



Πανεπιστήμιο Θεσσαλίας  
Πολυτεχνική Σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

**Πρόβλεψη Παραγωγής Ηλεκτρικής Ενέργειας  
από Φωτοβολταϊκά Ηλιακά Πάνελ  
με Χρήση Τεχνικών Μηχανικής Μάθησης**

**Διπλωματική Εργασία**

**ΑΛΕΞΙΟΥ ΠΑΠΑΪΩΑΝΝΟΥ**

**Επιβλέπων**  
Μιχαήλ Βασιλακόπουλος  
Αναπληρωτής Καθηγητής

Βόλος, Μάρτιος 2019





Πανεπιστήμιο Θεσσαλίας

Πολυτεχνική Σχολή

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

# **Πρόβλεψη Παραγωγής Ηλεκτρικής Ενέργειας από Φωτοβολταϊκά Ηλιακά Πάνελ με Χρήση Τεχνικών Μηχανικής Μάθησης**

## **Διπλωματική Εργασία**

### **ΑΛΕΞΙΟΥ ΠΑΠΑΪΩΑΝΝΟΥ**

Επιτροπή επίβλεψης

Επιβλέπων

Μιχαήλ Βασιλακόπουλος  
Αναπληρωτής Καθηγητής

Συνεπιβλέπων

Δημήτριος Μπαργιώτας  
Αναπληρωτής Καθηγητής

Συνεπιβλέπουσα

Ασπασία Δασκαλοπούλου  
Επίκουρη Καθηγήτρια

Βόλος, Μάρτιος 2019



Πανεπιστήμιο Θεσσαλίας  
Πολυτεχνική Σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή / της φοιτήτριας που την εκπόνησε. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Ο/Η συγγραφέας αυτής της εργασίας βεβαιώνει ότι κάθε βοήθεια την οποία είχε για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης βεβαιώνει ότι έχει αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται επακριβώς, είτε παραφρασμένες.



University of Thessaly  
Faculty of Engineering  
Department of Electrical & Computer Engineering

# **Prediction of Photovoltaic Solar Panel Energy-Production using Machine Learning Techniques**

## **Diploma Thesis**

**ALEXIOS PAPAIOANNOU**

**Supervisor**

Michael Vassilakopoulos  
Associate Professor

Volos, March 2019



# Περίληψη

Τα τελευταία χρόνια, οι εξελίξεις και οι αλλαγές στην παραγωγή και στην κατανάλωση ηλεκτρικής ενέργειας είναι ραγδαίες. Η μόλυνση του περιβάλλοντος, τα προβλήματα υγείας που προκαλούνται από τα ορυκτά καύσιμα αλλά και η εξάντλησή τους αποτελούν ορισμένους παράγοντες που προωθούν τη χρήση ανανεώσιμων μορφών ενέργειας στην παραγωγή. Η Ελλάδα αποτελεί μία χώρα η οποία βρίσκεται σε ένα ποσοστό περίπου στο 16% χρησιμοποίησης ΑΠΕ (Ανανεώσιμες Πηγές Ενέργειας) για παραγωγή ενέργειας. Το μεγαλύτερο ποσό ενέργειας παρέχεται από τα φωτοβολταϊκά πάρκα, συνεπώς η γνώση για το ποσό παραγωγής τους, παίζει σημαντικό ρόλο τόσο στην σταθερότητα του συστήματος όσο και στην διαμόρφωση της τιμής της kWh στο μέλλον, με το άνοιγμα τις αγορές ενέργειας που προσπαθεί να γίνει. Έτσι εργαζόμενοι πάνω σε ένα φωτοβολταϊκό πάρκο κατασκευάσαμε μία εφαρμογή που ο χρήστης έχει τη δυνατότητα μέσω των μετεωρολογικών προβλέψεων να γνωρίζει την ενέργεια που θα παράγει το πάρκο την επόμενη ημέρα. Για την κατασκευή του μοντέλου δοκιμάστηκαν 4 αλγόριθμοι μηχανικής μάθησης, και 3 υπομοντέλα σε κάθε αλγόριθμο, για κάθε μία από τις καιρικές συνθήκες όπως τις κατηγοριοποιήσαμε (ηλιοφάνεια, μερική νεφοκάλυψη, ολική νεφοκάλυψη) και στο τέλος επιλέχτηκαν σαν καλύτερη λύση ένα υβριδικό μοντέλο. Έτσι καταφέραμε να πετύχουμε ένα πολύ καλό αποτέλεσμα για τις ημέρες με ηλιοφάνεια που περιλαμβάνει σφάλμα 18.4 kWh σε συνολική παραγωγή περίπου 1000 kWh σε μία μέρα παραγωγής, και ένα επίσης μέτριο αποτέλεσμα για μέρες με πλήρη νεφοκάλυψη με σφάλμα περίπου 10.35 kWh σε ημερήσια παραγωγή 100 kWh. Τέλος μέσω αυτής της διπλωματικής αναδείχθηκε η ανάγκη μελέτης και πρόβλεψης των δεδομένων παραγωγής ηλεκτρικής ενέργειας από φωτοβολταϊκά πάνελ.

## Λέξεις Κλειδιά

Πρόβλεψη Ηλεκτρικής Ενέργειας, Μηχανική Μάθηση, Μετεωρολογικά Δεδομένα, Python, Tkinter





# Abstract

In recent years, developments and changes in electricity production and consumption have been strong. Environmental pollution, health problems caused by fossil fuels and their exhaustion are few factors for using more and more renewable forms of energy in production. Greece is a country that is about 16% of the use of RES (Renewable Energy Systems) for energy production. The largest amount of energy comes from photovoltaic parks, so knowledge of the amount of their production plays an important role in both the stability of the system and the formation of the kWh price in the future with the opening of the energy market. So, working on a photovoltaic park, an application was created that allows the user through the meteorological forecasts to know the energy the park will produce the next day. Four algorithms from machine learning and three submodels for each algorithm, for each of the weather conditions, such as categorized (sunshine, partial cloud cover, total cloud cover), and finally selected as the best solution a hybrid model. This achieves a very good result for sunny days that include an 18.4 kWh error in a total output of about 1000 kWh on a production day and a good result for days with full cloud cover with an error of about 10.35 kWh in a daily output of 100 kWh. Finally, through this diplomacy, the need for studying and forecasting the electricity production elements from photovoltaic panels was highlighted.

## Keywords

Prediction of Solar Energy, Machine Learning, Meteorological Data, Python, Tkinter



# Ευχαριστίες

Φτάνοντας λοιπόν στην ολοκλήρωση αυτής της πορείας μου, μέχρι την επίτευξη του διπλώματος θα ήθελα να ευχαριστήσω κάποιους ανθρώπους που ήταν δίπλα μου σε όλη αυτήν την προσπάθεια.

Πρώτα απ' όλα θα ήθελα να ευχαριστήσω τον κ. Βασιλακόπουλο για την πολύτιμη βοήθεια και καθοδήγησή που μου παρείχε όλο αυτό τον καιρό για την ολοκλήρωση της εργασίας αλλά και για την μεσολάβησή του για την απόκτηση των καιρικών δεδομένων. Επιπλέον θα ήθελα να ευχαριστήσω τον κ. Μπαργιώτα για τη συνεργασία και την πολύτιμη συμβολή του ως δεύτερος καθηγητής. Ακόμα θα ήθελα να ευχαριστήσω το Πανεπιστήμιο Θεσσαλίας για την αγορά των δεδομένων που χρησιμοποιήθηκαν στην εργασία καθώς χωρίς αυτά δεν θα μπορούσε να ολοκληρωθεί. Ένα μεγάλο ευχαριστώ θα ήθελα να πω ακόμα στην εταιρία Green Project του κ. Παπαστεργίου για την παροχή δεδομένων αλλά και την εταιρία Τσαραπατσάνη ΑΕ με ιδιοκτήτη τον κ. Ζιάκα, για την παραχώρηση του φωτοβολταϊκού πάρκου.

Σημαντική βοήθεια και στήριγμα μου όλα αυτά τα χρόνια ήταν οι φίλοι που έκανα τόσο στο Βόλο όσο και αυτοί από τα σχολικά μου χρόνια, που ήταν πάντα δίπλα μου σε ότι χρειάστηκα. Ένα λίγο μεγαλύτερο ευχαριστώ θα ήθελα να πω αρχικά στη Μυρτώ η οποία βρίσκεται πάντα δίπλα μου και με στηρίζει σε ότι και αν κάνω αλλά και για τη συνεισφορά της στα γραφιστικά της εφαρμογής και του εγγράφου. Και επίσης, στην Αποστολία η οποία ακολούθησε την πορεία της διπλωματικής και μεσολάβησε για την απόκτηση των δεδομένων παραγωγής.

Τελευταίοι και πιο σημαντικοί άνθρωποι, που θα ήθελα να ευχαριστήσω, που με στήριξαν και με στηρίζουν όλα αυτά τα χρόνια και χωρίς αυτούς δεν θα είχα πετύχει τίποτα από όσα έχω κάνει, είναι οι γονείς μου και τα αδέρφια μου.



# Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε ως τελευταίο βήμα για την απόκτηση του διπλώματος του τμήματος Ηλεκτρολόγων Μηχανικών & Μηχανικών Ηλεκτρονικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας στην πόλη του Βόλου υπό την επίβλεψη του καθηγητή κ. Μιχαήλ Βασιλακόπουλου. Το ερέθισμα για την ενασχόλησή μου με αυτό το θέμα ήταν αρχικά η προσπάθεια που γίνεται στην Ελλάδα για την ανάπτυξη των ΑΠΕ αλλά και το ενδιαφέρον μου τόσο για τη μελέτη των δεδομένων αλλά και για τη πρόβλεψη τους. Έτσι έχοντας στο κοντινό μου περιβάλλον ιδιοκτήτη ενός ηλιακού πάρκου η εύρεση δεδομένων παραγωγής ήταν εύκολη. Επιπλέον, η πολύτιμη βοήθεια του Πανεπιστημίου Θεσσαλίας με την αγορά των δεδομένων των καιρικών συνθηκών έκανε ακόμα πιο εύκολο αυτό το εγχείρημα μου. Ένας ακόμα παράγοντας που με ώθησε στην επιλογή αυτού του θέματος είναι και η προσπάθεια που έχει ξεκινήσει να γίνεται για το άνοιγμα της αγοράς της ηλεκτρικής ενέργειας και όχι μόνο. Έτσι θεώρησα ότι ήταν μία πολύ καλή ευκαιρία να κατασκευάσω ένα αρχικό πρότυπο μιας εφαρμογής που θα προβλέπει την παραγωγή της ηλεκτρικής ενέργειας κάνοντας έτσι την αρχή ίσως μιας καινούργιας επιχειρηματικής ιδέας.



# Περιεχόμενα

<b>Περίληψη</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Ευχαριστίες</b>	<b>v</b>
<b>Πρόλογος</b>	<b>vii</b>
<b>Περιεχόμενα</b>	<b>ix</b>
<b>Κατάλογος σχημάτων</b>	<b>xi</b>
<b>Κατάλογος πινάκων</b>	<b>xiii</b>
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Περιγραφή του προβλήματος . . . . .	1
1.2 Στόχος . . . . .	4
1.3 Οργάνωση του τόμου . . . . .	4
<b>2 Θεωρητικό Υπόβαθρο Και Βιβλιογραφική Ανασκόπηση</b>	<b>5</b>
2.1 Δεδομένα . . . . .	5
2.2 Βιβλιογραφική Ανασκόπηση . . . . .	7
2.3 Λειτουργία Φωτοβολταϊκών Πάνελ . . . . .	8
2.3.1 Δομή του φωτοβολταϊκού πάρκου . . . . .	9
<b>3 Θεωρητικό υπόβαθρο</b>	<b>11</b>
3.1 Εισαγωγή . . . . .	11
3.2 Μηχανική Μάθηση . . . . .	11
3.3 Σύνολο δεδομένων εκπαίδευσης και ελέγχου . . . . .	13
3.4 K-fold Cross Validation . . . . .	13
3.5 Υπερπροσαρμογή (overfitting) και υποπροσαρμογή (underfitting) . . . . .	13
3.6 Μέτρο απόδοσης ενός μοντέλου . . . . .	14
3.7 Χρονολογικές Σειρές (Time Series) . . . . .	15
3.7.1 Στατιστικά μεγέθη χρονοσειράς . . . . .	17

3.7.2	Αλγόριθμοι . . . . .	18
3.8	Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines) . . . . .	20
3.9	Gradient Boosted Regression Trees (GBRT) . . . . .	23
3.10	Τεχνητά Νευρωνικά Δίκτυα . . . . .	25
3.10.1	Πολυεπίπεδο δίκτυο Perceptron – MLP . . . . .	27
<b>4</b>	<b>Μεθοδολογία και Παρουσίαση Μοντέλων</b>	<b>31</b>
4.1	Εισαγωγή . . . . .	31
4.2	Μεθοδολογία . . . . .	31
4.3	Λογισμικά . . . . .	32
4.3.1	Python . . . . .	32
4.3.2	Keras & Tensorflow . . . . .	32
4.3.3	Anaconda . . . . .	33
4.3.4	Pycharm . . . . .	33
4.3.5	Tkinter . . . . .	33
4.4	Διαχείριση Δεδομένων . . . . .	33
4.5	Υπολογισμός Ημερήσιας Ενεργειακής Απόδοσης Ηλιακού Πάνελ . . . . .	35
4.6	Μοντέλα Πρόβλεψης . . . . .	36
4.6.1	Χρονολογικές Σειρές (Time Series) . . . . .	36
4.6.2	Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines) . . . . .	37
4.6.3	Gradient Boosted Regression Trees . . . . .	38
4.6.4	Πολυεπίπεδο δίκτυο Perceptron – MLP . . . . .	38
<b>5</b>	<b>Αποτελέσματα</b>	<b>41</b>
5.1	Χρονολογικές Σειρές (Time Series) . . . . .	41
5.2	Μηχανική Μάθηση . . . . .	42
5.2.1	Αποτελέσματα ηλιακής ακτινοβολίας (Solar Radiation) . . . . .	42
5.2.2	Αποτελέσματα παραγωγής ηλεκτρικής ενέργειας(kWh) . . . . .	46
<b>6</b>	<b>Εφαρμογή</b>	<b>49</b>
6.1	Λεπτομέρειες υλοποίησης . . . . .	49
6.2	Οδηγίες λήψης και εκτέλεσης της εφαρμογής . . . . .	52
<b>7</b>	<b>Επίλογος</b>	<b>55</b>
7.1	Σύνοψη και συμπεράσματα . . . . .	55
7.2	Μελλοντικές επεκτάσεις . . . . .	55
<b>I</b>	<b>Προδιαγραφές Φωτοβολταϊκών Πάνελ</b>	<b>57</b>
	<b>Βιβλιογραφία</b>	<b>61</b>
	<b>Συντομογραφίες</b>	<b>65</b>



# Κατάλογος σχημάτων

2.1	Φωτογραφία δορυφόρου της εγκατάστασης [11]	5
2.2	Λειτουργία φωτοβολταϊκού πάνελ [37]	9
2.3	Η εγκατάσταση του φωτοβολταϊκού πάρκου	10
2.4	SMA inverter [24]	10
3.1	Τρεις διαφορετικοί τρόποι μάθησης [17]	12
3.2	Διαδικασία K-fold Cross-Validation [6]	14
3.3	Παράδειγμα overfitting και underfitting [20]	14
3.4	Παράδειγμα χρονολογικής σειράς για διαφορά θερμοκρασίας [29]	16
3.5	Χρονολογική σειρά στάσιμη και μη στάσιμη [14]	18
3.6	Παράδειγμα διαγράμματος PACF και ACF	19
3.7	Γραφική απεικόνιση της ευθείας του αλγορίθμου SVM [27]	20
3.8	Γραφική απεικόνιση της ευθείας του αλγορίθμου SVR [26]	22
3.9	Μετατροπή με τη χρήση συνάρτησης πυρήνα [16]	24
3.10	Γραφική απεικόνιση του αλγορίθμου GBRT [15]	24
3.11	Δομή κόμβου νευρωνικού δικτύου	25
3.12	Γραφική απεικόνιση βηματικής συνάρτησης	26
3.13	Γραφική απεικόνιση γραμμικής συνάρτησης	27
3.14	Γραφική απεικόνιση σιγμοειδούς συναρτήσεως	28
3.15	Δομή νευρώνα νευρωνικού δικτύου	28
3.16	Νευρωνικό δίκτυο με απλή τροφοδότηση	29
3.17	Νευρωνικό δίκτυο με χρήση ErrorBack Propagation	29
4.1	Τιμές συσχέτισης του Shortwave Radiation με τις υπόλοιπες μεταβλητές	34
4.2	Διαχωρισμός Δεδομένων σε 3 Κατηγορίες	35
4.3	Απώλειες του φωτοβολταϊκού συστήματος [20]	36
4.4	Αποτελέσματα Dickey-Fuller test	36
4.5	ACF	37
4.6	PACF	37
4.7	Αναπαράσταση δέντρου GBRT (αναφέρεται στο μοντέλο που κατασκευάστηκε για τις ημέρες με ηλιοφάνεια)	38
4.8	Μεθοδολογία ανάπτυξης της διπλωματικής εργασίας	40

5.1	Αποτελέσματα (Solar Radiation) εκτέλεσης αλγορίθμου ARIMA για όλο το σύνολο δεδομένων . . . . .	41
5.2	Παραγόμενη Ηλεκτρική Ενέργεια με χρήση Χρονολογικών Σειρών . . . . .	42
5.3	Ημερήσια πρόβλεψη Solar Radiation με χρήση SVR . . . . .	43
5.4	Ημερήσια πρόβλεψη Solar Radiation με χρήση GBRT . . . . .	45
5.5	Ημερήσια πρόβλεψη Solar Radiation με χρήση MLP neural networks . . . . .	45
5.6	Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας με MLP . . . . .	47
5.7	Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας με SVR . . . . .	47
5.8	Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας με Gradient Boosting . . . . .	47
6.1	Αρχική επιφάνεια της εφαρμογής . . . . .	49
6.2	Πρώτη σελίδα της εφαρμογής, για εισαγωγή πληροφοριών . . . . .	50
6.3	Δεύτερη σελίδα της εφαρμογής, με ιστορικό της παραγωγής . . . . .	50
6.4	Τρίτη σελίδα της εφαρμογής, ζωντανές προβλέψεις καιρού . . . . .	51
6.5	Τρίτη σελίδα της εφαρμογής, πρόβλεψη για την επόμενη ημέρα . . . . .	52
6.6	Qr code με σύνδεσμο που οδηγεί στα αρχεία της εργασίας . . . . .	52

# Κατάλογος πινάκων

1.1	Ορυκτά καύσιμα [34]	1
1.2	Πρόβλεψη των ΑΠΕ για τα επόμενα χρόνια [34]	3
2.1	Ονόματα μεταβλητών και μονάδες μέτρησης	6
4.1	Παράμετροι SVR	38
4.2	Παράμετροι GBRT	38
4.3	Παράμετροι MLP	39
5.1	Μέτρηση απόδοσης και RMSE του αλγορίθμου SVR για τη πρόβλεψη του Solar Radiation	43
5.2	Μέτρηση RMSE του αλγορίθμου SVR για ημερήσια πρόβλεψη του Solar Radiation	44
5.3	Μέτρηση απόδοσης και RMSE του αλγορίθμου GBRT για τη πρόβλεψη του Solar Radiation	44
5.4	Μέτρηση RMSE του αλγορίθμου GBRT για ημερήσια πρόβλεψη του Solar Radiation	44
5.5	Μέτρηση απόδοσης και RMSE του αλγορίθμου MLP για τη πρόβλεψη του Solar Radiation	45
5.6	Μέτρηση RMSE για ημερήσια πρόβλεψη του Solar Radiation με τη χρήση MLP neural networks	46
5.7	Μετρήσεις του RMSE για τους 3 αλγορίθμους για τη παραγωγή ηλεκτρικής ενέργειας (kWh)	46
5.8	Μετρήσεις του RMSE για τους 3 αλγορίθμους παραγωγής ηλεκτρικής ενέργειας (για μία ημέρα)	48
6.1	Αναλυτική περιγραφή αρχείων στο github	53



# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Περιγραφή του προβλήματος

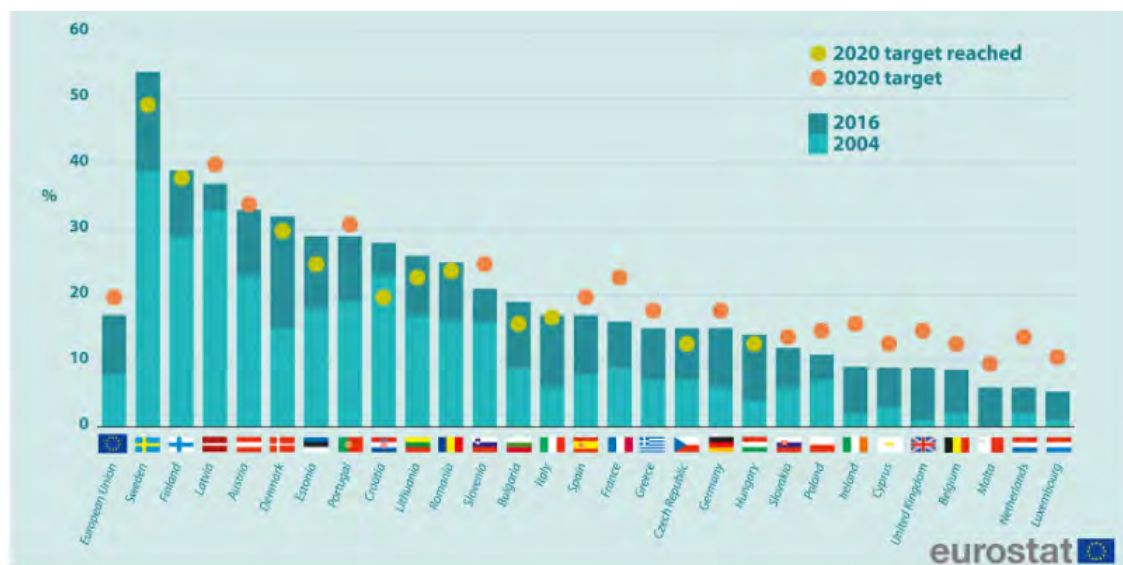
Τα τελευταία χρόνια η επικείμενη εξάντληση των ορυκτών καυσίμων σε συνδυασμό με την ψήφιση χαμηλότερων ορίων στις εκπομπές διοξειδίου του άνθρακα (CO<sub>2</sub>) από τα κράτη μέλη της Ευρωπαϊκής Ένωσης αλλά και από τις υπόλοιπες χώρες του κόσμου, έχουν προκαλέσει αρκετή ανησυχία τόσο στις κυβερνήσεις των κρατών όσο και στους απλούς πολίτες [33]. Σε αρκετές χώρες στον κόσμο ανάγκες για θέρμανση, μεταφοράς αλλά και παραγωγής ηλεκτρικής ενέργειας ικανοποιούνται κυρίως με την χρήση ορυκτών καυσίμων [12]. Αυτό, όπως είναι φυσικό είναι αρκετά ανησυχητικό διότι κάποιοι από τους βασικούς φυσικούς ενεργειακούς πόρους είναι πεπερασμένοι.

Πίνακας 1.1: Ορυκτά καύσιμα [34]

Καύσιμο	Χρόνια
Λιγνίτης	510
Γαϊάνθρακας	130
Πετρέλαιο	43
Φυσικό αέριο	65

Ακόμα πιο ανησυχητικές βέβαια είναι οι επιπτώσεις που προκαλούν ορισμένα ορυκτά καύσιμα τόσο στο περιβάλλον όσο και στην ανθρώπινη υγεία. Η ρύπανση του αέρα και των υδάτινων πόρων, το φαινόμενο του θερμοκηπίου και όξινη βροχή είναι μόνο μερικά από τα προβλήματα που προκαλούν τα ορυκτά καύσιμα. Η μεγαλύτερη σημασία όμως θα πρέπει να δοθεί στα προβλήματα υγείας που προκαλούν στον άνθρωπο καθώς αρκετές γενετικές ανωμαλίες και ασθένειες έχουν προκληθεί από διάφορες ουσίες των ορυκτών καυσίμων [32]. Συνεπώς όλοι αυτοί οι λόγοι, και πολλά άλλα, συμβάλλουν στη μετάβαση του κόσμου σε 100% χρήση ανανεώσιμων πηγών ενέργειας.

Φυσικά η εκπλήρωση όλων των αναγκών ενέργειας μόνο με την χρήση ανανεώσιμων πηγών ενέργειας είναι κάτι το οποίο σίγουρα για να επιτευχθεί χρειάζονται αρκετά χρόνια. Όμως από το διάγραμμα βλέπουμε (Σχήμα 1.1), ότι αρκετές χώρες έχουν καταφέρει να έχουν ένα υψηλό πο-



Σχήμα 1.1: Πορεία και στόχοι των Ευρωπαϊκών χωρών ως προς την χρήση των ΑΠΕ [8]

σοστό στη χρήση ανανεώσιμων πηγών ενέργειας. Σαν ανανεώσιμες πηγές λαμβάνονται η αιολική ενέργεια, η ηλιακή ενέργεια (θερμική, φωτοβολταϊκή και συμπυκνωμένη), η υδροηλεκτρική ενέργεια, η παλιρροϊκή ενέργεια, η γεωθερμική ενέργεια, τα βιοκαύσιμα και το ανανεώσιμο τμήμα των αποβλήτων [8]. Έτσι σύμφωνα με την Eurostat το μερίδιο της ενέργειας από ανανεώσιμες πηγές στην ακαθάριστη τελική κατανάλωση ενέργειας έχει σχεδόν διπλασιαστεί: από περίπου 8,5 % το 2004 σε 17,0 % το 2016, και σαν στόχο για το 2020 έχει να φτάσει το 20 %.

Στην Ελλάδα το ποσοστό χρήσης ανανεώσιμων πηγών ενέργειας βρίσκεται περίπου στο 16% σύμφωνα με την τελευταία πρόβλεψη της eurostat. Στόχος της είναι να φτάσει το 20% σύμφωνα με το νόμο που ψηφίστηκε από την Ευρωπαϊκή Ένωση. Όσον αφορά τις ανανεώσιμες πηγές ενέργειας που χρησιμοποιήθηκαν για την παραγωγή ηλεκτρισμού κατά το έτος 2016 ένα μεγάλο μέρος του, το 40% καλύφθηκε από αυτές. Με το μεγαλύτερο ρόλο σ' αυτό το ποσοστό να παίζει η παραγωγή ηλεκτρικής ενέργειας από φωτοβολταϊκά πάνελ με ποσοστό κοντά στο 30%. Το υπόλοιπο 60% των αναγκών καλύφθηκε από ορυκτά καύσιμα, πετρέλαιο και φυσικό αέριο [8].

Όπως είδαμε, σημαντικό ρόλο στην κάλυψη των αναγκών ηλεκτρικού ρεύματος από ΑΠΕ διαδραματίζει για χρόνια, αλλά και θα συνεχίσει και στο μέλλον η ηλιακή ενέργεια. Σύμφωνα με μελέτη της ΑΔΜΗΕ (Ανεξάρτητος Διαχειριστής Μεταφοράς Ηλεκτρικής Ενέργειας) για το μέλλον αυτό επαληθεύεται και φαίνεται στο Πίνακα 1.2.

