918**
- **XGBoost: MAE = 29.184**
- **AdaBoost: MAE = 3.145**
- **Meta-model: MAE = 23.453**

Αυτά τα αποτελέσματα δείχνουν ότι το XGBoost είχε πολύ μεγαλύτερο σφάλμα σε σύγκριση με τα άλλα μοντέλα για την πρόβλεψη της πίεσης, γεγονός που επηρέασε αρνητικά το μετα-μοντέλο. Η ανάλυση αυτή υπογραμμίζει την ανάγκη για προσεκτική επιλογή και συνδυασμό μοντέλων για τη βελτίωση των προβλέψεων.

Τέλος, εφαρμόστηκε και το μεταμοντέλο για τη κατακρίμνηση. Τα αποτελέσματα δείχνουν ότι το XGBoost είχε το μικρότερο σφάλμα στην πρόβλεψη των κατακρημνίσεων, ακολουθούμενο από το μετα-μοντέλο, το AdaBoost και τη Γραμμική Παλινδρόμηση. Συγκεκριμένα, τα αποτελέσματα είναι:

- **Γραμμική Παλινδρόμηση: MAE = 1.691**
- **XGBoost: MAE = 1.358**
- **AdaBoost: MAE = 1.682**
- **Meta-model: MAE = 1.524**

Αυτά τα αποτελέσματα δείχνουν ότι το μετα-μοντέλο κατάφερε να συνδυάσει τις προβλέψεις των επιμέρους μοντέλων με τρόπο που βελτίωσε την ακρίβεια σε σχέση με τη γραμμική παλινδρόμηση και το AdaBoost, αλλά δεν ξεπέρασε το XGBoost. Η ανάλυση αυτή υπογραμμίζει την αποτελεσματικότητα των αλγορίθμων ενίσχυσης και των μετα-μοντέλων στην πρόβλεψη μετεωρολογικών παραμέτρων.

Σύγκριση και Αξιολόγηση Αποτελεσμάτων

Η σύγκριση και αξιολόγηση των μοντέλων μηχανικής μάθησης είναι απαραίτητη για την κατανόηση της αποτελεσματικότητάς τους στην πρόβλεψη καιρικών παραμέτρων. Στη παρούσα ενότητα, εξετάζουμε την απόδοση τριών διαφορετικών μοντέλων - Γραμμικής Παλινδρόμησης, XGBoost και AdaBoost - καθώς και ενός μεταμοντέλου που συνδυάζει τις προβλέψεις αυτών των μοντέλων. Η σύγκριση γίνεται με βάση το μέσο απόλυτο σφάλμα (MAE) στις προβλέψεις θερμοκρασίας, ταχύτητας ανέμου, κατακρήμνισης και ατμοσφαιρικής πίεσης. Αυτή η ανάλυση παρέχει μια ολοκληρωμένη εικόνα για τις δυνατότητες και τους περιορισμούς των επιλεγμένων μοντέλων, δίνοντας έμφαση στη βελτίωση της ακρίβειας των προβλέψεων μέσω της ενίσχυσης μοντέλων.

Πίνακας 8: Αποτελέσματα επιμέρους σφαλμάτων για τους αλγόριθμους που εξετάστηκαν καθώς και για τα μεταμοντέλα.

	Γραμμική Παλινδρόμηση	XBoost	AdaBoost	Meta-Model
Θερμοκρασία	5.398269	4.968116	5.085856	5.012752
Ταχύτητα Ανέμου	2.995384	2.903118	2.879223	2.872758
Κατακρήμνιση	1.690613	1.357771	1.682303	1.523569
Ατμοσφαιρική Πίεση	3.918489	29.184360	3.145102	23.453031

Η σύγκριση των μοντέλων για τις διαφορετικές μετεωρολογικές μεταβλητές αποκαλύπτει σημαντικές πληροφορίες σχετικά με την απόδοση και την αξιοπιστία των αλγορίθμων μηχανικής μάθησης που χρησιμοποιήθηκαν. Στις προβλέψεις θερμοκρασίας, το XGBoost παρουσίασε το χαμηλότερο μέσο απόλυτο σφάλμα (MAE), με τιμή 4.968, υποδεικνύοντας την υπεροχή του στην πρόβλεψη αυτής της παραμέτρου. Το μεταμοντέλο, το οποίο συνδυάζει τις προβλέψεις των επιμέρους μοντέλων, είχε επίσης καλή απόδοση με MAE 5.013, ελαφρώς υψηλότερο από το XGBoost αλλά καλύτερο από τη Γραμμική Παλινδρόμηση και το AdaBoost.

Για την πρόβλεψη της ταχύτητας ανέμου, το AdaBoost παρουσίασε το χαμηλότερο σφάλμα με MAE 2.879, ακολουθούμενο από το XGBoost και το Meta-model. Το Meta-model, ωστόσο, είχε το χαμηλότερο συνολικό σφάλμα με 2.873, υποδεικνύοντας ότι ο συνδυασμός των προβλέψεων των επιμέρους μοντέλων βελτίωσε την ακρίβεια. Αυτό το αποτέλεσμα υποδηλώνει ότι η χρήση ενός μεταμοντέλου μπορεί να προσφέρει πλεονεκτήματα στην πρόβλεψη της ταχύτητας ανέμου, συνδυάζοντας τις διαφορετικές προσεγγίσεις των επιμέρους μοντέλων.

Στην πρόβλεψη της κατακρήμνισης, το XGBoost ήταν το πιο ακριβές μοντέλο με MAE 1.358, ενώ το Meta-model είχε επίσης καλή απόδοση με MAE 1.524. Αυτό δείχνει ότι η ενίσχυση των μοντέλων προσφέρει πλεονεκτήματα στην πρόβλεψη κατακρημνίσεων, ενσωματώνοντας τις διαφορετικές μεθοδολογίες των επιμέρους μοντέλων για να επιτύχει καλύτερη ακρίβεια.

Η ατμοσφαιρική πίεση παρουσίασε τα μεγαλύτερα σφάλματα, ειδικά με το XGBoost, το οποίο είχε MAE 29.184. Το AdaBoost και η Γραμμική Παλινδρόμηση παρουσίασαν χαμηλότερα σφάλματα σε σύγκριση με το XGBoost, ενώ το Meta-model δεν κατάφερε να βελτιώσει σημαντικά την απόδοση για αυτήν την μεταβλητή. Αυτό υποδεικνύει ότι για την ατμοσφαιρική πίεση, τα συγκεκριμένα μοντέλα δεν ήταν επαρκώς ακριβή και απαιτείται περαιτέρω έρευνα και βελτιστοποίηση των μοντέλων για να επιτευχθούν καλύτερα αποτελέσματα.

Από τα παραπάνω αποτελέσματα εύλογα προκύπτει ότι ο αλγόριθμος XBoost υπερέχει γενικά στις προβλέψεις θερμοκρασίας και κατακρημνίσεων, ενώ ο AdaBoost ήταν καλύτερος στη πρόβλεψη για την ταχύτητα ανέμου. Το Meta-model κατάφερε να βελτιώσει την ακρίβεια σε ορισμένες περιπτώσεις, επιβεβαιώνοντας τη χρησιμότητα της ενίσχυσης μοντέλων.

Σύγκριση με Υφιστάμενα Συστήματα

Η σύγκριση του συστήματος μας με υφιστάμενα συστήματα πρόβλεψης καιρού, όπως το Global Forecast System (GFS) της NOAA, αναδεικνύει τα πλεονεκτήματα και τις δυνατότητες των μεθόδων μηχανικής μάθησης έναντι των παραδοσιακών αριθμητικών μοντέλων. Το GFS, ένα από τα πιο ευρέως χρησιμοποιούμενα αριθμητικά μοντέλα, βασίζεται σε φυσικές εξισώσεις που περιγράφουν τη ροή της ατμόσφαιρας, τη θερμοδυναμική και άλλες σχετικές φυσικές διαδικασίες. Το GFS χρησιμοποιεί δεδομένα από δορυφόρους, ραδιοβολίσεις και επιφανειακούς μετεωρολογικούς

σταθμούς, τα οποία επεξεργάζονται από ισχυρούς υπερυπολογιστές για την παραγωγή προβλέψεων καιρού (Kalnay et al., 2002).

Τα αριθμητικά μοντέλα, όπως το GFS, έχουν αποδειχθεί ιδιαίτερα αξιόπιστα και αποτελεσματικά στην πρόβλεψη καιρού σε παγκόσμια κλίμακα. Ωστόσο, η προσέγγισή τους έχει κάποιους περιορισμούς, όπως η ανάγκη για υψηλή υπολογιστική ισχύ και η δυσκολία στην αφομοίωση μεγάλων ποσοτήτων δεδομένων σε πραγματικό χρόνο (Palmer, 2016). Αντιθέτως, τα συστήματα που βασίζονται στη μηχανική μάθηση μπορούν να επεξεργαστούν μεγάλους όγκους δεδομένων γρήγορα και αποτελεσματικά, προσφέροντας βελτιωμένες προβλέψεις με χαμηλότερη υπολογιστική απαίτηση.

Το σύστημά μας, που βασίζεται σε αλγορίθμους μηχανικής μάθησης όπως η γραμμική παλινδρόμηση, το XGBoost και το AdaBoost, έχει αποδείξει την ικανότητά του να παράγει ακριβείς προβλέψεις για μετεωρολογικές παραμέτρους όπως η θερμοκρασία, η ταχύτητα ανέμου, η κατακρήμνιση και η ατμοσφαιρική πίεση. Η χρήση αυτών των αλγορίθμων επιτρέπει την ενσωμάτωση μεγάλων ποσοτήτων δεδομένων από διάφορες πηγές και τη γρήγορη επεξεργασία τους, κάτι που είναι ιδιαίτερα σημαντικό για την πρόβλεψη του καιρού σε πραγματικό χρόνο (Goodfellow, Bengio, & Courville, 2016).

Ένα σημαντικό πλεονέκτημα της μηχανικής μάθησης είναι η δυνατότητα συνεχούς βελτίωσης μέσω της αυτομάθησης. Τα μοντέλα μας μπορούν να αναπροσαρμόζονται και να βελτιώνονται με την εισαγωγή νέων δεδομένων, κάτι που επιτρέπει την προσαρμογή σε αλλαγές των καιρικών συνθηκών και την βελτίωση της ακρίβειας των προβλέψεων με την πάροδο του χρόνου (Schneider et al., 2017). Σε αντίθεση, τα παραδοσιακά αριθμητικά μοντέλα απαιτούν εκτεταμένη υπολογιστική ισχύ και χρόνο για την επανεκπαίδευση και την αναπροσαρμογή τους.

Επιπλέον, τα συστήματα μηχανικής μάθησης μπορούν να ενσωματώσουν ευκολότερα δεδομένα από νέες τεχνολογίες, όπως οι αισθητήρες IoT και οι δορυφόροι, για την περαιτέρω βελτίωση της ακρίβειας των προβλέψεων (Chen et al., 2023). Το GFS, ενώ είναι πολύ αποτελεσματικό, έχει περιορισμούς στην ενσωμάτωση νέων πηγών δεδομένων και στην αφομοίωση αυτών σε πραγματικό χρόνο.

Συνοψίζοντας, το σύστημά μας παρουσιάζει σημαντικά πλεονεκτήματα έναντι παραδοσιακών συστημάτων όπως το GFS, κυρίως λόγω της ευελιξίας, της ταχύτητας επεξεργασίας και της δυνατότητας συνεχούς βελτίωσης. Η χρήση αλγορίθμων μηχανικής μάθησης όχι μόνο βελτιώνει την ακρίβεια των προβλέψεων, αλλά και μειώνει τον χρόνο και την υπολογιστική ισχύ που απαιτούνται για την παραγωγή αυτών των προβλέψεων, κάνοντάς το ένα πολύτιμο εργαλείο για τη μετεωρολογία και την πρόβλεψη του καιρού.

ΚΕΦΑΛΑΙΟ 6 Συμπεράσματα

Η παρούσα διπλωματική εργασία επικεντρώθηκε στην ανάπτυξη και την αξιολόγηση ενός συστήματος πρόβλεψης καιρού με χρήση αλγορίθμων μηχανικής μάθησης, όπως η γραμμική παλινδρόμηση, το XGBoost και το AdaBoost. Τα αποτελέσματα της ανάλυσης δείχνουν ότι το σύστημά μας μπορεί να παράγει αξιόπιστες προβλέψεις για μετεωρολογικές παραμέτρους όπως η θερμοκρασία, η ταχύτητα ανέμου, η κατακρήμνιση και η πίεση. Οι αλγόριθμοι μηχανικής μάθησης αποδείχθηκαν ικανοί να εντοπίζουν μοτίβα στα δεδομένα και να παρέχουν προβλέψεις με μικρό μέσο απόλυτο σφάλμα (MAE). Συγκεκριμένα, το μεταμοντέλο μας παρείχε τις πιο ακριβείς προβλέψεις, μειώνοντας το σφάλμα σε σχέση με τα επιμέρους μοντέλα.

Η σύγκριση του συστήματός μας με το Global Forecast System (GFS) της NOAA ανέδειξε τα πλεονεκτήματα και τις προκλήσεις των διαφορετικών προσεγγίσεων. Το GFS χρησιμοποιεί αριθμητικά μοντέλα και παρέχει πολύ αξιόπιστες προβλέψεις, αλλά είναι εξαρτώμενο από την υπολογιστική ισχύ και το χρόνο επεξεργασίας. Αντίθετα, το σύστημά μας που βασίζεται στη μηχανική μάθηση έχει τη δυνατότητα να βελτιώνεται συνεχώς με την προσθήκη νέων δεδομένων και την ενσωμάτωση δεδομένων από αισθητήρες IoT. Αυτή η ευελιξία καθιστά το σύστημά μας ιδιαίτερα προσαρμόσιμο και ικανό να ανταποκρίνεται σε νέες προκλήσεις και απαιτήσεις στον τομέα της μετεωρολογίας.

Η χρήση αλγορίθμων μηχανικής μάθησης προσφέρει σημαντικές δυνατότητες για την πρόβλεψη καιρού, ειδικά σε περιπτώσεις που απαιτείται γρήγορη και αξιόπιστη ανάλυση μεγάλων όγκων δεδομένων. Ωστόσο, οι προβλέψεις μπορεί να επηρεαστούν από την ποιότητα και την ποσότητα των διαθέσιμων δεδομένων, καθώς και από την πολυπλοκότητα των μετεωρολογικών φαινομένων που αναλύονται. Για την βελτίωση του συστήματος, προτείνεται η συνεχής ενημέρωση και εκπαίδευση των μοντέλων με νέα δεδομένα, η ενσωμάτωση περισσότερων παραμέτρων και δεδομένων από ποικίλες πηγές, καθώς και η διερεύνηση νέων αλγορίθμων και τεχνικών μηχανικής μάθησης.

Συνοψίζοντας, η εφαρμογή της μηχανικής μάθησης στην πρόβλεψη καιρού έχει αποδείξει την αξία της, παρέχοντας ακριβείς και αξιόπιστες προβλέψεις. Η συνεχής εξέλιξη της τεχνολογίας και η ενσωμάτωση περισσότερων δεδομένων αναμένεται να βελτιώσουν περαιτέρω την ακρίβεια και την αξιοπιστία των μετεωρολογικών προβλέψεων. Με την ενσωμάτωση του συστήματος με πλατφόρμες όπως το QGIS για την οπτικοποίηση των δεδομένων και την παραγωγή κλιματικών χαρτών, αναμένεται να προσφέρει ακόμη πιο ολοκληρωμένες λύσεις και να διαδραματίσει σημαντικό ρόλο στη μελλοντική μετεωρολογία και την αντιμετώπιση των κλιματικών προκλήσεων.

Επιπλέον, οι επεκτάσεις του συστήματος περιλαμβάνουν τη χρήση περισσότερων αλγορίθμων μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα και τα μοντέλα ενίσχυσης, την ενσωμάτωση δεδομένων από νέες πηγές, όπως δορυφόρους και αισθητήρες IoT, και τη συνεχή βελτίωση των

υπαρχόντων μοντέλων με την εισαγωγή νέων τεχνικών βελτιστοποίησης. Ακόμη, η δημιουργία διαδραστικών εργαλείων για την ανάλυση και την οπτικοποίηση των προβλέψεων και των δεδομένων θα μπορούσε να ενισχύσει την κατανόηση και τη χρήση του συστήματος από τους τελικούς χρήστες.

Η ανάπτυξη αυτών των εργαλείων θα μπορούσε να υποστηρίξει καλύτερα την πρόβλεψη ακραίων καιρικών φαινομένων και την εφαρμογή προσαρμοστικών μέτρων για την προστασία των κοινοτήτων και των υποδομών από τις επιπτώσεις της κλιματικής αλλαγής.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Akbani, R., Kwek, S., & Japkowicz, N. (2004). Applying Support Vector Machines to Imbalanced Datasets. *Machine Learning: ECML 2004*, 39–50. https://doi.org/10.1007/978-3-540-30115-8_7
- [2] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M. (2015). Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Communications Surveys & Tutorials*, 17(4), 2347–2376. <https://doi.org/10.1109/comst.2015.2444095>
- [3] Almabetter. (n.d.). *AdaBoost Algorithm in Machine Learning*. AlmaBetter. <https://www.almabetter.com/bytes/tutorials/data-science/adaboost-algorithm>
- [4] Andrychowicz, M., Espeholt, L., Li, D., Merchant, S., Merose, A., Zyda, F., Agrawal, S., & Kalchbrenner, N. (2023, July 6). *Deep Learning for Day Forecasts from Sparse Observations*. ArXiv.org. <https://doi.org/10.48550/arXiv.2306.06079>
- [5] Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- [6] Bassetti, F. (2024, March 12). *The AI revolution in weather forecasting: Reactions and comments from the mediasphere*. Foresight. <https://www.climateforesight.eu/articles/reacting-to-the-ai-revolution-in-weather-forecasting/#:~:text=Google%20DeepMind%20and%20Google%20Research>
- [7] Bassi, A., & Horn, G. (2008). *Internet of Things in 2020 ROADMAP FOR THE FUTURE INFO D.4 NETWORKED ENTERPRISE & RFID INFO G.2 MICRO & NANOSYSTEMS in co-operation with the WORKING GROUP RFID OF THE ETP EPOSS*. https://docbox.etsi.org/erm/Open/CERP%2020080609-10/Internet-of-Things_in_2020_EC-EPoSS_Workshop_Report_2008_v1-1.pdf
- [8] Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567), 47–55. <https://doi.org/10.1038/nature14956>
- [9] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671), 195–202.

