



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ

« ΛΟΓΙΣΤΙΚΗ ΚΑΙ ΕΛΕΓΚΤΙΚΗ »



Η ΧΡΗΣΗ ΤΗΣ ΤΕΧΝΟΛΟΓΙΑΣ ΤΩΝ ΜΑΖΙΚΩΝ

ΔΕΔΟΜΕΝΩΝ (BIG DATA) ΣΤΗ ΛΟΓΙΣΤΙΚΗ

ΧΡΥΣΟΥΛΑ Ν. ΚΑΤΣΑΜΑΚΑ, Α.Μ: 00105

ΕΠΙΒΛΕΠΩΝ: ΚΑΘΗΓΗΤΗΣ ΗΛΙΑΣ ΣΑΝΤΟΥΡΙΔΗΣ

ΒΟΛΟΣ, ΙΑΝΟΥΑΡΙΟΣ 2022



ΧΡΥΣΟΥΛΑ Ν. ΚΑΤΣΑΜΑΚΑ

Βεβαιώνω ότι είμαι συγγραφέας αυτής της μεταπτυχιακής διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην μεταπτυχιακή διπλωματική εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του Διατμηματικού Προγράμματος μεταπτυχιακών σπουδών στη «Λογιστική και Ελεγκτική» Τμήματα Οικονομικών Επιστημών – Χρηματοοικονομικής και Λογιστικής, Πανεπιστήμιο Θεσσαλίας, Βόλος 2022.

Ευχαριστίες

Ολοκληρώνοντας τη μεταπτυχιακή διπλωματική μου εργασία, δεν θα μπορούσα να παραλείψω να ευχαριστήσω τα άτομα που συνέβαλλαν στην εκπόνησή της.

Πρωτίστως, ευχαριστώ θερμά τον επιβλέποντα καθηγητή μου κύριο Σαντουρίδη Ηλία, ο οποίος πραγματικά με καθοδήγησε σε όλη την πορεία από την αρχή μέχρι το τέλος. Με υπομονή, κατανόηση, κατάλληλες υποδείξεις και διαρκή συμπαράσταση, ανεξαρτήτως ημέρας και ώρας, ήταν δίπλα μου μέχρι την ολοκλήρωση της εργασίας.

Συνεχίζοντας, ευχαριστώ ιδιαίτερα τον αδερφό μου Σωτήρη Κατσαμάκα για τη συμβολή του στην κατανόηση και χρήση, από μεριάς μου, του στατιστικού πακέτου SPSS το οποίο χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων και πινάκων που εμπεριέχονται στην παρούσα εργασία.

Τέλος, τίποτα δεν θα μπορούσε να πραγματοποιηθεί χωρίς τη στήριξη και την κατανόηση του συζύγου μου και των δύο παιδιών μου. Η συμπαράστασή τους και η υπομονή τους μου έδινε κουράγιο και δύναμη να συνεχίσω. Τους ευγνωμονώ για αυτή τους τη στάση, όχι μόνο κατά τη διάρκεια εκπόνησης της μεταπτυχιακής διπλωματικής μου εργασίας, αλλά καθ' όλη τη διάρκεια των μεταπτυχιακών μου σπουδών.

Αφιερωμένη στον σύζυγό μου, Νίκο Παγγιτούλη
και στα αγόρια μου Χρήστο και Νικόλα

Περιεχόμενα

Περίληψη.....	5
Abstract	6
Εισαγωγή.....	7
1. Εισαγωγή στα Big Data.....	10
1. 1. Ορισμός και χαρακτηριστικά των Big Data	15
1. 2. Προκλήσεις από τη χρήση των Big Data.....	21
1. 3. Δημιουργία και προστιθέμενη αξία των Big Data.....	24
1. 4. Απόκτηση δεδομένων Big Data.....	25
1. 5. Εφαρμογή των δεδομένων Big Data στην καθημερινότητα.....	29
1. 5. 1. Δεδομένα εταιριών	29
1. 5. 2. Δεδομένα IoT	29
1. 5. 3. Βιο-ιατρικά δεδομένα.....	31
2. Η εφαρμογή των Big Data στα οικονομικά και τη λογιστική	33
2. 1. Εφαρμογές στη χρηματοοικονομική	33
2. 2. Εφαρμογές στην ελεγκτική.....	38
2. 2. 1. Εξέταση της επάρκειας των Big Data.....	39
2. 2. 2. Εξέταση αξιοπιστίας των Big Data	40
2. 2. 3. Εξέταση συνάφειας των Big Data	41
2. 2. 4. Επίπεδα αποδεικτικών στοιχείων ελέγχου.....	42
2. 2. 5. Συνεχής έλεγχος	43
3. Μεθοδολογία	52
3. 1. Σχεδιασμός ερωτηματολογίου	52
3. 2. Δειγματοληψία.....	53
3. 3. Στατιστική ανάλυση	53
3. 4. Μέθοδοι στατιστικής ανάλυσης	59
4. Αποτελέσματα έρευνας	60
4. 1. Περιγραφική στατιστική και έλεγχος αξιοπιστίας.....	62
4. 2. Αναλύσεις παλινδρόμησης	71
5. Συμπεράσματα & συζήτηση.....	77