Η ερώτηση και η απορία πολλών βέβαια θα ήταν, γιατί η ηλιακή ενέργεια θα παίζει ιδιαίτερο ρόλο για την Ελλάδα, στο δρόμο προς την πλήρη εξάρτηση της χώρας από τις ΑΠΕ. Κυριότερος λόγος είναι το κλίμα που έχει η Ελλάδα. Με χαρακτηριστικό του, αρκετές μέρες ηλιοφάνειας τον χρόνο, σε συνδυασμό με θερμοκρασίες οι οποίες μόνο τους καλοκαιρινούς μήνες (Ιούνιο, Ιούλιο, Αύγουστο) ξεπερνάνε κατά πολύ τους 25 βαθμούς Κελσίου, θερμοκρασία η οποία θεωρείται ότι επιδρά αρνητικά στη παραγωγή ηλεκτρικής ενέργειας, η χρήση φωτοβολταϊκών πάνελ θα ήταν η ιδανική λύση για παραγωγή ενέργειας. Επιπλέον η μηδενική ρύπανση, η αθόρυβη λειτουργία

Πίνακας 1.2: Πρόβλεψη των ΑΠΕ για τα επόμενα χρόνια [34]

	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
	(MW)										
<b>Αιολικά</b>	2047	2300	2555	2810	3065	3220	3370	3520	3990	4160	4330
<b>Φωτοβολ- ταϊκά</b>	2444	2720	3000	3280	3560	3710	3860	4010	4256	4406	4556
<b>ΜΥΗΕ</b>	223	228	237	242	247	252	257	262	267	272	277
<b>Βιομάζα/ Βιοαέριο</b>	58	62	66	70	100	150	200	200	200	200	200
<b>ΣΗΘΥΑ</b>	100	104	108	110	110	125	125	125	125	125	125
<b>Ηλιοθερ- μικά</b>									50	50	50
<b>Σύνολο</b>	4872	5414	5966	6512	7082	7457	7812	8117	8888	9213	9538

τους, η μεγάλη διάρκεια ζωής (ακόμα και 30 χρόνια), η δυνατότητα επέκτασή τους ανάλογα με τις ανάγκες και τέλος η ελάχιστη συντήρηση τους καθιστά, εάν όχι τον μεγαλύτερο σίγουρα έναν σημαντικό πρωταγωνιστή στις ΑΠΕ, για την Ελλάδα.

Άλλος ένας λόγος που έχει οδηγήσει τα φωτοβολταϊκά πάνελ να αποτελούν ιδιαίτερα σημαντικό κομμάτι των ΑΠΕ είναι η συνεχόμενη βελτίωση της τεχνολογίας των υλικών τους, καθώς ο συντελεστής απόδοσής τους διαρκώς αυξάνεται, προσφέροντας συνεχώς όλο μεγαλύτερες ποσότητες ηλεκτρικού ρεύματος με αρκετά μειωμένο κόστος. Συγκεκριμένα, αναφέροντας και σε πολλές περιπτώσεις μείωση κόστους παραγωγής ηλεκτρικού ρεύματος κατά 73% μεταξύ των ετών 2010 και 2017 [22]. Έτσι η μείωση του κόστους και η αύξηση της απόδοσης τους, έχουν κάνει τα φωτοβολταϊκά πάνελ εναλλακτική λύση ως ΑΠΕ σε πολλές χώρες [1].

Σημαντικός παράγοντας για την απόδοση των φωτοβολταϊκών πάνελ αποτελούν οι καιρικές συνθήκες, και ιδιαίτερα κατά τους χειμερινούς μήνες που ο καιρός χαρακτηρίζεται από αρκετές μεταβολές, με την παραγωγή ηλεκτρικής ενέργειας να είναι αρκετά ασταθής. Έτσι η βαθεία κατανόηση των χαρακτηριστικών του καιρού που επηρεάζουν την παραγωγή ηλεκτρικού ρεύματος μας δίνουν μια σημαντική γνώση για να αντιμετωπίσουμε αυτήν την αστάθεια. Αυτή η αντιμετώπιση της αστάθειας είναι ιδιαίτερα σημαντική για τους παραγωγούς των ΑΠΕ σε πολλές Ευρωπαϊκές χώρες, ακόμα και στην Ελλάδα, καθώς έχει ξεκινήσει η λειτουργία μιας ενιαίας ευρωπαϊκής αγοράς ηλεκτρικής ενέργειας, το λεγόμενο «Μοντέλο Στόχο» (Target Model).

Βασική επιδίωξη του Target Model, είναι η εξισορρόπηση της προσφοράς και ζήτησης ηλεκτρικής ενέργειας σε πραγματικό χρόνο [28]. Έτσι όταν σε μία Ευρωπαϊκή χώρα δημιουργηθεί ένα έλλειμμα ενέργειας αυτό αντισταθμίζεται με το πλεόνασμα μίας άλλης, καταφέροντας έτσι επάρκεια ανεφοδιασμού για όλες τις χώρες και φυσικά χαμηλότερο ενεργειακό κόστος. Πάνω σε αυτό το μοντέλο, στην Ελλάδα δημιουργείται το Ελληνικό Χρηματιστήριο Ενέργειας που διαχειρίζεται μία ή και περισσότερες αγορές ενέργειας (ηλεκτρικής ενέργειας, φυσικού αερίου και περιβαλλοντικές αγορές). Συγκεκριμένα στις αγορές ηλεκτρικής ενέργειας περιλαμβάνονται:

- **Η αγορά επόμενης ημέρας**, όπου οι συμμετέχοντες θα αγοράζουν και θα πωλούν ενέργεια για την επόμενη ημέρα.
- **Η ενδοημερήσια αγορά**, όπου οι συμμετέχοντες θα μπορούν να αγοράζουν και να πωλούν ενέργεια για την ίδια ημέρα.
- **Η αγορά εξισορρόπησης**.

Στις παραπάνω διαδικασίες δικαίωμα συμμετοχής έχει κάθε φυσικό και νομικό πρόσωπο που δικαιούται να κάνει συναλλαγές σε μία ή περισσότερες αγορές ενέργειας.

## 1.2 Στόχος

Στόχος μας λοιπόν, με αυτήν την διπλωματική είναι να προβλέψουμε την παραγωγή ηλεκτρικής ενέργειας ενός φωτοβολταϊκού πάρκου για τις επόμενες 24 ώρες. Αυτό το πετυχαίνουμε συνδυάζοντας τα δεδομένα παραγωγής που έχουμε από το ηλιακό πάρκο με τις καιρικές συνθήκες που θα επικρατήσουν στην περιοχή τις επόμενες 24 ώρες αλλά και τα ιστορικά δεδομένα του καιρού που υπάρχουν για εκείνη την τοποθεσία. Με αυτήν την πρόβλεψή μας βοηθάμε τους έμπορους ηλεκτρικής ενέργειας να γνωρίζουν κατά προσέγγιση την ποσότητα ηλεκτρικής ενέργειας που θα παραγάγουν διαμορφώνοντας έτσι και την κατάλληλη τιμή της. Επιπλέον σημαντικό ρόλο κατέχει αυτή η πρόβλεψη και σε επιχειρήσεις οι οποίες χρησιμοποιούν την ενέργεια που παράγουν από ηλιακά πάρκα για αυτοκατανάλωση, καταφέρνοντας έτσι να την διαχειριστούν σωστά αποθηκευοντάς την σε μπαταρίες και να την χρησιμοποιούν σε ώρες και μέρες χαμηλής παραγωγής ή υψηλής ζήτησης.

## 1.3 Οργάνωση του τόμου

Στο **Κεφάλαιο 1**, υπάρχει η εισαγωγή της διπλωματικής, όπου αναλύουμε το πρόβλημα που υπάρχει και περιγράφουμε το στόχο της διπλωματικής.

Στο **Κεφάλαιο 2**, βρίσκεται η περιγραφή των δεδομένων που χρησιμοποιήθηκαν, μια βιβλιογραφική ανασκόπηση, η λειτουργία των φωτοβολταϊκών πάνελ καθώς και μία περιγραφή του φωτοβολταϊκού πάρκου που γίνεται η έρευνα.

Στο **Κεφάλαιο 3**, περιλαμβάνονται όλες οι θεωρητικές γνώσεις πάνω στους αλγορίθμους μηχανικής μάθησης που χρησιμοποιήθηκαν καθώς επίσης και θεωρία με την οποία μετράμε την απόδοση αυτών των αλγορίθμων.

Στο **Κεφάλαιο 4**, περιγράφουμε την μεθοδολογία με την οποία κατασκευάστηκαν τα μοντέλα πρόβλεψης αλλά και οι διάφορες πλατφόρμες και γλώσσες προγραμματισμού που χρησιμοποιήθηκαν.

Στο **Κεφάλαιο 5**, έχουμε συμπεριλάβει τα αποτελέσματα των μοντέλων πρόβλεψης καθώς και κάποιες συγκρίσεις που έγιναν.

Στο **Κεφάλαιο 6**, βρίσκεται μία αναλυτική περιγραφή της εφαρμογής και ο τρόπος εκτέλεσής της. Και τέλος στο **Κεφάλαιο 7**, μπορεί κανένας να βρει τα συμπεράσματα αυτής της διπλωματικής καθώς και μελλοντικές επεκτάσεις της εργασίας.



## Κεφάλαιο 2

# Θεωρητικό Υπόβαθρο Και Βιβλιογραφική Ανασκόπηση

Αυτό το κεφάλαιο θα περιλαμβάνει πληροφορίες σχετικά με τα δεδομένα που χρησιμοποιήθηκαν για την κατασκευή των μοντέλων, ορισμένες συγγενικές εργασίες και έρευνες που έχουν πραγματοποιηθεί και στο τέλος θα περιλαμβάνει μία σύντομη περιγραφή της λειτουργίας των φωτοβολταϊκών πάνελ καθώς επίσης και την δομή του ηλιακού πάρκου.

### 2.1 Δεδομένα

Το ηλιακό πάρκο από το οποίο χρησιμοποιήθηκαν τα δεδομένα παραγωγής για την κατασκευή των μοντέλων βρίσκεται στους Αγίους Θεοδώρους Τρικάλων. Η εγκατάσταση περιλαμβάνει ένα σύνολο 638 ηλιακών πάνελ διαστάσεων  $1,64 \text{ m}^2$  με παραγωγή 235 Watt το καθένα και έχει συνολική παραγωγή 150 kWatt. Το μοντέλο των φωτοβολταϊκών πάνελ είναι η aleo s18 235Watt, περισσότερα χαρακτηριστικά των πάνελ παρουσιάζονται στο Παράρτημα Ι.



Σχήμα 2.1: Φωτογραφία δορυφόρου της εγκατάστασης [11]

Τα δεδομένα που χρησιμοποιήθηκαν έχουν αντληθεί από δύο πηγές και περιλαμβάνουν πληροφορίες ανά ώρα για παραγωγή ηλεκτρικής ενέργειας αλλά και καιρικές συνθήκες στην περιοχή. Για την ωριαία παραγωγή ηλεκτρικής ενέργειας του πάρκου χρησιμοποιήθηκαν δεδομένα από το sunnyportal.

Το Sunnyportal πρόκειται για μία διαδικτυακή εφαρμογή της εταιρίας SMA (System Mess and Anlagentechnik) Solar Technology AG η οποία χρησιμοποιείται για τον έλεγχο και τη διαχείριση φωτοβολταϊκών εγκαταστάσεων. Έτσι περιλαμβάνει πλήρη εποπτεία της εγκατάστασης με πληροφορίες που δίνει σχετικά με την κατάσταση των φωτοβολταϊκών πάνελ (θερμοκρασία), πληροφορίες για την κατάσταση των inverter, πληροφορίες για τις καιρικές συνθήκες της περιοχής και τέλος διαθέτει και μία βάση δεδομένων παραγωγής ηλεκτρικής ενέργειας μέσω της οποίας μπορεί να δει κανένας την ημερήσια, μηνιαία αλλά και ετήσια παραγωγή.

Αντίστοιχα για τα δεδομένα καιρού σε ωριαία βάση αγοράστηκαν από το πανεπιστήμιο Θεσσαλίας από τη ιστοσελίδα Meteoblue. Τα δεδομένα που χρησιμοποιήθηκαν στην ανάλυση του προβλήματος είναι από το έτος 2017 έως και 2018 .

Η Meteoblue πρόκειται για μία ιστοσελίδα που ασχολείται με μετεωρολογικές προβλέψεις ανά τον κόσμο. Επίσης διαθέτει και ένα ιστορικό καιρικών δεδομένων που φτάνει μέχρι το 1960, με αναλυτικά ωριαία δεδομένα που δεν περιορίζονται μόνο στα βασικά στοιχεία των καιρικών συνθηκών (θερμοκρασία, άνεμος, υγρασία...) αλλά και σε πιο εξειδικευμένα όπως η ηλιακή ακτινοβολία.

Τα δεδομένα που αντλήθηκαν από το Sunnyportal αναφέρονται στην συνολική ενεργειακή απόδοση ανά ώρα και έχουν μονάδα μέτρησης την kWh.

Τα δεδομένα καιρού που αγοράστηκαν από το Meteoblue περιλαμβάνουν 28 μεταβλητές του καιρού. Στον παρακάτω πίνακα (Πίνακας 2.1) αναφέρονται μόνο οι 6 μεταβλητές που χρησιμοποιήθηκαν για τη κατασκευή των μοντέλων.

Πίνακας 2.1: Ονόματα μεταβλητών και μονάδες μέτρησης

Όνομα	Μονάδα Μέτρησης
Temperature	°C
Relative Humidity	hPa
Total Cloud Cover	%
Sunshine Duration	min
Shortwave Radiation	$W/m^2$
Wind Speed	km/h

- **Temperature (Θερμοκρασία):** Αναφέρεται στην θερμοκρασία που επικρατεί στην συγκεκριμένη περιοχή σε ύψος 2m από τη γη και έχει μονάδα μέτρησης του βαθμούς Κελσίου.
- **Relative Humidity (Σχετική υγρασία):** Η μονάδα μέτρησης της σχετικής υγρασία είναι το hPa το οποίο ισούται με 100 Pa, και αναφέρεται στην σχετική υγρασία που επικρατεί στην συγκεκριμένη περιοχή. Η μέτρησή της έχει γίνει σε ύψος 2m από την γη.

- **Total Cloud Cover (Συνολικό ποσό νεφοκάλυψης):** Μετριέται σε ποσοστό επί της 100 και αναφέρεται στο ποσό επί της 100 ο ουρανός είναι καλυμμένος από σύννεφα.
- **Sunshine Duration (Διάρκεια ηλιοφάνειας):** Σε αντίθεση με την προηγούμενη μεταβλητή αυτή μας δείχνει την διάρκεια ηλιοφάνειας μέσα στη μέρα. Μονάδα μέτρησης είναι τα λεπτά.
- **Shortwave Radiation (Ηλιακή ακτινοβολία):** Η ποσότητα της ηλιακής ακτινοβολίας που φτάνει στα ανώτερα στρώματα της ατμόσφαιρας μέχρι να φτάσει στην επιφάνεια της γης υφίσταται απώλειες και μεταβολές, που οφείλονται κυρίως στην διάχυση και απορρόφηση. Το ποσό της ακτινοβολίας που φτάνει στην γη ονομάζεται άμεση ηλιακή ακτινοβολία, ενώ εκείνο το ποσό που διαχέεται από την ατμόσφαιρα και φτάνει στη γη ονομάζεται διάχυτη ηλιακή ακτινοβολία. Το άθροισμα αυτών των δυο αποτελεί την ολική ηλιακή ακτινοβολία (Global Solar Radiation)[25]. Μονάδα μέτρησης είναι W/m<sup>2</sup>.
- **Wind Speed (Ταχύτητα ανέμου):** Αναφέρεται στην ταχύτητα του ανέμου σε ύψος περίπου 80m. Η μονάδα μέτρησης είναι τα Km/h.

## 2.2 Βιβλιογραφική Ανασκόπηση

Για την ανάπτυξη του τελικού μοντέλου που χρησιμοποιήθηκε στην εφαρμογή αλλά και για όλα τα υπόλοιπα μοντέλα που δοκιμάστηκαν αρκετές γνώσεις και επιλογές παραμέτρων προήλθαν από προηγούμενες έρευνες. Έτσι το κύριο και βασικότερο κομμάτι του μοντέλου πρόβλεψης ήταν η επιλογή των μετεωρολογικών δεδομένων που θα χρησιμοποιούσαμε στην διπλωματική. Σύμφωνα με την έρευνα του J. Antonanzas [2] παρατηρήσαμε ότι η πρότασή του ήταν να χωριστεί το έτος σε καιρικές περιόδους (Χειμώνας, Άνοιξη, Καλοκαίρι) και το μοντέλο να εκπαιδευτεί με δεδομένα ίδιας καιρικής περιόδου για μερικά χρόνια πριν, καθώς οι κλιματικές συνθήκες μπορεί να ποικίλουν κατά τη διάρκεια ενός ετήσιου κύκλου, και έτσι ένα μοντέλο μπορεί να αποδίδει καλύτερα αποτελέσματα όταν εκπαιδευτεί σε μια καιρική εποχή και όχι σε αρκετές. Εμείς παρόλο που είχαμε τα καιρικά δεδομένα για αρκετά χρόνια πίσω, μας έλειπαν δεδομένα παραγωγής του ηλιακού πάρκου. Έτσι αποφασίσαμε να αναπτύξουμε 3 ξεχωριστά μοντέλα με κριτήριο διαχωρισμού το ποσοστό νεφοκάλυψης, προσπαθώντας έτσι να φτιάξουμε 3 διαφορετικές καιρικές περιόδους (ηλιοφάνεια, μερική νεφοκάλυψη, ολική νεφοκάλυψη).

Το επόμενο κομμάτι ήταν η επιλογή των μεταβλητών του καιρού που θα χρησιμοποιούσαμε. Σύμφωνα με την έρευνα του Y. Atsushi η πρόταση του ήταν η χρήση 5 μεταβλητών του καιρού (θερμοκρασία, ταχύτητα ανέμου, πίεση, σχετική υγρασία και ηλιακή ακτινοβολία) [3]. Μία άλλη έρευνα του Haidar Almohri, πρότεινε πάλι τη χρησιμοποίηση 5 μεταβλητών κρατώντας ίδια με την προηγούμενη έρευνα τη θερμοκρασία, ταχύτητα ανέμου και ηλιακή ακτινοβολία, προσθέτοντας την θερμοκρασία του πάνελ και την ηλιακή ακτινοβολία που έφτανε πάνω σ' αυτό. Η δικιά μας πρόταση είναι να κρατήσουμε κάποιες από τις μεταβλητές των παραπάνω ερευνών όπως η θερμοκρασία, η σχετική υγρασία, η ταχύτητα του ανέμου και η ηλιακή ακτινοβολία και να προσθέσουμε σαν μεταβλητή τη διάρκεια της ηλιοφάνειας και την συνολική νεφοκάλυψη. Επιπλέον σημαντικό είναι να αναφερθεί ότι και στις δύο παραπάνω έρευνες τα μοντέλα κατασκευάστηκαν

για την πρόβλεψη πρώτα της ηλιακής ακτινοβολίας και όχι απευθείας την παραγωγή της ηλεκτρικής ενέργειας. Η ίδια λογική ακολουθήθηκε και στη δικιά μας περίπτωση.

Τέλος για την επιλογή των αλγορίθμων που χρησιμοποιήσαμε οι διαθέσιμες επιλογές ήταν πολλές καθώς κάθε έρευνα χρησιμοποιούσε αρκετούς. Έτσι επιλέξαμε σαν πιο αποδοτικούς αρχικά τον SVR (Support Vector Regression) τον οποίον χρησιμοποιεί στην έρευνα του ο Jie Shi [23], ο οποίος χωρίζοντας το σύνολο το δεδομένων σε 4 κατηγορίες, μία για κάθε μία από τις 4 καιρικές συνθήκες συννεφιά, ομίχλη, βροχή και ήλιο και προβλέποντας πρώτα την ηλιακή ακτινοβολία και μετά την ηλεκτρική παραγωγή κατάφερε να πετύχει ένα μοντέλο του οποίου το RMSE (Root Mean Square Error) για τις 4 κατηγορίες φτάνει 2.10 MW σε συνολική ισχύς 486MW. Στη συνέχεια για την επιλογή του gradient boosting regression trees, έγινε ύστερα από μελέτη της έρευνας του Caroline Persson ο οποίος χρησιμοποιώντας σαν δεδομένα κυρίως τα ποσοστά νεφοκάλυψης αλλά και την πιθανότητα βροχής κατάφερε να φτιάξει ένα μοντέλο με σφάλμα που είναι ανάμεσα στο 0.100 και 0.137 για εγκατάσταση ισχύος 9.66 kW [21]. Σημαντική διαφορά του μοντέλου που πρότεινε μέσω της έρευνάς του ο Caroline Persson σε σχέση με τη δικιά μας είναι η διαφορά του χρονικού ορίου της πρόβλεψης, καθώς εμείς προβλέπουμε για τις επόμενες 24 ώρες ενώ το μοντέλο του για 1 με 6 ώρες. Για την επιλογή του μοντέλου των νευρωνικών δικτύων χρησιμοποιήθηκε η έρευνα του Cristian-Dragos Dumitru, ο οποίος προτείνει τη χρήση των MLP (Multilayer Perceptron) νευρωνικών δικτύων [7]. Αξιοσημείωτη και εδώ είναι η διαφορά στην μεταβλητή που προβλέπει, μέσω του μοντέλου, η οποία είναι κατευθείαν η ισχύς που θα παραγάγει το φωτοβολταϊκό σύστημα και όχι η ηλιακή ακτινοβολία. Τέλος η επιλογή του μοντέλου ARIMA (Autoregressive Integrated Moving Average) έγινε ύστερα από ανάγνωση της έρευνας του Rich H. Inman, ο οποίος το χρησιμοποίησε για τη πρόβλεψη της ηλιακής ακτινοβολίας, αποτέλεσμα το οποίο θα μπορούσε να χρησιμοποιήσει κάποιος για να υπολογίσει μετά την παραγωγή ηλεκτρικής ενέργειας [13]. Σ' αυτή την έρευνα χρησιμοποίησε και άλλες μεθόδους για τη κατασκευή μοντέλων πρόβλεψης και κατέληξε ότι η πρόβλεψη της ηλιακής ενέργειας μέσω των χρονολογικών σειρών είναι χρήσιμη όταν δεν έχουμε αρκετές πληροφορίες σχετικά με δεδομένα καιρού. Μια ακόμη διαφορά που θα πρέπει να τονιστεί είναι ότι για την κατασκευή του μοντέλου στην έρευνα του Rich H. Inman, Caroline Persson και του Jie Shi η πρόβλεψη της ηλιακής ακτινοβολίας και της ηλεκτρικής ενέργειας αντίστοιχα, έγιναν με χρήση δεδομένων που δεν αφορούσαν μόνο μία περιοχή όπως στη δικιά μας περίπτωση αλλά για περισσότερες από μία. Έτσι τα μοντέλα τους δεν αναφερόταν μόνο σε ένα φωτοβολταϊκό πάρκο αλλά για ένα σύνολο φωτοβολταϊκών πλακών.

### 2.3 Λειτουργία Φωτοβολταϊκών Πάνελ

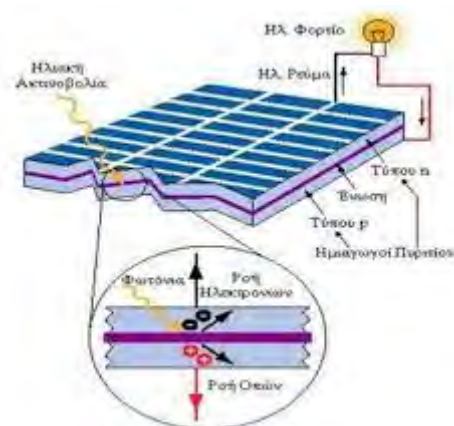
Η αρχή λειτουργίας των φωτοβολταϊκών πάνελ στηρίζεται στο φωτοβολταϊκό φαινόμενο σύμφωνα με το οποίο όταν δύο υλικά (ημιαγωγοί προσμίξεων διαφορετικής πολικότητας ιόντων) βρεθούν σε επαφή και εκτεθούν σε ηλιακή ακτινοβολία θα παραγάγουν ηλεκτρικό ρεύμα [37].

Οι επαφές αυτές αποτελούν τα φωτοβολταϊκά στοιχεία και εν συνεχεία πλήθος τέτοιων στοιχείων μπορεί να μας δώσει ένα φωτοβολταϊκό πάνελ. Υπάρχουν αρκετοί τύποι φωτοβολταϊκών πλαισίων.

- Μονοκρυσταλλικού πυριτίου (c-Si)

- Πολυκρυσταλλικού πυριτίου (m-Si)
- Άμορφου πυριτίου (a-Si)
- Χαλκοπυριτών CIS / CIGS

Σε αυτά τα υλικά (ημιαγωγοί) όταν προσπίπτει πάνω τους ηλιακή ακτινοβολία τα φωτόνια της ακτινοβολίας είτε ανακλώνται, είτε διαπερνούν το στοιχείο ή απορροφώνται από αυτό. Στην παραγωγή βοηθούν μόνο τα φωτόνια που απορροφώνται καθώς αυξάνουν την ενέργεια ορισμένων ηλεκτρονίων του φωτοβολταϊκού στοιχείου αναγκάζοντας τα να μετακινηθούν σε άλλες θέσεις. Έτσι δημιουργείται μια συνεχής κίνηση ηλεκτρονίων που ουσιαστικά αποτελεί το παραγόμενο ρεύμα.



Σχήμα 2.2: Λειτουργία φωτοβολταϊκού πάνελ [37]

Ανάλογα με τον τύπο του πλαισίου υπάρχουν συγκεκριμένες ονομαστικές τιμές απόδοσης. Έτσι για πάνελ Μονοκρυσταλλικού πυριτίου οι αποδόσεις κυμαίνονται από 14,5% έως 21%, για Πολυκρυσταλλικού πυριτίου 13% έως 14,5%, για Άμορφου πυριτίου 7% και τέλος Χαλκοπυριτών από 7% έως 14% [37].

Σε όλους τους παραπάνω τύπους φωτοβολταϊκών βλέπουμε ότι σαν βάση χρησιμοποιείται το πυρίτιο, και συγκεκριμένα το πυρίτιο είναι η βάση για το 90% της παγκόσμιας παραγωγής φωτοβολταϊκών. Αυτό οφείλεται στο ότι πολλοί επιστήμονες άρχισαν από την δεκαετία του 60' να ασχολούνται με την ανάλυση των ιδιοτήτων του πυριτίου. Επιπλέον μεγάλες βιομηχανίες και κυβερνήσεις έδωσαν αρκετή σημασία στην ανάπτυξη της τεχνολογίας του πυριτίου αναπτύσσοντας τόσο την τεχνολογία όσο και τις υποδομές όσον αφορά την επεξεργασία του πυριτίου. Κάτι το οποίο έχει σαν αποτέλεσμα σήμερα την ευρεία χρήση του σε πολλούς τομείς των ηλεκτρονικών.

### 2.3.1 Δομή του φωτοβολταϊκού πάρκου

Το φωτοβολταϊκό πάρκο όπως αναφέραμε, βρίσκεται στην περιοχή των Αγίων Θεοδώρων Τρικάλων και ανήκει στην εταιρεία Τσαραπατσάνης ΑΕ με ιδιοκτήτες τον κ. Ζιάκα και κ. Τσαραπατσάνη. Την διαχείριση του ηλιακού πάρκου έχει αναλάβει η εταιρία Green Project του κ. Παπαστεργίου από την οποία μας δόθηκε η πρόσβαση στα δεδομένα παραγωγής.

Η εγκατάσταση έχει συνολική ισχύς 150 kWatt και αποτελείται από ένα σύνολο 638 ηλιακών πάνελ μάρκας aleo S18 με συνολική ισχύς το καθένα 235 Watt. Το σύνολο των 638 ηλιακών πάνελ ενωμένα ανά 16 αποτελούν ένα ενιαίο, με το σύνολό τους να ανέρχεται στα 40.