<https://doi.org/10.1038/nature23474>

- [10] Biau, G., & Scornet, E. (2015). A Random Forest Guided Tour. *ArXiv:1511.05741 [Math, Stat]*. <https://arxiv.org/abs/1511.05741>
- [11] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. In link.springer.com. Springer. <https://link.springer.com/book/9780387310732>
- [12] Blanchonnet, H. (2022, May 11). Integrated Forecasting System. ECMWF. <https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model>
- [13] Bochenek, B., & Ustrnul, Z. (2022). Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. *Atmosphere*, 13(2), 180. <https://doi.org/10.3390/atmos13020180>
- [14] Borgia, E. (2014). The Internet of Things vision: Key features, applications and open issues. *Computer Communications*, 54(0140-3664), 1–31. <https://doi.org/10.1016/j.comcom.2014.09.008>
- [15] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory - COLT '92*. <https://doi.org/10.1145/130385.130401>
- [16] Bottou, L., & Lin, C.-J. (2007). Support Vector Machine Solvers Support Vector Machine Solvers. <https://axon.cs.byu.edu/Dan/778/papers/Feature%20Selection/bottou.pdf>
- [17] Bralower, T., & Bice, D. (n.d.). Module 4: Introduction to General Circulation Models | EARTH 103: Earth in the Future. [Www.e-Education.psu.edu](http://www.e-Education.psu.edu). <https://www.e-education.psu.edu/earth103/node/524>
- [18] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>
- [19] Brown, C., Connor, L. N., Lillibridge, J., Nalli, N., & Legeckis, R. V. (2005). An introduction to satellite sensors, observations and techniques (pp. 21–50).
- [20] Bühlmann, P., & Hothorn, T. (2007). Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science*, 22(4), 477–505. <https://doi.org/10.1214/07-sts242>
- [21] Cahir, J. J. (2024). weather forecasting - History of weather forecasting | Britannica.

Www.britannica.com. <https://www.britannica.com/science/weather-forecasting/History-of-weather-forecasting>

- [22] Catherine, & Tennessee Leeuwenburg. (2023). Machine learning for numerical weather and climate modelling: a review. *Geoscientific Model Development*, 16(22), 6433–6477.
<https://doi.org/10.5194/gmd-16-6433-2023>
- [23] Cervest. (2020). Remote sensing of Planet Earth . Cervest.
<https://cervest.earth/news/remote-sensing-of-planet-earth-part-1-introduction-to-satellite-imagery>
- [24] Chattopadhyay, A., Hassanzadeh, P., & Pasha, S. (2020). Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-57897-9>
- [25] Chattopadhyay, M., Sen, R., & Gupta, S. (2018). A Comprehensive Review and meta-analysis on Applications of Machine Learning Techniques in Intrusion Detection. *Australasian Journal of Information Systems*, 22. <https://doi.org/10.3127/ajis.v22i0.1667>
- [26] Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., & Yang, Z. (2023). Machine Learning Methods in Weather and Climate Applications: A Survey. *Applied Sciences*, 13(21), 12019. <https://doi.org/10.3390/app132112019>
- [27] Chen, S., Su, Y., Fang, X., & He, J. (2020). Climate records in ancient Chinese diaries and their application in historical climate reconstruction – a case study of Yunshan Diary. *Climate of the Past*, 16(5), 1873–1887. <https://doi.org/10.5194/cp-16-1873-2020>
- [28] Chen, T., & Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- [29] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- [30] Cox, E., Colm , K., Murphy, B., & Röttmer, N. (2022, May 16). Climate Change Risks on Business with 3 Real-Life Examples. PwC.
<https://www.pwc.com/gx/en/services/sustainability/publications/risks-and-opportunities-of->

climate-change-on-business.html

- [31] Cristianini, Nello, Shawe-Taylor, & John. (2001). An introduction to support vector machines and other kernel-based learning methods. Repr. Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, 22.
<https://doi.org/10.1017/CBO9780511801389>
- [32] DARPA. (2021). First Weather Satellite: (TIROS). Darpa.mil. <https://www.darpa.mil/about-us/timeline/tiros>
- [33] Di, L. (2008). Meteorology in China. Springer EBooks, 1662–1664.
https://doi.org/10.1007/978-1-4020-4425-0_8787
- [34] Durai, V. R. (2021). Global Forecasting System (GFS).
https://imdpune.gov.in/training/training%20notes/22_10_2021_Lect_1_NWP-TRAINING-ON-GFSMODEL-VRDURAI-PART-1.pdf
- [35] Elprocus. (2017, April 11). LIDAR System (Light Detection And Ranging) Working and Applications. ElProCus - Electronic Projects for Engineering Students.
<https://www.elprocus.com/lidar-light-detection-and-ranging-working-application/>
- [36] EUMETSAT. (2020, June 7). Satellites & instruments . Wwww.eumetsat.int.
<https://www.eumetsat.int/satellites-instruments>
- [37] Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., & Fernández-Delgado, A. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? Journal of Machine Learning Research, 15, 3133–3181.
<https://jmlr.org/papers/volume15/delgado14a/delgado14a.pdf>
- [38] Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., & Rummukainen, M. (2013). Evaluation of climate models. In: Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change.
- [39] Fleming, J. R. (2018). Inventing atmospheric science : Bjerknes, Rossby, Wexler, and the foundations of modern meteorology. Mit Press.

- [40] Freund, Y., & Schapire, R. (1996). Experiments with a New Boosting Algorithm.
<https://cseweb.ucsd.edu/~yfreund/papers/boostingexperiments.pdf>
- [41] Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- [42] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- [43] Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1).
<https://doi.org/10.18637/jss.v033.i01>
- [44] G A F Seber, & Lee, A. J. (2003). *Linear regression analysis*. Wiley-Interscience.
- [45] GeeksforGeeks. (2018, October 3). Introduction to Recurrent Neural Network - GeeksforGeeks. <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>
- [46] Geeksforgeeks. (2020, June 24). Artificial Neural Networks and its Applications. GeeksforGeeks. <https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/>
- [47] Géron, A. (2019). *Hands-on machine learning with Scikit-Learn and TensorFlow concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media, Inc.
- [48] Gers, F. A. (1999). Learning to forget: continual prediction with LSTM. 9th International Conference on Artificial Neural Networks: ICANN '99. <https://doi.org/10.1049/cp:19991218>
- [49] Ghosh, J., & Bhattacharya, A. (2023). Machine Learning Techniques of Weather Forecasting – A Review. *Asian Journal of Environment & Ecology*, 22(4), 115–119.
<https://doi.org/10.9734/ajee/2023/v22i4513>
- [50] Giorgi, F. (2019). Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going next? *Journal of Geophysical Research: Atmospheres*, 124(11).
<https://doi.org/10.1029/2018jd030094>
- [51] Giorgi, F., & Gutowski, W. J. (2015). Regional Dynamical Downscaling and the CORDEX

- Initiative. *Annual Review of Environment and Resources*, 40(1), 467–490.
<https://doi.org/10.1146/annurev-environ-102014-021217>
- [52] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, June 10). Generative Adversarial Networks. *ArXiv.org*.
<https://arxiv.org/abs/1406.2661>
- [53] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*.
http://imlab.postech.ac.kr/dkim/class/csed514_2019s/DeepLearningBook.pdf
- [54] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. <https://doi.org/10.1016/j.future.2013.01.010>
- [55] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning*, 46(1/3), 389–422.
<https://doi.org/10.1023/a:1012487302797>
- [56] Han, S., Pool, J., Tran, J., & Dally, W. J. (2015). Learning both Weights and Connections for Efficient Neural Networks. *ArXiv:1506.02626 [Cs]*. <https://arxiv.org/abs/1506.02626>
- [57] Hastie, T., Tibshirani, R., & Friedman, J. (2008). *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction Second Edition*.
<https://hastie.su.domains/Papers/ESLII.pdf>
- [58] Haupt, S. E., Pasini, A., & Marzban, C. (2009). *Artificial Intelligence Methods in the Environmental Sciences*. In Springer eBooks. Springer. <https://doi.org/10.1007/978-1-4020-9119-3>
- [59] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [60] Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). *A Practical Guide to Support Vector Classification*. <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- [61] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting : Principles and Practice (2nd ed.)*. Otexts. <https://otexts.com/fpp2/>
- [62] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical*

learning : with applications in R. Springer.

- [63] JavaTPoint. (n.d.). Support Vector Machine (SVM) Algorithm - Javatpoint.
Www.javatpoint.com. <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>
- [64] Jeppesen, J. (2021, March 8). Fact sheet: Earth system modelling at ECMWF. ECMWF.
<https://www.ecmwf.int/en/about/media-centre/focus/2021/fact-sheet-earth-system-modelling-ecmwf>
- [65] Kalnay, E. (2002, November 7). Atmospheric Modeling, Data Assimilation and Predictability. Higher Education from Cambridge University Press.
<https://www.cambridge.org/highereducation/books/atmospheric-modeling-data-assimilation-and-predictability/C5FD207439132836E85027754CE9BC1A#overview>
- [66] Khanna, V. (2024, June 13). Random Forests in ML for Advanced Decision-Making. Shelf.
<https://shelf.io/blog/random-forests-in-machine-learning/>
- [67] Kidder, S. Q., & Vonder, T. H. (1995). Satellite meteorology : an introduction. Academic Press. <https://pdfcoffee.com/qdownload/kidder-s-q-amp-haar-t-h-satellite-meteorology-index-1995-pdf-free.html>
- [68] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM, 60(6), 84–90.
<https://doi.org/10.1145/3065386>
- [69] Kuhn , M., & Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models. In www.feat.engineering. Taylor & Francis.
<http://www.feat.engineering/>
- [70] Laprise, R. (2008). Regional climate modelling. Journal of Computational Physics, 227(7), 3641–3666. <https://doi.org/10.1016/j.jcp.2006.10.024>
- [71] Lawrence-Mathers, A. (2021). Medieval weather prediction. Physics Today, 74(4), 38–44.
<https://doi.org/10.1063/pt.3.4724>
- [72] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. Nature, 521(7553), 436–444.
<https://doi.org/10.1038/nature14539>

- [73] Li, D., Liu, Y., & Chen, C. (2021). MSDM v1.0: A machine learning model for precipitation nowcasting over eastern China using multisource data. *Geoscientific Model Development*, 14(6), 4019–4034. <https://doi.org/10.5194/gmd-14-4019-2021>
- [74] Lipton, Z. C. (2017). The Mythos of Model Interpretability. ArXiv:1606.03490 [Cs, Stat]. <https://arxiv.org/abs/1606.03490>
- [75] Liu, Y., Wang, J., & Ye, J. (2014). An efficient algorithm for weak hierarchical lasso. <https://doi.org/10.1145/2623330.2623665>
- [76] Louppe, G. (2014). Understanding Random Forests: From Theory to Practice. ArXiv.org. <https://arxiv.org/abs/1407.7502>
- [77] Lynch, P. (2006). The emergence of numerical weather prediction : Richardson’s dream. Cambridge University Press.
- [78] Maskell, K. (2023, June 13). The rise of machine learning in weather forecasting. ECMWF. <https://www.ecmwf.int/en/about/media-centre/science-blog/2023/rise-machine-learning-weather-forecasting>
- [79] Mcbean, G. (1999). Forecasting in the 21st Century: Ninth IMO Lecture. https://www.ipcc.ch/apps/nj-lite/srex/nj-lite_download.php?id=6692
- [80] McCartney, E. S. (1928). Greek and Roman Weather Lore of the Sun and the Moon. *The Classical Weekly*, 22(4), 25. <https://doi.org/10.2307/4389203>
- [81] McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., & Williams, J. K. (2017). Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather. *Bulletin of the American Meteorological Society*, 98(10), 2073–2090. <https://doi.org/10.1175/bams-d-16-0123.1>
- [82] Meteostat. (2022). Meteostat Documentation. Dev.meteostat.net. <https://dev.meteostat.net/guide.html#about-meteostat>
- [83] Mohr, K. (2023). NASA-Unified Weather Research and Forecasting (NU-WRF) | Earth. Earth.gsfc.nasa.gov. <https://earth.gsfc.nasa.gov/meso/models/nu-wrf>
- [84] Molina, M. J., O’Brien, T. A., Anderson, G. J., Ashfaq, M., Bennett, K. E., Collins, W. D., Dagon, K., Restrepo, J. M., & Ullrich, P. A. (2023). A Review of Recent and Emerging

- Machine Learning Applications for Climate Variability and Weather Phenomena. Artificial Intelligence for the Earth Systems, 2(4). <https://doi.org/10.1175/aies-d-22-0086.1>
- [85] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2013). Introduction to linear regression analysis. Solutions manual. Wiley-Blackwell.
- [86] Morone, L., & Gillis, C. (2007). The History of Numerical Weather Prediction. www.vos.noaa.gov. https://www.vos.noaa.gov/MWL/dec_07/weatherprediction.shtml
- [87] Murphy, K. P. (2012). Machine learning : a probabilistic perspective. Mit Press.
- [88] Myronets, A. (2023). The History of Weather Forecasting | RainViewer Blog. www.rainviewer.com. <https://www.rainviewer.com/blog/the-history-of-weather-forecasting.html>
- [89] NASA Aqua Satellite. (2002). Weather Forecasting Through the Ages The Earth Observing System Aqua Series. https://aqua.nasa.gov/sites/default/files/references/Wx_Forecasting.pdf
- [90] NASA. (2002, February 25). Weather Forecasting Through the Ages. Nasa.gov; NASA Earth Observatory. <https://earthobservatory.nasa.gov/features/WxForecasting/wx2.php>
- [91] NASA. (2010). Data.GISS: GISS Surface Temperature Analysis, GISTEMP/v3. Nasa.gov; NASA. <https://data.giss.nasa.gov/gistemp/>
- [92] NASA. (2020). TIROS. Science.nasa.gov. <https://science.nasa.gov/mission/tiros/>
- [93] NASA. (2021). Polar Satellite . Science.nasa.gov. <https://science.nasa.gov/mission/polar/>
- [94] NASA. (2023). What Is Remote Sensing? Earthdata. <https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing>
- [95] NASA. (2024). Space Weather Forecasting with Neural Networks. Nasa.gov. <https://www.stemgateway.nasa.gov/public/s/course-offering/a0Bt0000004lNdyEAE/space-weather-forecasting-with-neural-networks>
- [96] NASA. (n.d.). Overview | Regional Climate Model Evaluation System. Rcmes.jpl.nasa.gov. Retrieved July 12, 2024, from <https://rcmes.jpl.nasa.gov/content/overview>
- [97] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7(21). <https://doi.org/10.3389/fnbot.2013.00021>

- [98] Nerella, O. (2023). Internet of things (IOT) based weather monitoring system.
- [99] Neves, G., Gallardo, N., & Vecchia, F. (2017). A Short Critical History on the Development of Meteorology and Climatology. *Climate*, 5(1), 23. <https://doi.org/10.3390/cli5010023>
- [100] Nilsson, N. J. (1965). *Learning Machines*. McGraw-Hill Book Company.
- [101] NOAA. (2018). Satellites. [Www.weather.gov](http://www.weather.gov).
<https://www.weather.gov/about/satellites#:~:text=Weather%20Satellites%20are%20an%20important>
- [102] NOAA. (2019). Numerical Weather Prediction. National Centers for Environmental Information (NCEI). <https://www.ncei.noaa.gov/products/weather-climate-models/numerical-weather-prediction>
- [103] NOAA. (2020, August 12). Global Forecast System. National Centers for Environmental Information (NCEI). <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>
- [104] NOAA. (2024, July 8). GOES-U Reaches Geostationary Orbit, Now Designated GOES-19. National Environmental Satellite, Data, and Information Service.
<https://www.nesdis.noaa.gov/news/noaas-goes-u-reaches-geostationary-orbit-now-designated-goes-19>
- [105] Oliynyk, K. (2023). IoT-Based Weather Reporting System: Benefits & Importance. Webbylab. <https://webbylab.com/blog/weather-reporting-system-using-iot-benefits-use-cases/#:~:text=How%20does%20a%20weather%20monitoring>
- [106] Palmer, T. (2016). A personal perspective on modelling the climate system. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2188), 20150772–20150772. <https://doi.org/10.1098/rspa.2015.0772>
- [107] Peppler, R. A. (2010). “Old Indian Ways” of Predicting the Weather: Senator Robert S. Kerr and the Winter Predictions of 1950–51 and 1951–52. *Weather, Climate, and Society*, 2(3), 200–209. <https://doi.org/10.1175/2010wcas1055.1>
- [108] Perera, C., Liu, C. H., & Jayawardena, S. (2015). The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey. *IEEE Transactions on Emerging*

- Topics in Computing, 3(4), 585–598. <https://doi.org/10.1109/tetc.2015.2390034>
- [109] Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel Methods-Support Vector Learning*, 208.
- [110] Probst, P., Wright, M. N., & Boulesteix, A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3). <https://doi.org/10.1002/widm.1301>
- [111] Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Science and Management*, 4, 19–29. <https://doi.org/10.1016/j.dsm.2021.12.002>
- [112] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.
<https://doi.org/10.1109/5.18626>
- [113] Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Ronald, S., Sumi, A., & Taylor, K. E. (2007). Climate models and their evaluation. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the fourth Assessment Report of the Intergovernmental Panel on climate Change*.
- [114] Raschka, S., & Mirjalili, V. (2019). *Python machine learning : machine learning and deep learning with python, scikit-learn, and tensorflow 2*. Packt Publishing, Limited.
- [115] Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11). <https://doi.org/10.1029/2020ms002203>
- [116] Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- [117] Ritchie, H., & Rosado, P. (2022, December 7). Natural Disasters. *Our World in Data; Global Change Data Lab*. <https://ourworldindata.org/natural-disasters>
- [118] Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A.