Σχήμα 2.3: Η εγκατάσταση του φωτοβολταϊκού πάρκου

Χαρακτηριστικό της ηλιακής εγκατάστασης, είναι η χρήση συστημάτων ιχνηλάτησης της πορείας του ήλιου στα φωτοβολταϊκά πάνελ. Τα συστήματα αυτά που ονομάζονται και ηλιοπαρακολουθητές ή τρακερ (tracker) είναι σίγουρα πιο περίπλοκα από μια απλή στήριξη, όμως προσφέρουν αυξημένες αποδόσεις περίπου 30% [35]. Υπάρχουν δύο ειδών τράκερ, μονού άξονα και διπλού άξονα. Το πάρκο της έρευνάς μας διαθέτει διπλού άξονα, με κινήσεις που γίνονται προς όλες τις κατευθύνσεις, δηλαδή αλλαγές στον άξονα τις Ανατολής-Δύσης αλλά και αλλαγές στην κλίση ως προς την οριζόντιο. Έτσι μέσω αυτού του συστήματος υπολογίζεται ότι υπάρχει αύξηση στην απόδοση 25-40 % σε σχέση με τα πάνελ σε σταθερή βάση [35].

Τέλος ο τύπος του μετατροπέα (inverter) που χρησιμοποιείται στο πάρκο είναι της εταιρίας SMA και συγκεκριμένα είναι το Sunny Tripower 10000TL με ονομαστική ισχύς 10 KW και με τον συνολικό τους αριθμό να ανέρχεται στα 15.



Σχήμα 2.4: SMA inverter [24]

## Κεφάλαιο 3

# Θεωρητικό υπόβαθρο

### 3.1 Εισαγωγή

Σε αυτό το κεφάλαιο είναι σκόπιμο να αναφέρουμε με το τι πραγματεύεται η μηχανική μάθηση (Machine Learning) αλλά και κάποιες βασικές έννοιες όσον αφορά το θεωρητικό υπόβαθρο που κρύβεται πίσω από τους αλγορίθμους μηχανικής μάθησης που χρησιμοποιούνται στα μοντέλα πρόβλεψης. Επίσης θα αναφερθούμε στις διάφορες τεχνικές που υπάρχουν για την επεξεργασία των δεδομένων και τέλος θα αναφέρουμε τους τρόπους με τους οποίους μετράμε την απόδοση των αλγορίθμων μηχανικής μάθησης

### 3.2 Μηχανική Μάθηση

Η μηχανική μάθηση αποτελεί πεδίο της επιστήμης των υπολογιστών και ορίζεται ως η διαδικασία μέσα από την οποία ένα υπολογιστικό σύστημα, μέσα από την συνεχόμενη εκτέλεση μιας εντολής, βελτιώνει την απόδοση του χωρίς να χρειάζεται εκ νέου προγραμματισμό [36]. Αυτό πετυχαίνεται με την χρήση διαφόρων μαθηματικών και στατιστικών μοντέλων. Έτσι τα συστήματα με την χρήση των παραπάνω μαθηματικών μοντέλων είναι σε θέση να προβλέψουν μελλοντικές συνθήκες.

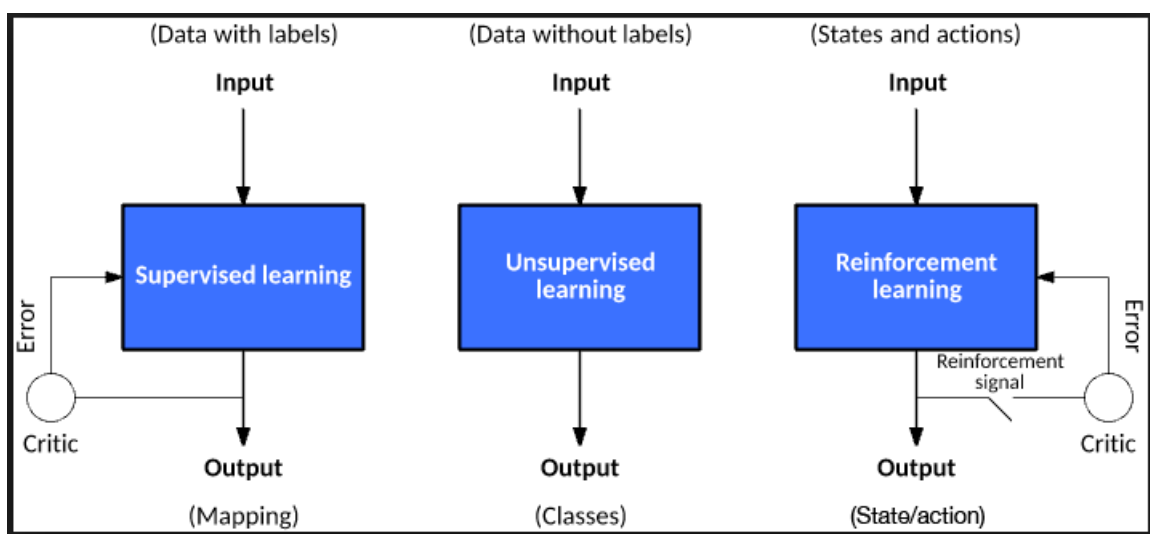
Στον τομέα της μηχανικής μάθησης έχουν αναπτυχθεί τρεις τρόποι μάθησης οι οποίοι μπορούν να συσχετισθούν και με τρόπους που μαθαίνει ένας άνθρωπος. Η επιτηρούμενη μάθηση (Supervised Learning), η μη επιτηρούμενη μάθηση (Unsupervised Learning) και τέλος η ενισχυτική μάθηση (Reinforcement Learning).

- **Επιτηρούμενη μάθηση (Supervised Learning)**, είναι η διαδικασία με την οποία κατά την κατασκευή του αλγορίθμου τα δεδομένα που έχουμε στη διάθεσή μας περιέχουν μέσα και το πραγματικό αποτέλεσμα το οποίο θα πρέπει να μας δώσει ο αλγόριθμος. Έτσι ο αλγόριθμος μπορεί να έχει σαν "δάσκαλο" τα σωστά αποτελέσματα και εμείς να μπορούμε να μετρήσουμε την απόδοσή του. Αυτή η μάθηση χρησιμοποιείται σε προβλήματα Ταξινόμησης (Classification), Πρόγνωσης (Prediction) και Διερμηνείας (Interpretation).
- **Μη επιτηρούμενη μάθηση (Unsupervised Learning)**, είναι η διαδικασία στην οποία κατά

την κατασκευή του μοντέλου, το πρόγραμμα δεν γνωρίζει την τιμή της επιθυμητής εξόδου. Έτσι σ' αυτήν την μορφή μάθησης το πρόγραμμα ψάχνει να βρει κάποια συσχέτιση ή κάποια κοινά χαρακτηριστικά μεταξύ των μεταβλητών του προβλήματος. Αυτή η μέθοδος χρησιμοποιείται σε προβλήματα Ανάλυσης Συσχετισμών (Association Analysis) και Ομαδοποίησης (Clustering)

- **Ενισχυτική μάθηση (Reinforcement Learning)**, ορίζεται ως η διαδικασία μέσα από την οποία ο αλγόριθμος μαθαίνει μία συγκεκριμένη στρατηγική ενεργειών που προκύπτει από την αλληλεπίδρασή του με το περιβάλλον. Αυτή η μάθηση χρησιμοποιείται για καθοδήγηση οχημάτων και ρομπότ.

Όλοι οι παραπάνω μέθοδοι μηχανικής μάθησης εφαρμόζονται πάνω σε ένα πλήθος δεδομένων τα οποία για την καλύτερη εκπαίδευση του μοντέλου αλλά και τον έλεγχο τους από τον προγραμματιστή χωρίζονται σε δύο σύνολα. Το ένα είναι για την εκπαίδευση του μοντέλου και το άλλο για την αξιολόγησή του. Υπάρχουν αρκετές αναλογίες με τις οποίες κάποιος μπορεί να χωρίσει το σύνολο, μία συνηθισμένη επιλογή αναλογίας είναι τα 2/3 να χρησιμοποιούνται για μάθηση και το υπόλοιπο 1/3 για αξιολόγηση του μοντέλου. Τέλος αυτό το σύνολο δεδομένων για την καλύτερη απόδοση των μοντέλων θα πρέπει να τροφοδοτείται συχνά με καινούργια δεδομένα.



Σχήμα 3.1: Τρεις διαφορετικοί τρόποι μάθησης [17]

Μια ακόμη κατηγοριοποίηση που σχετίζεται με το αποτέλεσμα που επιδιώκουμε με τη χρήση αλγορίθμων μηχανικής μάθησης είναι η παρακάτω.

- **Ταξινόμηση (Classification)**, με αυτή τη μέθοδο προσπαθούμε να ταξινομήσουμε τα διαθέσιμα δεδομένα σε δύο ή και περισσότερες κλάσεις. Αυτή η μέθοδος ανήκει στην επιτηρούμενη μάθηση. Ο έλεγχος για το αν ένα χαρτονόμισμα είναι πλαστό ή όχι είναι ένα παράδειγμα ταξινόμησης που περιέχει δύο κλάσεις, τη κλάση που περιέχει τα γνήσια χαρτονόμισμα και εκείνη με τα πλαστά.



- **Παλινδρόμηση (Regression)**, μέθοδο την οποία τη χρησιμοποιούμε για πρόβλεψη συνεχών τιμών, όπως για παράδειγμα η πρόβλεψη της τιμής μίας μετοχής στο χρηματιστήριο. Και αυτή η μέθοδος ανήκει στην επιτηρούμενη μάθηση.
- **Συσταδοποίηση (Clustering)**, με την οποία πάλι προσπαθούμε να κατηγοριοποιήσουμε τα δεδομένα σε ομάδες όμως αυτές δεν είναι γνωστές, σε αντίθεση με την ταξινόμηση. Όπως είναι φυσικό αυτή η μέθοδος ανήκει στην μη επιτηρούμενη μάθηση.
- **Μείωσης διαστασιμότητας (Dimensionality reduction)**, μέθοδος η οποία χρησιμοποιείται για την απλοποίηση των δεδομένων, μετατρέποντας τον χώρο σε λιγότερες διαστάσεις.

### 3.3 Σύνολο δεδομένων εκπαίδευσης και ελέγχου

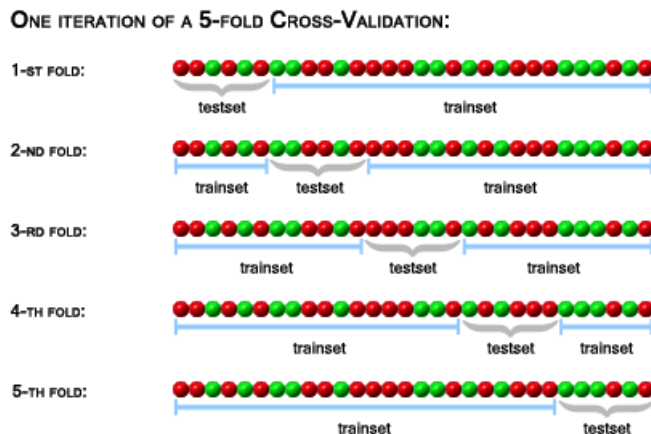
Αρκετές θεωρίες έχουν αναπτυχθεί σχετικά με το ποιος είναι ο πιο κατάλληλος συνδυασμός μεταξύ συνόλου δεδομένων για εκπαίδευση και ελέγχου. Το πιο συχνό σε χρήση είναι ο διαχωρισμός του συνόλου δεδομένων σε 75% δεδομένα για εκπαίδευση του αλγορίθμου και τα υπόλοιπα 25% για τον έλεγχο του. Συχνά όμως θα δούμε να χρησιμοποιείται και μικρότερο ποσοστό για την εκπαίδευση του αλγορίθμου με κύριο στόχο την αποφυγή της υπερπροσαρμογής (overfitting). Έτσι ο αλγόριθμος χρησιμοποιεί τα δεδομένα εκπαίδευσης για να κατασκευάσει το μοντέλο και στη συνέχεια τα δεδομένα ελέγχου μέσω των οποίων προκύπτει και το πόσο αποδοτικός είναι ο αλγόριθμος.

### 3.4 K-fold Cross Validation

Αφού τα δεδομένα έχουν χωριστεί σε δύο υποσύνολα ένα για εκπαίδευση και ένα για έλεγχο στην συνέχεια ο αλγόριθμος θα πρέπει να υπολογίσει τις παραμέτρους του μοντέλου. Για τον λόγο αυτό θα χρησιμοποιήσει τα δεδομένα εκπαίδευσης. Όμως ο τρόπος που θα έχουν χωριστεί τα δεδομένα, όντας τυχαίος κατά πάσα πιθανότητα δεν θα μας δώσει το καλύτερο δυνατό αποτέλεσμα για την πρόβλεψη, και ο λόγος είναι ότι τα δεδομένα εκπαίδευσης μπορεί να μην είναι τα κατάλληλα. Για να λύσουμε αυτό το πρόβλημα θα χρησιμοποιήσουμε τη μέθοδο Cross Validation και συγκεκριμένα την K-fold Cross Validation. Όπως θα δούμε στο Σχήμα 3.2 πιο κάτω, με τη μέθοδο αυτή κατασκευάζουμε ανάλογα με την τιμή που έχουμε δώσει στην παράμετρο  $k$ , τόσα σύνολα δεδομένων, από δεδομένα εκπαίδευσης και ελέγχου χωρισμένα με τυχαίο τρόπο. Έτσι, για παράδειγμα αν έχουμε ορίσει την μεταβλητή  $k$  ίσον με 5, θα δημιουργηθούν 5 σύνολα δεδομένων με το καθένα να είναι χωρισμένα με τυχαίο αλλά διαφορετικό τρόπο κάθε φορά σε δεδομένα εκπαίδευσης και ελέγχου.

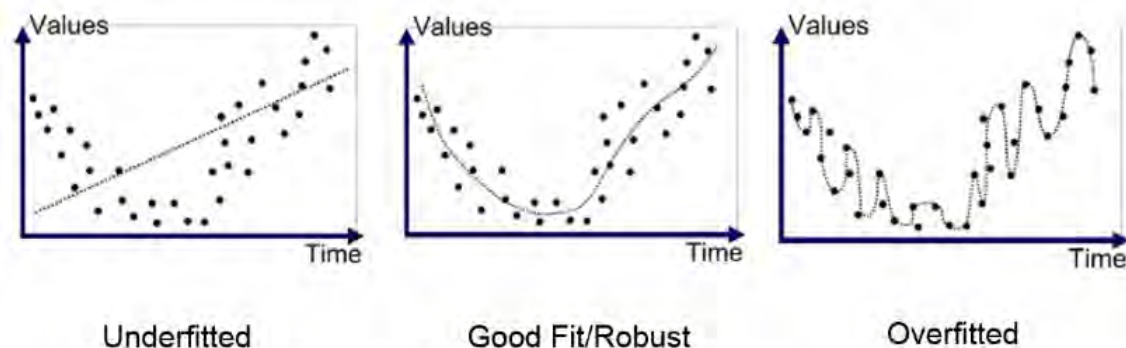
### 3.5 Υπερπροσαρμογή (overfitting) και υποπροσαρμογή (underfitting)

Ένα αρκετά συχνό φαινόμενο που παρατηρείτε στα μοντέλα μηχανικής μάθησης είναι αυτό της υπερπροσαρμογής (overfitting). Με τον όρο αυτόν χαρακτηρίζουμε ένα μοντέλο όταν το μοντέλο έχει "μάθει" πάρα πολύ καλά τα δεδομένα εκπαίδευσης με αποτέλεσμα να μην μπορεί να



Σχήμα 3.2: Διαδικασία K-fold Cross-Validation [6]

κάνει καλές προβλέψεις για τα δεδομένα ελέγχου. Έτσι έχει καταλήξει να μαθαίνει τα πάντα από τα δεδομένα εκπαίδευσης ακόμα και τον θόρυβο. Αυτό συχνά συμβαίνει όταν δεν έχει γίνει καλός διαχωρισμός του συνόλου δεδομένων. Στην αντίθετη περίπτωση τώρα, όταν ένας αλγόριθμος δεν έχει "μάθει" καλά τα δεδομένα εκπαίδευσης τότε έχουμε το φαινόμενο της υποπροσαρμογής (underfitting) το οποίο μπορεί να αντιμετωπιστεί συνήθως με την εισαγωγή περισσότερων δεδομένων εκπαίδευσης. Αρκετές συζητήσεις έχουν γίνει σχετικά με το αν το φαινόμενο της υπερπροσαρμογής για το αν είναι καλό να υπάρχει και οι απόψεις δίστανται.



Σχήμα 3.3: Παράδειγμα overfitting και underfitting [20]

### 3.6 Μέτρο απόδοσης ενός μοντέλου

Ένας τρόπος για να αξιολογήσουμε εάν ένα μοντέλο είναι καλό είναι να μετρήσουμε την απόδοσή του ως προς τα αποτελέσματά του. Αυτό πετυχαίνεται με τον υπολογισμό του σφάλματος, το οποίο δεν είναι τίποτα άλλο παρά η διαφορά μεταξύ της πραγματικής και της τιμής πρόβλεψης που δίνει το μοντέλο. Η παραπάνω μέθοδος εφαρμόζεται σε μοντέλα επιτηρούμενης μάθησης και είναι φυσικό ότι όσο πιο μικρό είναι το σφάλμα τόσο καλύτερη είναι η απόδοση του μοντέλου. Μερικά παραδείγματα μαθηματικών μοντέλων που υπολογίζουν την απόδοση ενός μοντέλου είναι

τα παρακάτω.

- **Μέσο Τετραγωνικό Σφάλμα (MSE) (Mean Square Error)**, το οποίο μετράει τον μέσο όρο των τετραγώνων των σφαλμάτων μεταξύ των πραγματικών τιμών και εκείνων που προβλέπει το μοντέλο.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

- **Ρίζα Μέσου Τετραγωνικού Σφάλματος (RMSE)**, το οποίο είναι παρόμοιο με MSE μόνο που περιλαμβάνει όλη η έκφραση από μία τετραγωνική ρίζα.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

- **Μέσο Απόλυτο Σφάλμα (MAE) (Mean Absolute Error)**, το οποίο μετράει τα αθροίσματα των διαφορών μεταξύ πραγματικής και προβλεπόμενης τιμής διαιρεμένο με το πλήθος όλων των παρατηρήσεων.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

- **Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (MAPE) (Mean Absolute Percentage Error)**, αναφέρεται στο ποσοστιαίο σφάλμα και ορίζεται ως η απόλυτη τιμή της διαφορά μεταξύ της πραγματικής και προβλεπόμενης τιμής προς την πραγματική τιμή και όλο διαιρούμενο με το πλήθος των παρατηρήσεων.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

Στις παραπάνω εκφράσεις το  $\hat{Y}_i$  αναφέρεται στις προβλεπόμενες τιμές ενώ το  $Y_i$  στις πραγματικές τιμές. Τέλος το  $n$  αναφέρεται στο σύνολο των παρατηρήσεων.

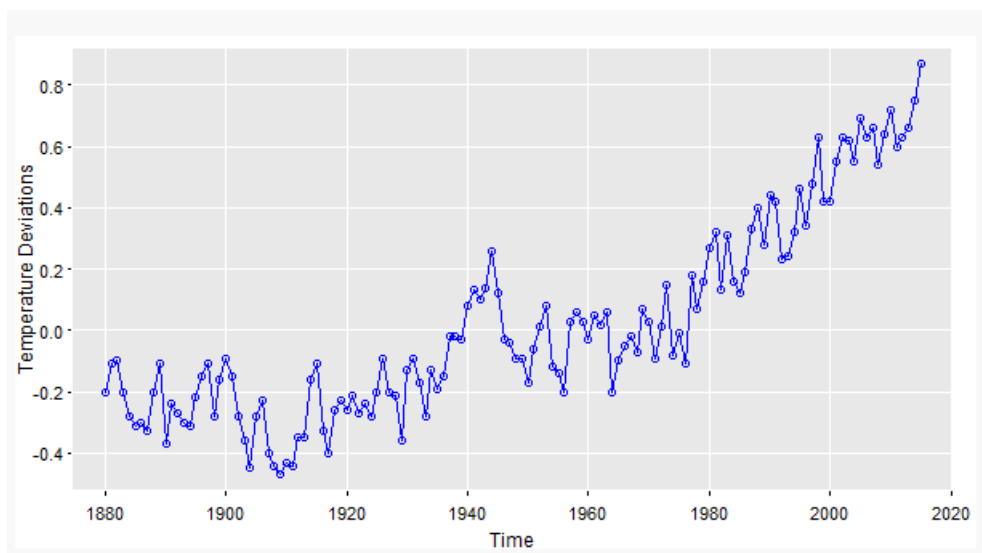
### 3.7 Χρονολογικές Σειρές (Time Series)

Με τον όρο χρονολογική σειρά αναφερόμαστε σε μία σειρά από παρατηρήσεις που γίνονται σε ορισμένες χρονικές στιγμές ή περιόδους και ισαπέχουν μεταξύ τους. Αποτελεί μια στοχαστική διαδικασία δηλαδή κάθε παρατήρηση αποτελεί μια τυχαία μεταβλητή.

Δεδομένα χρονολογικών σειρών συναντάμε σε πολλές επιστήμες όπως οικονομικές, κοινωνικές, ιατρική κτλ. Τέτοια δεδομένα μπορεί να είναι η τιμή της θερμοκρασίας για κάθε μία ώρα σε μία περιοχή, ή η τιμή μιας μετοχής ανά 5 λεπτά για μία μέρα. Βασικό χαρακτηριστικό όλων των παραπάνω είναι εξάρτηση μεταξύ των χρονολογικών σειρών.

Σκοπός της μελέτης των χρονολογικών σειρών είναι ο εντοπισμός αλληλεξαρτήσεων που υπάρχουν μεταξύ των παρατηρήσεων. Με απώτερο στόχο την δημιουργία μοντέλων πρόβλεψης και στην συνέχεια μείωση της αβεβαιότητας και την καλύτερη εκτίμηση του τελικού αποτελέσματος.

Στο παρακάτω σχήμα (Σχήμα 3.4) βλέπουμε την σχηματική αναπαράσταση μιας χρονολογικής σειράς που τις διαφορές της θερμοκρασίας στο πέρασμα του χρόνου



Σχήμα 3.4: Παράδειγμα χρονολογικής σειράς για διαφορά θερμοκρασίας [29]

Οι χρονολογικές σειρές μπορούν να περιέχουν ένα ή και περισσότερα από τα παρακάτω στοιχεία

- **Τάση (Trend).** Με τον όρο αυτόν αναφερόμαστε στην συνολική κατεύθυνση που ακολουθεί η χρονολογική σειρά. Αυτή μπορεί να είναι είτε αύξουσα όταν αυξάνονται οι τιμές σε συνάρτηση με το χρόνο είτε φθίνουσα όταν μειώνονται.
- **Εποχική Συνιστώσα ή Εποχικότητα (Seasonality),** μοιάζει αρκετά με την τάση που αναφέραμε παραπάνω όμως διαφέρει στο ότι θα πρέπει να επαναλαμβάνεται η χρονολογική σειρά περιοδικά στο χρόνο (π.χ. εβδομαδιαία, μηνιαία, κτλ).

Αξίζει να σημειωθεί σ' αυτό το σημείο ότι η τάση και η εποχικότητα μπορούν να συνυπάρχουν. Για παράδειγμα οι κατανάλωση φυσικού αερίου σε μια περιοχή μπορεί να αυξάνεται με την πάροδο των ετών όμως μπορεί ταυτόχρονα να υπάρχει και μία εποχικότητα κατά τους χειμερινούς μήνες.

- **Τυχαία Συνιστώσα (Random component, noise).** Αναφέρεται στις μεταβολές της χρονολογικής σειράς, στην πάροδο του χρόνου που δεν αποτελούν μέρος των δύο προηγούμενων και που μπορούν να αποδοθούν σε θόρυβο ή σε άλλα τυχαία γεγονότα.

### 3.7.1 Στατιστικά μεγέθη χρονοσειράς

- **Μέση τιμή ή αναμενόμενη τιμή.** Αναφέρεται στην μέση τιμή των παρατηρήσεων της χρονολογικής σειράς και δίνεται από την παρακάτω σχέση

$$\mu_t = E(Y_t) = \int_{-\infty}^{+\infty} y f_{y_t}(y_t) dy$$

Όπου  $y_t$  η παρατήρηση της χρονικής σειράς στην συγκεκριμένη χρονική στιγμή  $t$ . Με τη χρήση της μέσης τιμής είναι εύκολο να βρούμε και την τάση μιας χρονικής σειράς (αυξητική ή φθίνουσα) σε ένα χρονικό διάστημα.

- **Αυτοσυνδιακύμανση (Autocovariance).** Η αυτοσυνδιακύμανση μετράει την συνδιακύμανση μεταξύ δύο παρατηρήσεων που βρίσκονται στην ίδια χρονολογική σειρά σε κάποια απόσταση μεταξύ τους. Με τον όρο συνδιακύμανση αναφερόμαστε σε μία στατιστική τιμή η οποία μας δείχνει τι μεταβολές θα επιφέρει η αλλαγή σε μια συνεχής τυχαία μεταβλητή όταν μία άλλη μεταβάλλεται. Η αυτοσυνδιακύμανση συμβολίζεται με  $\gamma_k$  και ορίζεται.

$$\gamma_k = Cov(y_t, y_{t+k}) = E[y_t - E(y_t)][y_{t+k} - E(y_{t+k})]$$

- **Αυτοσυσχέτιση (Autocorrelation).** Ο συντελεστής αυτοσυσχέτισης είναι ένας στατιστικός συντελεστής που μας βοηθάει στην ανάλυση των χρονολογικών σειρών και έχει σαν στόχο τον καθορισμό της τυχειότητας ή μη της χρονολογικής σειράς. Ορίζεται ως:

$$\rho_k = \frac{Cov(y_t, y_{t+k})}{\sqrt{Var(y_t)}\sqrt{Var(y_{t+k})}}$$

Οι τιμές που μπορεί να πάρει η αυτοσυσχέτιση είναι από  $[-1, 1]$  και μας δείχνει τον βαθμό συσχέτισης μεταξύ δύο μεταβλητών. Όσο πιο κοντά είναι στο 1 και το -1 τόσο μεγαλύτερη συσχέτιση υπάρχει. Δημιουργώντας έτσι το γράφημα της αυτοσυσχέτισης σε συνάρτηση με την καθυστέρηση  $j$  μπορούμε να βρούμε τον συντελεστής αυτοσυσχέτισης (autocorrelation function, ACF) τον οποίον θα χρειαστούμε παρακάτω για να εξηγήσουμε τους συντελεστές που θα χρησιμοποιήσουμε για να δημιουργήσουμε τα μοντέλα.

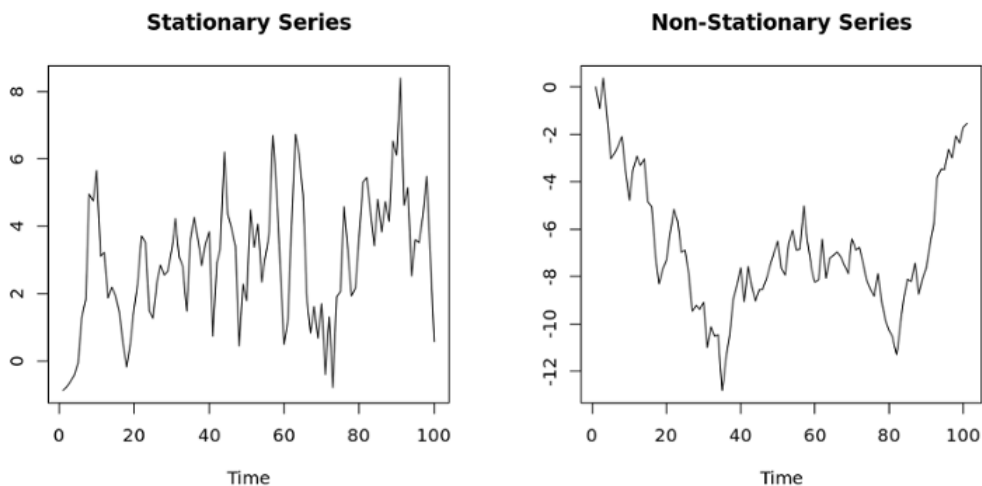
- **Στασιμότητα (Stationarity):** Μία χρονολογική σειρά είναι στάσιμη όταν οι στατιστικές της ιδιότητες δεν αλλάζουν (μέση τιμή, διακύμανση) [38]. Η ιδιότητα της στασιμότητας είναι αρκετά σημαντική καθώς είναι απαραίτητη προϋπόθεση για την εφαρμογή ορισμένων εργαλείων ανάλυσης χρονολογικών σειρών. Αρκετές χρονολογικές σειρές που βρίσκουμε είναι μη στάσιμες και έτσι είναι δύσκολο να μελετηθούν. Γι' αυτόν τον λόγο χρησιμοποιούμε κατάλληλες τεχνικές για να μετατρέψουμε αυτές τις σειρές σε στάσιμες για να χρησιμοποιήσουμε αντίστοιχες μεθόδους ανάλυσης. Για μια στάσιμη χρονολογική σειρά θα ισχύουν:

$$E(y_t) = \mu_y$$

$$Var(y_t) = E[y_t - E(y_t)]^2 = \sigma_y^2$$

$$\text{Cov}(y_t, y_{t+k}) = \text{Cov}(y_{t+m}, y_{t+k+m}) = \gamma_k$$

Οι δυο πρώτες ισότητες δηλώνουν ότι σε μία στάσιμη χρονολογική σειρά η μέση τιμή και η διακύμανση δεν μεταβάλλονται όπως αναφέραμε και πιο πάνω ενώ το τελευταίο μας δείχνει ότι η συνδιακύμανση μεταξύ δύο τιμών  $y_t$  που απέχουν μεταξύ τους διάστημα  $k$ , είναι συναρτήσει μόνο του  $k$ .



Σχήμα 3.5: Χρονολογική σειρά στάσιμη και μη στάσιμη [14]

### 3.7.2 Αλγόριθμοι

#### Αυτοπαλίνδρομα Υποδείγματα AR(p) (Autoregressive Model)

Σε μία χρονολογική σειρά σημαντικό μέρος ενός μοντέλου πρόβλεψης είναι η προηγούμενη ή προηγούμενες τιμές της σειράς. Έτσι προέκυψε το μοντέλο AR το οποίο για να προβλέψει την τωρινή τιμή χρησιμοποιεί προηγούμενες τιμές του συνόλου δεδομένων. Έτσι ορίζεται από τον παρακάτω τύπο.

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + \epsilon_t$$

Όπου  $a_0$  είναι η αρχική τιμή,  $a_{1-p}$  οι συντελεστές του μοντέλου,  $Y_{t-1} - Y_{t-p}$  οι προηγούμενες τιμές και  $\epsilon_t$  ο τυχαίος θόρυβος.

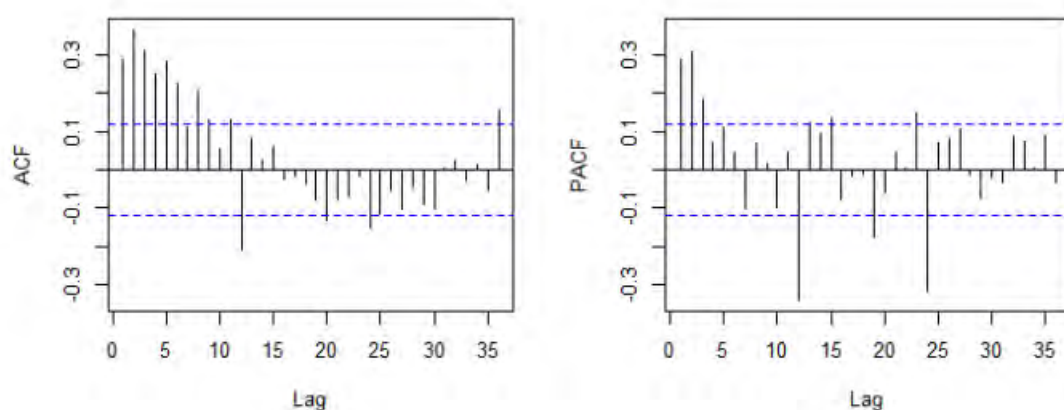
#### Υποδείγματα Κινητού Μέσου MA(q) (Moving Average)

Ένα άλλο σημαντικό μοντέλο πρόβλεψης είναι το υπόδειγμα κινητού μέσου ή MA, στο οποίο η τρέχουσα τιμή προκύπτει από το άθροισμα των προηγούμενων τιμών του τυχαίου θορύβου πολλαπλασιαζόμενο αντίστοιχα με έναν συντελεστή  $\theta_q$ . Έτσι ορίζεται από τον παρακάτω τύπο.

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Όπου  $\mu$  η μέση τιμή της χρονολογικής σειράς,  $\epsilon_t - \epsilon_{t-1}$  ο τυχαίος θόρυβος κάθε παρατήρησης και  $\theta_1 - \theta_{t-q}$  οι συντελεστές του μοντέλου.

Για την επιλογή της τιμής των  $p$  και  $q$  των παραπάνω μοντέλων χρησιμοποιήσαμε τις γραφικές παραστάσεις της συνάρτησης μερικής αυτοσυσχέτισης (PACF) (Partial Autocorrelation Function) και την γραφική παράσταση της συνάρτησης αυτοσυσχέτισης (ACF). Έτσι για την εύρεση του  $p$  μέσω του PACF παρατηρούμε πότε οι τιμές των χρονικών καθυστερήσεων (lags) στο Σχήμα 3.6, πέσουν για πρώτη φορά κάτω από το άνω επίπεδο εμπιστοσύνης (upper confidence interval). Το ίδιο κάνουμε για να βρούμε το  $q$  χρησιμοποιώντας το διάγραμμα του ACF. Στα δύο διαγράμματα (Σχήμα 3.6) βλέπουμε ένα παράδειγμα των συναρτήσεων PACF και ACF.



Σχήμα 3.6: Παράδειγμα διαγράμματος PACF και ACF

#### Μεικτό Αυτοπαλίνδρομο-Κινητού Μέσου Υπόδειγμα ARMA( $p,q$ ) (Autoregressive Moving Average Model)

Πρόκειται για έναν συνδυασμό των δυο παραπάνω μοντέλων (AR) (MA). Η γενική μορφή ενός ARMA μοντέλου είναι ARMA( $p,q$ ) όπου το  $p$  αναφέρεται στους αυτοπαλινδρους όρους και το  $q$  στους όρους κινητού μέσου. Έτσι ορίζεται από τον παρακάτω τύπο.

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

Όπου  $a_0$  είναι η αρχική τιμή,  $a_{1-p}$  οι συντελεστές του μοντέλου,  $Y_{t-1} - Y_{t-p}$  οι προηγούμενες τιμές,  $\epsilon_t$  ο τυχαίος θόρυβος,  $\epsilon_t - \epsilon_{t-1}$  ο τυχαίος θόρυβος κάθε παρατήρησης και  $\theta_1 - \theta_{t-q}$  οι συντελεστές του μοντέλου.

#### Μεικτό Αυτοπαλίνδρομο Ολοκληρωμένο Υπόδειγμα Κινητού Μέσου ARIMA( $p,q,d$ )

Όπως έχουμε αναφέρει και παραπάνω αρκετές χρονολογικές σειρές δεν είναι στάσιμες και έτσι για να χρησιμοποιήσουμε κάποια στατιστικά μοντέλα θα πρέπει να τις μετατρέψουμε σε στάσιμες. Όταν μία χρονολογική σειρά μετατρέπεται σε στάσιμη παίρνοντας τις πρώτες διαφορές η σειρά αυτή ονομάζεται ολοκληρωμένη πρώτης τάξης και συμβολίζεται με  $I(1)$ . Αν παίρναμε τις δεύτερες διαφορές θα ονομαζόταν ολοκληρωμένη δεύτερης σειράς και ούτω καθεξής. Έτσι με

το  $d$  συμβολίζουμε τον αριθμό των διαφορών που έχει υποστεί η σειρά για να γίνει στάσιμη και συμβολίζεται με  $I(d)$ .

Οι τιμές  $p$  και  $q$  αναφέρονται στους αυτοπαλινδρους όρους και στους όρους κινητού μέσου αντίστοιχα. Έτσι ένα μοντέλο ARIMA( $p,q,d$ ) αναπαρίσταται με την παρακάτω μορφή [18]:

$$a(B)(1 - B)^d y_t = \delta + \Theta(B)\epsilon_t$$

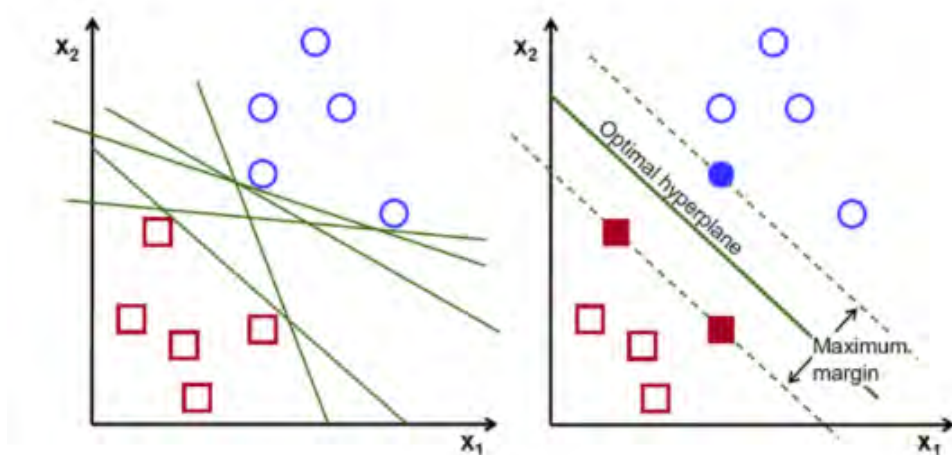
Όπου  $a(B)$  οι συντελεστές του μοντέλου AR,  $\Theta(B)$  οι συντελεστές του μοντέλου MA,  $\delta$  η αρχική τιμή και  $\epsilon_t$  ο τυχαίος θόρυβος.

### 3.8 Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines)

Μια από τις πιο διάσημες αλλά και πολυσυζητημένες μέθοδοι είναι οι μηχανές διανυσματικής υποστήριξης ή με τη συντομογραφία SVM (Support Vector Machines). Η μέθοδος αυτή εφευρέθηκε το 1963 από τους Vladimir N. Vapnik και Alexey Ya. Chervonenkis [8] και έγινε αρκετά γνωστή σαν μέθοδο περίπου το 1990 [5]. Αρχικά αυτή η μέθοδος είχε αρχίσει να χρησιμοποιείται μόνο για προβλήματα ταξινόμησης όμως την πορεία των χρόνων αναπτύχθηκαν μοντέλα SVM και για την αντιμετώπιση προβλημάτων παλινδρόμησης.

#### Ταξινόμηση

Οι μηχανές διανυσματικής υποστήριξης όπως αναφέραμε και παραπάνω στην αρχή άρχισαν να χρησιμοποιούνται για διαχωρισμό των δεδομένων σε δύο κατηγορίες ή κλάσεις. Βασικός τους στόχος είναι η δημιουργία ενός υπερεπιπέδου έτσι ώστε να μεγιστοποιείται το περιθώριο (margin) μεταξύ των δύο κλάσεων .



Σχήμα 3.7: Γραφική απεικόνιση της ευθείας του αλγορίθμου SVM [27]

Όπως βλέπουμε στο αριστερό διάγραμμα του σχήματος (Σχήμα 3.7) για να διαχωρίσουμε τα δεδομένα σε δύο κλάσεις υπάρχουν αρκετοί τρόποι, ο αλγόριθμος SVM όμως μας βοηθάει να



βρούμε εκείνη την γραμμή που θα κάνει την καλύτερη δυνατή διαχώριση των δεδομένων. Επομένως έχουμε δύο ευθείες που ορίζουν τον χώρο και ονομάζονται boundary lines, την ευθεία που βρίσκεται στον χώρο που έχουν δημιουργήσει αυτές οι δύο ευθείες που ονομάζεται optimal hyperplane και τέλος τα διανύσματα υποστήριξης, support vectors τα οποία μας βοηθούν στην δημιουργία του μοντέλου.

Η εξίσωση της ευθείας του θα χωρίσει το χώρο σε δύο κλάσεις είναι  $wx + b = 0$ . Στόχος μας όπως αναφέραμε παραπάνω είναι να βρούμε εκείνη την ευθεία που θα μεγιστοποιήσει το περιθώριο μεταξύ των δύο κλάσεων. Γι' αυτόν τον λόγο θα χρησιμοποιήσουμε δυο διανύσματα υποστήριξης  $wx + b = + - 1$  και θα πρέπει να βρούμε την μεταξύ τους απόσταση παίρνοντας την κάθετη απόσταση. Έτσι προκύπτει:

$$d = \frac{2}{\|w\|}$$

Όπως αναφέραμε στόχος μας είναι η μεγιστοποίηση της απόστασης  $d$  οπότε ισοδύναμα μπορούμε να ελαχιστοποιήσουμε το  $d = \frac{2}{\|w\|}$ .

$$\min \frac{(\|w\|)^2}{2}, y_i(wx_i + b) \geq 1, \forall x_i$$

Με τον περιορισμό της εξίσωσης που θέσαμε ορίζουμε ότι τα δεδομένα της κλάσης 1 θα πρέπει να είναι στα δεξιά του υπερεπιπέδου ενώ της δεύτερης κλάσης στα αριστερά.

Ο παραπάνω τύπος θα μπορούσε να εφαρμοστεί σε ένα ιδανικό πρόβλημα ταξινόμησης που όλα τα στοιχεία θα ταξινομούνται πλήρως πράγμα το οποίο δύσκολα συναντάται στον πραγματικό κόσμο. Σ' αυτή την περίπτωση αναζητάμε ένα υπερεπίπεδο «χαλαρού» περιθωρίου, δηλαδή μια επιφάνεια απόφασης η οποία διαχωρίζει τα δεδομένα εκπαίδευσης κάνοντας τα λιγότερα λάθη. Την ονομάζουμε  $\xi_i$  και είναι μεγαλύτερη ή ίση με το μηδέν ανάλογα τις λανθασμένες κατηγοριοποιήσεις. Έτσι η τελική μορφή του θα είναι.

$$\min \left( \frac{(\|w\|)^2}{2} + C \sum_{i=1}^N \xi_i \right), y_i(wx_i + b) \geq 1 - \xi_i, \forall x_i, \xi_i \geq 0$$

Όπου  $C$  είναι ένας συντελεστής που καθορίζει πόσο αυστηροί θέλουμε να είμαστε με τα σφάλματα.

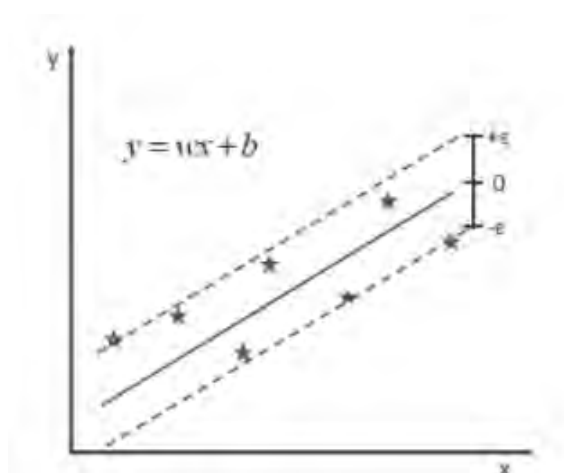
### Παλινδρόμηση

Ο σκοπός της παλινδρόμησης σε ένα πρόβλημα στον διδιάστατο χώρο είναι η δημιουργία μιας ευθείας με στόχο την ελαχιστοποίηση των σφαλμάτων μεταξύ των πραγματικών τιμών και αυτών της ευθείας. Αυτό το πρόβλημα μπορεί να λυθεί με την χρήση της μεθόδου της γραμμικής παλινδρόμησης όμως όταν τα δεδομένα δεν θα είναι γραμμικά δεν μπορεί να εφαρμοστεί αυτήν η

μέθοδος. Έτσι την θέση της έρχεται να πάρει η μέθοδος των Μηχανών Διανυσματικής Υποστήριξης για παλινδρόμηση (SVR).

Σκοπός αυτής της μεθόδου όπως και παραπάνω είναι η μεγιστοποίηση του περιθωρίου μεταξύ των παρατηρήσεων και του υπερεπιπέδου. Αυτή η μέθοδος χρησιμοποιεί τα ίδια στοιχεία με την SVM όμως με μερικές τροποποιήσεις. Αρχικά, στη περίπτωση της SVR τα διανύσματα υποστήριξης έχουν τον ρόλο ενός περιθωρίου ανοχής (epsilon). Τα δυο αυτά διανύσματα όπως και παραπάνω, ορίζουν τον χώρο μέσα στον οποίο θα βρίσκεται το υπερεπίπεδο που θα μοντελοποιεί τις παρατηρήσεις. Η μορφή του μοντέλου είναι αρκετά παρόμοια με της ταξινόμησης.

$$\min \frac{(\|w\|)^2}{2}, y_i - wx_i - b \leq \epsilon, wx_i + b - y_i \leq \epsilon$$



Σχήμα 3.8: Γραφική απεικόνιση της ευθείας του αλγορίθμου SVR [26]

Όπως και στην περίπτωση της ταξινόμησης το αρχικό μοντέλο που περιγράψαμε είναι για ιδανικές συνθήκες πράγμα το οποίο δύσκολα το συναντάμε στον πραγματικό κόσμο. Γι' αυτόν τον λόγο υπάρχει ένα ζεύγος μεταβλητών  $\xi_i$  και  $\xi_i^*$ , οι οποίες μας δείχνουν πόσο απέχουν οι παρατηρήσεις από τα διανύσματα υποστήριξης. Όπως βλέπουμε και στο Σχήμα 3.8, εάν οι παρατηρήσεις βρίσκονται μέσα στο χώρο που ορίζουν τα διανύσματα υποστήριξης τότε οι δύο αυτές μεταβλητές είναι μηδέν. Έτσι η μορφή του μοντέλου ύστερα από αυτήν την προσθήκη θα γίνει:

$$\min \left( \frac{(\|w\|)^2}{2} + C \sum_{i=1}^N \xi_i + \xi_i^* \right),$$

$$y_i - wx_i - b \leq \epsilon + \xi_i, \forall x_i, wx_i + b - y_i \leq \epsilon + \xi_i^*, \forall x_i, \xi_i, \xi_i^* \geq 0$$

### Πυρήνας (Kernel)

Η χρήση της SVR όπως αναφέραμε και παραπάνω για μοντέλο παλινδρόμησης υπερέχει έναντι της γραμμικής παλινδρόμησης διότι μπορεί να χρησιμοποιηθεί και για μη γραμμικά μοντέλα. Αυτό

το πετυχαίνει με τη χρήση συναρτήσεων πυρήνα (kernel), τα οποία μετατρέπουν τα δεδομένα του προβλήματος σε χώρο μεγαλύτερης διάστασης από αυτόν που βρίσκονται. Αυτή η μέθοδος μπορεί να χρησιμοποιηθεί και για μοντέλα ταξινόμησης. Έτσι η μορφή των συναρτήσεων πυρήνα θα είναι [30]:

$$K(x_i, x_j) = \phi(x_i)\phi(x_j)$$

Οι πιο γνωστές συναρτήσεις πυρήνα (kernel) είναι:

- Γραμμική (Linear)

$$K(x_i, y_i) = x_i y_i + \gamma$$

- Πολυωνυμική (Polynomial)

$$K(x_i, y_i) = (x_i y_i + 1)^p$$

- Ακτινωτή (Radial Basis Function) (RBF) (Radial Basis Function)

$$K(x_i, y_i) = e^{-\gamma(x_i - x_j)^2}$$

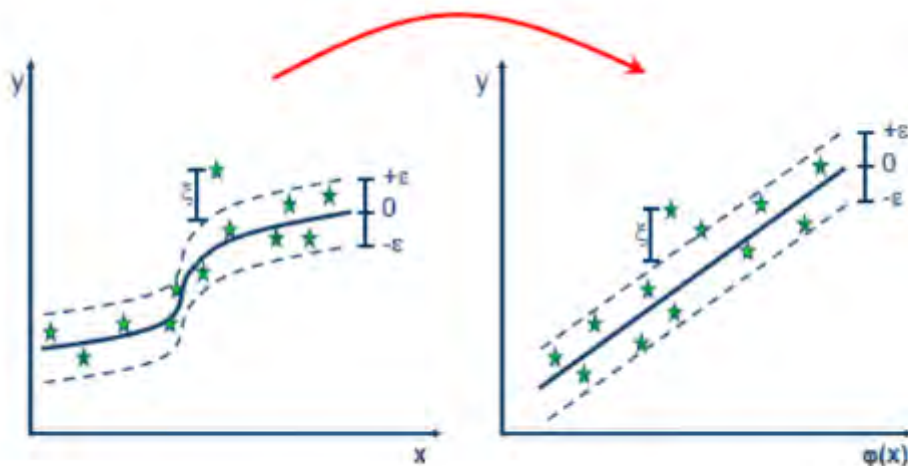
- Γκαουσιανή (Gaussian kernel)

$$K(x_i, y_i) = e^{-\frac{1}{2\sigma^2}(x_i - y_i)^2}$$

### 3.9 Gradient Boosted Regression Trees (GBRT)

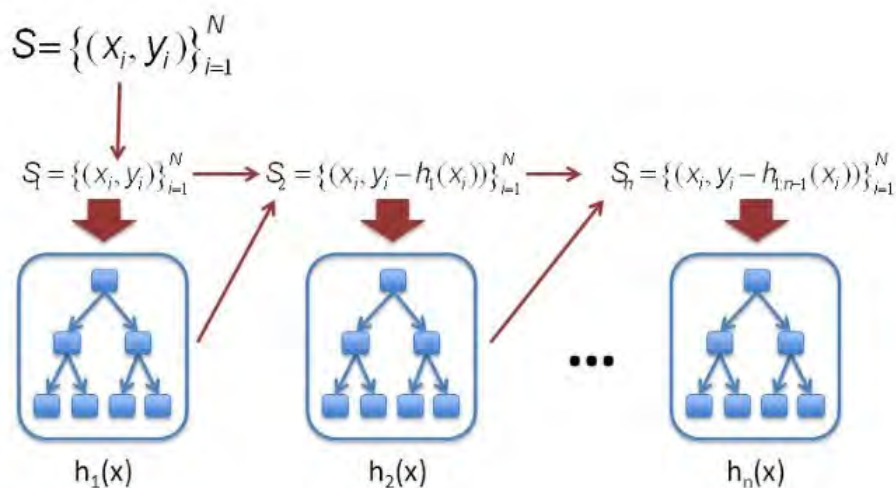
Οι αλγόριθμοι boosting αποτελούνται από μια οικογένεια ισχυρών τεχνικών μηχανικής μάθησης που έχουν αποδείξει ότι δίνουν αρκετά καλά αποτελέσματα σε ένα ευρύ φάσμα πρακτικών εφαρμογών. Είναι ιδιαίτερα προσαρμόσιμοι στις ανάγκες της εφαρμογής, καθώς η μάθηση γίνεται σε σχέση με τις διαφορετικές συναρτήσεις κόστους [19].

Οι αλγόριθμοι boosting χρησιμοποιούνται τόσο για προβλήματα παλινδρόμησης όσο και ταξινόμησης. Η λειτουργία τους στηρίζεται στην δημιουργία ενός μοντέλου πρόβλεψης που θα στηρίζεται σε ένα σύνολο αδύναμων μοντέλων πρόβλεψης. Σαν αδύναμο μοντέλο ορίζουμε εκείνο του οποίου η πρόβλεψή του είναι λίγο καλύτερη από το τυχαίο αποτέλεσμα. Η βασική ιδέα του boosting είναι η σύνδεση κάθε παρατήρησης με ένα βάρος. Οπότε, έτσι κάθε μοντέλο αποτελείται από ένα σύνολο βαρών για το κάθε υπό μοντέλο, το οποίο ανάλογα με το αποτέλεσμα που δίνει το μοντέλο προσαρμόζεται.



Σχήμα 3.9: Μετατροπή με τη χρήση συνάρτησης πυρήνα [16]

Όπως αναφέραμε παραπάνω τα βάρη προσαρμόζονται ανάλογα με το αποτέλεσμα και στο τέλος προκύπτει ένα αθροιστικό μοντέλο, το οποίο περιέχει τα αποτελέσματα κάθε υπό μοντέλου μαζί με ένα βάρος που έχει λάβει. Το πλεονέκτημα αυτού του μοντέλου είναι ότι δεν χρειάζεται αρκετή προσαρμογή καθώς το μόνο που θα πρέπει να το προσέχουμε είναι η επιλογή και μόνο αδύναμων μοντέλων πρόβλεψης [9].



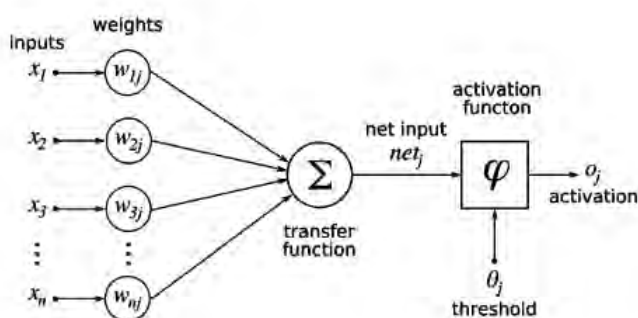
Σχήμα 3.10: Γραφική απεικόνιση του αλγορίθμου GBRT [15]

### 3.10 Τεχνητά Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αποτελούν μία σχετικά καινούργια περιοχή στον τομέα της μηχανικής μάθησης καθώς η μελέτη και η έρευνά τους έχει αρχίσει τα τελευταία 40 χρόνια. Όμως παρόλα αυτά η μελέτη τους έχει προχωρήσει αρκετά και σήμερα αποτελούν μία από τις καλύτερες λύσεις για την επίλυση προβλημάτων μηχανικής μάθησης και κυρίως για την κατασκευή μοντέλων πρόβλεψης. Είναι εφαρμόσιμα σχεδόν σε κάθε κατάσταση όπου έχουμε ύπαρξη εξαρτημένων και ανεξάρτητων μεταβλητών, ακόμα και όταν αυτή η σχέση μεταξύ αυτών των μεταβλητών είναι δύσκολη να βρεθεί και αποτυπωθεί με τους άλλους τρόπους.

Η βασική τους λειτουργία μοιάζει πάρα πολύ με την λειτουργία των νευρώνων του ανθρώπινου εγκεφάλου, σε μία πιο απλοϊκή μορφή βέβαια. Αποτελούνται από στοιχεία τα οποία είναι διασυνδεδεμένα μεταξύ τους, όπως οι νευρώνες του ανθρώπινου εγκεφάλου, δέχονται ερεθίσματα από την είσοδο και σύμφωνα με αυτά τα ερεθίσματα μαθαίνουν και αντιδρούν. Να τονιστεί επίσης ότι τα νευρωνικά δίκτυα χρησιμοποιούνται για προβλήματα ταξινόμησης και παλινδρόμησης.

Όπως και στον ανθρώπινο εγκέφαλο έτσι και στα νευρωνικά δίκτυα υπάρχει ένα διασυνδεδεμένο δίκτυο από υπολογιστικούς κόμβους (νευρώνες, νευρώνια). Ένα παράδειγμα κόμβου φαίνεται στο παρακάτω Σχήμα 3.11.



Σχήμα 3.11: Δομή κόμβου νευρωνικού δικτύου

Υπάρχουν τρεις τύποι νευρώνων: οι νευρώνες εισόδου (input layer), οι νευρώνες εξόδου (output layer) και οι υπολογιστικοί νευρώνες ή κρυμμένοι νευρώνες (hidden layer).