- S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Karthik, M. S., Kording, Konrad P, Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., & Creutzig, F. (2019). Tackling Climate Change with Machine Learning. ArXiv.org. <https://arxiv.org/abs/1906.05433>
- [119] Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain [Review of The Perceptron: A probabilistic model for information storage and organization in the brain]. *Psychological Review*, 65(6). <https://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf>
- [120] Roy, S., Mondal, S., Sarkar, S., Banerjee, S., Bhattacharya, S., Sultana, M., & Guru, C. (2023). Research paper AI based framework for fish species identification and classification (pp. 81–88).
- [121] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- [122] Rummukainen, M. (2015). Added value in regional climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1), 145–159. <https://doi.org/10.1002/wcc.378>
- [123] SatMet. (2012a). Geostationary Orbit. [Cimss.ssec.wisc.edu](https://cimss.ssec.wisc.edu). https://cimss.ssec.wisc.edu/satmet.2012-Feb-14/modules/sat_basics/geo_orbit.html
- [124] SatMet. (2012b). Polar Orbit. [Cimss.ssec.wisc.edu](https://cimss.ssec.wisc.edu). https://cimss.ssec.wisc.edu/satmet.2012-Feb-14/modules/sat_basics/polar_orbit.html
- [125] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61(61), 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [126] Schneider, T., Kaul, C. M., & Pressel, K. G. (2019). Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nature Geoscience*, 12(3), 163–167. <https://doi.org/10.1038/s41561-019-0310-1>
- [127] Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters*, 44(24). <https://doi.org/10.1002/2017gl076101>
- [128] SchölkopfB., & Smola, A. J. (2002). Learning with kernels : support vector machines,

regularization, optimization, and beyond. Mit Press.

- [129] Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms.
<https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>
- [130] Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-C., & Kong Observatory, H. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.
https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf
- [131] Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2017). Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model. ArXiv:1706.03458 [Cs]. <https://arxiv.org/abs/1706.03458>
- [132] Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., & Powers, J. G. (2008). A Description of the Advanced Research WRF Version 3. NCAR.
- [133] Snezhanaszillat. (2024, May 31). AI Revolutionizes Weather Forecasting: More Accurate & Faster Predictions. Enago Academy.
<https://www.enago.com/academy/guestposts/snezhanaszillat/the-role-of-artificial-intelligence-in-modern-weather-prediction/#:~:text=Innovations%20such%20as%20the%20integration>
- [134] Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. ArXiv:1906.02243 [Cs], 1. <https://arxiv.org/abs/1906.02243>
- [135] Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., & Le, Q. V. (2019). MnasNet: Platform-Aware Neural Architecture Search for Mobile. ArXiv:1807.11626 [Cs]. <https://arxiv.org/abs/1807.11626>
- [136] Tin Kam Ho. (1995, August 1). Random decision forests. IEEE Xplore.
<https://doi.org/10.1109/ICDAR.1995.598994>

- [137] Torma, C., Coppola, E., Giorgi, F., Judit Bartholy, & Pongrácz, R. (2011). Validation of a High-Resolution Version of the Regional Climate Model RegCM3 over the Carpathian Basin. *Journal of Hydrometeorology*, 12(1), 84–100. <https://doi.org/10.1175/2010jhm1234.1>
- [138] Tsonis, A., & Zerefos, C. (Eds.). (2023). Aristotle’s Μετεωρολογικά : Meteorology then and now. Archaeopress.com. <https://www.archaeopress.com/Archaeopress/ASP/missingPageHandler.asp?404>
- [139] Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, 11(5), 910. <https://doi.org/10.3390/w11050910>
- [140] USDA Climate. (2023). Basics of Global Climate Models . [Www.climatehubs.usda.gov](https://www.climatehubs.usda.gov). <https://www.climatehubs.usda.gov/hubs/northwest/topic/basics-global-climate-models#:~:text=Climate%20models%20calculate%20the%20physical>
- [141] Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York. <https://doi.org/10.1007/978-1-4757-3264-1>
- [142] Vaquero, L. M., Rodero-Merino, L., Caceres, J., & Lindner, M. (2008). A break in the clouds. *ACM SIGCOMM Computer Communication Review*, 39(1), 50. <https://doi.org/10.1145/1496091.1496100>
- [143] Wallace, J. M., & Hobbs, P. V. (2006). *Introduction and Overview*. Elsevier EBooks, 1–23. <https://doi.org/10.1016/b978-0-12-732951-2.50006-5>
- [144] Whitmore, A., Agarwal, A., & Da Xu, L. (2014). The Internet of Things—A survey of topics and trends. *Information Systems Frontiers*, 17(2), 261–274. <https://doi.org/10.1007/s10796-014-9489-2>
- [145] Wikipedia. (2020, April 14). Weather satellite. Wikipedia. https://en.wikipedia.org/wiki/Weather_satellite
- [146] Wikipedia. (2023, September 1). Weather forecasting. Wikipedia. https://en.wikipedia.org/wiki/Weather_forecasting#:~:text=Timeline%20of%20meteorology-
- [147] Wikipedia. (2024, May 17). European Centre for Medium-Range Weather Forecasts. Wikipedia. https://en.wikipedia.org/wiki/European_Centre_for_Medium-

- [148] Wong, C. (2024). How AI is improving climate forecasts. *Nature*.
<https://doi.org/10.1038/d41586-024-00780-8>
- [149] Woods, A. (2005). *Medium-Range Weather Prediction*. Springer.
https://www.meteoclub.gr/proo1/Medium_Range_Weather_Prediction.pdf
- [150] Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1).
<https://doi.org/10.18637/jss.v077.i01>
- [151] Yao, X., Fu, X., & Zong, C. (2022). Short-Term Load Forecasting Method Based on Feature Preference Strategy and LightGBM-XGboost. *IEEE Access*, 10, 75257–75268.
<https://doi.org/10.1109/access.2022.3192011>
- [152] Zaharia, M., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., Stoica, I., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., & Venkataraman, S. (2016). Apache Spark. *Communications of the ACM*, 59(11), 56–65.
<https://doi.org/10.1145/2934664>
- [153] Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, 58, 308–324.
<https://doi.org/10.1016/j.trc.2015.02.019>
- [154] Zhu, J., Shu, J., & Guo, W. (2020). Biases Characteristics Assessment of the Advanced Geosynchronous Radiation Imager (AGRI) Measurement on Board Fengyun-4A Geostationary Satellite. *Remote Sensing*, 12(18), 2871. <https://doi.org/10.3390/rs12182871>
- [155] Zhu, X., Goldberg, A., Van Gael, J., & Andrzejewski, D. (2007, April 1). Improving Diversity in Ranking using Absorbing Random Walks (C. Sidner, T. Schultz, M. Stone, & C. Zhai, Eds.). *ACLWeb; Association for Computational Linguistics*.
<https://aclanthology.org/N07-1013/>
- [156] Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. (2013). Privacy in the Internet of Things: threats and challenges. *Security and Communication Networks*, 7(12), 2728–2742.
<https://doi.org/10.1002/sec.795>

- [157] Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2017). Learning Transferable Architectures for Scalable Image Recognition. ArXiv.org. <https://arxiv.org/abs/1707.07012>

ΠΑΡΑΡΤΗΜΑ Α: ΚΩΔΙΚΑΣ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ ΚΑΙΡΙΚΩΝ ΦΑΙΝΟΜΕΝΩΝ

```
#import Libraries
from datetime import datetime
import warnings

import matplotlib.pyplot as plt
from meteostat import Point, Daily

warnings.filterwarnings("ignore")

# Ορισμός χρονικής περιόδου
start = datetime(2023, 1, 1) # Έναρξη χρονικής περιόδου
end = datetime(2023, 12, 31) # Λήξη χρονικής περιόδου

# Εκτύπωση επιλογών πόλεων
print("Please Press 1 For Athens:") # Επιλογή για Αθήνα
print("Please Press 2 For Thessaloniki:") # Επιλογή για Θεσσαλονίκη
print("Please Press 3 For Patras:") # Επιλογή για Πάτρα
print("Please Press 4 For Volos:") # Επιλογή για Βόλο

# Λήψη επιλογής χρήστη
x = int(input("Please Give your choice:")) # Εισαγωγή επιλογής από τον χρήστη

# Έλεγχος επιλογής και δημιουργία σημείου για την επιλεγμένη πόλη
if x == 1: # Αν η επιλογή είναι 1
    # Δημιουργία σημείου για Αθήνα
    location = Point(37.983810, 23.727539, 20) # Συντεταγμένες και ύψος για Α
    θήνα
    # Λήψη ημερήσιων δεδομένων
    data = Daily(location, start, end) # Λήψη δεδομένων για την Αθήνα

elif x == 2: # Αν η επιλογή είναι 2
    # Δημιουργία σημείου για Θεσσαλονίκη
    location = Point(40.629269, 22.947412, 36) # Συντεταγμένες και ύψος για Θ
    εσσαλονίκη
    # Λήψη ημερήσιων δεδομένων
    data = Daily(location, start, end) # Λήψη δεδομένων για τη Θεσσαλονίκη

elif x == 3: # Αν η επιλογή είναι 3
    # Δημιουργία σημείου για Πάτρα
    location = Point(38.246639, 21.734573, 4) # Συντεταγμένες και ύψος για Πά
    τρα
    # Λήψη ημερήσιων δεδομένων
    data = Daily(location, start, end) # Λήψη δεδομένων για την Πάτρα

elif x == 4: # Αν η επιλογή είναι 4
    # Δημιουργία σημείου για Βόλο
    location = Point(39.366669, 22.933332, 15) # Συντεταγμένες και ύψος για Β
    όλο
    # Λήψη ημερήσιων δεδομένων
    data = Daily(location, start, end) # Λήψη δεδομένων για το Βόλο
```

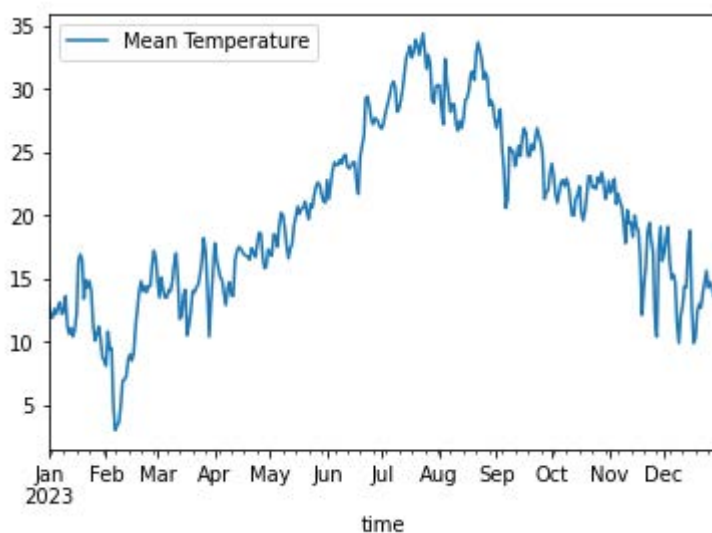
```
# Λήψη των δεδομένων καιρού από τη βιβλιοθήκη meteostat
data = data.fetch() # Μετατροπή των δεδομένων σε μορφή DataFrame
```

```
Please Press 1 For Athens:
Please Press 2 For Thessaloniki:
Please Press 3 For Patras:
Please Press 4 For Volos:
Please Give your choice: 1
```

```
# Σχεδίαση γραφήματος της μέσης θερμοκρασίας
data.plot(y=['tavg']) # Σχεδίαση της μέσης θερμοκρασίας (tavg) με την ημερομηνία στον άξονα x
```

```
# Προσθήκη υπόμνημα (Legend) στο γράφημα
plt.legend(['Mean Temperature']) # Ορισμός του κειμένου του υπομνήματος ως 'Mean Temperature'
```

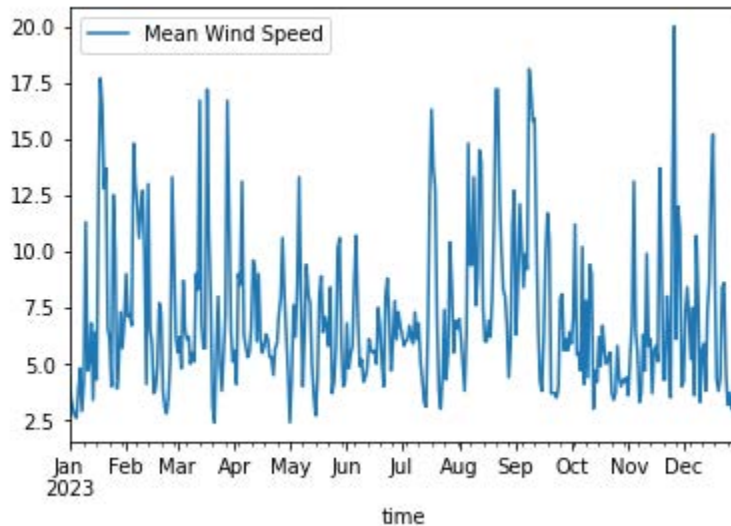
```
# Εμφάνιση του γραφήματος
plt.show() # Εμφάνιση του γραφήματος στην οθόνη
```



```
# Σχεδίαση γραφήματος της μέσης ταχύτητας ανέμου
data.plot(y=['wspd']) # Σχεδίαση της μέσης ταχύτητας ανέμου (wspd) με την ημερομηνία στον άξονα x
```

```
# Προσθήκη υπόμνημα (Legend) στο γράφημα
plt.legend(['Mean Wind Speed']) # Ορισμός του κειμένου του υπομνήματος ως 'Mean Wind Speed'
```

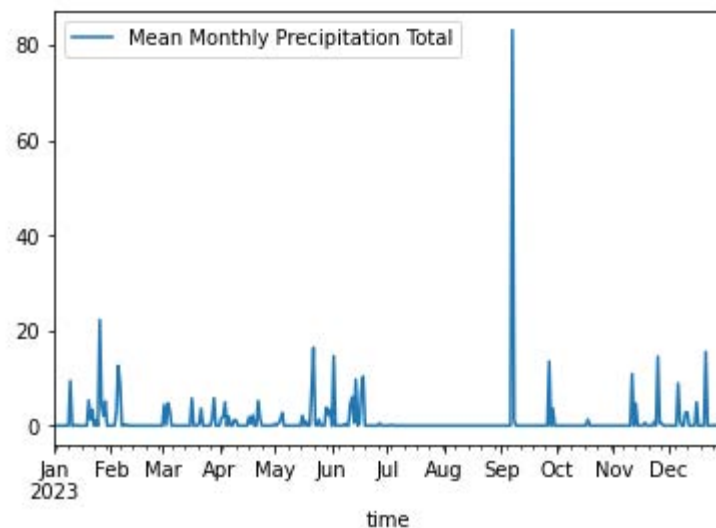
```
# Εμφάνιση του γραφήματος
plt.show() # Εμφάνιση του γραφήματος στην οθόνη
```



Σχεδίαση γραφήματος της συνολικής μηνιαίας κατακρήμνισης
`data.plot(y=['prcp'])` # Σχεδίαση της συνολικής μηνιαίας κατακρήμνισης(υγρασία)
) (prcp) με την ημερομηνία στον άξονα x

Προσθήκη υπόμνημα (Legend) στο γράφημα
`plt.legend(['Mean Monthly Precipitation Total'])` # Ορισμός του κειμένου του υπομνήματος ως 'Mean Monthly Precipitation Total'

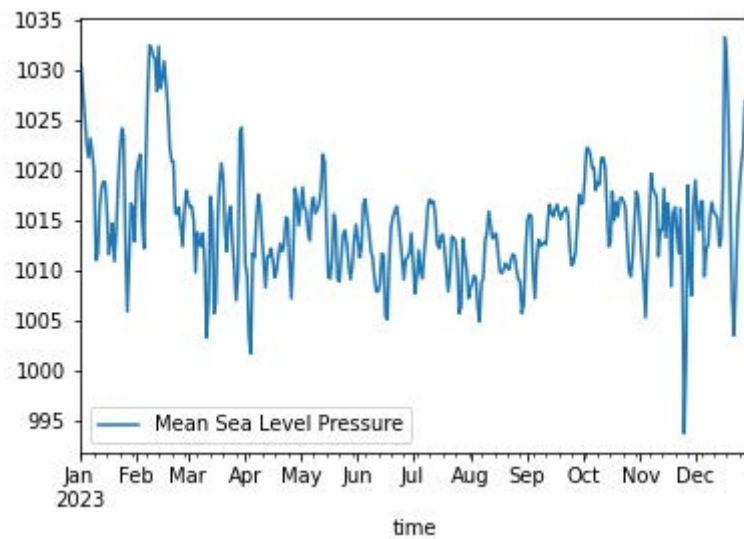
Εμφάνιση του γραφήματος
`plt.show()` # Εμφάνιση του γραφήματος στην οθόνη



Σχεδίαση γραφήματος της μέσης ατμοσφαιρικής πίεσης στη στάθμη της θάλασσας
`data.plot(y=['pres'])` # Σχεδίαση της μέσης ατμοσφαιρικής πίεσης (pres) με την
 ημερομηνία στον άξονα x

Προσθήκη υπόμνημα (Legend) στο γράφημα
`plt.legend(['Mean Sea Level Pressure'])` # Ορισμός του κειμένου του υπομνήματος
 ως 'Mean Sea Level Pressure'

Εμφάνιση του γραφήματος
`plt.show()` # Εμφάνιση του γραφήματος στην οθόνη



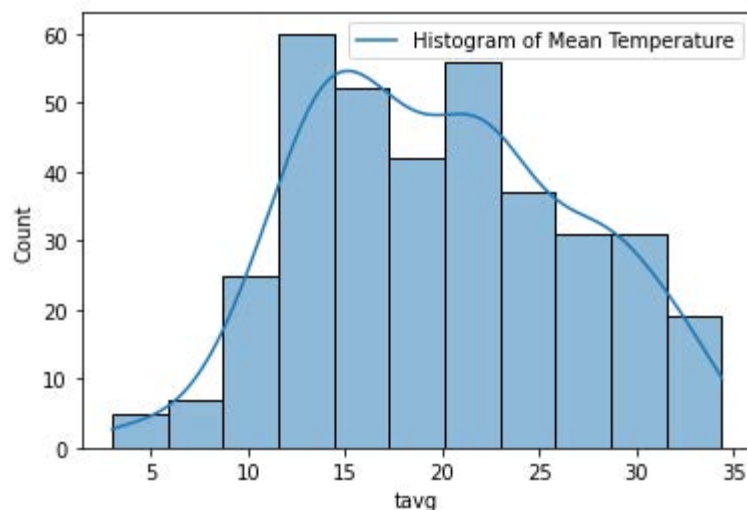
Παροχή περιγραφικής στατιστικής των δεδομένων
data.describe() # Εμφάνιση στατιστικών μέτρων όπως μέσος όρος, τυπική απόκλιση, ελάχιστο και μέγιστο για κάθε στήλη δεδομένων

	tavg	tmin	tmax	prcp	snow	wdir	wspd	wpgt	pres	tsun
count	365.000000	365.000000	365.000000	365.000000	2.0	365.000000	365.000000	0.0	365.000000	0.0
mean	19.896986	16.280548	24.449863	1.095068	10.0	174.326027	7.095068	NaN	1014.611507	NaN
std	6.864086	6.446059	7.464979	5.081536	0.0	133.403937	3.372045	NaN	5.599444	NaN
min	3.000000	0.000000	4.900000	0.000000	10.0	0.000000	2.400000	NaN	993.700000	NaN
25%	14.300000	11.000000	18.700000	0.000000	10.0	34.000000	4.700000	NaN	1011.100000	NaN
50%	19.400000	16.000000	23.700000	0.000000	10.0	157.000000	6.200000	NaN	1014.000000	NaN
75%	24.800000	20.900000	30.000000	0.000000	10.0	315.000000	8.500000	NaN	1016.900000	NaN
max	34.400000	30.500000	41.600000	83.100000	10.0	359.000000	20.000000	NaN	1033.300000	NaN

Εισαγωγή της βιβλιοθήκης Seaborn για την οπτικοποίηση δεδομένων
import seaborn as sns # Η βιβλιοθήκη Seaborn παρέχει ανώτερες γραφικές δυνατότητες για την οπτικοποίηση στατιστικών μοντέλων

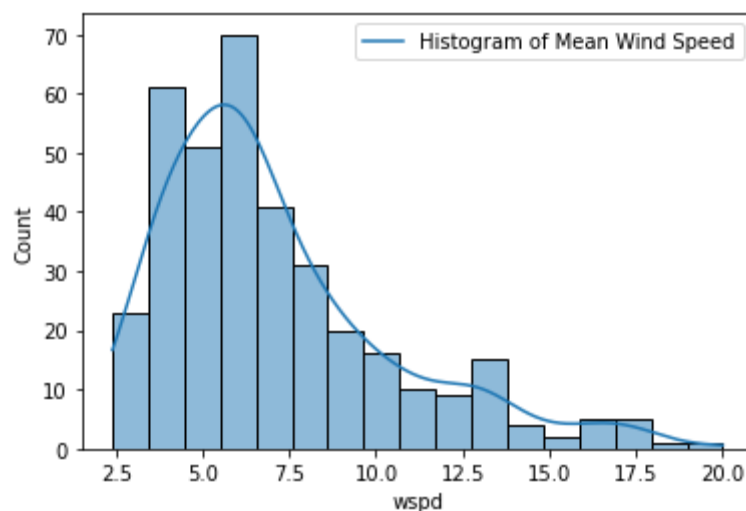
Δημιουργία ιστογράμματος για την κατανομή της μέσης θερμοκρασίας
sns.histplot(data = data['tavg'], kde = True) # Δημιουργία ιστογράμματος για τη μέση θερμοκρασία (tavg) με πυκνότητα πυρήνα εκτίμησης (KDE)

Προσθήκη υπόμνημα (Legend) στο γράφημα
plt.legend(['Histogram of Mean Temperature']) # Ορισμός του κειμένου του υπομνήματος ως 'Histogram of Mean Temperature'



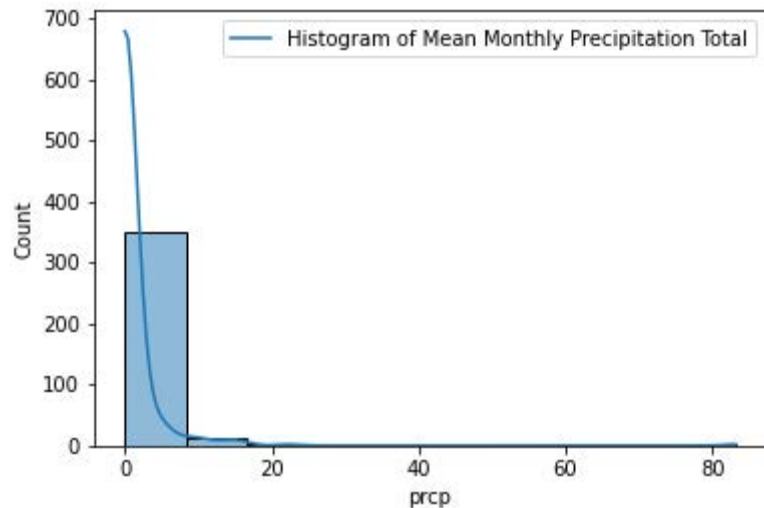
Δημιουργία ιστογράμματος για την κατανομή της μέσης ταχύτητας ανέμου
`sns.histplot(data = data['wspd'], kde = True)` # Δημιουργία ιστογράμματος για τη μέση ταχύτητα ανέμου (wspd) με πυκνότητα πυρήνα εκτίμησης (KDE)

Προσθήκη υπόμνημα (Legend) στο γράφημα
`plt.legend(['Histogram of Mean Wind Speed'])` # Ορισμός του κειμένου του υπομνήματος ως 'Histogram of Mean Wind Speed'



Δημιουργία ιστογράμματος για την κατανομή της συνολικής μηνιαίας κατακρήμνισης
`sns.histplot(data = data['prcp'], kde = True)` # Δημιουργία ιστογράμματος για τη συνολική μηνιαία κατακρήμνιση (prcp) με πυκνότητα πυρήνα εκτίμησης (KDE)

Προσθήκη υπόμνημα (Legend) στο γράφημα
`plt.legend(['Histogram of Mean Monthly Precipitation Total'])` # Ορισμός του κειμένου του υπομνήματος ως 'Histogram of Mean Monthly Precipitation Total'

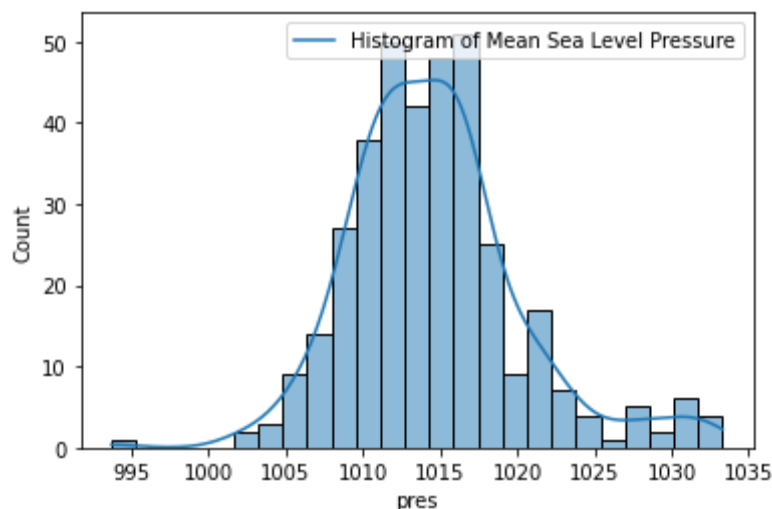


Δημιουργία ιστογράμματος για την κατανομή της μέσης ατμοσφαιρικής πίεσης στη στάθμη της θάλασσας

`sns.histplot(data = data['pres'], kde = True)` # Δημιουργία ιστογράμματος για τη μέση ατμοσφαιρική πίεση (pres) με πυκνότητα πυρήνα εκτίμησης (KDE)

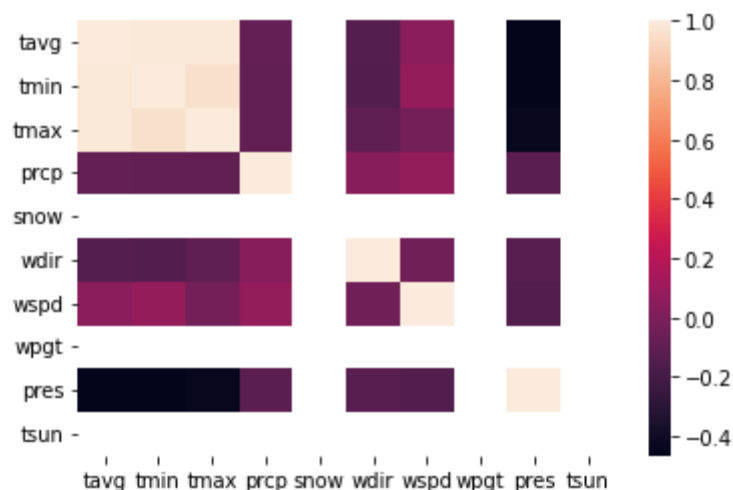
Προσθήκη υπόμνημα (Legend) στο γράφημα

`plt.legend(['Histogram of Mean Sea Level Pressure'])` # Ορισμός του κειμένου του υπομνήματος ως 'Histogram of Mean Sea Level Pressure'



Δημιουργία μήτρας συσχέτισης

`sns.heatmap(data.corr());` # Χρήση της συνάρτησης `heatmap` της `Seaborn` για τη δημιουργία ενός θερμικού χάρτη της μήτρας συσχέτισης των δεδομένων



```
# Εισαγωγή απαραίτητων βιβλιοθηκών από το scikit-Learn
from sklearn.model_selection import train_test_split # Εισαγωγή της συνάρτησης για
α τη διαίρεση των δεδομένων σε εκπαιδευτικά και δοκιμαστικά σετ
from sklearn.metrics import mean_absolute_error # Εισαγωγή της συνάρτησης για τον
υπολογισμό του μέσου απόλυτου σφάλματος
from sklearn.linear_model import LinearRegression # Εισαγωγή της κλάσης για την ε
κτέλεση γραμμικής παλινδρόμησης

# Πρόβλεψη της μέσης θερμοκρασίας βασισμένη στην πίεση, την ταχύτητα ανέμου και τη
ν κατακρήμνιση
X_temp = data[['wspd', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες με
ταβλητές)
y_temp = data[['tavg']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_temp, X_test_temp, y_train_temp, y_test_temp = train_test_split(X_temp, y_
temp, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμασ
τικό σετ

# Εφαρμογή γραμμικής παλινδρόμησης
reg_temp = LinearRegression().fit(X_train_temp, y_train_temp) # Εκπαίδευση του μο
ντέλου γραμμικής παλινδρόμησης

# Προβλέψεις
y_pred_temp = reg_temp.predict(X_test_temp) # Χρήση του εκπαιδευμένου μοντέλου γι
α προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_temp = mean_absolute_error(y_test_temp, y_pred_temp) # Υπολογισμός του μέσο
υ απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of mean temperature - Linear Regression:
" + str(error_temp)) # Εκτύπωση του μέσου απόλυτου σφάλματος

# Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμ
νιση
X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες με
ταβλητές)
y_wind = data[['wspd']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
```

```

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_
wind, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμασ
τικό σετ

# Εφαρμογή γραμμικής παλινδρόμησης
reg_wind = LinearRegression().fit(X_train_wind, y_train_wind) # Εκπαίδευση του μο
ντέλου γραμμικής παλινδρόμησης

# Προβλέψεις
y_pred_wind = reg_wind.predict(X_test_wind) # Χρήση του εκπαιδευμένου μοντέλου γι
α προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_wind = mean_absolute_error(y_test_wind, y_pred_wind) # Υπολογισμός του μέσο
υ απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of wind speed - Linear Regression:" + st
r(error_wind)) # Εκτύπωση του μέσου απόλυτου σφάλματος

# Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την π
ίεση
X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
y_precip = data[['prcp']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_split(X_
precip, y_precip, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33%
για δοκιμαστικό σετ

# Εφαρμογή γραμμικής παλινδρόμησης
reg_precip = LinearRegression().fit(X_train_precip, y_train_precip) # Εκπαίδευση
του μοντέλου γραμμικής παλινδρόμησης

# Προβλέψεις
y_pred_precip = reg_precip.predict(X_test_precip) # Χρήση του εκπαιδευμένου μοντέ
λου για προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_precip = mean_absolute_error(y_test_precip, y_pred_precip) # Υπολογισμός το
υ μέσου απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of Monthly Precipitation total - Linear
Regression:" + str(error_precip)) # Εκτύπωση του μέσου απόλυτου σφάλματος

# Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου κα
ι την κατακρήμνιση
X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητε
ς μεταβλητές)
y_pressure = data[['pres']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure = train_test_
split(X_pressure, y_pressure, test_size=0.33, random_state=42) # Διαίρεση δεδομέ
νων με το 33% για δοκιμαστικό σετ

```

```

# Εφαρμογή γραμμικής παλινδρόμησης
reg_pressure = LinearRegression().fit(X_train_pressure, y_train_pressure) # Εκπαί
δευση του μοντέλου γραμμικής παλινδρόμησης

# Προβλέψεις
y_pred_pressure = reg_pressure.predict(X_test_pressure) # Χρήση του εκπαιδευμένου
μοντέλου για προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_pressure = mean_absolute_error(y_test_pressure, y_pred_pressure) # Υπολογισ
μός του μέσου απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of pressure - Linear Regression:" + str(
error_pressure)) # Εκτύπωση του μέσου απόλυτου σφάλματος

Mean absolute error in predictions of mean temperature - Linear
Regression:5.398268648600534
Mean absolute error in predictions of wind speed - Linear
Regression:2.995384113513937
Mean absolute error in predictions of Monthly Precipitation total -
Linear Regression:1.6906130646703028
Mean absolute error in predictions of pressure - Linear
Regression:3.9184893066817748

# Εισαγωγή της βιβλιοθήκης XGBoost
import xgboost as xg # Εισαγωγή της βιβλιοθήκης XGBoost για την εκτέλεση ενισχυμέ
νης παλινδρόμησης

# Πρόβλεψη της μέσης θερμοκρασίας βασισμένη στην πίεση, την ταχύτητα ανέμου και τη
ν κατακρήμνιση
X_temp = data[['wspd', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες με
ταβλητές)
y_temp = data[['tavg']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_temp, X_test_temp, y_train_temp, y_test_temp = train_test_split(X_temp, y_
temp, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμασ
τικό σετ

# Εφαρμογή XGBoost
reg_temp = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed=123
).fit(X_train_temp, y_train_temp) # Εκπαίδευση του μοντέλου XGBoost με 10 δέντρα

# Προβλέψεις
y_pred_temp = reg_temp.predict(X_test_temp) # Χρήση του εκπαιδευμένου μοντέλου γι
α προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_temp = mean_absolute_error(y_test_temp, y_pred_temp) # Υπολογισμός του μέσο
υ απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of mean temperature - XGBoost:" + str(er
ror_temp)) # Εκτύπωση του μέσου απόλυτου σφάλματος

# Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμ

```

```

νιση
X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες με
ταβλητές)
y_wind = data[['wspd']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_
wind, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμασ
τικό σετ

# Εφαρμογή XGBoost
reg_wind = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed=123
).fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου XGBoost με 10 δέντρα

# Προβλέψεις
y_pred_wind = reg_wind.predict(X_test_wind) # Χρήση του εκπαιδευμένου μοντέλου γι
α προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_wind = mean_absolute_error(y_test_wind, y_pred_wind) # Υπολογισμός του μέσο
υ απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of wind speed - XGBoost:" + str(error_wi
nd)) # Εκτύπωση του μέσου απόλυτου σφάλματος

# Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την π
ίεση
X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες
μεταβλητές)
y_precip = data[['prcp']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_split(X_
precip, y_precip, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33%
για δοκιμαστικό σετ

# Εφαρμογή XGBoost
reg_precip = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed=1
23).fit(X_train_precip, y_train_precip) # Εκπαίδευση του μοντέλου XGBoost με 10 δ
έντρα

# Προβλέψεις
y_pred_precip = reg_precip.predict(X_test_precip) # Χρήση του εκπαιδευμένου μοντέ
λου για προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_precip = mean_absolute_error(y_test_precip, y_pred_precip) # Υπολογισμός το
υ μέσου απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of Monthly Precipitation total - XGBoost
:" + str(error_precip)) # Εκτύπωση του μέσου απόλυτου σφάλματος

# Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου κα
ι την κατακρήμνιση
X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητε
ς μεταβλητές)
y_pressure = data[['pres']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

```

```

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure = train_test_
split(X_pressure, y_pressure, test_size=0.33, random_state=42) # Διαίρεση δεδομέν
ων με το 33% για δοκιμαστικό σετ

# Εφαρμογή XGBoost
reg_pressure = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed
=123).fit(X_train_pressure, y_train_pressure) # Εκπαίδευση του μοντέλου XGBoost μ
ε 10 δέντρα

# Προβλέψεις
y_pred_pressure = reg_pressure.predict(X_test_pressure) # Χρήση του εκπαιδευμένου
μοντέλου για προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_pressure = mean_absolute_error(y_test_pressure, y_pred_pressure) # Υπολογισ
μός του μέσου απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of pressure - XGBoost:" + str(error_pres
sure)) # Εκτύπωση του μέσου απόλυτου σφάλματος

Mean absolute error in predictions of mean temperature -
XGBoost:4.968115572495893
Mean absolute error in predictions of wind speed -
XGBoost:2.903117922317883
Mean absolute error in predictions of Monthly Precipitation total -
XGBoost:1.3577709051603375
Mean absolute error in predictions of pressure -
XGBoost:29.184360371739412

# Εισαγωγή της βιβλιοθήκης για τον ενισχυτή παλινδρόμησης AdaBoost
from sklearn.ensemble import AdaBoostRegressor # Εισαγωγή της κλάσης AdaBoostRegr
essor για την εκτέλεση ενισχυμένης παλινδρόμησης

# Πρόβλεψη της μέσης θερμοκρασίας βασισμένη στην πίεση, την ταχύτητα ανέμου και τη
ν κατακρήμνιση
X_temp = data[['wspd', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες με
ταβλητές)
y_temp = data['tavg'] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_temp, X_test_temp, y_train_temp, y_test_temp = train_test_split(X_temp, y_
temp, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμασ
τικό σετ

# Εφαρμογή AdaBoost
reg_temp = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_temp, y_
train_temp) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα

# Προβλέψεις
y_pred_temp = reg_temp.predict(X_test_temp) # Χρήση του εκπαιδευμένου μοντέλου γι
α προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_temp = mean_absolute_error(y_test_temp, y_pred_temp) # Υπολογισμός του μέσο