- **Νευρώνες εισόδου (input layer)**

Σκοπός των νευρώνων εισόδου είναι να εισάγουν στο δίκτυο τα αρχικά δεδομένα ή τα αποτελέσματα ενός προηγούμενου δικτύου. Συνήθως ο αριθμός των νευρώνων εισόδου είναι ίσος με τα χαρακτηριστικά (attributes) του συνόλου δεδομένων που θέλουμε να εισάγουμε. Σε αυτόν τον νευρώνα δεν εκτελείται καμία άλλη διαδικασία παρά μόνο η προώθηση των δεδομένων εισόδου στους κρυμμένους νευρώνες (hidden layer).

- **Κρυμμένοι νευρώνες (hidden layer)**

Στους κρυμμένους νευρώνες εισέρχονται τα δεδομένα από τους νευρώνες εισόδου και εφαρμόζονται κάποιοι μετασχηματισμοί πάνω σε αυτά. Αρχικά πολλαπλασιάζεται το καθένα με ένα συγκεκριμένο βάρος και στην συνέχεια όλα μαζί αθροίζονται και προκύπτει ένας αριθμός. Ο αριθμός των κρυμμένων νευρώνων ποικίλει και υπάρχουν αρκετές θεωρίες για το πως μπορεί να επιλεγεί ο αριθμός των νευρώνων, όμως στο δικό μας μοντέλο έχει προκύψει από δοκιμές διαφορετικών αριθμών κρυμμένων νευρώνων.

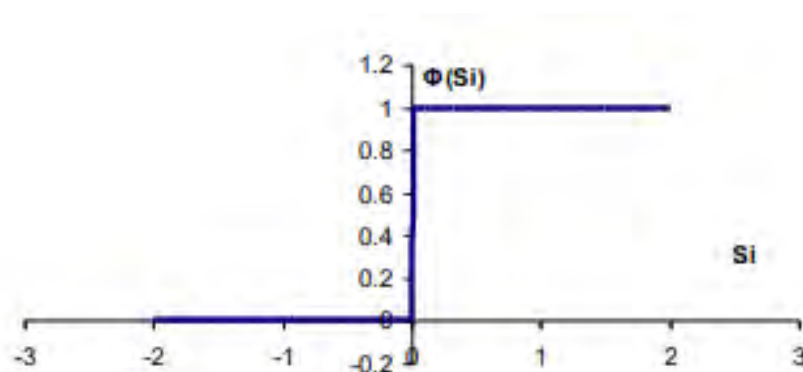
- **Νευρώνες εξόδου (output layer)**

Στην συνέχεια τα δεδομένα από τους κρυμμένους νευρώνες περνάνε πρώτα από την συνάρτηση ενεργοποίησης (activation function) η οποία θα ενεργοποιηθεί όταν το άθροισμα των δεδομένων ξεπεράσει το όριο που έχουμε θέσει. Και στην συνέχεια θα προκύψει η έξοδος του νευρωνικού δικτύου. Η έξοδος είναι ένα αριθμητικό αποτέλεσμα και ανάλογα το πρόβλημα που θέλουμε να λύσουμε αυτή θα είναι είτε κάποιος συνεχής αριθμός για μοντέλο παλινδρόμησης ή κάποιος αριθμός που θα ανήκει σε κάποια κατηγορία για μοντέλο κατηγοριοποίησης.

Όπως αναφέραμε παραπάνω τα δεδομένα πριν βγουν από την έξοδο του νευρωνικού δικτύου περνάνε πρώτα από συνάρτηση ενεργοποίησης (activation function). Υπάρχουν διάφορες περιπτώσεις συναρτήσεων ενεργοποίησης, οι πιο γνωστές είναι [31]:

- Βηματική συνάρτηση (step function) ή συνάρτηση κατωφλίου (threshold function). Σε αυτήν την συνάρτηση όπως φαίνεται και στην παρακάτω αναπαράσταση (Σχήμα 3.12) της συνάρτησης, αν η έξοδος της συνάρτησης είναι μεγαλύτερη του ορίου (bias) τότε η έξοδος της συνάρτησης ενεργοποίησης είναι 1, διαφορετικά είναι 0.

$$\Phi(S) = \begin{cases} 1, S > 0 \\ 0, S \leq 0 \end{cases}$$

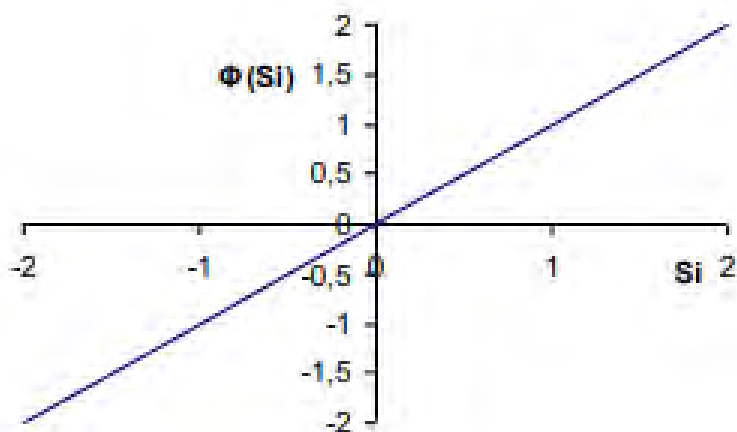


Σχήμα 3.12: Γραφική απεικόνιση βηματικής συνάρτησης

- Γραμμική συνάρτηση ενεργοποίησης (linear function). Η γραμμική συνάρτηση χρησιμοποιείται κυρίως για επίλυση προβλημάτων γραμμικής παλινδρόμησης και όπως και όλες μπαίνει πριν τον νευρώνα εξόδου. Μπορεί να έχει την παρακάτω μορφή:

$$\Phi(S_i) = \lambda S_i$$

ή και οποιαδήποτε άλλη γραμμική συνάρτηση.



Σχήμα 3.13: Γραφική απεικόνιση γραμμικής συνάρτησης

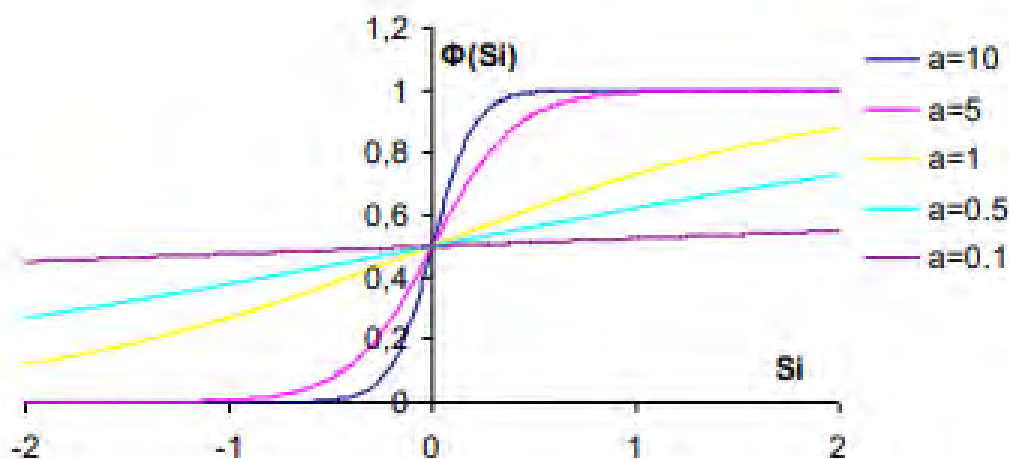
- Σιγμοειδής συνάρτηση (sigmoid ή logistics function). Πρόκειται για μία μη γραμμική συνάρτηση η οποία χρησιμοποιείται όταν έχουμε να επιλύσουμε μη γραμμικά προβλήματα. Είναι από τις πιο γνωστές χρησιμοποιούμενες σε νευρωνικά δίκτυα πολλών επιπέδων. Η μορφή της είναι:

$$\Phi(S_i) = \frac{1}{1 + e^{(-aS_i)}}$$

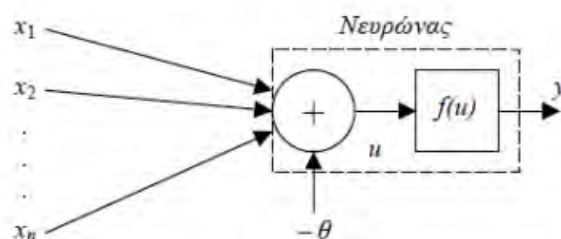
### 3.10.1 Πολυεπίπεδο δίκτυο Perceptron – MLP

Ένα από τους πιο γνωστούς ίσως και ο πιο γνωστός τύπος νευρωνικών δικτύων είναι τα πολυεπίπεδα νευρωνικά δίκτυα ή αλλιώς με συντομογραφία MLP [4]. Αρχικά το μοντέλο του αισθητήρα ή Perceptron αποτελεί την μία απλή μορφή νευρωνικού δικτύου που μπορεί να μελετηθεί καθώς αποτελείται από έναν μόνο νευρώνα, το οποίο έχει συγκεκριμένο αριθμό εισόδων και παράγει μία μόνο έξοδο όπως φαίνεται στο σχήμα παρακάτω.

Οπότε ένα πολυεπίπεδο δίκτυο αποτελείται από δύο και περισσότερους αισθητήρες (Perceptrons). Στο Σχήμα 3.15 παρακάτω βλέπουμε ένα MLP νευρωνικό δίκτυο με ένα hidden layer. Αποτελείται από το επίπεδο εισόδου, το επίπεδο των κρυμμένων νευρώνων, που μπορεί να είναι ένα ή και περισσότερα και το επίπεδο εξόδου. Χαρακτηριστικό αυτών των νευρωνικών δικτύων είναι ότι κάθε κόμβος τροφοδοτείται από την έξοδο του προηγούμενου επιπέδου, αλλά και ότι είναι



Σχήμα 3.14: Γραφική απεικόνιση σιγμοειδούς συναρτήσεως



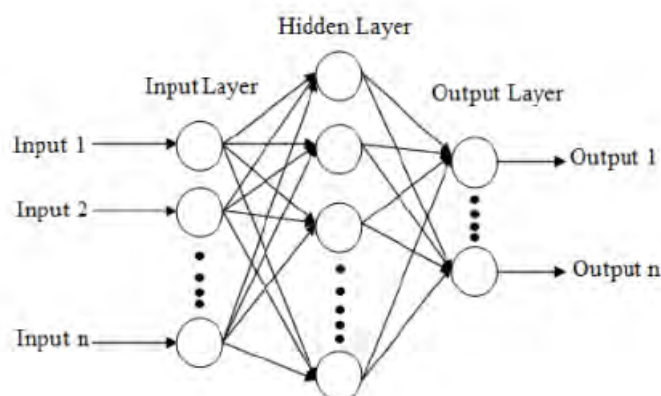
Σχήμα 3.15: Δομή νευρώνα νευρωνικού δικτύου

ένα πλήρως διασυνδεδεμένο δέντρο. Δηλαδή κάθε νευρώνας σε ένα επίπεδο είναι διασυνδεδεμένος με όλους τους νευρώνες του προηγούμενου επιπέδου.

Άλλα χαρακτηριστικά που έχουν τα MLP νευρωνικά δίκτυα είναι, η χρησιμοποίηση σε κάθε νευρώνα του κρυφού επιπέδου νευρώνων, σαν συνάρτηση ενεργοποίησης μίας μη γραμμική. Σύννηθες είναι να χρησιμοποιούμε την Σιγμοειδής συνάρτηση (sigmoid ή logistics function). Ο λόγος που την χρησιμοποιούμε είναι το ότι είναι συνεχής συνάρτηση άρα και παραγωγίσιμη σε αντίθεση με τη βηματική συνάρτηση που έχει δυαδικές τιμές. Και ο λόγος που θέλουμε να είναι παραγωγίσιμη η συνάρτηση είναι ότι όλες οι μέθοδοι βελτιστοποίησης που θα χρησιμοποιήσουμε εμπλέκουν μέσα τους παραγώγους.

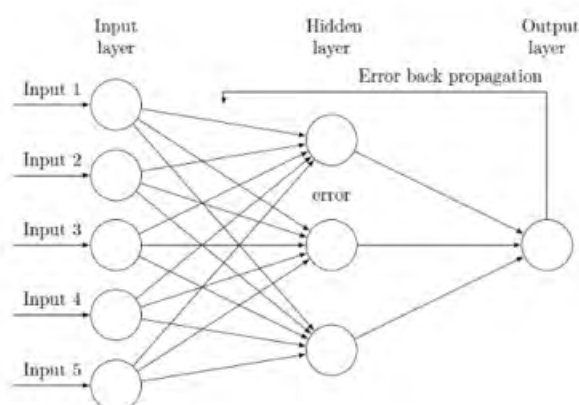
Η εκπαίδευση αυτών των MLP γίνεται με τη διαδικασία ρύθμισης των βαρών μέχρι να ικανοποιηθεί κάποιο από τα κριτήρια βελτιστοποίησης που έχουμε. Στόχος μας είναι το μοντέλο να μάθει από τα δεδομένα που το έχουμε δώσει για εκπαίδευση, έτσι ώστε να μπορεί να προβλέψει τις τιμές των επόμενων δεδομένων που θα του δώσουμε. Ο αλγόριθμος που χρησιμοποιούμε σε τέτοιου είδους νευρωνικά δίκτυα είναι ο αλγόριθμος ανάστροφης διάδοσης σφάλματος (Error Back Propagation) [10]. Βασικό χαρακτηριστικό αυτού του αλγόριθμου είναι ότι εκπαιδεύεται με





Σχήμα 3.16: Νευρωνικό δίκτυο με απλή τροφοδότηση

επίβλεψη δηλαδή υπάρχει ένας στόχος που θέλουμε να φτάσει ο αλγόριθμος.



Σχήμα 3.17: Νευρωνικό δίκτυο με χρήση ErrorBack Propagation

Η διαδικασία εκπαίδευσης ξεκινάει με την τυχαία αρχικοποίηση των βαρών κάθε νευρώνα. Στην συνέχεια η διαδικασία προχωράει με δύο περάσματα του δικτύου ένα προς τα εμπρός και ένα προς τα πίσω. Στην διαδικασία προς τα εμπρός κάθε στοιχείο της εισόδου περνάει μέσα από τα κρυμμένα επίπεδα και στην συνέχεια καταλήγει στην έξοδο του δικτύου δίνοντας μία τιμή στην εξαρτημένη μεταβλητή. Σε αυτό το πέραςμα τα βάρη παραμένουν σταθερά. Αφού περάσουν όλα τα στοιχεία της εισόδου και δώσουν ένα αποτέλεσμα περνώντας από το επίπεδο εξόδου τότε μπορούμε να υπολογίσουμε το σφάλμα του μοντέλου αφαιρώντας τις τιμές που έχουμε βρει από τις πραγματικές τιμές χρησιμοποιώντας μια συνάρτηση κόστους. Έπειτα για να υπολογίσουμε πόσο επηρεάζει κάθε βάρος στο σφάλμα θα πρέπει να υπολογίσουμε τις μερικές παραγώγους του τελικού κόμβου και στη συνέχεια να εφαρμόσουμε τον αλγόριθμο της ελάττωσης της παραγώγου ( gradient descent). Η μορφή της συνάρτησης που χρησιμοποιείται για την ανανέωση των βαρών είναι η εξής παρακάτω:

$$w = w - \eta \nabla_w J(w)$$

Στην συνέχεια αυτήν η πληροφορία μεταφέρεται στο αμέσως προηγούμενο επίπεδο όπου και εκεί θα πρέπει να υπολογίσουμε τις μερικές παραγώγους όμως θα χρησιμοποιήσουμε και την προηγούμενη πληροφορία που γνωρίζουμε για τα βάρη και έτσι θα καθορίσουμε την τιμή τους. Και συνεχίζεται αυτή η διαδικασία μέχρι να φτάσουμε στο επίπεδο εισόδου το οποίο προφανώς δεν θα αλλάξει.

Η παραπάνω διαδικασία θα σταματήσει μόλις ολοκληρωθεί ο αριθμός των επαναλήψεων που έχει ορίσει ο χρήστης, ή όταν το σφάλμα του δικτύου είναι μικρότερο από κάποιο όριο που έχει δώσει πάλι ο χρήστης.

## Κεφάλαιο 4

# Μεθοδολογία και Παρουσίαση Μοντέλων

### 4.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα παρουσιάσουμε την μεθοδολογία που εφαρμόσαμε ώστε να επιλύσουμε το πρόβλημα της πρόβλεψης της παραγωγής ηλεκτρικής ενέργειας μέσω φωτοβολταϊκών πάνελ για το επόμενο 24ωρο. Επίσης σ' αυτό το κεφάλαιο θα συμπεριλάβουμε τα μοντέλα που κατασκευάστηκαν καθώς επίσης και τα λογισμικά που χρησιμοποιήθηκαν για την μελέτη και την επίλυση του προβλήματος.

### 4.2 Μεθοδολογία

Στο πρώτο στάδιο της κατασκευής του μοντέλου έγινε άντληση των δεδομένων από τους δύο ιστοτόπους και ένωση των δύο συνόλων δεδομένων σε ένα για να είναι πιο εύκολη η μελέτη τους. Στη συνέχεια έγινε μελέτη της μεταξύ τους συσχέτισης με τη χρήση της συνάρτησης `corr()`. Αυτή η συνάρτηση αποτελεί μέρος μια βιβλιοθήκης συναρτήσεων που βρίσκεται στην `python` και ονομάζεται `Pandas`. Από αυτήν την διαδικασία από τα αρχικά 24 δεδομένα που είχαμε επιλέξαμε τα 6 (Temperature, Relative Humidity, Total Cloud Cover, Sunshine Duration, Wind Speed, Short-wave Radiation) με στόχο να προβλέψουμε την τιμή της ηλιακής ακτινοβολίας. Αυτή η επιλογή έγινε σύμφωνα με το ποιες μεταβλητές του συνόλου δεδομένων είχαν μεγαλύτερη συσχέτιση σε σχέση με τη ηλιακή ακτινοβολία αλλά και το ποιες μεταβλητές του καιρού θα την επηρέαζαν αν το σκεφτόμασταν λογικά.

Σε αυτό το σημείο κρίθηκε σκόπιμο να γίνει χρήση και μιας επιπλέον μεταβλητής την οποία δεν περιείχε το σύνολο δεδομένων. Αυτή η μεταβλητή ονομάστηκε `Calc radiation` και προέρχεται από τη βιβλιοθήκη της `Python` την `Pysolar`[18]. Η βιβλιοθήκη αυτή μας δίνει στοιχεία για την τιμή της ηλιακής ακτινοβολίας που δέχεται μία περιοχή. Η τιμή που μας δίνει προέρχεται μέσω κάποιων μαθηματικών υπολογισμών, δίνοντας μας την ηλιακή ακτινοβολία που δέχεται μια περιοχή χωρίς όμως να λαμβάνει υπόψιν καιρικές συνθήκες όπως ηλιοφάνεια, θερμοκρασία και τους παραπάνω παράγοντες του συνόλου δεδομένων.

Για την καλύτερη πρόβλεψη της ηλιακής ακτινοβολίας που φτάνει στα πάνελ του ηλιακού πάρκου, τα καιρικά δεδομένα που αναφέραμε παραπάνω χωρίστηκαν σε 3 κατηγορίες ανάλογα με το ποσοστό νεφοκάλυψης, τις ώρες που τα ηλιακά πάνελ παράγουν ενέργεια. Έτσι για την πρόβλεψη ανάλογα με το ποσοστό νεφοκάλυψης που μας δίνει ο μετεωρολογικός σταθμός χρησιμοποιούμε και το αντίστοιχο μοντέλο.

Στη συνέχεια επιλέχθηκαν και αναπτύχθηκαν 4 μοντέλα μηχανικής μάθησης. Πρώτα χρησιμοποιήθηκαν οι μηχανές διανυσματικής υποστήριξης, στη συνέχεια η χρονολογικές σειρές, ο αλγόριθμος boosting για παλινδρόμηση και τέλος τα τεχνητά νευρωνικά δίκτυα και συγκεκριμένα το Πολυεπίπεδο δίκτυο Perceptron (MLP).

Τέλος για τον υπολογισμό της παραγόμενης ηλεκτρικής ενέργειας του ηλιακού πάρκου κατασκευάστηκε μία συνάρτηση η οποία χρησιμοποιεί την τιμή της ηλιακής ακτινοβολίας που προβλέπεται από τις παραπάνω τεχνικές και λαμβάνοντας υπόψη κάποιους παράγοντες (τύπος πάνελ, ηλικία πάνελ, τύπος inverter κτλ), που θα αναφερθούν παρακάτω εκτενέστερα, βγάζουμε το τελικό αποτέλεσμα.

Για την πρόβλεψη της παραγόμενης ηλεκτρικής ενέργειας την επόμενη ημέρα από την εφαρμογή μας χρησιμοποιήθηκε ένα υβριδικό μοντέλο. Καθώς για κάθε μία από τις κατηγορίες των καιρικών συνθηκών, όπως τις χωρίσαμε πιο πάνω, χρησιμοποιείται διαφορετικό μοντέλο. Έτσι ύστερα από δοκιμές βρέθηκε ότι, για την κατηγορία της ηλιοφάνειας καλύτερο μοντέλο ήταν αυτό με τη χρήση MLP νευρωνικών δικτύων, για τις μέρες με μερική νεφοκάλυψη το μοντέλο με τη χρήση του GBRT και για τις μέρες ολικής νεφοκάλυψης το μοντέλο του SVR.

## 4.3 Λογισμικά

### 4.3.1 Python

Η γλώσσα που χρησιμοποιήθηκε για την μελέτη και κατασκευή των μοντέλων ήταν η Python. Η Python πρόκειται για μία ισχυρή γλώσσα προγραμματισμού η οποία έχει αρκετές χρήσεις και χρησιμοποιείται αρκετά συχνά στον τομέα της ανάλυσης δεδομένων. Γι' αυτόν τον λόγο έχουν δημιουργηθεί αρκετές πλατφόρμες για την φιλοξενία αυτής της γλώσσας, που βοηθούν στην ανάλυση δεδομένων. Μερικές από τις βασικές βιβλιοθήκες που χρησιμοποιήθηκαν είναι η Pandas η οποία περιέχει ένα ευρύ φάσμα συναρτήσεων για μελέτη συνόλων δεδομένων, η NumPy η οποία περιέχει από συναρτήσεις για μαθηματικές πράξεις μέχρι και συναρτήσεις για διαχείριση και μελέτη πινάκων και η Matplotlib η οποία χρησιμοποιείται για αναπαράσταση γραφικών παραστάσεων.

### 4.3.2 Keras & Tensorflow

Για την ανάπτυξη των νευρωνικών δικτύων χρησιμοποιήθηκε ένα API (Application Programming Interface) (Διασύνδεση προγραμματισμού εφαρμογών) το Keras το οποίο πρόκειται για ένα πολύ ισχυρό API που χρησιμοποιείται κυρίως για την κατασκευή νευρωνικών δικτύων. Σαν API ορίζουμε ένα ενδιάμεσο λογισμικό που επιτρέπει την επικοινωνία μεταξύ δύο εφαρμογών. Για τη λειτουργία όμως του Keras προϋποθέτει την ύπαρξη ενός εκ των TensorFlow ή Theano. Εμείς

χρησιμοποιήσαμε τον TensorFlow όπου πρόκειται για μία βιβλιοθήκη ανοικτού κώδικα η οποία χρησιμοποιείται για έρευνα και παραγωγή μοντέλων.

### 4.3.3 Anaconda

Η πλατφόρμα που χρησιμοποιήσαμε για την μελέτη και ανάπτυξη των μοντέλων ονομάζεται Anaconda και πρόκειται για μία δωρεάν πλατφόρμα όπου περιέχει διάφορα προγράμματα για την μελέτη δεδομένων. Οι γλώσσες που φιλοξενεί είναι η Python και η R. Η αρχική εφαρμογή που χρησιμοποιήσαμε για την μελέτη των δεδομένων είναι το Jupyter Notebook. Επίσης η αρχική προσέγγιση των μοντέλων πρόβλεψης έγινε εκεί και στη συνέχεια μεταφέρθηκαν σε μία άλλη εφαρμογή το Pycharm.

### 4.3.4 Pycharm

Πρόκειται για μια πλατφόρμα η οποία δημιουργήθηκε από την Τσέχικη εταιρία JetBrains και έχει αναπτυχθεί ειδικά για να προγραμματίζεις σε Python. Προσφέρει αρκετές ευκολίες στον προγραμματισμό με Python όπως αυτόματη συμπλήρωση και παρατηρήσεις διόρθωσης. Αυτή η πλατφόρμα χρησιμοποιήθηκε για την κατασκευή της εφαρμογής.

### 4.3.5 Tkinter

Η Tkinter πρόκειται για μία βιβλιοθήκη που χρησιμοποιείται για την κατασκευή παραθυρικών εφαρμογών στην Python. Ουσιαστικά με την Tkinter δημιουργούμε παράθυρα που περιέχουν γραφικά στοιχεία widgets όπως labels, buttons, entry texts, menu, scroll bars. Τα widgets δέχονται γεγονότα από ενέργειες του χρήστη ανταποκρίνονται με κλήση συναρτήσεων.

## 4.4 Διαχείριση Δεδομένων

Τα δεδομένα όπως αναφέραμε και πιο πάνω αποτελούνται από δύο ξεχωριστά σύνολα δεδομένων, το πρώτο που περιέχει καιρικά δεδομένα (Meteoblue) και το δεύτερο περιέχει τα δεδομένα παραγωγής (SunnyPortal). Τα δύο σύνολα δεδομένων ενώθηκαν μεταξύ τους για την καλύτερη διαχείρισή τους. Προτού γίνει αυτή η ένωση τα δύο σύνολα καθαρίστηκαν, καθώς έπρεπε να γίνουν κάποιες αλλαγές τόσο στις επικεφαλίδες αλλά και στο περιεχόμενό τους (κενές τιμές, αλλαγή κάποιων χαρακτήρων...).

Στην συνέχεια για την επίτευξη του καλύτερου αποτελέσματος κρίναμε σκόπιμο να χρησιμοποιήσουμε και μια ακόμα μεταβλητή, που την ονομάσαμε Calc Rad, η οποία προέρχεται από μία βιβλιοθήκη της Python την Pysolar. Αυτή η βιβλιοθήκη μέσω της συνάρτησης `get_radiation`, δίνοντας της σαν παραμέτρους το γεωγραφικό πλάτος και μήκος της τοποθεσίας που μας ενδιαφέρει, μας επιστρέφει την ποσότητα της ηλιακής ακτινοβολίας που φτάνει σε εκείνο το σημείο της επιφάνειας της γης που έχουμε δηλώσει παραπάνω. Σημαντικό είναι να επισημάνουμε ότι αυτό το ποσό της ηλιακής ακτινοβολίας που υπολογίζει η συνάρτηση γίνεται χωρίς να λαμβάνονται υπόψιν καιρικές συνθήκες.

Αφού έχουμε ολοκληρώσει το τελικό σύνολο δεδομένων, θα πρέπει να γίνει και επιλογή ορισμένων μεταβλητών καθώς δεν είναι όλα χρήσιμα για την κατασκευή του μοντέλου. Αυτή η επιλογή έγινε με τη χρήση της συνάρτησης `corr` η οποία μας δίνει τις συσχετίσεις μεταξύ των μεταβλητών και συγκεκριμένα χρησιμοποιήθηκε ο συντελεστής Pearson που μας δίνει την συσχέτιση μεταξύ δύο ανεξάρτητων μεταβλητών. Έτσι επιλέχθηκαν 6 μεταβλητές οι οποίες είχαν την μεγαλύτερη συσχέτιση με την μεταβλητή Shortwave Radiation που θέλουμε να προβλέψουμε. Επίσης η επιλογή τους έγινε και με κριτήριο, το ποια στοιχεία του καιρού, θα μπορούσαν να επηρεάσουν την ηλιακή ακτινοβολία που θα έφτανε στο φωτοβολταϊκό πάνελ. Στο Σχήμα 4.1 βλέπουμε τις τιμές της συνάρτησης `corr` για την μεταβλητή Shortwave Radiation.