```

υ απόλυτου σφάλματος

Εκτύπωση του σφάλματος

```
print("Mean absolute error in predictions of mean temperature - AdaBoost:" + str(error_temp)) # Εκτύπωση του μέσου απόλυτου σφάλματος
```

Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμνιση

```
X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες μεταβλητές)
```

```
y_wind = data[['wspd']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
```

Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ

```
X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_wind, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
```

Εφαρμογή AdaBoost

```
reg_wind = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα
```

Προβλέψεις

```
y_pred_wind = reg_wind.predict(X_test_wind) # Χρήση του εκπαιδευμένου μοντέλου για προβλέψεις στο δοκιμαστικό σετ
```

Υπολογισμός σφάλματος

```
error_wind = mean_absolute_error(y_test_wind, y_pred_wind) # Υπολογισμός του μέσου απόλυτου σφάλματος
```

Εκτύπωση του σφάλματος

```
print("Mean absolute error in predictions of wind speed - AdaBoost:" + str(error_wind)) # Εκτύπωση του μέσου απόλυτου σφάλματος
```

Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την πίεση

```
X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες μεταβλητές)
```

```
y_precip = data[['prcp']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)
```

Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ

```
X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_split(X_precip, y_precip, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ
```

Εφαρμογή AdaBoost

```
reg_precip = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_precip, y_train_precip) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα
```

Προβλέψεις

```
y_pred_precip = reg_precip.predict(X_test_precip) # Χρήση του εκπαιδευμένου μοντέλου για προβλέψεις στο δοκιμαστικό σετ
```

Υπολογισμός σφάλματος

```
error_precip = mean_absolute_error(y_test_precip, y_pred_precip) # Υπολογισμός του μέσου απόλυτου σφάλματος
```

Εκτύπωση του σφάλματος

```
print("Mean absolute error in predictions of Monthly Precipitation total - AdaBoost:" + str(error_precip)) # Εκτύπωση του μέσου απόλυτου σφάλματος
```

```

# Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την κατακρήμνιση
X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητες μεταβλητές)
y_pressure = data[['pres']] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure = train_test_split(X_pressure, y_pressure, test_size=0.33, random_state=42) # Διαίρεση δεδομένων με το 33% για δοκιμαστικό σετ

# Εφαρμογή AdaBoost
reg_pressure = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_pressure, y_train_pressure) # Εκπαίδευση του μοντέλου AdaBoost με 10 δέντρα

# Προβλέψεις
y_pred_pressure = reg_pressure.predict(X_test_pressure) # Χρήση του εκπαιδευμένου μοντέλου για προβλέψεις στο δοκιμαστικό σετ

# Υπολογισμός σφάλματος
error_pressure = mean_absolute_error(y_test_pressure, y_pred_pressure) # Υπολογισμός του μέσου απόλυτου σφάλματος

# Εκτύπωση του σφάλματος
print("Mean absolute error in predictions of pressure - AdaBoost:" + str(error_pressure)) # Εκτύπωση του μέσου απόλυτου σφάλματος

Mean absolute error in predictions of mean temperature -
AdaBoost:5.0858560549248235
Mean absolute error in predictions of wind speed -
AdaBoost:2.879222720818808
Mean absolute error in predictions of Monthly Precipitation total -
AdaBoost:1.6823025461273449
Mean absolute error in predictions of pressure -
AdaBoost:3.1451020838958046

# Συνάρτηση για συνδυασμό προβλέψεων
def combine_predictions(y_pred_lr, y_pred_xg, y_pred_ada, error_lr, error_xg, error_ada):
    # Υπολογισμός του συνολικού σφάλματος από όλα τα μοντέλα
    total_error = error_lr + error_xg + error_ada

    # Υπολογισμός του βάρους για το μοντέλο Linear Regression
    weight_lr = error_lr / total_error

    # Υπολογισμός του βάρους για το μοντέλο XGBoost
    weight_xg = error_xg / total_error

    # Υπολογισμός του βάρους για το μοντέλο AdaBoost
    weight_ada = error_ada / total_error

    # Συνδυασμός των προβλέψεων πολλαπλασιάζοντας κάθε πρόβλεψη με το αντίστοιχο βάρος της
    combined_predictions = (weight_lr * y_pred_lr) + (weight_xg * y_pred_xg) + (weight_ada * y_pred_ada)

```

```

# Επιστροφή των συνδυασμένων προβλέψεων
return combined_predictions

import matplotlib.pyplot as plt # Βιβλιοθήκη για δημιουργία γραφημάτων
import xgboost as xg # Εισαγωγή της βιβλιοθήκης XGBoost
from sklearn.metrics import mean_absolute_error # Συνάρτηση για υπολογισμό του μέ
σου απόλυτου σφάλματος
from sklearn.linear_model import LinearRegression # Κλάση για εκτέλεση γραμμικής
παλινδρόμησης
from sklearn.ensemble import AdaBoostRegressor # Κλάση για εκτέλεση ενισχυμένης π
αλινδρόμησης (AdaBoost)

# Γραμμική Παλινδρόμηση
# Εκπαίδευση του μοντέλου γραμμικής παλινδρόμησης με τα εκπαιδευτικά δεδομένα
reg_temp_lr = LinearRegression().fit(X_train_temp, y_train_temp)
# Πρόβλεψη με το εκπαιδευμένο μοντέλο γραμμικής παλινδρόμησης
y_pred_temp_lr = reg_temp_lr.predict(X_test_temp)
# Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
error_temp_lr = mean_absolute_error(y_test_temp, y_pred_temp_lr)
print("Mean absolute error in predictions of mean temperature - Linear Regression:
", error_temp_lr)

# XGBoost
# Εκπαίδευση του μοντέλου XGBoost με τα εκπαιδευτικά δεδομένα
reg_temp_xg = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed=
123).fit(X_train_temp, y_train_temp)
# Πρόβλεψη με το εκπαιδευμένο μοντέλο XGBoost
y_pred_temp_xg = reg_temp_xg.predict(X_test_temp)
# Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
error_temp_xg = mean_absolute_error(y_test_temp, y_pred_temp_xg)
print("Mean absolute error in predictions of mean temperature - XGBoost:", error_t
emp_xg)

# AdaBoost
# Εκπαίδευση του μοντέλου AdaBoost με τα εκπαιδευτικά δεδομένα
reg_temp_ada = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_temp
, y_train_temp)
# Πρόβλεψη με το εκπαιδευμένο μοντέλο AdaBoost
y_pred_temp_ada = reg_temp_ada.predict(X_test_temp)
# Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
error_temp_ada = mean_absolute_error(y_test_temp, y_pred_temp_ada)
print("Mean absolute error in predictions of mean temperature - AdaBoost:", error_
temp_ada)

# Συνδυασμός προβλέψεων για θερμοκρασία
# Συνδυασμός των προβλέψεων με βάση τα σφάλματα τους για τη δημιουργία ενός μετα-
μοντέλου
combined_pred_temp = combine_predictions(y_pred_temp_lr, y_pred_temp_xg, y_pred_te
mp_ada, error_temp_lr, error_temp_xg, error_temp_ada)
# Υπολογισμός του μέσου απόλυτου σφάλματος για τις συνδυασμένες προβλέψεις
combined_error_temp = mean_absolute_error(y_test_temp, combined_pred_temp)
print("Mean absolute error in predictions of mean temperature - Meta-
model:", combined_error_temp)

# Δημιουργία γραφημάτων για θερμοκρασία
# Δημιουργία γραφήματος για τη σύγκριση των πραγματικών τιμών με τις προβλέψεις το
υ μοντέλου γραμμικής παλινδρόμησης
plt.figure(figsize=(10, 6))

```

```
plt.plot(y_test_temp.values, label='Actual', color='black')
plt.plot(y_pred_temp_lr, label='Linear Regression Predictions', linestyle='dashed'
)
plt.title('Temperature Predictions - Linear Regression')
plt.xlabel('Samples')
plt.ylabel('Temperature')
plt.legend()
plt.show()
```

*# Δημιουργία γραφήματος για τη σύγκριση των πραγματικών τιμών με τις προβλέψεις το
υ μοντέλου XGBoost*

```
plt.figure(figsize=(10, 6))
plt.plot(y_test_temp.values, label='Actual', color='black')
plt.plot(y_pred_temp_xg, label='XGBoost Predictions', linestyle='dotted')
plt.title('Temperature Predictions - XGBoost')
plt.xlabel('Samples')
plt.ylabel('Temperature')
plt.legend()
plt.show()
```

*# Δημιουργία γραφήματος για τη σύγκριση των πραγματικών τιμών με τις προβλέψεις το
υ μοντέλου AdaBoost*

```
plt.figure(figsize=(10, 6))
plt.plot(y_test_temp.values, label='Actual', color='black')
plt.plot(y_pred_temp_ada, label='AdaBoost Predictions', linestyle='dashdot')
plt.title('Temperature Predictions - AdaBoost')
plt.xlabel('Samples')
plt.ylabel('Temperature')
plt.legend()
plt.show()
```

*# Δημιουργία γραφήματος για τη σύγκριση των πραγματικών τιμών με τις συνδυασμένες
προβλέψεις του μετα-μοντέλου*

```
plt.figure(figsize=(10, 6))
plt.plot(y_test_temp.values, label='Actual', color='black')
plt.plot(combined_pred_temp, label='Meta-model Predictions', linestyle='solid')
plt.title('Temperature Predictions - Meta-model')
plt.xlabel('Samples')
plt.ylabel('Temperature')
plt.legend()
plt.show()
```

Mean absolute error in predictions of mean temperature - Linear Regression:

5.398268648600534

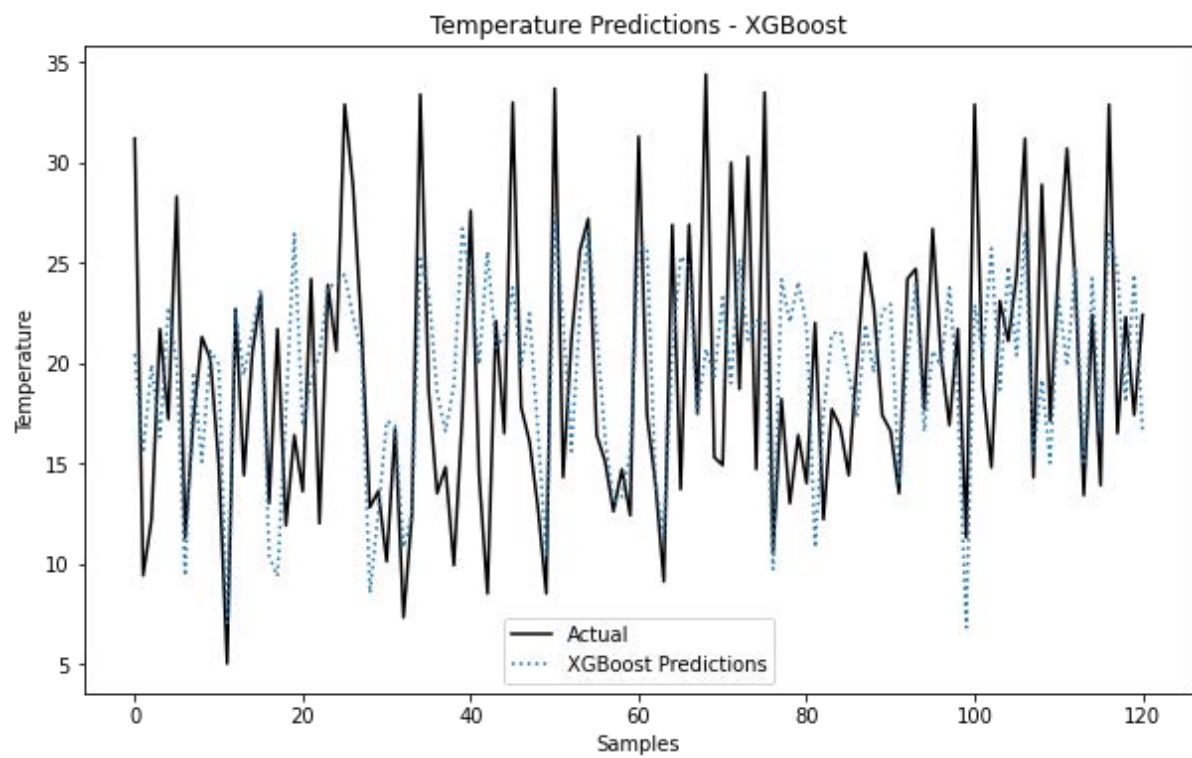
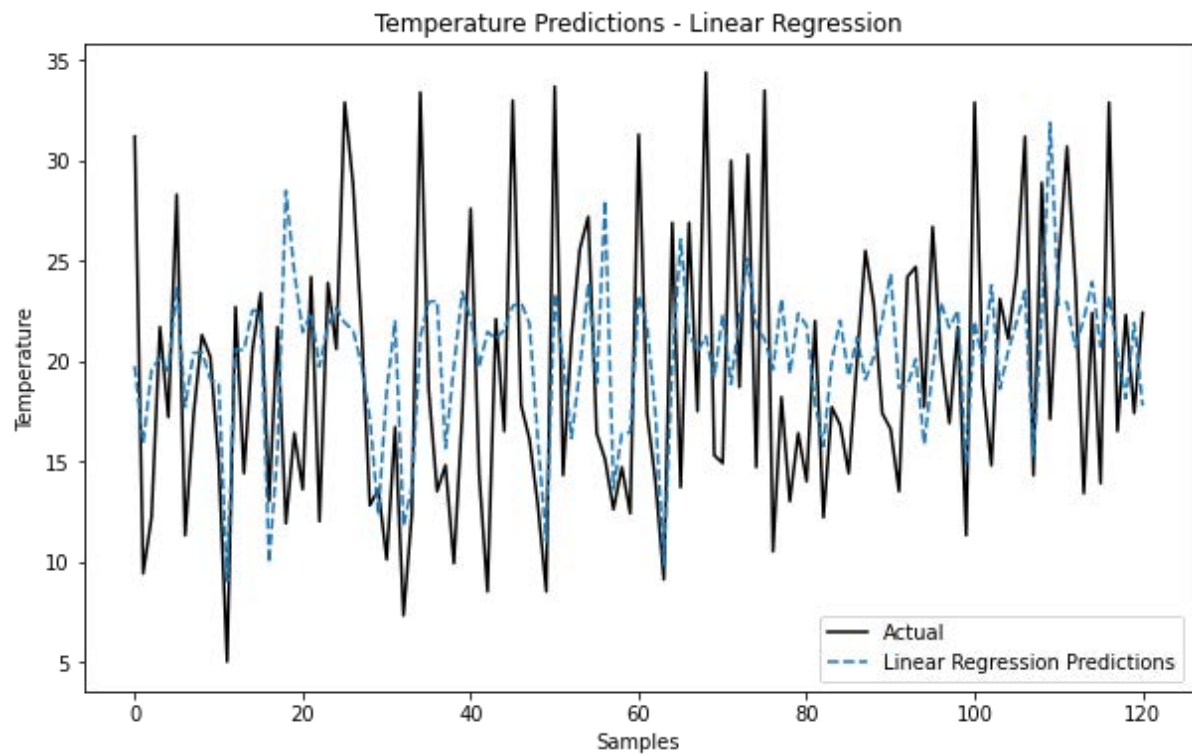
Mean absolute error in predictions of mean temperature - XGBoost:

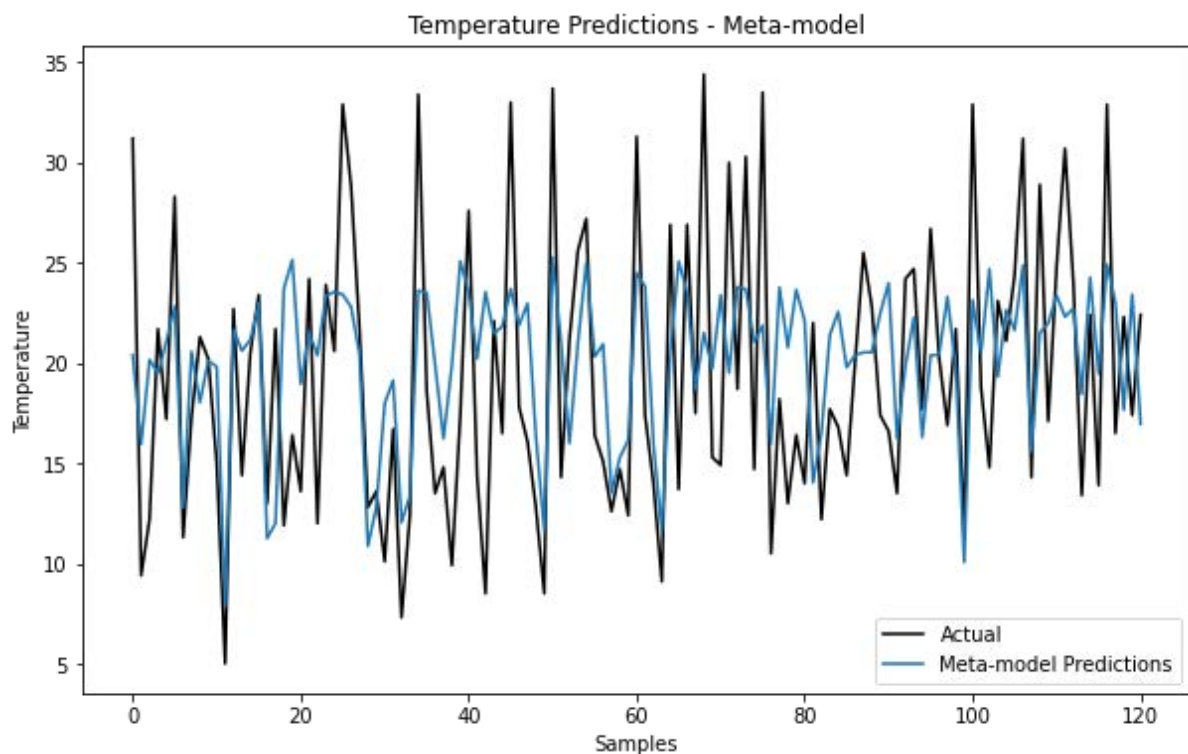
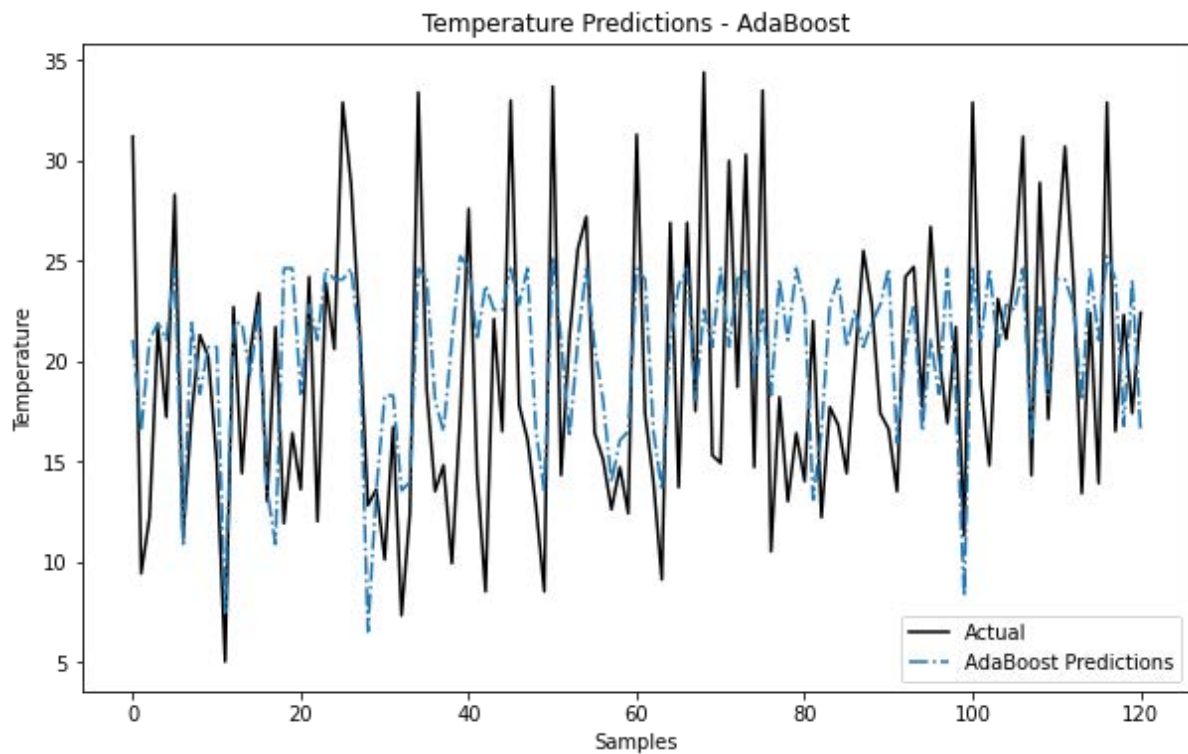
4.968115572495893

Mean absolute error in predictions of mean temperature - AdaBoost:

5.0858560549248235

Mean absolute error in predictions of mean temperature - Meta-model:





```
import matplotlib.pyplot as plt # Βιβλιοθήκη για δημιουργία γραφημάτων
import xgboost as xg # Εισαγωγή της βιβλιοθήκης XGBoost
from sklearn.model_selection import train_test_split # Συνάρτηση για διαίρεση δεδομένων σε εκπαιδευτικά και δοκιμαστικά σετ
from sklearn.metrics import mean_absolute_error # Συνάρτηση για υπολογισμό του μέσου απόλυτου σφάλματος
from sklearn.linear_model import LinearRegression # Κλάση για εκτέλεση γραμμικής παλινδρόμησης
from sklearn.ensemble import AdaBoostRegressor # Κλάση για εκτέλεση ενισχυμένης παλινδρόμησης (AdaBoost)
```

```

# Πρόβλεψη ταχύτητας ανέμου βασισμένη στην θερμοκρασία, την πίεση και την κατακρήμνιση
X_wind = data[['tavg', 'prcp', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτητες μεταβλητές)
y_wind = data['wspd'] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_wind, X_test_wind, y_train_wind, y_test_wind = train_test_split(X_wind, y_wind, test_size=0.33, random_state=42)

# Γραμμική Παλινδρόμηση
reg_wind_lr = LinearRegression().fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου γραμμικής παλινδρόμησης με τα εκπαιδευτικά δεδομένα
y_pred_wind_lr = reg_wind_lr.predict(X_test_wind) # Πρόβλεψη με το εκπαιδευμένο μοντέλο γραμμικής παλινδρόμησης
error_wind_lr = mean_absolute_error(y_test_wind, y_pred_wind_lr) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of wind speed - Linear Regression:", error_wind_lr)

# XGBoost
reg_wind_xg = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed=123).fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου XGBoost με τα εκπαιδευτικά δεδομένα
y_pred_wind_xg = reg_wind_xg.predict(X_test_wind) # Πρόβλεψη με το εκπαιδευμένο μοντέλο XGBoost
error_wind_xg = mean_absolute_error(y_test_wind, y_pred_wind_xg) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of wind speed - XGBoost:", error_wind_xg)

# AdaBoost
reg_wind_ada = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_wind, y_train_wind) # Εκπαίδευση του μοντέλου AdaBoost με τα εκπαιδευτικά δεδομένα
y_pred_wind_ada = reg_wind_ada.predict(X_test_wind) # Πρόβλεψη με το εκπαιδευμένο μοντέλο AdaBoost
error_wind_ada = mean_absolute_error(y_test_wind, y_pred_wind_ada) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of wind speed - AdaBoost:", error_wind_ada)

# Συνδυασμός προβλέψεων για Ταχύτητα Ανέμου
combined_pred_wind = combine_predictions(y_pred_wind_lr, y_pred_wind_xg, y_pred_wind_ada, error_wind_lr, error_wind_xg, error_wind_ada) # Συνδυασμός των προβλέψεων με βάση τα σφάλματα τους για τη δημιουργία ενός μετα-μοντέλου
combined_error_wind = mean_absolute_error(y_test_wind, combined_pred_wind) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις συνδυασμένες προβλέψεις
print("Mean absolute error in predictions of wind speed - Meta-model:", combined_error_wind)

# Δημιουργία γραφημάτων για Ταχύτητα Ανέμου
plt.figure(figsize=(10, 6))
plt.plot(y_test_wind.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_wind_lr, label='Linear Regression Predictions', linestyle='dashed') # Προβλέψεις γραμμικής παλινδρόμησης
plt.title('Wind Speed Predictions - Linear Regression')
plt.xlabel('Samples')
plt.ylabel('Wind Speed')

```

```

plt.legend()
plt.show()

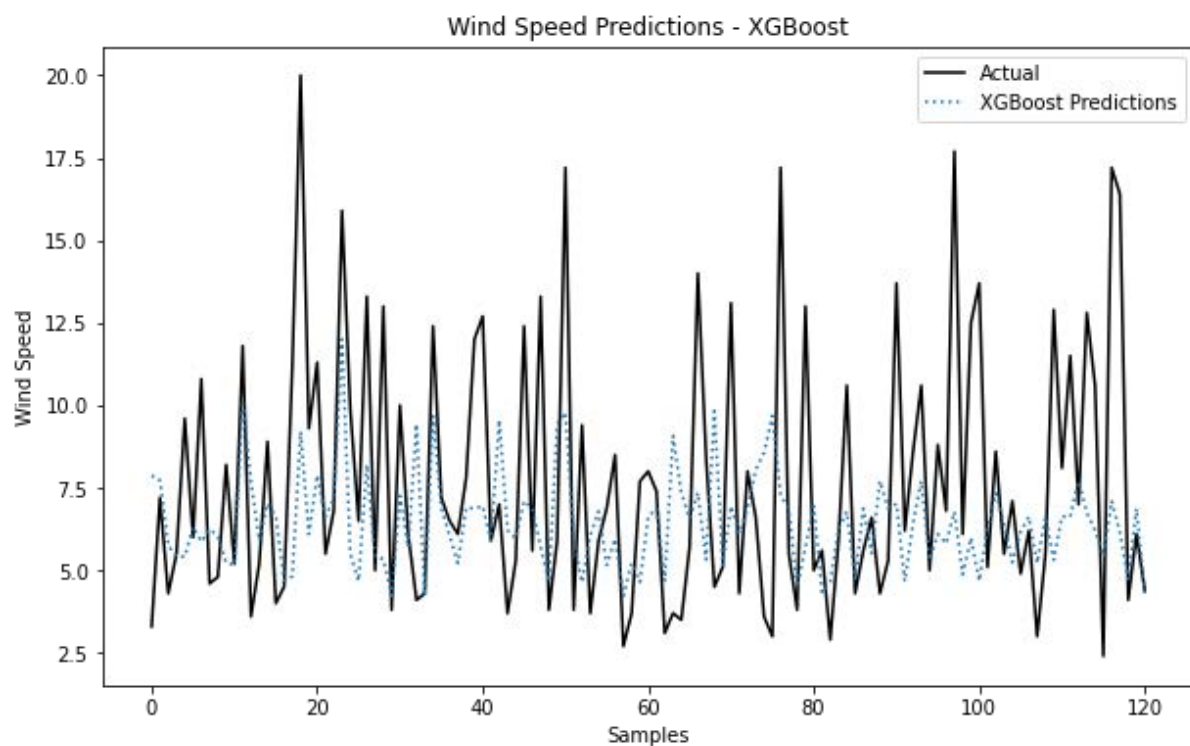
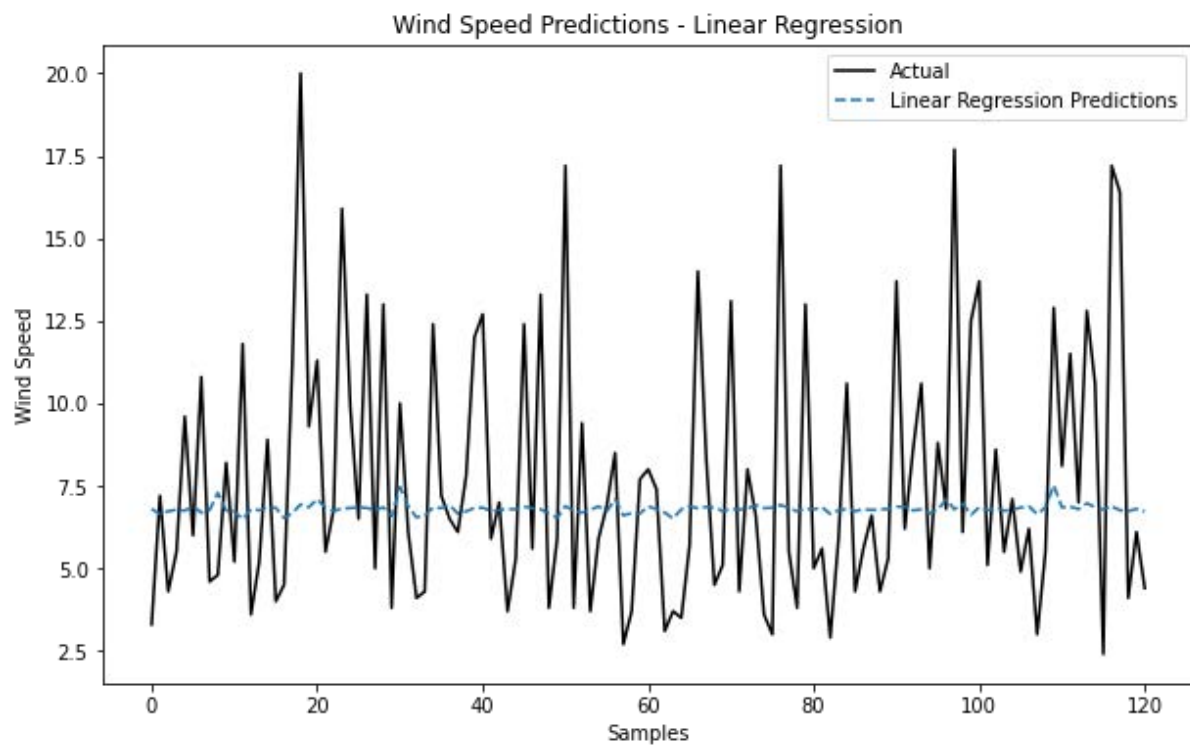
plt.figure(figsize=(10, 6))
plt.plot(y_test_wind.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_wind_xg, label='XGBoost Predictions', linestyle='dotted') # Προβλ
έψεις XGBoost
plt.title('Wind Speed Predictions - XGBoost')
plt.xlabel('Samples')
plt.ylabel('Wind Speed')
plt.legend()
plt.show()

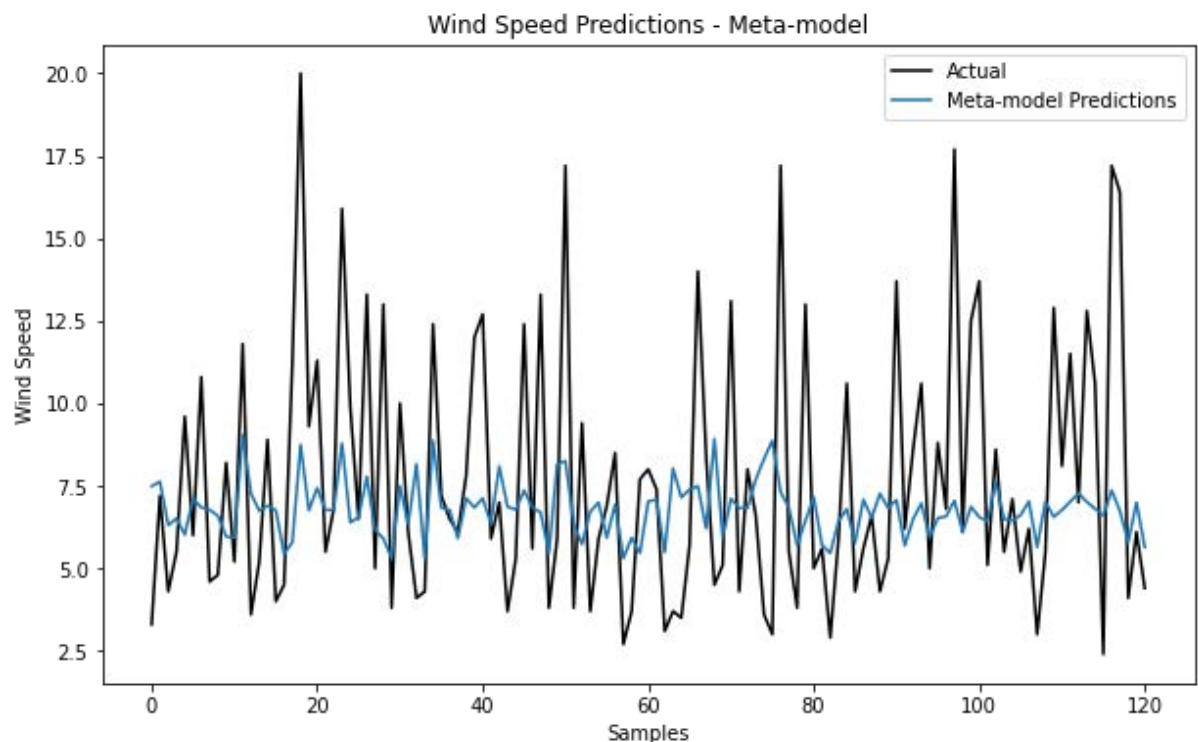
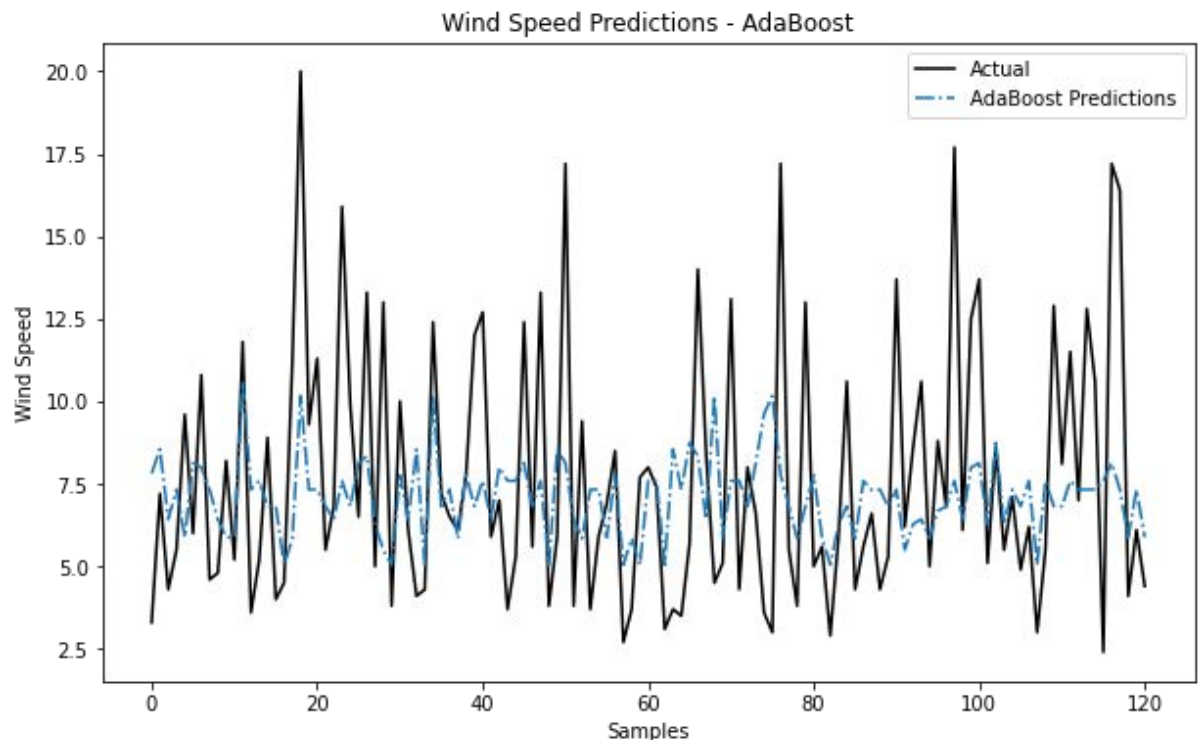
plt.figure(figsize=(10, 6))
plt.plot(y_test_wind.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_wind_ada, label='AdaBoost Predictions', linestyle='dashdot') # Πρ
οβλέψεις AdaBoost
plt.title('Wind Speed Predictions - AdaBoost')
plt.xlabel('Samples')
plt.ylabel('Wind Speed')
plt.legend()
plt.show()

plt.figure(figsize=(10, 6))
plt.plot(y_test_wind.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(combined_pred_wind, label='Meta-
model Predictions', linestyle='solid') # Προβλέψεις μετα-μοντέλου
plt.title('Wind Speed Predictions - Meta-model')
plt.xlabel('Samples')
plt.ylabel('Wind Speed')
plt.legend()
plt.show()

Mean absolute error in predictions of wind speed - Linear Regression:
2.995384113513937
Mean absolute error in predictions of wind speed - XGBoost: 2.903117922317883
Mean absolute error in predictions of wind speed - AdaBoost: 2.879222720818808
Mean absolute error in predictions of wind speed - Meta-model: 2.872

```





```
import matplotlib.pyplot as plt # Βιβλιοθήκη για δημιουργία γραφημάτων
import xgboost as xg # Εισαγωγή της βιβλιοθήκης XGBoost
from sklearn.model_selection import train_test_split # Συνάρτηση για διαίρεση δεδομένων σε εκπαιδευτικά και δοκιμαστικά σετ
from sklearn.metrics import mean_absolute_error # Συνάρτηση για υπολογισμό του μέσου απόλυτου σφάλματος
from sklearn.linear_model import LinearRegression # Κλάση για εκτέλεση γραμμικής παλινδρόμησης
from sklearn.ensemble import AdaBoostRegressor # Κλάση για εκτέλεση ενισχυμένης παλινδρόμησης (AdaBoost)
```

```

# Πρόβλεψη ατμοσφαιρικής πίεσης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και την κατακρήμνιση
X_pressure = data[['wspd', 'prcp', 'tavg']] # Επιλογή χαρακτηριστικών (ανεξάρτητες μεταβλητές)
y_pressure = data['pres'] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_pressure, X_test_pressure, y_train_pressure, y_test_pressure = train_test_split(X_pressure, y_pressure, test_size=0.33, random_state=42)

# Γραμμική Παλινδρόμηση
reg_pressure_lr = LinearRegression().fit(X_train_pressure, y_train_pressure) # Εκπαίδευση του μοντέλου γραμμικής παλινδρόμησης με τα εκπαιδευτικά δεδομένα
y_pred_pressure_lr = reg_pressure_lr.predict(X_test_pressure) # Πρόβλεψη με το εκπαιδευμένο μοντέλο γραμμικής παλινδρόμησης
error_pressure_lr = mean_absolute_error(y_test_pressure, y_pred_pressure_lr) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of pressure - Linear Regression:", error_pressure_lr)

# XGBoost
reg_pressure_xg = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10, seed=123).fit(X_train_pressure, y_train_pressure) # Εκπαίδευση του μοντέλου XGBoost με τα εκπαιδευτικά δεδομένα
y_pred_pressure_xg = reg_pressure_xg.predict(X_test_pressure) # Πρόβλεψη με το εκπαιδευμένο μοντέλο XGBoost
error_pressure_xg = mean_absolute_error(y_test_pressure, y_pred_pressure_xg) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of pressure - XGBoost:", error_pressure_xg)

# AdaBoost
reg_pressure_ada = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_pressure, y_train_pressure) # Εκπαίδευση του μοντέλου AdaBoost με τα εκπαιδευτικά δεδομένα
y_pred_pressure_ada = reg_pressure_ada.predict(X_test_pressure) # Πρόβλεψη με το εκπαιδευμένο μοντέλο AdaBoost
error_pressure_ada = mean_absolute_error(y_test_pressure, y_pred_pressure_ada) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of pressure - AdaBoost:", error_pressure_ada)

# Συνδυασμός προβλέψεων για Πίεση
combined_pred_pressure = combine_predictions(y_pred_pressure_lr, y_pred_pressure_xg, y_pred_pressure_ada, error_pressure_lr, error_pressure_xg, error_pressure_ada)
# Συνδυασμός των προβλέψεων με βάση τα σφάλματα τους για τη δημιουργία ενός μετα-μοντέλου
combined_error_pressure = mean_absolute_error(y_test_pressure, combined_pred_pressure) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις συνδυασμένες προβλέψεις
print("Mean absolute error in predictions of pressure - Meta-model:", combined_error_pressure)

# Δημιουργία γραφημάτων για Πίεση
plt.figure(figsize=(10, 6))
plt.plot(y_test_pressure.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_pressure_lr, label='Linear Regression Predictions', linestyle='dashed') # Προβλέψεις γραμμικής παλινδρόμησης
plt.title('Pressure Predictions - Linear Regression')

```

```

plt.xlabel('Samples')
plt.ylabel('Pressure')
plt.legend()
plt.show()

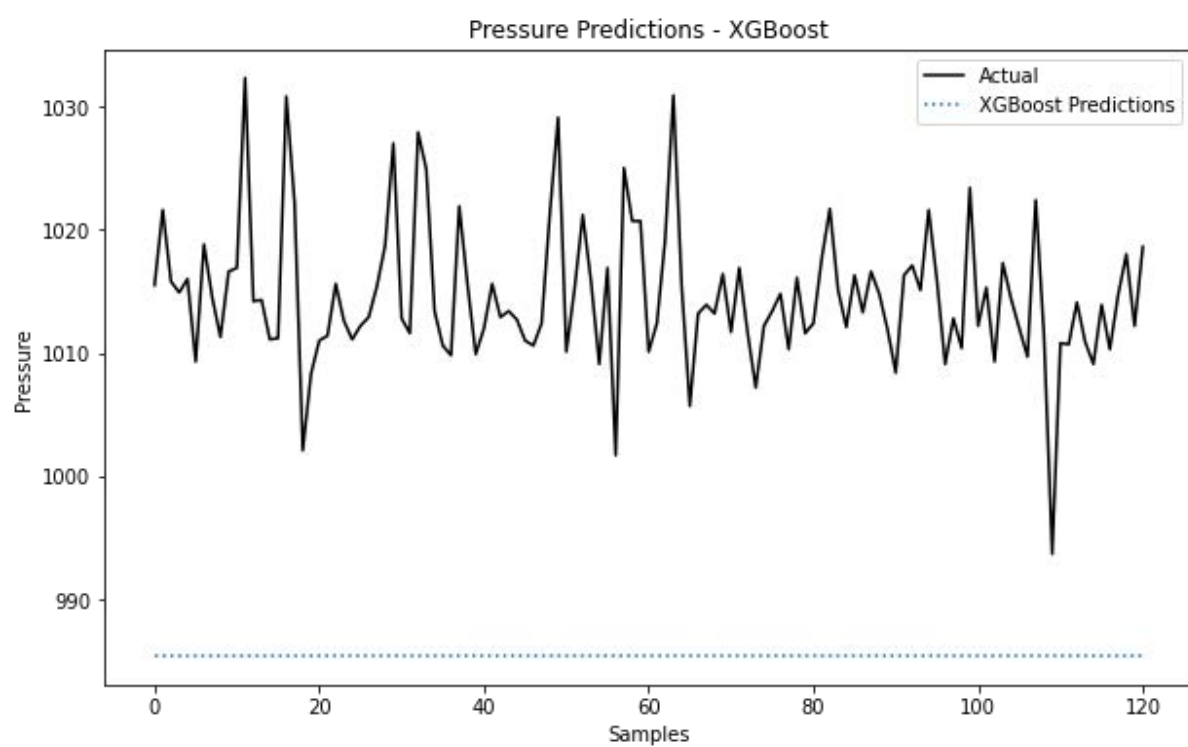
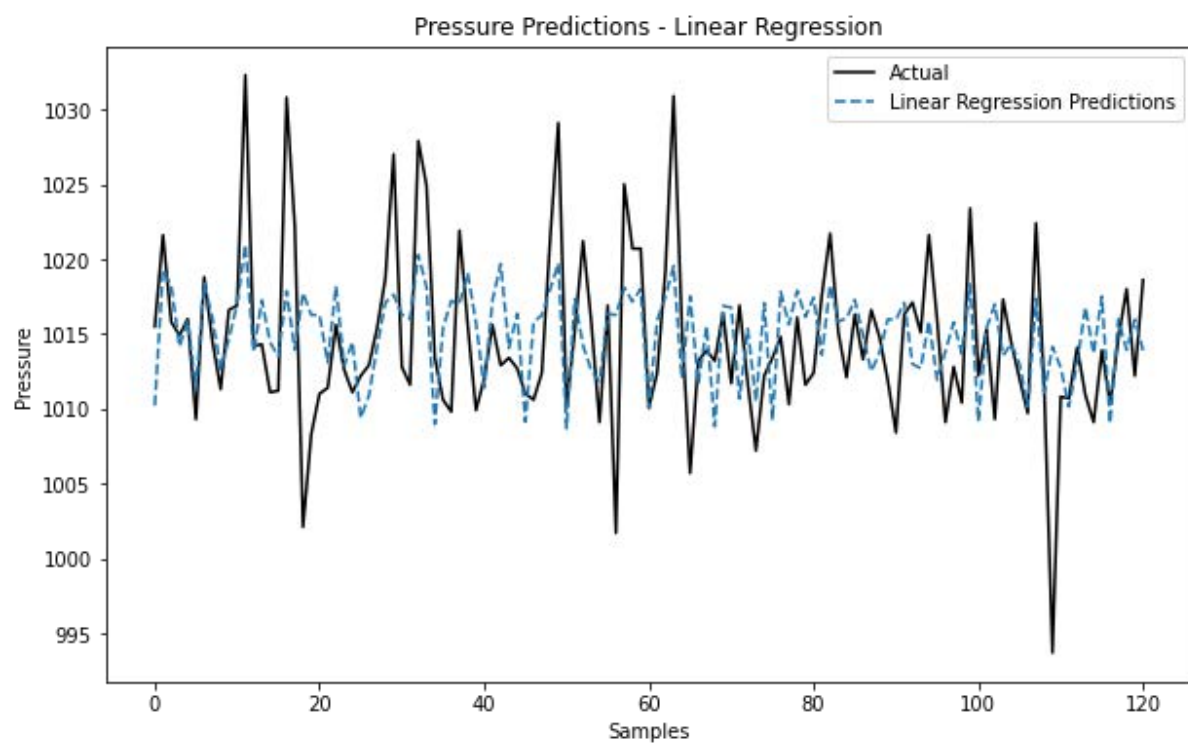
plt.figure(figsize=(10, 6))
plt.plot(y_test_pressure.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_pressure_xg, label='XGBoost Predictions', linestyle='dotted') # Προβλέψεις XGBoost
plt.title('Pressure Predictions - XGBoost')
plt.xlabel('Samples')
plt.ylabel('Pressure')
plt.legend()
plt.show()

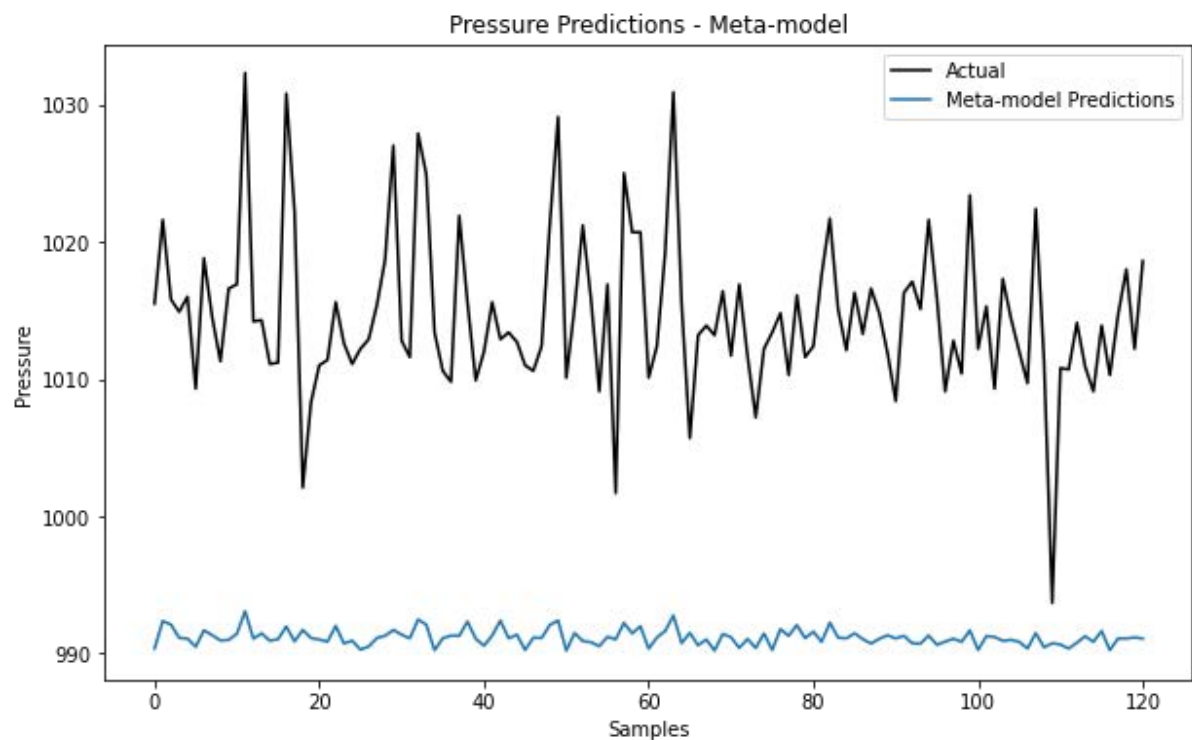
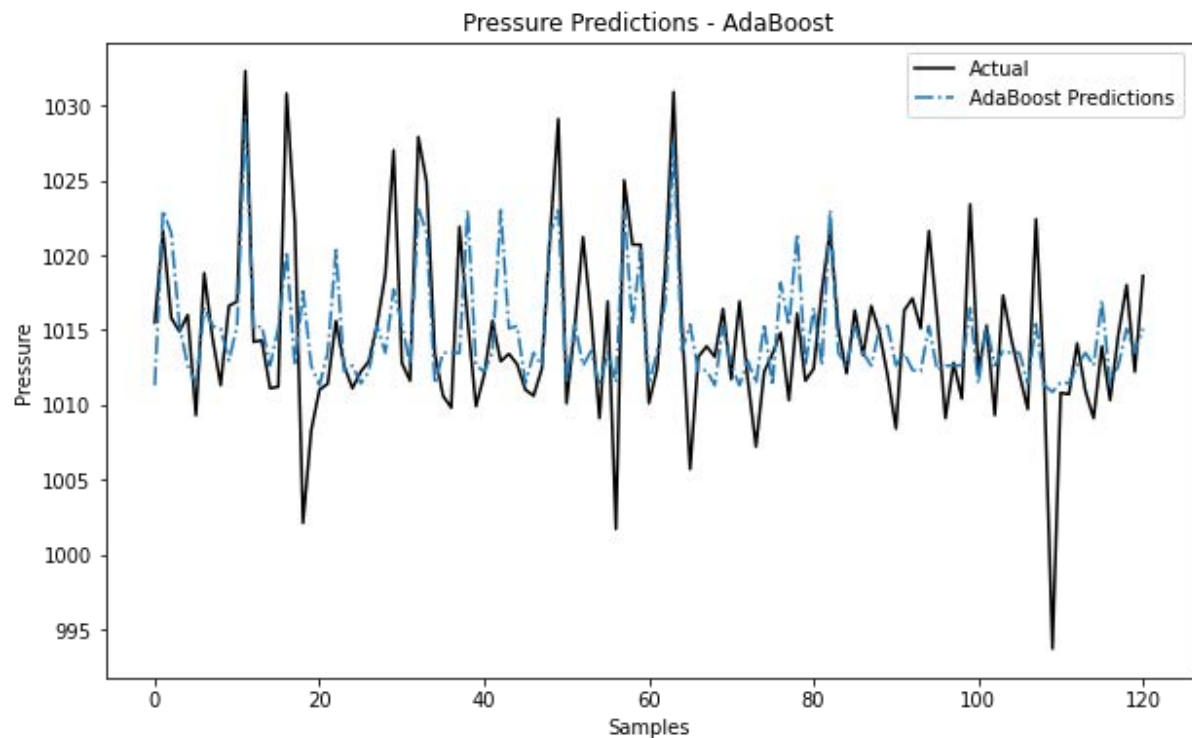
plt.figure(figsize=(10, 6))
plt.plot(y_test_pressure.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_pressure_ada, label='AdaBoost Predictions', linestyle='dashdot') # Προβλέψεις AdaBoost
plt.title('Pressure Predictions - AdaBoost')
plt.xlabel('Samples')
plt.ylabel('Pressure')
plt.legend()
plt.show()

plt.figure(figsize=(10, 6))
plt.plot(y_test_pressure.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(combined_pred_pressure, label='Meta-model Predictions', linestyle='solid') # Προβλέψεις μετα-μοντέλου
plt.title('Pressure Predictions - Meta-model')
plt.xlabel('Samples')
plt.ylabel('Pressure')
plt.legend()
plt.show()

Mean absolute error in predictions of pressure - Linear Regression:
3.9184893066817748
Mean absolute error in predictions of pressure - XGBoost: 29.184360371739412
Mean absolute error in predictions of pressure - AdaBoost: 3.1451020838958046
Mean absolute error in predictions of pressure - Meta-model: 23.453