	Shortwave_Radiation
Year	0.142864
Month	-0.013725
Day	-0.014773
Hour	0.269965
Minute	NaN
Temperature	0.588828
Relative_Humidity	-0.644037
Mean_Sea_Level_Pressure	-0.101482
Total_Precipitation	-0.071398
Total_Precipitation.1	-0.043526
Snowfall_Amount_(high_resolution)	-0.017040
Snowfall_Amount_(low_resolution)	-0.035427
Total_Cloud_Cover	-0.201885
High_Cloud_Cover	-0.106783
Medium_Cloud_Cover	-0.114710
Low_Cloud_Cover	-0.205018
Sunshine_Duration	0.729524
Shortwave_Radiation	1.000000
Wind_Speed_10m	-0.091405
Wind_Direction_10m	-0.202390
Wind_Speed_80m	-0.294376
Wind_Direction_80m	-0.212288
Wind_Speed_900mb	-0.210035
Wind_Direction_900_mb	-0.026958
Wind_Gust	-0.108582
Kwh	0.794531

Σχήμα 4.1: Τιμές συσχέτισης του Shortwave Radiation με τις υπόλοιπες μεταβλητές

Αφού έχουμε ολοκληρώσει τη διαδικασία της δημιουργίας και επεξεργασίας του συνόλου δεδομένων στη συνέχεια χωρίζουμε αυτό το σύνολο δεδομένων σε 3 υποσύνολα με κριτήριο το ποσοστό νεφοκάλυψης της κάθε ημέρας. Έτσι προκύπτουν 3 κατηγορίες, οι ημέρες ηλιοφάνειας, οι ημέρες μερικής νεφοκάλυψης και οι ημέρες πλήρους νεφοκάλυψης. Στην τελευταία κατηγορία προφανώς ανήκουν και οι ημέρες με βροχή ή χιονόπτωση. Επίσης στην κατηγορία μερικής νεφοκάλυψης για την επίτευξη καλύτερου αποτελέσματος έχουν συμπεριληφθεί και οι μέρες πλήρους νεφοκάλυψης, δηλαδή αυτές οι μέρες υπάρχουν και στα δύο σύνολα δεδομένων. Στο Σχήμα 4.2 φαίνονται ενδεικτικά ο αριθμός των δεδομένων όπως αυτά έχουν καταταξιολογηθεί.

Ύστερα επειδή τα δεδομένα μας είναι σε διαφορετικές τάξεις μεγέθους για να λειτουργήσει σωστά η διαδικασία θα πρέπει να τα μετατρέψουμε και να τα κάνουμε όλα σε ίδια κλίμακα. Έτσι για τον λόγο αυτόν, χρησιμοποιούμε την συνάρτηση `StandardScaler`, η οποία πρόκειται για μία συνάρτηση της `Sklearn` που στόχο έχει τη μετατροπή των δεδομένων σε ίδια κλίμακα για την καλύτερη εκτέλεση των αλγορίθμων.

Τελειώνοντας την διαδικασία της επεξεργασίας των δεδομένων, χωρίζουμε το σύνολο δεδομένων σε δύο μέρη: Το σύνολο για εκπαίδευση του αλγορίθμου και αυτό για τον έλεγχο του. Σαν



Σχήμα 4.2: Διαχωρισμός Δεδομένων σε 3 Κατηγορίες

ποσοστό για τα δύο υπο μέρη διαλέξαμε το 70% για εκπαίδευση και το υπόλοιπο 30% για έλεγχο.

## 4.5 Υπολογισμός Ημερήσιας Ενεργειακής Απόδοσης Ηλιακού Πάνελ

Ο ακριβής υπολογισμός της ενεργειακής απόδοσης ενός ηλιακού πάνελ είναι συχνά μία δύσκολη διαδικασία. Για τον λόγο αυτόν έχουν αναπτυχθεί διάφορες εφαρμογές που μέσω περίπλοκων μαθηματικών τύπων μας δίνουν το αποτέλεσμα. Εμείς για την παρούσα διπλωματική θα χρησιμοποιήσουμε έναν πιο απλό τύπο που θα αναλύσουμε στη συνέχεια.

Για την παραγωγή της ηλεκτρικής ενέργειας μέσω των φωτοβολταϊκών στοιχείων υπεύθυνη είναι η ηλιακή ακτινοβολία (Solar Radiation) η οποία φτάνει στην επιφάνεια του ηλιακού πάνελ. Σημαντικός παράγοντας για την ποσότητα μετατροπής της σε ηλεκτρική ενέργεια παίζει ο προσανατολισμός του πάνελ. Στη δικιά μας περίπτωση τα ηλιακά πάνελ του πάρκου μέσω ενός μοτέρ είναι προγραμματισμένα να περιστρέφονται ακολουθώντας εκείνη την κατεύθυνση όπου θα έχουν την μέγιστη παραγωγή ηλεκτρικής ενέργειας. Ένας άλλος σημαντικός παράγοντας που επηρεάζει την παραγωγή είναι η θερμοκρασία του περιβάλλοντος. Όταν η θερμοκρασία του ηλιακού πάνελ ξεπερνάει τους 25 βαθμούς Κελσίου τότε η απόδοση τους μειώνεται. Έτσι στο μοντέλο που αναπτύξαμε για τον υπολογισμό της ηλεκτρικής ενέργειας έχει γίνει πρόβλεψη να υπάρχει και ένας συντελεστής ο οποίος αν η θερμοκρασία του περιβάλλοντος ξεπερνάει τους 25 βαθμούς να επηρεάζει την παραγωγή, μειώνοντάς την. Τέλος σημαντικό ρόλο στην τελική παραγόμενη ποσότητα παίζουν και οι απώλειες του συστήματος όπως απώλειες μετατροπής της ενέργειας, απώλειες λόγω των inverter, η παλαιότητα των φωτοβολταϊκών στοιχείων κ.α.

Έτσι ο μαθηματικός τύπος που θα χρησιμοποιήσουμε για τον υπολογισμό της παραγόμενης ηλεκτρικής ενέργειας είναι ο παρακάτω.

$$E = A * r * R * PR$$

Όπου E είναι η παραγόμενη ηλεκτρική ενέργεια εκείνη την στιγμή (kWh), A είναι το μέγεθος του ηλιακού πάνελ ( $m^2$ ), r είναι ο συντελεστής απόδοσης του συγκεκριμένου ηλιακού πάνελ (%), R είναι η ποσότητα της ηλιακής ακτινοβολίας που φτάνει στο ηλιακό πάνελ ( $W/m^2$ ), και τέλος το

PR το οποίο είναι ένας συντελεστής που μετράει τις απώλειες του συστήματος.



Σχήμα 4.3: Απώλειες του φωτοβλοταϊκού συστήματος [20]

## 4.6 Μοντέλα Πρόβλεψης

### 4.6.1 Χρονολογικές Σειρές (Time Series)

Το πρώτο μοντέλο που μελετήσαμε είναι αυτό με τη χρήση χρονολογικών σειρών. Συγκεκριμένα ύστερα από μελέτες και δοκιμές των αλγορίθμων που εμπεριέχουν οι χρονολογικές σειρές καταλήξαμε στην χρήση του Μεικτού Αυτοπαλίνδρομου Ολοκληρωμένου Υποδείγματος Κινητού Μέσου ή ARIMA(p,q,d). Χαρακτηριστική διαφορά του ARIMA σε σχέση με τον ARMA είναι η δυνατότητα μετατροπής μιας χρονολογικής σειράς από μη στάσιμη σε στάσιμη. Και αυτό το πετυχαίνει με τη χρήση της παραμέτρου d, με την οποία ορίζουμε τον αριθμό των διαφορών που χρειαζόμαστε έτσι ώστε να γίνει στάσιμη.

Το πρώτο βήμα είναι να δούμε αν η χρονολογική σειρά μας είναι στάσιμη. Αυτός ο έλεγχος επιτυγχάνεται με την χρήση μίας στατιστικής μεθόδου που ονομάζεται Dickey–Fuller test. Η μέθοδος αυτήν στηρίζεται στην χρήση του στατιστικού ελέγχου υποθέσεων, σύμφωνα με τον οποίον θέτουμε ως μηδενική υπόθεση (H0) εκείνη της οποίας αμφισβητούμε το αποτέλεσμα της και ελέγχουμε μέσω ενός τυχαίου δείγματος που παίρνουμε από τον πληθυσμό αν υπάρχουν αρκετές αποδείξεις για την απόρριψή της, έναντι της εναλλακτικής(H1).

```

Results of Dickey-Fuller Test:
Test Statistic          -5.955952e+00
p-value                 2.093123e-07
#Lags Used              2.800000e+01
Number of Observations Used 8.731000e+03
Critical Value (1%)     -3.431099e+00
Critical Value (5%)     -2.861871e+00
Critical Value (10%)    -2.566946e+00
dtype: float64

```

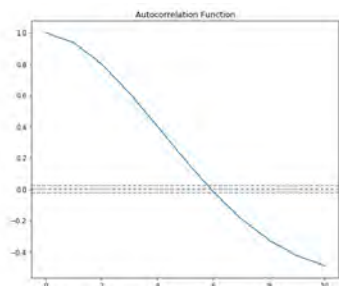
Σχήμα 4.4: Αποτελέσματα Dickey-Fuller test

Έτσι στο Σχήμα 4.4 βλέπουμε ότι η τιμή του Test Statistic είναι μικρότερη από την τιμή της μεταβλητής Critical Value (1%) που σημαίνει ότι χρονολογική σειρά είναι στάσιμη με ποσοστό

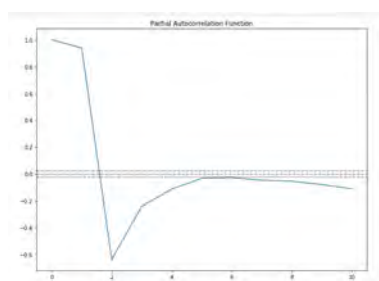


εμπιστοσύνης 99%.

Επόμενο κομμάτι του μοντέλου είναι ο καθορισμός των μεταβλητών  $p$  και  $q$  του μοντέλου. Αυτό θα γίνει με την χρήση δύο γραφικών παραστάσεων. Της συνάρτησης μερικής αυτοσυσχέτισης (PACF) και την γραφική παράσταση της συνάρτησης αυτοσυσχέτισης (ACF).



Σχήμα 4.5: ACF



Σχήμα 4.6: PACF

Από τα σχήματα Σχήμα 4.5 και Σχήμα 4.6 ελέγχουμε το σημείο που διαπερνά για πρώτη φορά το άνω όριο εμπιστοσύνης η γραφική μας παράσταση και από αυτό το σημείο πάνω στον άξονα  $x$  παίρνουμε την τιμή του  $p$  και  $q$ , από τις αντίστοιχες γραφικές παραστάσεις.

Έτσι οι παράμετροι που επιλέξαμε είναι  $p = 2$ ,  $q = 6$  και  $d = 0$  καθώς η χρονολογική μας σειρά είναι στάσιμη.

Ύστερα από εκτέλεση του μοντέλου που κατασκευάσαμε είδαμε ότι, όταν υπάρχουν μέρες με μεγάλες διακυμάνσεις στις καιρικές συνθήκες το μοντέλο δεν δίνει καλό αποτέλεσμα, σε αντίθεση με τις μέρες που επικρατεί αρκετή ηλιοφάνεια. Γενικά επειδή χρησιμοποιεί τις προηγούμενες μέρες το μοντέλο για κάνει προβλέψεις όταν επικρατούν σχετικά σταθερές καιρικές συνθήκες (μέρες με συνεχόμενη ηλιοφάνεια ή συννεφιά), χωρίς πολλές εναλλαγές τότε έχουμε ένα καλό αποτέλεσμα.

#### 4.6.2 Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines)

Το επόμενο μοντέλο κατασκευάστηκε με την χρήση των μηχανών διανυσματικής υποστήριξης. Σ' αυτό το μοντέλο η πρώτη παράμετρος ρύθμισης είναι το κόστος  $C$ , που καθορίζει πόσο αυστηροί θέλουμε να είμαστε με τα σφάλματα. Για το δικό μας μοντέλο επιλέχθηκε η τιμή 10. Επόμενη παράμετρος είναι η επιλογή πυρήνα (kernel) που για τον δικό μας μοντέλο χρησιμοποιήθηκε η Ακτινωτή (Radial Basis Function (RBF)). Και τέλος το Epsilon ( $\epsilon$ ) το οποίο καθορίζει τα όρια που θα έχει το Hyper-plane και σαν τιμή επιλέχθηκε το 0.1. Για την καλύτερο δυνατό καθορισμό των παραπάνω παραμέτρων κατασκευάστηκε και χρησιμοποιήθηκε μία συνάρτηση η οποία σαν οδηγό της το τετραγωνικό σφάλμα επέλεξε τον καλύτερο δυνατό συνδυασμό των παρακάτω διαθέσιμων παραμέτρων. Στόχος της παραπάνω διαδικασίας ήταν η καλύτερη πρόβλεψη του μοντέλου αλλά και η αποφυγή της υπερπροσαρμογής (overfitting).

Στην συνέχεια όπως αναφέραμε και στην μεθοδολογία τα δεδομένα έχουν χωριστεί σε 3 κατηγορίες ανάλογα με το ποσοστό νεφοκάλυψης και η κάθε κατηγορία χωρίστηκε σε δεδομένα για εκπαίδευση και δεδομένα για δοκιμή του αλγορίθμου. Τέλος η συνάρτηση εκτέλεσης του SVR προήλθε από την Sklearn βιβλιοθήκη της Python.

Πίνακας 4.1: Παράμετροι SVR

Παράμετροι	Τιμές
Kernel	rbf
C	10
epsilon	0.1

### 4.6.3 Gradient Boosted Regression Trees

Για την κατασκευή αυτού του μοντέλου χρησιμοποιήθηκαν εκ νέου οι συναρτήσεις της Sklearn βιβλιοθήκη της Python. Οι παράμετροι που επηρεάζουν την αποτελεσματικότητα του μοντέλου είναι ο αριθμός των δέντρων που θα κατασκευαστούν και το μέγιστο βάθος των δέντρων. Εμείς ύστερα από δοκιμές καταλήξαμε να επιλέξουμε σαν μέγιστο βάθος το 5 και σαν σύνολο δέντρων τα 50.

Πίνακας 4.2: Παράμετροι GBRT

Παράμετροι	Τιμές
Αριθμός Δέντρων	50
Μέγιστο βάθος	5

Χρησιμοποιήθηκαν και περισσότερος αριθμός δέντρων αλλά και μεγαλύτερο βάθος όμως τα αποτελέσματα των δοκιμών έδειξε το μοντέλο να έχει εκπαιδευτεί πολύ καλά με τα δεδομένα που ήταν για εκπαίδευση οδηγώντας το έτσι σε overfitting, με αποτέλεσμα στις δοκιμές που έγιναν στο σύνολο δεδομένων που ήταν για δοκιμή να μην έχουμε τα επιθυμητά αποτελέσματα.

Στο Σχήμα 4.7 μπορούμε να διακρίνουμε την μορφή ενός εκ των δέντρων του αλγορίθμου.



Σχήμα 4.7: Αναπαράσταση δέντρου GBRT (αναφέρεται στο μοντέλο που κατασκευάστηκε για τις ημέρες με ηλιοφάνεια)

### 4.6.4 Πολυεπίπεδο δίκτυο Perceptron – MLP

Για το επόμενο μοντέλο χρησιμοποιήθηκαν τα νευρωνικά δίκτυα και συγκεκριμένα το πολυεπίπεδο δίκτυο Perceptron ή MLP. Οι παράμετροι ρύθμισης αυτού του μοντέλου ήταν αρκετοί, αυτοί όμως που έπαιζαν τον μεγαλύτερο ρόλο ήταν ο αριθμός των κρυφών επιπέδων (hidden layer), οι

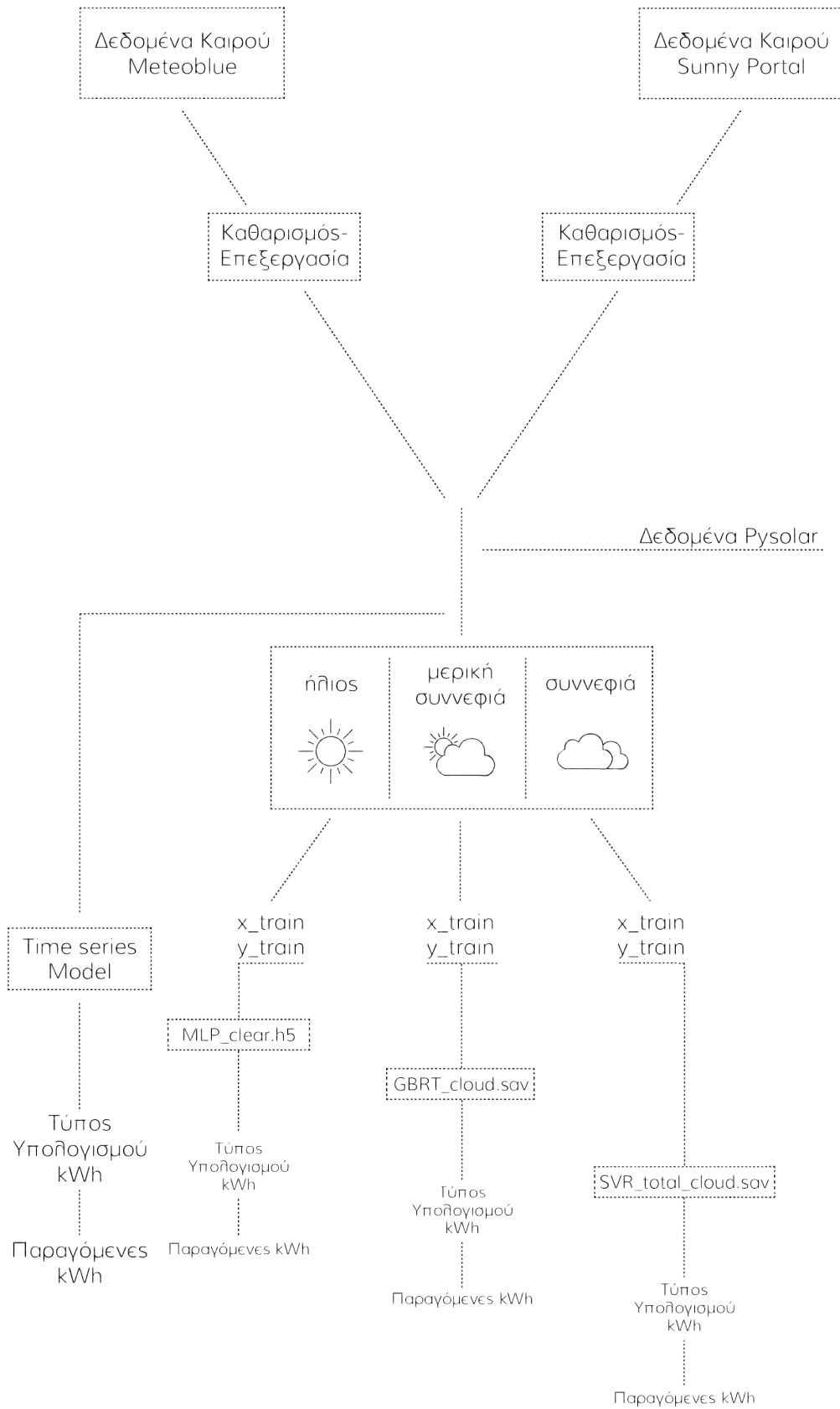
συναρτήσεις πυρήνα που χρησιμοποιήθηκαν και τέλος οι συναρτήσεις ενεργοποίησης (activation function). Για μία ακόμη φορά η επιλογή αυτών των παραμέτρων έγινε δοκιμάζοντας τις διαθέσιμες τιμές παραμέτρων που είχαμε για προβλήματα εύρεσης συνεχών τιμών.

Πίνακας 4.3: Παράμετροι MLP

Παράμετροι	Τιμές
Αριθμός Κόμβων	6
Συνάρτηση Πυρήνα	normal
Συνάρτηση ενεργοποίησης	relu
Κρυφά επίπεδα(Hidden Layer)	2

Επιπλέον δοκιμάστηκε να αυξηθεί ο αριθμός των κρυφών επιπέδων (hidden layer), από 2 σε 3 αλλά και 4, όμως παρατηρώντας την τιμή του RMSE είδαμε ότι η μείωση του δεν ήταν αξιοσημείωτη, ενώ ταυτόχρονα ο χρόνος και η πολυπλοκότητα του μοντέλου αυξανόταν. Επιπλέον δοκιμάστηκε και αλλαγή της συνάρτησης ενεργοποίησης, όμως και πάλι τα αποτελέσματα δεν ήταν τόσο καλά όσο αυτής που χρησιμοποιήσαμε τελικά (relu).

Τέλος για την επίτευξη του καλύτερου δυνατού αποτελέσματος μέσω του μοντέλου αυτού χρησιμοποιήθηκε μια συνάρτηση βελτιστοποίησης (optimizer) η adam. Η συγκεκριμένη συνάρτηση χρησιμοποιείται αρκετά συχνά σε προβλήματα πρόβλεψης συνεχών τιμών.



Σχήμα 4.8: Μεθοδολογία ανάπτυξης της διπλωματικής εργασίας

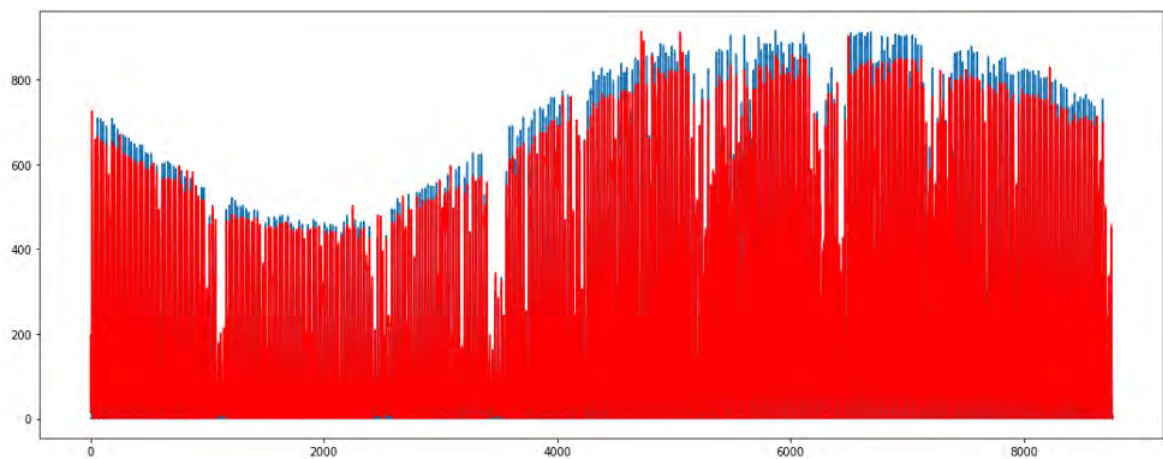
## Κεφάλαιο 5

# Αποτελέσματα

Στο κεφάλαιο αυτό θα συγκρίνουμε τα αποτελέσματα όλων των αλγορίθμων μηχανικής εκμάθησης και χρονολογικών σειρών για τα ιστορικά δεδομένα του 2017-2018. Επίσης θα γίνει και μία συγκριτική μελέτη για την απόδοση των αλγορίθμων στις 3 διαφορετικές κατηγορίες καιρικών συνθηκών όπως χωρίστηκαν πιο πάνω.

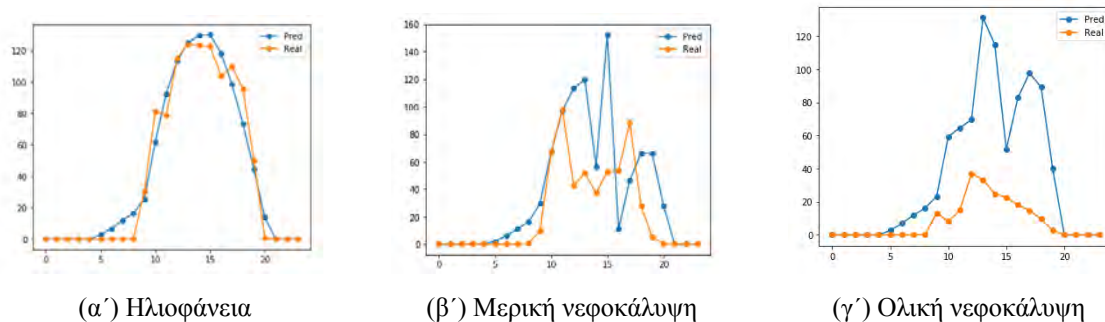
### 5.1 Χρονολογικές Σειρές (Time Series)

Όπως αναφέραμε και παραπάνω το μοντέλο που θα χρησιμοποιήσουμε για να προβλέψουμε την ηλιακή ακτινοβολία είναι το ARIMA με παραμέτρους  $(2,6,0)(p,q,d)$ . Έτσι στην συνέχεια χωρίζοντας το σύνολο δεδομένων σε 70% για εκμάθηση του αλγορίθμου και τα άλλα 30% για έλεγχο, προέκυψε το παρακάτω σχήμα (Σχήμα 5.1)..



Σχήμα 5.1: Αποτελέσματα (Solar Radiation) εκτέλεσης αλγορίθμου ARIMA για όλο το σύνολο δεδομένων

Στο Σχήμα 5.1 βλέπουμε τις 8760 εγγραφές, που είναι οι 365 ημέρες που περιλαμβάνονται στο σύνολο δεδομένων. Με μπλε χρώμα είναι οι πραγματικές τιμές της ηλιακής ακτινοβολίας (Solar radiation) κ με κόκκινο οι προβλέψεις.



Σχήμα 5.2: Παραγόμενη Ηλεκτρική Ενέργεια με χρήση Χρονολογικών Σειρών

Το RMSE που μας δίνει αυτό το μοντέλο είναι 62.6255 για το δοκιμαστικό σύνολο τιμών, το οποίο όμως δεν είναι αντιπροσωπευτικό για όλες τις περιπτώσεις (ηλιοφάνεια, μερική νεφοκάλυψη, ολική νεφοκάλυψη). Καθώς με την χρήση των χρονολογικών σειρών έχουμε προβλέψει την ηλιακή ακτινοβολία που θα φτάσει στα φωτοβολταϊκά πάνελ, εμάς αυτό που μας ενδιαφέρει είναι η ποσότητα της ηλεκτρικής ενέργειας που θα παραγάγει το πάνελ. Έτσι χρησιμοποιούμε τον τύπο που αναφέραμε στο 4ο κεφάλαιο για τον υπολογισμό της.

Στα 3 διαγράμματα (Σχήμα 5.2) πιο πάνω φαίνονται οι περιπτώσεις με ηλιοφάνεια, με μερική νεφοκάλυψη (<30%) και με ολική νεφοκάλυψη (>80%). Η αντίστοιχες τιμές των RMSE είναι 9.5132, 35.4522, 42.8314.

Έτσι, βλέπουμε ότι καθώς αυξάνεται το ποσοστό νεφοκάλυψης τόσο αυξάνεται και η ρίζα του τετραγωνικού σφάλματος. Αυτό συμβαίνει κυρίως στις μέρες όπου υπάρχουν απότομες αλλαγές στον καιρό καθώς η χρονολογική σειρά δεν μπορεί να προσαρμόσει άμεσα το μοντέλο για αυτές τις μέρες. Και αυτό συμβαίνει διότι οι παράμετροι που παίρνει το μοντέλο δεν είναι τα καιρικά δεδομένα, έτσι δεν μπορεί να γνωρίζει τις καιρικές συνθήκες που θα επικρατούν την επόμενη ημέρα, καθώς τα δεδομένα που επηρεάζουν το αποτέλεσμα είναι η παραγωγή της προηγούμενης ημέρας. Συνεπώς αυτό το μοντέλο λειτουργεί αρκετά καλά για χρονική περίοδο όπου οι καιρικές συνθήκες δεν έχουν μεγάλες αλλαγές, συνεχόμενες μέρες ηλιοφάνειας ή νεφοκάλυψης. Τέτοιες μέρες πιο συχνά συναντώνται κατά τους καλοκαιρινούς μήνες.

## 5.2 Μηχανική Μάθηση

### 5.2.1 Αποτελέσματα ηλιακής ακτινοβολίας (Solar Radiation)

Όπως αναφέραμε και πιο πάνω τα μοντέλα πρόβλεψης που έχουμε κάνει, αρχικά προβλέπουν την ηλιακή ακτινοβολία που φτάνει στα φωτοβολταϊκά πάνελ και στη συνέχεια υπολογίζουν την παραγωγή της ηλεκτρικής ενέργειας. Με αυτόν τον τρόπο κατασκευάζουμε ένα πιο γενικό μοντέλο καθώς τα ίδια αποτελέσματα ηλιακής ακτινοβολίας μπορούν να χρησιμοποιηθούν και για άλλο ηλιακό πάρκο που βρίσκεται στη γύρω περιοχή. Έτσι κατασκευάζοντας τα 3 μοντέλα με τους αλγορίθμους μηχανικής μάθησης, SVR, GBRT και MLP Neural Network προκύπτουν τα παρακάτω αποτελέσματα.

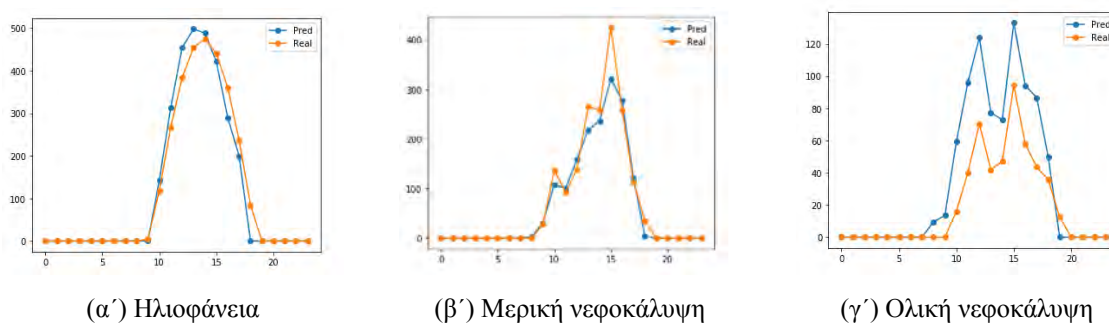
## SVR

Με τη χρήση του αλγορίθμου SVR βλέπουμε στο Πίνακα 5.1 να έχουμε μια πολύ καλή απόδοση του αλγορίθμου για τις ημέρες με ηλιοφάνεια σε ποσοστό 91.5% και με σφάλμα που ξεπερνάει λίγο τις 398 μονάδες. Επίσης βλέπουμε τις ημέρες με ολική νεφοκάλυψη, που είναι και οι πιο δύσκολες για πρόβλεψη, έχουμε ένα ποσοστό επιτυχίας κοντά στο 74%, με σφάλμα 184 μονάδες. Επιπλέον παρατηρούμε ότι και στις μέρες με μερική νεφοκάλυψη τα αποτελέσματα επιτυχίας βρίσκονται περίπου στο 78%. Αξίζει να σημειωθεί ότι όλες οι μετρήσεις έγιναν μόνο για τις ώρες που το ηλιακό πάρκο παράγει ηλεκτρική ενέργεια (8:00 - 20:00).

Πίνακας 5.1: Μέτρηση απόδοσης και RMSE του αλγορίθμου SVR για τη πρόβλεψη του Solar Radiation

Καιρικές Συνθήκες	accuracy(%)	RMSE (scaled)	RMSE
Ηλιοφάνεια(5640 εγγραφές)	91.50	0.2861	398.0061
Μερική Νεφοκάλυψη(3096 εγγραφές)	77.64	0.4393	256.2578
Ολική Νεφοκάλυψη(1008 εγγραφές)	73.70	0.4588	184.1343

Στη συνέχεια επιλέγοντας τυχαία 3 ημέρες μία για κάθε μία από τις κατηγορίες καιρικών συνθηκών (ηλιοφάνεια, μερική νεφοκάλυψη, ολική νεφοκάλυψη) στο Σχήμα 5.3 φαίνονται σχηματικά τα αποτελέσματα τις πρόβλεψης και στο Πίνακα 5.2 φαίνονται τα αποτελέσματα των σφάλματα (RMSE) κάθε ημέρας. Σημαντικό εδώ είναι να παρατηρήσουμε στο Σχήμα 5.3 το (γ), τη περίπτωση της πλήρης νεφοκάλυψης, που παρόλο ότι το RMSE είναι σχετικά υψηλό έχουμε καταφέρει να πετύχουμε τις εναλλαγές που κάνει ο καιρός μέσα στην ημέρα.



Σχήμα 5.3: Ημερήσια πρόβλεψη Solar Radiation με χρήση SVR

Πίνακας 5.2: Μέτρηση RMSE του αλγορίθμου SVR για ημερήσια πρόβλεψη του Solar Radiation

Καιρικές Συνθήκες	RMSE (scaled)	RMSE
Ηλιοφάνεια(24 εγγραφές)	0.3138	362.5132
Μερική Νεφοκάλυψη(24 εγγραφές)	0.3653	246.4688
Ολική Νεφοκάλυψη(24 εγγραφές)	0.6212	196.4922

### Gradient Boosted Regression Trees

Το επόμενο μοντέλο ήταν αυτό του Gradient Boosted Regression Trees. Στο μοντέλο αυτό βλέπουμε ότι τα αποτελέσματά του είναι αρκετά κοντά με τα αποτελέσματα του SVR, με αξιοσημείωτο στοιχείο την ελάχιστη καλύτερα αποτελεσματικότητά του σε ημέρες ολικής νεφοκάλυψης με ποσοστό επιτυχίας να φτάνει το 75.8 και σφάλμα RMSE κοντά στις 181 μονάδες. Επίσης για τις ημέρες με μερική νεφοκάλυψη παρατηρούμε ένα καλό αποτέλεσμα λίγο πάνω από 79%, με αποτέλεσμα να τον επιλέγουμε για την πρόβλεψη εκείνων των ημερών (μερικής νεφοκάλυψης). Και εδώ πρέπει να σημειωθεί ότι οι μετρήσεις έγιναν τις ώρες παραγωγής των ηλιακών πάνελ.

Πίνακας 5.3: Μέτρηση απόδοσης και RMSE του αλγορίθμου GBRT για τη πρόβλεψη του Solar Radiation

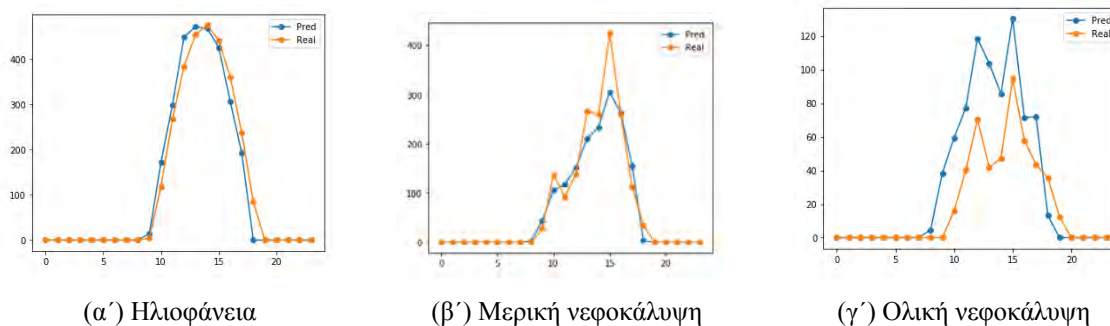
Καιρικές Συνθήκες	accuracy(%)	RMSE (scaled)	RMSE
Ηλιοφάνεια(5640 εγγραφές)	91.86	0.2408	397.1035
Μερική Νεφοκάλυψη(3096 εγγραφές)	79.51	0.4229	255.8658
Ολική Νεφοκάλυψη(1008 εγγραφές)	75.85	0.3738	181.7924

Ακολουθώντας επιλέγοντας τις ίδιες ημέρες με το προηγούμενο παράδειγμα, μία για κάθε μία από τις κατηγορίες καιρικών συνθηκών (ηλιοφάνεια, μερική νεφοκάλυψη, ολική νεφοκάλυψη) τα αποτελέσματα της πρόβλεψης φαίνονται σχηματικά (Σχήμα 5.4) πιο κάτω και στο Πίνακα 5.4 φαίνονται τα σφάλματα (RMSE) κάθε ημέρας. Και εδώ βλέπουμε να έχουμε πετύχει τις εναλλαγές του καιρού μέσα στη μέρα τόσο για την ημέρα της μερικής νεφοκάλυψης όσο και για την ημέρα της πλήρους νεφοκάλυψης.

Πίνακας 5.4: Μέτρηση RMSE του αλγορίθμου GBRT για ημερήσια πρόβλεψη του Solar Radiation

Καιρικές Συνθήκες	RMSE (scaled)	RMSE
Ηλιοφάνεια(24 εγγραφές)	0.3108	362.3736
Μερική Νεφοκάλυψη(24 εγγραφές)	0.3520	246.3213
Ολική Νεφοκάλυψη(24 εγγραφές)	0.3738	181.7924





Σχήμα 5.4: Ημερήσια πρόβλεψη Solar Radiation με χρήση GBRT

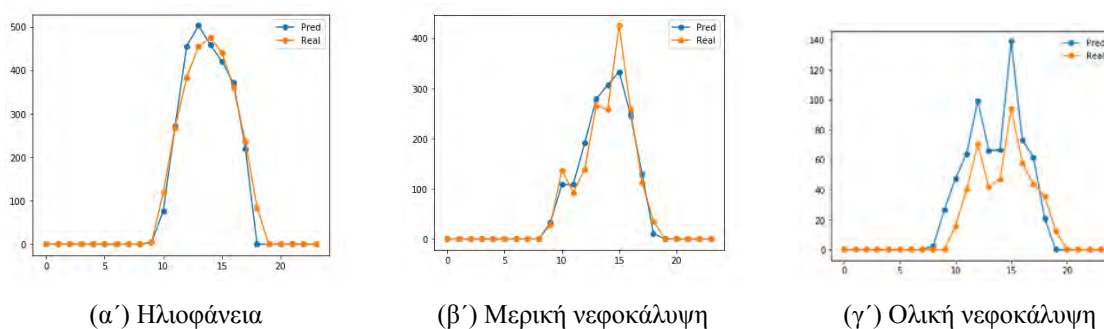
### MLP Neural Networks

Το τελευταίο μοντέλο που χρησιμοποιήσαμε ήταν αυτό με τη χρήση των MLP νευρωνικών δικτύων. Τα αποτελέσματα τα βλέπουμε στο Πίνακα 5.5 και παρατηρούμε ότι τα σφάλματα που δίνει ο αλγόριθμος για κάθε μία από τις κατηγορίες είναι λίγο πιο μεγάλα από τους προηγούμενους αλγορίθμους, για όλες τις κατηγορίες εκτός από τις ημέρες μερικής νεφοκάλυψης. Και εδώ οι μετρήσεις έγιναν μόνο τις ώρες παραγωγής των φωτοβολταϊκών πάνελ.

Πίνακας 5.5: Μέτρηση απόδοσης και RMSE του αλγορίθμου MLP για τη πρόβλεψη του Solar Radiation

Καιρικές Συνθήκες	RMSE (scaled)	RMSE
Ηλιοφάνεια(5640 εγγραφές)	0.3289	398.9376
Μερική Νεφοκάλυψη(3096 εγγραφές)	0.4018	248.8490
Ολική Νεφοκάλυψη(1008 εγγραφές)	0.5072	185.4267

Ακολουθώντας επιλέγοντας τις ίδιες ημέρες με το προηγούμενο παράδειγμα, μία για κάθε μία από τις κατηγορίες καιρικών συνθηκών (ηλιοφάνεια, μερική νεφοκάλυψη, ολική νεφοκάλυψη) τα αποτελέσματα της πρόβλεψης φαίνονται σχηματικά (Σχήμα 5.5) πιο κάτω και στον Πίνακα 5.6 φαίνονται τα σφάλματα (RMSE) κάθε ημέρας.



Σχήμα 5.5: Ημερήσια πρόβλεψη Solar Radiation με χρήση MLP neural networks

Πίνακας 5.6: Μέτρηση RMSE για ημερήσια πρόβλεψη του Solar Radiation με τη χρήση MLP neural networks

Καιρικές Συνθήκες	RMSE (scaled)	RMSE
Ηλιοφάνεια(24 εγγραφές)	0.2639	361.2405
Μερική Νεφοκάλυψη(24 εγγραφές)	0.4018	248.8490
Ολική Νεφοκάλυψη(24 εγγραφές)	0.6934	193.0591

### 5.2.2 Αποτελέσματα παραγωγής ηλεκτρικής ενέργειας(kWh)

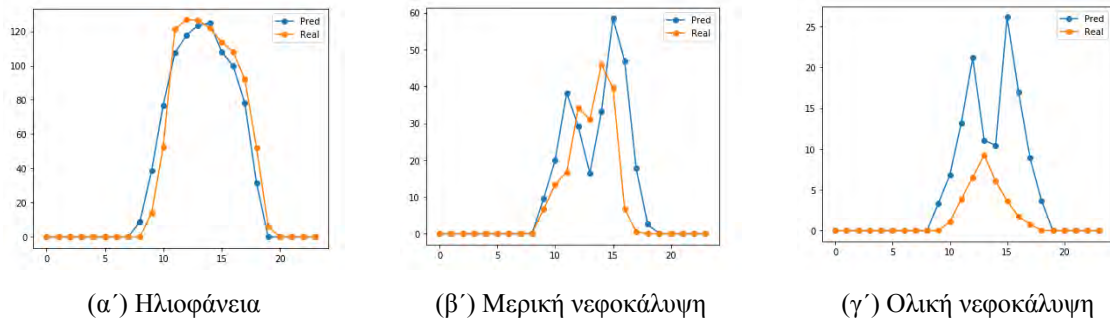
Μετά την πρόβλεψη της ηλιακής ενέργειας (Solar Radiation), θέλουμε να υπολογίσουμε την παραγόμενη ποσότητα ηλεκτρικής ενέργειας που θα μας δώσουν τα φωτοβολταϊκά πάνελ. Αυτή η ποσότητα όπως αναφέραμε και πιο πάνω θα προκύψει από έναν τύπο γινομένου περιλαμβάνοντας μέσα και την τιμή της ηλιακής ενέργειας που έχουμε προβλέψει από την παραπάνω διαδικασία. Και εδώ ο υπολογισμός του RMSE έχει γίνει λαμβάνοντας υπόψιν μόνο τις ώρες που παράγουν ηλεκτρική ενέργεια τα πάνελ. Έτσι τα συνολικά μέσα τετραγωνικά σφάλματα για κάθε μία περίπτωση φαίνονται στον παρακάτω πίνακα (Πίνακας 5.7).

Πίνακας 5.7: Μετρήσεις του RMSE για τους 3 αλγορίθμους για τη παραγωγή ηλεκτρικής ενέργειας (kWh)

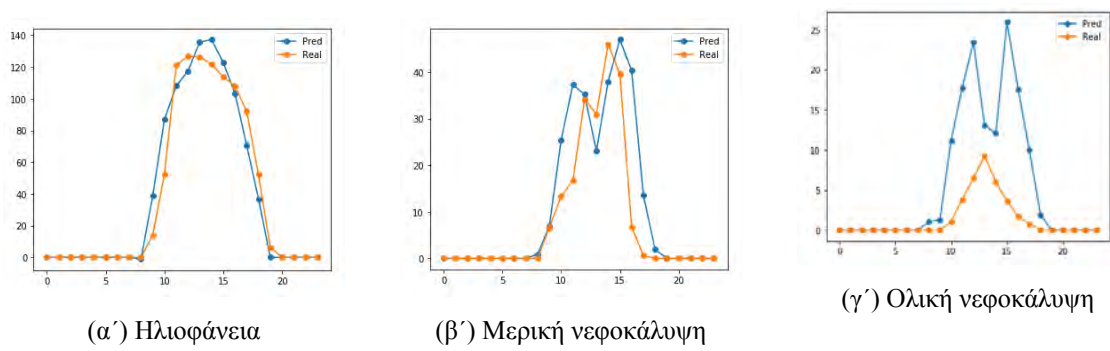
Καιρικές Συνθήκες	MLP Neural	SVR	Gradient Boosting
Ηλιοφάνεια(5640 εγγραφές)	44.1536	48.5086	48.4996
Μερική Νεφοκάλυψη(3096 εγγραφές)	34.9080	34.8829	34.8563
Ολική Νεφοκάλυψη(1008 εγγραφές)	29.1209	27.7179	28.2479

Παρατηρούμε ότι ανά περίπτωση των καιρικών συνθηκών υπάρχει και διαφορετικός αλγόριθμος που δίνει καλύτερο αποτέλεσμα. Έτσι για τις μέρες ηλιοφάνειας χρησιμοποιήθηκε ο αλγόριθμος των MLP νευρωνικών δικτύων, για τις μέρες με μερική νεφοκάλυψη ο GBRT και ο SVR για τις μέρες με ολική νεφοκάλυψη.

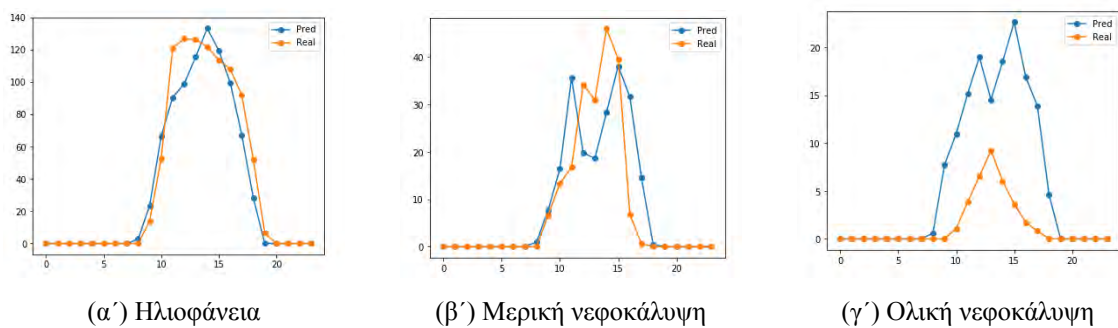
Στην συνέχεια έχουν επιλεγεί 3 ημέρες του σύνολο τιμών, για τις 3 διαφορετικές κατηγορίες καιρικών συνθηκών και θα δείξουμε τα αποτελέσματα που έχουν δώσει τα 3 μοντέλα ανά περίπτωση.



Σχήμα 5.6: Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας με MLP



Σχήμα 5.7: Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας με SVR



Σχήμα 5.8: Πρόβλεψη Παραγόμενης Ηλεκτρικής Ενέργειας με Gradient Boosting

Από τα σχήματα πιο πάνω (Σχήμα 5.6), (Σχήμα 5.7), (Σχήμα 5.8) βλέπουμε ότι έχουμε μία πολύ καλή πρόβλεψη για την ημέρα με ηλιοφάνεια με RMSE κοντά στις 18 μονάδες, πράγμα το οποίο επιτεύχθηκε με τη χρήση των MLP νευρωνικών δικτύων. Μια σχετικά καλή για την ημέρα με μερική νεφοκάλυψη σε διάφορες ώρες μέσα στη μέρα, με RMSE κοντά στις 22 μονάδες με τη χρήση του αλγορίθμου GBRT. Και τέλος την ημέρα με αρκετή νεφοκάλυψη έχουμε μια πρόβλεψη με RMSE λίγο πάνω από 10 μονάδες με τη χρήση του αλγορίθμου SVR, το οποίο είναι σχετικά καλό αν αναλογιστεί κανένας ότι η συνολική παραγωγή μέσα στη μέρα φτάνει τις 50 παραγωγή.

Πίνακας 5.8: Μετρήσεις του RMSE για τους 3 αλγορίθμου παραγωγής ηλεκτρικής ενέργειας (για μία ημέρα)

Καιρικές Συνθήκες	MLP Neural	SVR	Gradient Boosting
Ηλιοφάνεια(24 εγγραφές)	18.4201	18.3582	19.9119
Μερική Νεφοκάλυψη(24 εγγραφές)	19.3922	22.1094	17.9611
Ολική Νεφοκάλυψη(24 εγγραφές)	10.3559	11.1528	12.8136

## Κεφάλαιο 6

# Εφαρμογή

Σε αυτό το κεφάλαιο θα παρουσιάσουμε τις τεχνικές λεπτομέρειες της εφαρμογής, τις επιμέρους λειτουργίες της αλλά και το που μπορεί κάποιος να βρει την εφαρμογή για να την κατεβάσει και να την εκτελέσει.

### 6.1 Λεπτομέρειες υλοποίησης

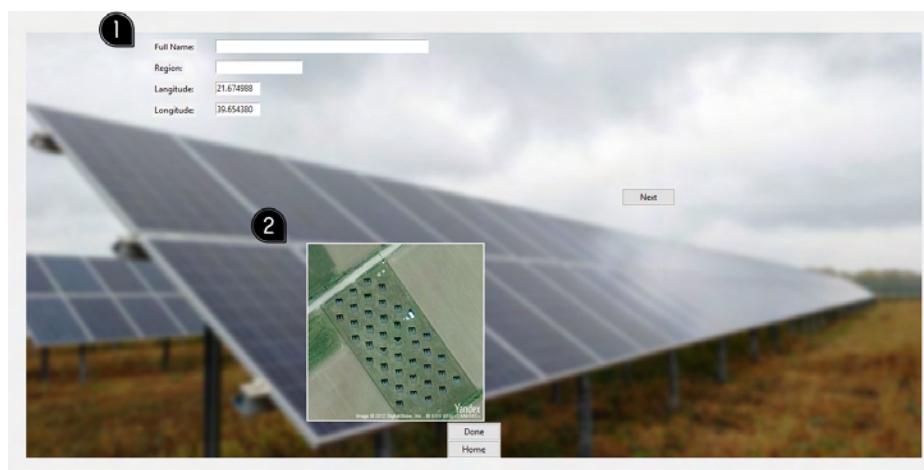
Η εφαρμογή αναπτύχθηκε σε γλώσσα python με τη χρήση της βιβλιοθήκης Tkinter, βιβλιοθήκη η οποία χρησιμοποιείται για την κατασκευή gui. Τα ιστορικά δεδομένα που χρησιμοποιήθηκαν προήλθαν από το Meteoblue ενώ τα δεδομένα των καιρικών συνθηκών που επικρατούν εκείνη την στιγμή που ο χρήστης δοκιμάζει την εφαρμογή αντλούνται από το Darksky, ιστότοπο ο οποίος περιέχει τόσο καιρικές προβλέψεις όσο και ιστορικά δεδομένα.



Σχήμα 6.1: Αρχική επιφάνεια της εφαρμογής

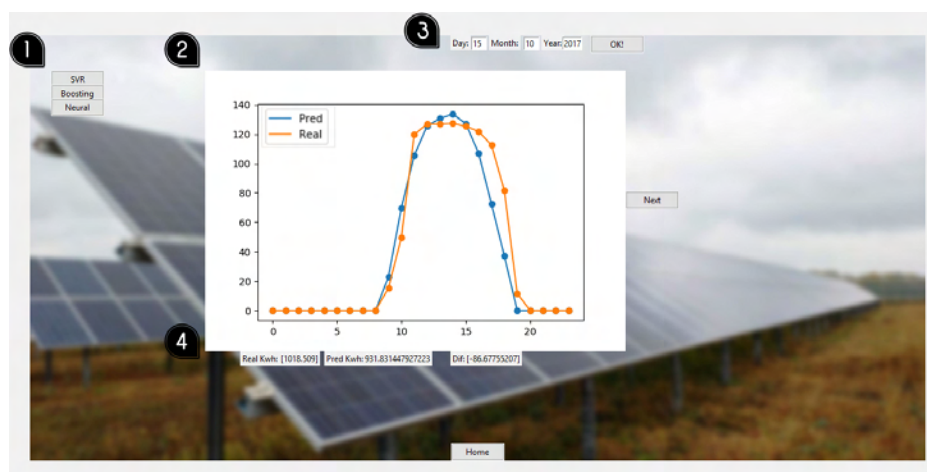
Στην αρχική επιφάνεια της εφαρμογής βλέπουμε ο χρήστης να έχει 3 επιλογές. Η πρώτη αφορά ορισμένα στοιχεία της εγκατάστασης όπως όνομα του ιδιοκτήτη, περιοχή εγκατάστασης και τις συντεταγμένες του ηλιακού πάρκου. Η δεύτερη επιλογή αναφέρεται σε ιστορικά στοιχεία παραγωγής για τη συγκεκριμένη εγκατάσταση καθώς επίσης και συγκριτικό της πραγματικής παραγω-

γής με αυτή που πρόβλεψαν οι αλγόριθμοι. Και η τελευταία αφορά τις τωρινές καιρικές συνθήκες που επικρατούν στην περιοχή και στη συνέχεια η πρόβλεψη που κάνει για την επόμενη ημέρα. Έχει χαρακτηριστεί σαν beta διότι τα στοιχεία του καιρού που χρησιμοποιούμε προέρχονται από διαφορετικό ιστότοπο, από αυτόν που έγινε η ανάπτυξη όλων των μοντέλων. Η επιλογή αυτή έγινε λόγω του υψηλού κόστους αγοράς του API από το μετεωρολογικό ιστότοπο Metoblue.



Σχήμα 6.2: Πρώτη σελίδα της εφαρμογής, για εισαγωγή πληροφοριών

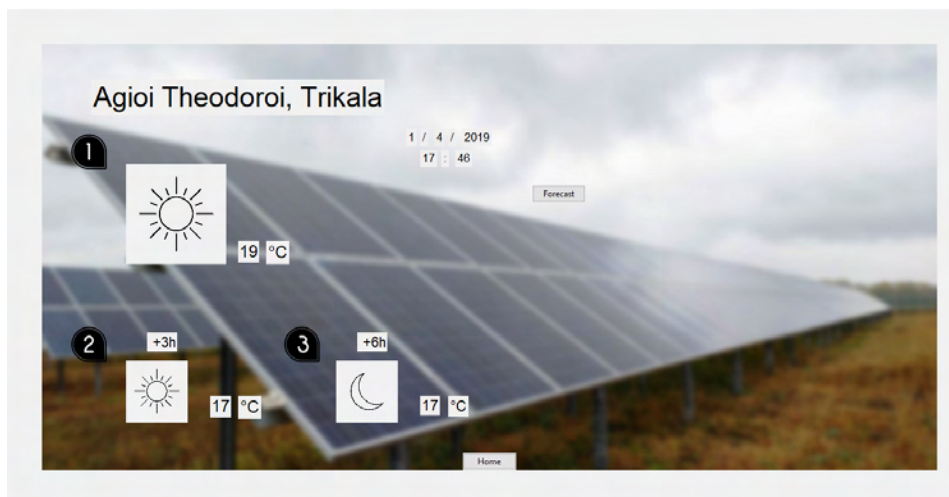
Στη πρώτη σελίδα ο χρήστης θα βρει 4 πεδία προς συμπλήρωση το πρώτο αφορά το ονοματεπώνυμό του (Σχήμα 6.2, 1), το δεύτερο την περιοχή και τα άλλα δύο αφορούν τις συντεταγμένες που βρίσκεται το φωτοβολταϊκό πάρκο. Τοποθετώντας τα όλα και πατώντας το 'Done' τότε μέσω ενός API της εταιρείας Yandex ανατρέχει με τις συντεταγμένες που έχει δώσει ο χρήστης, βρίσκει την τοποθεσία του φωτοβολταϊκού πάρκου και εμφανίζει τη φωτογραφία του δορυφόρου από την συγκεκριμένη τοποθεσία (Σχήμα 6.2, 2). Τα δύο τελευταία πεδία είναι προ συμπληρωμένα (από την τοποθεσία του πάρκου που γίνεται η εφαρμογή) για λόγους ευκολίας.