```





```
import matplotlib.pyplot as plt # Βιβλιοθήκη για δημιουργία γραφημάτων
import xgboost as xg # Εισαγωγή της βιβλιοθήκης XGBoost
from sklearn.model_selection import train_test_split # Συνάρτηση για διαίρεση
# δεδομένων σε εκπαιδευτικά και δοκιμαστικά σετ
from sklearn.metrics import mean_absolute_error # Συνάρτηση για υπολογισμό το
# υ μέσου απόλυτου σφάλματος
from sklearn.linear_model import LinearRegression # Κλάση για εκτέλεση γραμμι
# κής παλινδρόμησης
from sklearn.ensemble import AdaBoostRegressor # Κλάση για εκτέλεση ενισχυμέν
# ης παλινδρόμησης (AdaBoost)
```

```

# Πρόβλεψη κατακρήμνισης βασισμένη στην θερμοκρασία, την ταχύτητα ανέμου και τ
ην πίεση
X_precip = data[['wspd', 'tavg', 'pres']] # Επιλογή χαρακτηριστικών (ανεξάρτη
τες μεταβλητές)
y_precip = data['prcp'] # Επιλογή στόχου (εξαρτημένη μεταβλητή)

# Διαχωρισμός δεδομένων σε εκπαιδευτικό και δοκιμαστικό σετ
X_train_precip, X_test_precip, y_train_precip, y_test_precip = train_test_spli
t(X_precip, y_precip, test_size=0.33, random_state=42)

# Γραμμική Παλινδρόμηση
reg_precip_lr = LinearRegression().fit(X_train_precip, y_train_precip) # Εκπα
ίδευση του μοντέλου γραμμικής παλινδρόμησης με τα εκπαιδευτικά δεδομένα
y_pred_precip_lr = reg_precip_lr.predict(X_test_precip) # Πρόβλεψη με το εκπα
ιδευμένο μοντέλο γραμμικής παλινδρόμησης
error_precip_lr = mean_absolute_error(y_test_precip, y_pred_precip_lr) # Υπολ
ογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of precipitation - Linear Regression
:", error_precip_lr)

# XGBoost
reg_precip_xg = xg.XGBRegressor(objective='reg:squarederror', n_estimators=10,
seed=123).fit(X_train_precip, y_train_precip) # Εκπαίδευση του μοντέλου XGBo
ost με τα εκπαιδευτικά δεδομένα
y_pred_precip_xg = reg_precip_xg.predict(X_test_precip) # Πρόβλεψη με το εκπα
ιδευμένο μοντέλο XGBoost
error_precip_xg = mean_absolute_error(y_test_precip, y_pred_precip_xg) # Υπολ
ογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of precipitation - XGBoost:", error_
precip_xg)

# AdaBoost
reg_precip_ada = AdaBoostRegressor(random_state=0, n_estimators=10).fit(X_train_precip, y_train_precip) # Εκπαίδευση του μοντέλου AdaBoost με τα εκπαιδευτικά δεδομένα
y_pred_precip_ada = reg_precip_ada.predict(X_test_precip) # Πρόβλεψη με το εκπαιδευμένο μοντέλο AdaBoost
error_precip_ada = mean_absolute_error(y_test_precip, y_pred_precip_ada) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις προβλέψεις
print("Mean absolute error in predictions of precipitation - AdaBoost:", error_precip_ada)

# Συνδυασμός προβλέψεων για Κατακρήμνιση
combined_pred_precip = combine_predictions(y_pred_precip_lr, y_pred_precip_xg, y_pred_precip_ada, error_precip_lr, error_precip_xg, error_precip_ada) # Συνδυασμός των προβλέψεων με βάση τα σφάλματα τους για τη δημιουργία ενός μετα-μοντέλου
combined_error_precip = mean_absolute_error(y_test_precip, combined_pred_precip) # Υπολογισμός του μέσου απόλυτου σφάλματος για τις συνδυασμένες προβλέψεις
print("Mean absolute error in predictions of precipitation - Meta-model:", combined_error_precip)

# Δημιουργία γραφημάτων για Κατακρήμνιση
plt.figure(figsize=(10, 6))
plt.plot(y_test_precip.values, label='Actual', color='black') # Πραγματικές τ

```

```

ιμέσ
plt.plot(y_pred_precip_lr, label='Linear Regression Predictions', linestyle='dashed') # Προβλέψεις γραμμικής παλινδρόμησης
plt.title('Precipitation Predictions - Linear Regression')
plt.xlabel('Samples')
plt.ylabel('Precipitation')
plt.legend()
plt.show()

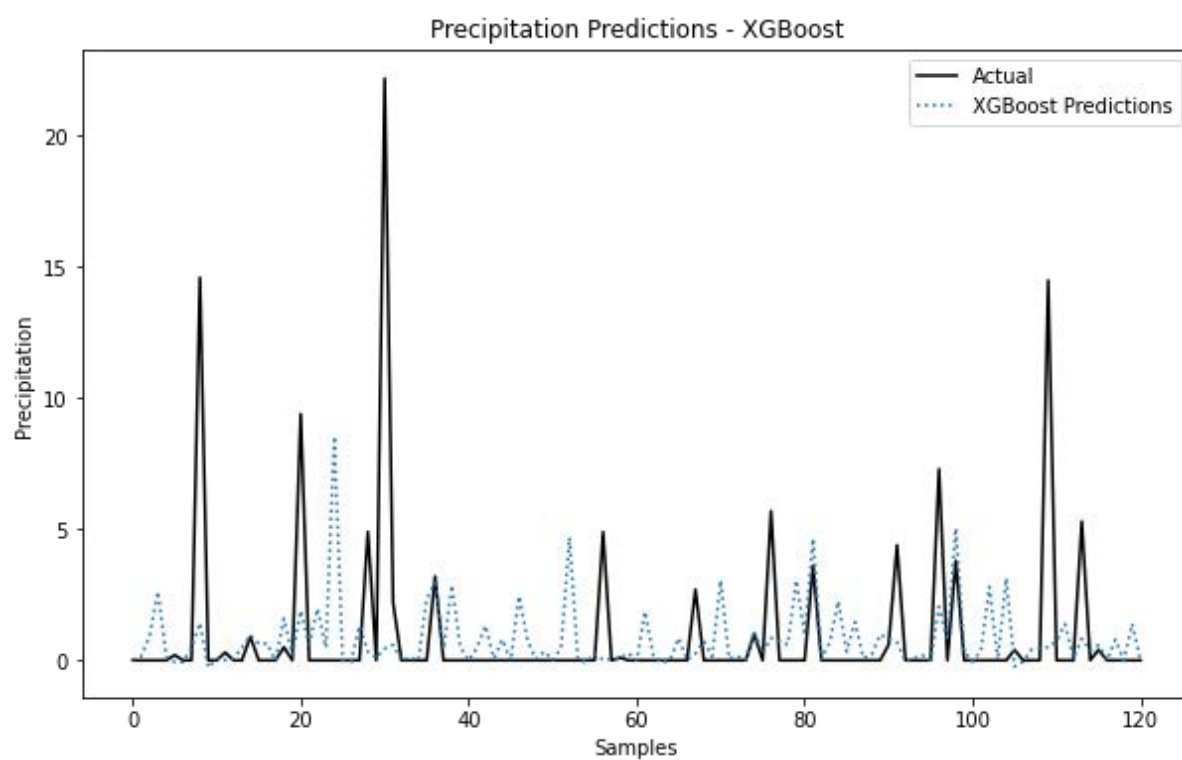
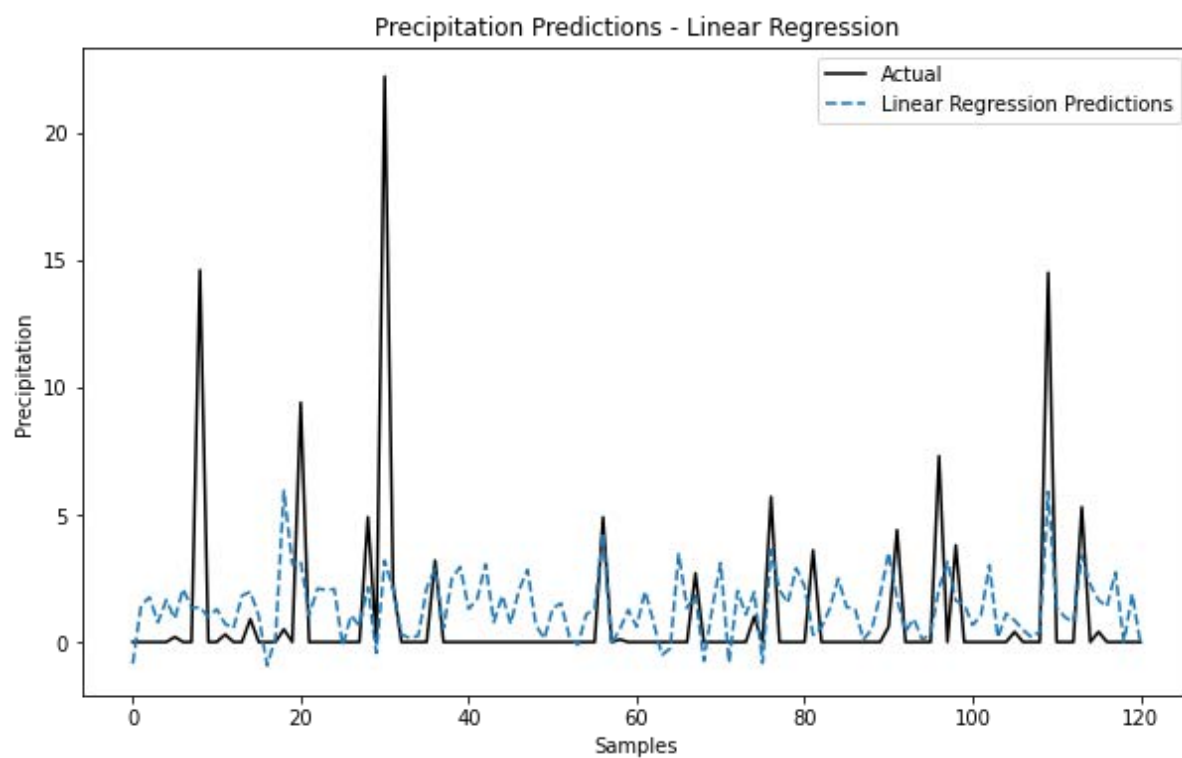
plt.figure(figsize=(10, 6))
plt.plot(y_test_precip.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_precip_xg, label='XGBoost Predictions', linestyle='dotted') # Προβλέψεις XGBoost
plt.title('Precipitation Predictions - XGBoost')
plt.xlabel('Samples')
plt.ylabel('Precipitation')
plt.legend()
plt.show()

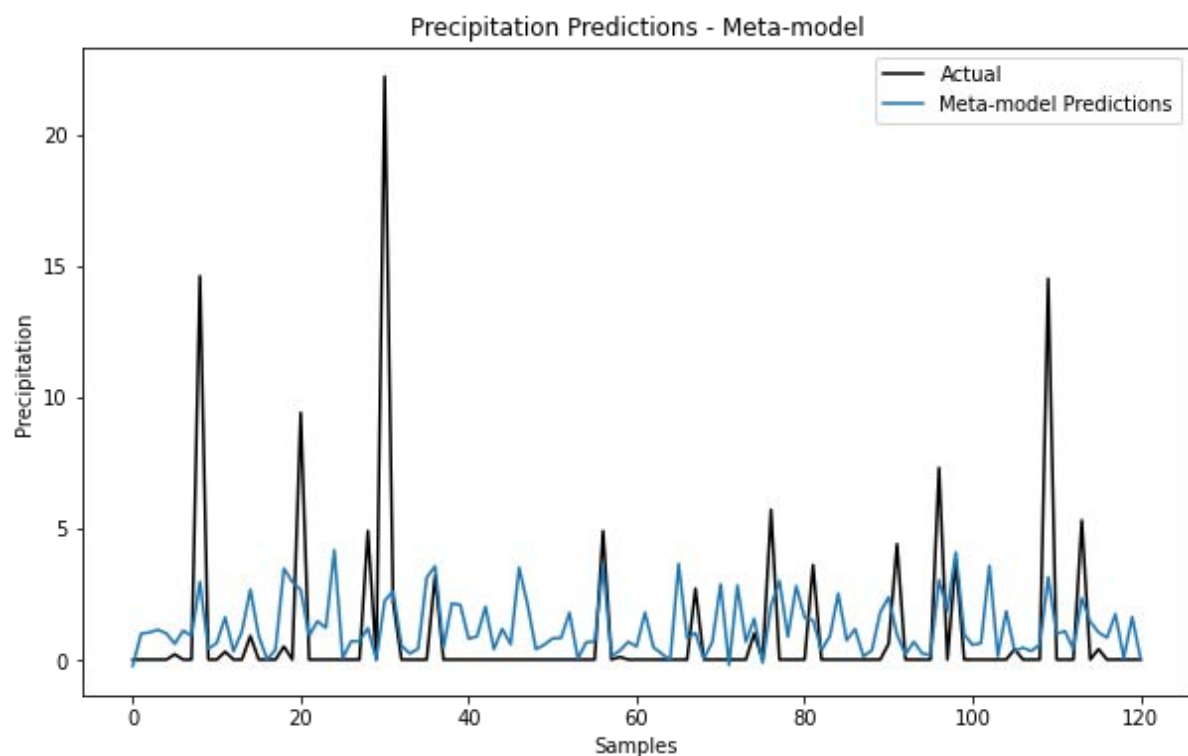
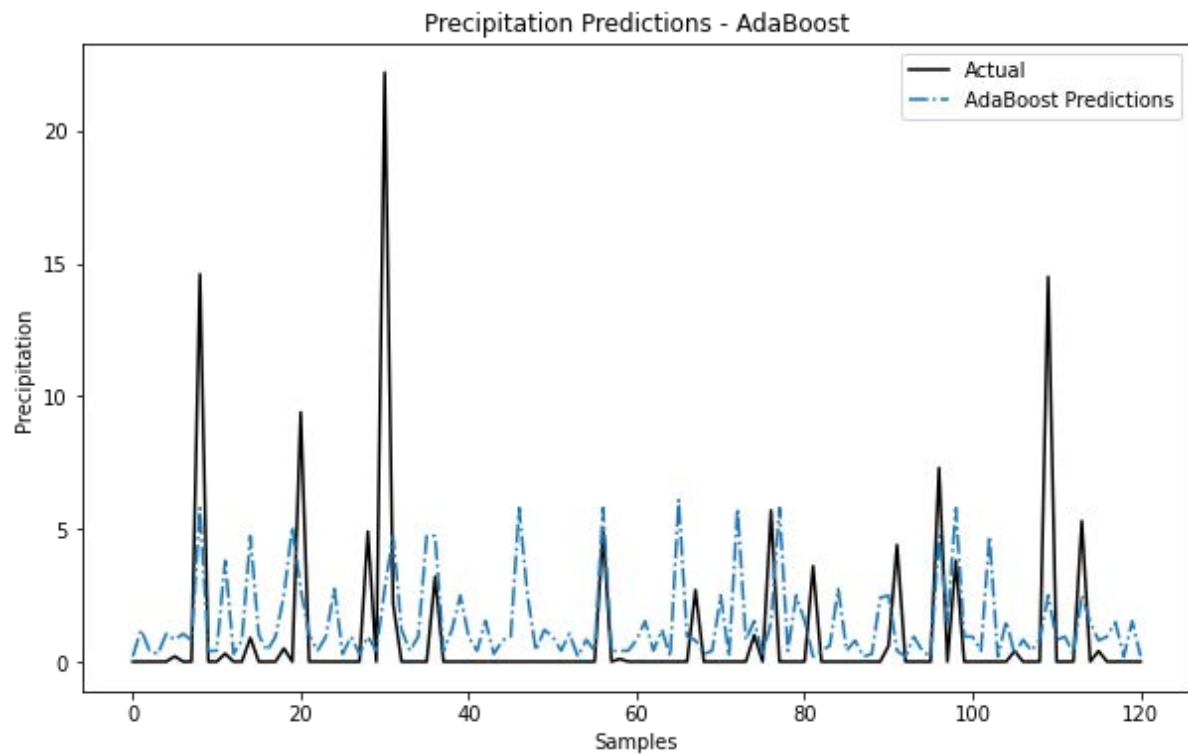
plt.figure(figsize=(10, 6))
plt.plot(y_test_precip.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(y_pred_precip_ada, label='AdaBoost Predictions', linestyle='dashdot') # Προβλέψεις AdaBoost
plt.title('Precipitation Predictions - AdaBoost')
plt.xlabel('Samples')
plt.ylabel('Precipitation')
plt.legend()
plt.show()

plt.figure(figsize=(10, 6))
plt.plot(y_test_precip.values, label='Actual', color='black') # Πραγματικές τιμές
plt.plot(combined_pred_precip, label='Meta-model Predictions', linestyle='solid') # Προβλέψεις μετα-μοντέλου
plt.title('Precipitation Predictions - Meta-model')
plt.xlabel('Samples')
plt.ylabel('Precipitation')
plt.legend()
plt.show()

Mean absolute error in predictions of precipitation - Linear Regression:
1.6906130646703028
Mean absolute error in predictions of precipitation - XGBoost:
1.3577709051603375
Mean absolute error in predictions of precipitation - AdaBoost:
1.6823025461273449
Mean absolute error in predictions of precipitation - Meta-model:
1.523568590818857

```





```
import pandas as pd
import matplotlib.pyplot as plt

# Συλλογή των σφαλμάτων για κάθε μοντέλο και κάθε μεταβλητή
errors = {
    'Variable': ['Temperature', 'Wind Speed', 'Precipitation', 'Pressure'],
    'Linear Regression': [error_temp_lr, error_wind_lr, error_precip_lr, error_pressure_lr],
    'XGBoost': [error_temp_xg, error_wind_xg, error_precip_xg, error_pressure_xg],
    'AdaBoost': [error_temp_ada, error_wind_ada, error_precip_ada, error_pressure_ada]
```

```

ada],
    'Meta-
model': [combined_error_temp, combined_error_wind, combined_error_precip, combined
_error_pressure]
}

# Δημιουργία DataFrame
errors_df = pd.DataFrame(errors)

# Εκτύπωση των σφαλμάτων
print(errors_df)

# Δημιουργία bar chart για τη σύγκριση των σφαλμάτων
errors_df.set_index('Variable').plot(kind='bar', figsize=(12, 8))
plt.title('Comparison of Model Errors for Different Variables')
plt.ylabel('Mean Absolute Error')
plt.xlabel('Variable')
plt.xticks(rotation=0)
plt.legend(title='Models')
plt.show()

```

	Variable	Linear Regression	XGBoost	AdaBoost	Meta-model
0	Temperature	5.398269	4.968116	5.085856	5.012752
1	Wind Speed	2.995384	2.903118	2.879223	2.872758
2	Precipitation	1.690613	1.357771	1.682303	1.523569
3	Pressure	3.918489	29.184360	3.145102	23.453031