Σχήμα 6.3: Δεύτερη σελίδα της εφαρμογής, με ιστορικό της παραγωγής

Στην δεύτερη σελίδα ο χρήστης έχει την επιλογή βάζοντας την ημερομηνία που θέλει, από το σύνολο δεδομένων που υπάρχει διαθέσιμο (Σχήμα 6.3, 3), να δει ένα συγκριτικό διάγραμμα μεταξύ

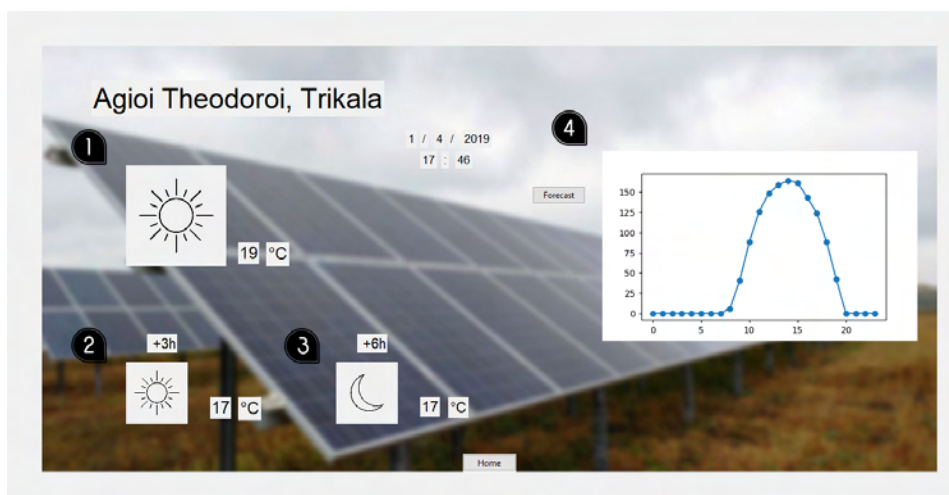
της παραγόμενης ενέργειας σε σύγκριση με αυτήν που έχει προβλέψει ο κάθε αλγόριθμος (Σχήμα 6.3, 2). Έτσι από τα αριστερά μπορεί να επιλέξει μπορεί να επιλέξει ένα εκ των 3 μοντέλων (Σχήμα 6.3, 1) και να δει στο διάγραμμα στο κέντρο τη σύγκριση. Επιπλέον κάτω από το διάγραμμα, ο χρήστης μπορεί να δει τις συνολικές τιμές παραγωγής, της ημέρας, τόσο για την προβλεπόμενη όσο και για την πραγματική τιμή (Σχήμα 6.3, 4). Τέλος μπορεί να δει και την διαφορά μεταξύ αυτών(προβλεπόμενη - πραγματική) (Σχήμα 6.3, 4).



Σχήμα 6.4: Τρίτη σελίδα της εφαρμογής, ζωντανές προβλέψεις καιρού

Στην τελευταία σελίδα ο χρήστης θα μπορεί να δει ζωντανά τις καιρικές συνθήκες καθώς και την θερμοκρασία που επικρατούν στην τοποθεσία που έχει βάλει στην πρώτη καρτέλα, τόσο αυτήν τη στιγμή, όσο και για μετά από 3 και 6 ώρες (Σχήμα 6.4, 1, 2, 3). Όλα αυτά τα στοιχεία προέρχονται μέσω του API που έχουμε πάρει από τον ιστότοπο darksky. Έτσι στη συνέχεια πατώντας ο χρήστης το 'forecast' η εφαρμογή ανάλογα με τις καιρικές συνθήκες που θα επικρατήσουν την επόμενη ημέρα επιλέγει ένα από τα 3 διαθέσιμα μοντέλα που υπάρχουν και κάνει την πρόβλεψη (Σχήμα 6.5, 4). Το αποτέλεσμα της πρόβλεψης αποθηκεύεται και σε ένα αρχείο από όπου ο χρήστης μπορεί να δει τις ακριβείς τιμές της παραγωγής.

Όπως αναφέραμε και πιο πάνω αυτή η καρτέλα έχει χαρακτηριστεί ως beta διότι οι καιρικές τιμές που δίνουμε στο μοντέλο μέσω του API, προέρχονται από διαφορετικό ιστότοπο με αποτέλεσμα να δημιουργούνται κάποιες αποκλίσεις σε σχέση με τη σωστή τιμή πρόβλεψης. Φυσικά οι μεταβλητές που χρησιμοποιούνται από τον ιστότοπο darksky παραμένουν ίδιες και έχουν γίνει μόνο μερικές αλλαγές όσον αφορά τις μονάδες μέτρησης και την κλίμακά τους. Έτσι η θερμοκρασία αλλάχθηκε από Faraday σε Celsius, για την ηλιοφάνεια ανάλογα με την κατάσταση του καιρού δόθηκε μία τιμή που αφορούσε την διάρκεια ηλιοφάνειας ανά ώρα και τέλος η υγρασία και η νεφοκάλυψη μετατράπηκαν σε ποσοστά.



Σχήμα 6.5: Τρίτη σελίδα της εφαρμογής, πρόβλεψη για την επόμενη ημέρα

## 6.2 Οδηγίες λήψης και εκτέλεσης της εφαρμογής

Ο κώδικας της παραπάνω εφαρμογής καθώς επίσης και το σύνολο δεδομένων που χρησιμοποιήσαμε βρίσκονται ανεβασμένα στο Github και είναι διαθέσιμα σε όποιον θέλει να τα δοκιμάσει, κατεβάζοντάς τα από τον παρακάτω σύνδεσμο.

<https://github.com/alexpap13/Prediction-of-Photovoltaic-Solar-Energy-Production>



Σχήμα 6.6: Qr code με σύνδεσμο που οδηγεί στα αρχεία της εργασίας

Αναλυτικά τα αρχεία που βρίσκονται μέσα στο Github φαίνονται στον ακόλουθο πίνακα (Πίνακας 6.1)

Για την εκτέλεση της εφαρμογής απαιτείται σύνδεση στο διαδίκτυο διότι τα API που χρησιμοποιούμε αντλούν πληροφορίες από το διαδίκτυο, επιπλέον θα πρέπει να έχουν εγκατασταθεί όλες οι βιβλιοθήκες που χρησιμοποιούνται στο πρόγραμμα, οι οποίες είναι tkinter, pandas, numpy, pickle, sklearn, matplotlib.pyplot, requests, datetime, PIL. Επιπλέον θα πρέπει να έχει γίνει εγκατάσταση του tensorflow και του kerras.



Πίνακας 6.1: Αναλυτική περιγραφή αρχείων στο github

Όνομα	Περιγραφή
1-10-17#1-10-18.csv	Βάση δεδομένων csv
App_solar_pred.py	Εφαρμογή Python
models.rar	Φάκελος με το σύνολο των μοντέλων
photos.rar	Φάκελος με το σύνολο των φωτογραφιών
weather_icon.rar	Φάκελος με το σύνολο των φωτογραφιών για τη περιγραφή του καιρού

Για να εκτελέσει κανείς την εφαρμογή θα πρέπει ο χρήστης να κατεβάσει όλα τα αρχεία και να τα τοποθετήσει όλα στον ίδιο φάκελο. Στη συνέχεια μέσω της γραμμής εντολών μπορεί να δώσει την εντολή **python App\_solar\_pred.py** και να αρχίσει να τρέχει η εφαρμογή.



## Κεφάλαιο 7

# Επίλογος

### 7.1 Σύνοψη και συμπεράσματα

Συνοψίζοντας όλα τα παραπάνω και βλέποντας τα αποτελέσματα, παρατηρούμε ότι δεν υπάρχει κάποιος αλγόριθμος από τους 4 που να κάνει πολύ καλές προβλέψεις και για τις 3 κατηγορίες καιρικών φαινομένων. Καθώς ο SVR δίνει καλύτερα αποτελέσματα από τους υπόλοιπους για την κατηγορία ολικής νεφοκάλυψης, ο GBRT δίνει καλύτερα για τις ημέρες με μερική νεφοκάλυψη και ο MLP για ημέρες με ηλιοφάνεια. Ο πιο αποδοτικός τρόπος για να έχουμε το καλύτερο δυνατό αποτέλεσμα ήταν να τους συνδυάσουμε και τους 3 μαζί. Ο καθένας για τις καιρικές συνθήκες που αποδίδει καλύτερα. Έτσι βλέποντας στις μέρες με ηλιοφάνεια να έχουμε μία πολύ καλή πρόβλεψη, σε αντίθεση με ημέρες πλήρης νεφοκάλυψης μπορούμε να συμπεράνουμε ότι το μοντέλο μας θα δούλευε καλύτερα σε περιοχές με κλίμα γενικά σταθερό και με χωρίς πολλές καιρικές εναλλαγές. Έτσι παρατηρώντας το κλίμα της Ελλάδας θα λέγαμε ότι το μοντέλο μας θα δούλευε αρκετά καλά τους καλοκαιρινούς μήνες, φυσικά πλην εκείνων των ημερών που θα υπήρχε αρκετή νεφοκάλυψη, δίνοντας αποτελέσματα αρκετά κοντά με τα πραγματικά. Με αποτέλεσμα οπότε, αυτή η γνώση που θα προέρχεται από τη πρόβλεψη να είναι είναι χρήσιμη τόσο για την αύξηση των κερδών εταιριών παραγωγής και διαχείρισης ενέργειας, καθώς θα είναι σε θέση να γνωρίζουν τα ποσά ενέργειας που θα παράγουν την επόμενη μέρα με μεγάλη ακρίβεια, αλλά και για διάφορες επιχειρήσεις που έχουν την ενέργεια από τα ηλιακά πάνελ για αυτοκατανάλωση, να έχουν μία σταθερότητα στο σύστημά τους.

Όσον αφορά την εφαρμογή είναι πλήρως λειτουργική δίνοντας στο χρήστη την πρόβλεψη της επόμενης ημέρας, με ένα γραφικό περιβάλλον το οποίο είναι εύκολο να χρησιμοποιηθεί από τον οποιοδήποτε και από αποτελέσματα τα οποία είναι εύκολο να μελετηθούν αναλυτικά μέσω του αρχείου που δημιουργείται μόλις γίνει η πρόβλεψη.

### 7.2 Μελλοντικές επεκτάσεις

Όσον αφορά τις μελλοντικές επεκτάσεις που θα μπορούσαν να γίνουν στην διπλωματική, αυτές ξεκινάνε τόσο από τα καιρικά δεδομένα όσο και από τα μοντέλα πρόβλεψης. Έτσι για τα καιρικά δεδομένα μπορούμε να χρησιμοποιήσουμε δεδομένα που θα προέρχονται από περισσότερους

ιστότοπους για να μπορέσουμε έτσι να απαλείψουμε όσο το δυνατό τα σφάλματα που υπάρχουν στις καιρικές προβλέψεις. Το ιδανικότερο βέβαια θα ήταν να εγκαταστήσουμε τον δικό μας μετεωρολογικό σταθμό στο φωτοβολταϊκό πάρκο λαμβάνοντας έτσι, ακριβώς ότι πληροφορίες θα θέλαμε και με μεγάλη ακρίβεια καθώς ο σταθμός θα βρισκόταν ακριβώς στην τοποθεσία που θα θέλαμε. Όσον αφορά τώρα για τα μοντέλα, στην αύξηση της απόδοσής τους θα βοηθούσε η κατασκευή ενός μοντέλου που θα χρησιμοποιούσε φωτογραφίες δορυφόρων για την παρακολούθηση της πορείας της νεφοκάλυψης για να είναι σε θέση να προβλέπει ακόμα και τις απότομες εναλλαγές της νεφοκάλυψης, καθώς από τα παραπάνω μοντέλα που κατασκευάσαμε και δοκιμάσαμε είδαμε ότι τα μεγαλύτερα σφάλματα υπήρχαν στις ημέρες με νεφοκάλυψη. Τώρα όσον αφορά την εφαρμογή θα μπορούσε να αλλαχθεί το γραφικό της περιβάλλον και να γίνει πιο διαδραστικό και θα μπορούσε να ξανά γραφτεί σαν web application για να μπορεί να έχει ο οποιοσδήποτε εύκολη πρόσβαση σε αυτήν και να μην χρειάζεται να κατεβάζει αρχεία.



## Παράρτημα Ι

# Προδιαγραφές Φωτοβολταϊκών Πάνελ

GR   BR	Φωτοβολταϊκά πλαίσια aleo	S_1B
		<p>Κατηγορίες απόδοσης 235   240 W</p>
	<p><b>Καλοί λόγοι</b></p>	
	<ul style="list-style-type: none"> <li>→ 10 χρόνια εγγύηση προϊόντος</li> <li>→ 25 χρόνια εγγύηση απόδοσης</li> <li>→ Θετική ταξινόμηση</li> <li>→ Τα καλύτερα εξαρτήματα σύμφωνα με τα γερμανικά πρότυπα ποιότητας</li> <li>→ Επαγγελματική εξυπηρέτηση πελατών</li> <li>→ Πιστοποίηση κατά IEC/EN 61215 (Ed. 2)</li> <li>→ Πιστοποίηση κατά IEC/EN 61730, κατηγορία προστασίας II</li> <li>→ Γερμανικό περιοδικό καταναλωτών Öko-Test: «Πολύ καλά»</li> </ul>	
		

## aleo s\_1e

Ηλεκτρικά χαρακτηριστικά (STC)		S 18. 235		S 18. 240	
Όνομαστική ισχύς	$P_{\text{nom}}$ [W]	235		240	
Όνομαστική τάση	$U_{\text{nom}}$ [V]	29,3		29,5	
Όνομαστικό ρεύμα	$I_{\text{nom}}$ [A]	8,01		8,12	
Τάση άνω φορτίου	$U_{\text{oc}}$ [V]	36,8		37,0	
Ρεύμα βραχυκύκλισης	$I_{\text{sc}}$ [A]	8,54		8,65	
Απόδοση	$\eta$ [%]	14,3		14,6	
Επιφάνεια ανά μονάδα ισχύος	$A_p$ [m <sup>2</sup> /kW <sub>p</sub> ]	6,99		6,89	

Ηλεκτρικά χαρακτηριστικά σε τυπικές συνθήκες δοκιμών (STC): 1000 W/m<sup>2</sup>, 25°C, AM 1,5

Ηλεκτρικά χαρακτηριστικά (NOCT)		S 18. 235		S 18. 240	
Ισχύς	$P_{\text{nom}}$ [W]	169		173	
Τάση	$U_{\text{nom}}$ [V]	27,9		28,3	
Ρεύμα	$I_{\text{nom}}$ [A]	6,05		6,11	
Τάση άνω φορτίου	$U_{\text{oc}}$ [V]	33,8		34,0	
Ρεύμα βραχυκύκλισης	$I_{\text{sc}}$ [A]	6,67		6,72	
Απόδοση	$\eta$ [%]	12,8		13,1	

Ηλεκτρικά χαρακτηριστικά σε ονομαστικές συνθήκες λειτουργίας κυψελών: 800 W/m<sup>2</sup>, 20°C, AM 1,5, άνεμος 1 m/s

NOCT: 48°C (ονομαστική θερμοκρασία λειτουργίας κυψελών)

Άλλα ηλεκτρικά χαρακτηριστικά		
Μείωση της απόδοσης STC Από 1000 W/m <sup>2</sup> σε 200 W/m <sup>2</sup>	[K]	< 6
Εύρος κατηγορίας (βελτική ταξινόμηση)	[W]	-0/+4,99
Ακρίβεια μέτρησης $P_{\text{nom}}$ σε STC	[K]	-3/+3
Αναχή λοιπών ηλεκτρικών τιμών	[K]	-10/+10

## Φυσικά

Μηχανική καταπόνηση πλαστικών	[Pa]	5400
Μέγιστη τάση αντισήματος	[V <sub>oc</sub> ]	1000
Ρεύμα επιστροφής	$I_{\text{L}}$ [A]	15

Μηχανική καταπόνηση κατά IEC/EN 61215

## Συντελεστές θερμοκρασίας

1. Συντελεστής θερμοκρασίας	$\alpha$ [ $I_{\text{sc}}$ ]	[%/K]	+0,04
2. Συντελεστής θερμοκρασίας	$\beta$ [ $U_{\text{oc}}$ ]	[%/K]	-0,34
3. Συντελεστής θερμοκρασίας	$\gamma$ [ $P_{\text{nom}}$ ]	[%/K]	-0,46

Διεύθυνση
aleo solar AG Gewerbegebiet Nord Krummer Weg 1 17 291 Pörschau Γερμανία

## Διεύθυνση εγκαταστάσεων

aleo solar Ελλάς Ερεδών Αξιοματικών 3 15351 Παλλήνη   Αττική Ελλάδα T +30 2 10 66 57 293 info@aleo-solar.gr www.aleo-solar.gr
---

Βαθμός ανοδικότητας με βάση τον ανακτό επιφάνεια των πλακών

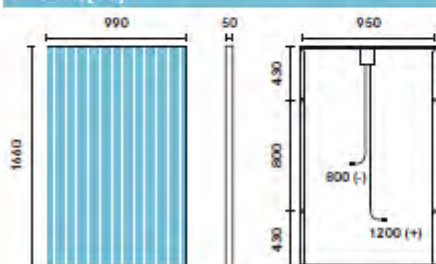
## Βασικά στοιχεία

Μήκος x πλάτος x ύψος	[mm <sup>3</sup> ]	1660 x 990 x 50
Βάρος	[kg]	21
Αριθμός κυψελών		60
Μέγεθος κυψελών	[mm <sup>2</sup> ]	156 x 156
Υλικό κυψελών		Πολυκρυσταλλικό Si
Μπροστινό κάλυμμα		Ηλεκτρομαγνητικό (TSG)
Πίσω κάλυμμα		Πολυμερής μεμβράνη
Υλικό πλαστικού		Χρώμα αλουμινίου
Μήκος καλωδίου	[mm]	1200 (+), 800 (-)
Φίς		Κατηγορία MC3
Κατηγορία IP		IP65
Διοδοί by-pass		3

Το επάillon πληροφορεί σχετικά με τις εγκαταστάσεις στο πρόβλημα και τις υπηρεσίες μας, μπορείτε να επικοινωνήσετε μαζί μας οποιαδήποτε στιγμή ή να αναζητήσετε στην ιστοσελίδα μας στη διεύθυνση [www.aleo-solar.gr](http://www.aleo-solar.gr)

© aleo solar AG | 06/2016 | Με κάθε επιφώλη για παρόν λάθη και ανεπάρκειες

## Διαστάσεις [mm]



## Ο εξοικονομημένος ανταποδοτικός της αλεα της περιεχόμενης

**avelion**  
s o l a r r o o f

Μισαυλή 44, 153 44, Γέρακας  
Τηλ. 217 7000 600, Fax 217 7000 601  
info@avelion.gr www.avelion.gr





# Βιβλιογραφία

- [1] Jose R. Andrade και Ricardo J. Bessa. Improving renewable energy forecasting with a grid of numerical weather predictions. *IEEE Transactions on Sustainable Energy*, 8(4):1571–1580, 2017.
- [2] J. Antonanzas, N. Osorio, R. Escobar, F.J. Martinez de Pison R. Urraca και F. Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar Energy*, 136(6):78–111, 2016.
- [3] Y. Atsushi, S. Tomonobu και S. Hideomi. Application of Neural Network to 24-hour-Ahead Generating Power Forecasting for PV System. *IEEE*, σελίδα 6, 2008.
- [4] Jason Brownlee. *Deep learning with python*. eBook, Melbourne, Australia, 1ή έκδοση, 2018.
- [5] Jason Brownlee. *Machine Learning Mastery With Python*. eBook, Melbourne, Australia, 1ή έκδοση, 2018.
- [6] Cross-Validation. <http://genome.tugraz.at/proclassify/help/pages/xv.html>. Ημερομηνία πρόσβασης: 10-11-2018.
- [7] Cristian Dragos Dumitru, Adrian Gligor και Calin Enachescu. Solar Photovoltaic Energy Production Forecast Using Neural Networks. *9th International Conference Interdisciplinarity in Engineering*, 22(4):808 – 815, 2015.
- [8] Eurostat Statistics-explained. [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Figure\\_1-Share\\_of\\_energy\\_from\\_renewable\\_sources\\_2004-2016.png](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Figure_1-Share_of_energy_from_renewable_sources_2004-2016.png). Ημερομηνία πρόσβασης: 13-1-2019.
- [9] Gentle Introduction Gradient Boosting Algorithm Machine Learning. <https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/>. Ημερομηνία πρόσβασης: 30-11-2018.
- [10] I. Goodfellow, Y. Bengio και A. Courville. Deep Learning. *MIT Press*, σελίδα 196, 2016.
- [11] Google maps. <https://www.google.com/maps/@39.618364,21.6756477,2164m/data=!3m1!1e3>. Ημερομηνία πρόσβασης: 7-4-2019.

- [12] M. Höök. *Depletion and Decline Curve Analysis in Crude Oil Production*. Uppsala, Available at: <http://www.diva-portal.org/smash/get/diva2:338111/FULLTEXT01>, 1ή έκδοση, 2019.
- [13] Rich H. Inman, Hugo T.C. Pedro και Carlos F.M. Coimbra. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6):535 – 576, 2013.
- [14] Introduction to forecasting with arima. <https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials>. Ημερομηνία πρόσβασης: 30-12-2018.
- [15] E. Isaksson. *Solar Power Forecasting with Machine Learning Techniques*. KTH ROYAL INSTITUTE OF TECHNOLOGY SCHOOL OF ENGINEERING SCIENCES, 2018.
- [16] Kernel. <https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f>. Ημερομηνία πρόσβασης: 10-11-2018.
- [17] Models machine learning. <https://developer.ibm.com/articles/cc-models-machine-learning/>. Ημερομηνία πρόσβασης: 20-12-2018.
- [18] Douglas C. Montgomery. *Introduction to Time Series Analysis and Forecasting*. Wiley, Canada, 1ή έκδοση, 2008.
- [19] Alexey Natekin και Alois Knoll. Gradient boosting machines, a tutorial. *Frontiers Media SA*, σελίδα 25, 2013.
- [20] Overfitting - Underfitting. <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>. Ημερομηνία πρόσβασης: 10-11-2018.
- [21] C. Perssona, P. Bachera, T. Shigab και T. Madsen. Multi-site solar power forecasting using gradient boosted regressiontrees. *Social Systems Research Domain, Toyota Central R&D Labs*, σελίδα 14, 2017.
- [22] Renewable power generation costs in 2017. Τεχνική αναφορά με αριθμό, International Renewable Energy Agency, Abu Dhabi, 2018.
- [23] J. Shi, W. J. Lee, Y. Liu, Y. Yang και P. Wang. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *PIEEE Transactions on Industry Applications*, σελίδα 6, 2012.
- [24] SMA solar inverters. <https://www.sma.de/en/products/solarinverters/sunny-tripower-10000tlee-jp.html>. Ημερομηνία πρόσβασης: 30-11-2018.
- [25] Solar. <http://www.geo.auth.gr/courses/gmc/1000/solar.html>. Ημερομηνία πρόσβασης: 3-2-2019.

- [26] Support vector machine. [http://www.saedsayad.com/support\\_vector\\_machine\\_reg.html](http://www.saedsayad.com/support_vector_machine_reg.html). Ημερομηνία πρόσβασης: 30-11-2018.
- [27] Support vector machine introduction to machine learning algorithms. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>. Ημερομηνία πρόσβασης: 30-11-2018.
- [28] Target Model. <https://energypress.gr/news/target-model-0>. Ημερομηνία πρόσβασης: 7-4-2019.
- [29] tsgraphics. <https://www.stat.pitt.edu/stoffer/tsa4/tsgraphics.htm>. Ημερομηνία πρόσβασης: 20-12-2018.
- [30] Understanding support vector machine. <https://towardsdatascience.com/understanding-support-vector-machine-part-2-kernel-trick-mercercs-theorem-e1e6848c6c4d>. Ημερομηνία πρόσβασης: 30-11-2018.
- [31] Understanding Activation Functions in Neural Networks. <https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0>. Ημερομηνία πρόσβασης: 30-11-2018.
- [32] B. Wisner και J. Adams. *Environmental health in emergencies and disasters*. World Health Organization, Switzerland, 1ή έκδοση, 2002.
- [33] Απόφαση αριθ. 406/2009/ΕΚ του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 23ης Απριλίου 2009, περί των προσπαθειών των κρατών μελών να μειώσουν τις οικείες εκπομπές αερίων θερμοκηπίου, ώστε να τηρηθούν οι δεσμεύσεις της Κοινότητας για μείωση των εκπομπών αυτών μέχρι το 2020. σελίδα 136–148, 2009.
- [34] Καλύβα Β. ΑΞΙΟΠΟΙΗΣΗ ΤΗΣ ΗΛΙΑΚΗΣ ΕΝΕΡΓΕΙΑΣ ΚΑΙ ΤΩΝ ΑΛΛΩΝ ΑΠΕ ΓΙΑ ΤΗΝ ΕΝΕΡΓΕΙΑΚΗ ΖΗΤΗΣΗ ΤΟ 2050. Μεταπτυχιακή διπλωματική εργασία, ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ, 2015.
- [35] Χ. Μαμάκης και Α. Σκλήβας. Μελέτη και εγκατάσταση Φ/Β πάρκου. Πτυχιακή εργασία, ΤΕΙ Δυτικής Ελλάδας, 2017.
- [36] Μηχανική Μάθηση. [https://repository.kallipos.gr/bitstream/11419/3382/1/02\\_chapter\\_04.pdf](https://repository.kallipos.gr/bitstream/11419/3382/1/02_chapter_04.pdf). Ημερομηνία πρόσβασης: 20-11-2018.
- [37] Φωτοβολταϊκά. <https://www.4green.gr/news/data/fwtoboltaika/88687.asp>. Ημερομηνία πρόσβασης: 7-4-2019.
- [38] Χρονοσειρές. <http://users.auth.gr/~dkugiu/Teach/DataAnalysis/Chr6.pdf>. Ημερομηνία πρόσβασης: 30-12-2018.



# Συντομογραφίες

ΑΠΕ	Ανανεώσιμες πηγές ενέργειας
RES	Renewable Energy Systems
ΑΔΜΗΕ	Ανεξάρτητος Διαχειριστής Μεταφοράς Ηλεκτρικής Ενέργειας
SMA	System Mess and Anlagentechnik
kWh	Kilo Watt Hour
w/m <sup>2</sup>	watt/square meter
m	meter
Pa	Pascal
Km/h	Kilometer/hour
RMSE	Root Mean Square Error
SVR	Support Vector Regression
SVM	Support Vector Machines
MW	Mega Watt
κ.α.	και άλλα
κ.λ.π.	και τα λοιπά
MLP	Multilayer Perceptron
MSE	Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ACF	Autocorrelation Function
AR	Autoregressive Model
MA	Moving Average
PACF	Partial Autocorrelation Function
ARMA	Autoregressive Moving Average Model
ARIMA	Autoregressive Integrated Moving Average
API	Application Programming Interface
RBF	Radial Basis Function
GBRT	Gradient Boosted Regression Trees