6. Βιβλιογραφικές Αναφορές	79
ΠΑΡΑΡΤΗΜΑ Α - Ερωτηματολόγιο.....	93
ΠΑΡΑΡΤΗΜΑ Β - Πίνακες του SPSS.....	96
ΠΑΡΑΡΤΗΜΑ Γ - Γενικά στατιστικά ανά ερώτηση	104

Περίληψη

Στόχος της παρούσας Διπλωματικής εργασίας είναι να διερευνήσει τη σχέση της χρήσης τεχνολογίας Μαζικών Δεδομένων με τον κλάδο της Λογιστικής και την πρόθεση χρήσης της εν λόγω τεχνολογίας από τους Έλληνες λογιστές. Η εργασία χρησιμοποιεί ποσοτική έρευνα, με συλλογή δεδομένων βάσει ερωτηματολογίου. Ο σχεδιασμός του ερωτηματολογίου βασίστηκε στο TAM (Technology Acceptance Model) και η διάθεσή του πραγματοποιήθηκε μόνο με ηλεκτρονικά μέσα, σε όλη την ελληνική επικράτεια. Για την επίτευξη του στόχου, διενεργήθηκαν αναλύσεις γραμμικής παλινδρόμησης με σκοπό τον έλεγχο λογικών υποθέσεων. Από τα αποτελέσματα προκύπτει πως η Πρόθεση Χρήσης εργαλείων Μαζικών Δεδομένων από τους λογιστές εξηγείται σε μεγάλο ποσοστό από την Αντιλαμβανόμενη Ευκολία Χρήσης και την Αντιλαμβανόμενη Χρησιμότητα, με την τελευταία να υπερισχύει στον βαθμό επιρροής. Άλλοι παράγοντες που εξετάστηκαν κι επηρεάζουν την πρόθεση χρήσης της τεχνολογίας Μαζικών Δεδομένων είναι οι ακόλουθοι: η Συμβατότητα, οι Διαστάσεις Μαζικών Δεδομένων, η Αυτό-αποτελεσματικότητα, η Συνάφεια Εργασίας, η Πολυπλοκότητα και το Άγχος. Αν και βιβλιογραφικά προκύπτει πως η χρήση της Τεχνολογίας Μαζικών Δεδομένων από τους λογιστές είναι σε πιο πρώιμο στάδιο από ότι θα έπρεπε, τα αποτελέσματα της έρευνας προμηνύουν την επικείμενη αλλαγή παρουσιάζοντας την τάση της πλειοψηφίας των Ελλήνων λογιστών να αποδέχεται τη χρήση της τεχνολογίας Μαζικών Δεδομένων.

Λέξεις κλειδιά: Μαζικά Δεδομένα, Έλληνες Λογιστές, TAM (Technology Acceptance Model), Πρόθεση Χρήσης

Abstract

The aim of this Thesis is to explore the relation between the use of Big Data technology and accounting, and the intention of Greek accountants to use this technology. This paper uses quantitative research, by collecting data based on a questionnaire. The creation of the questionnaire was based on TAM (Technology Acceptance Model) and its sending was carried out only by electronic means, throughout the Greek territory. In order for the goal of this Thesis to be achieved, linear regression analyzes were performed in order to control logical hypotheses. The results show that the Intention of Use of Big Data tools is explained to a large extent by Perceived Ease of Use and Perceived Usefulness, with the latter prevailing in the degree of influence. Other factors that have been considered and influence the Intention to Use Big Data technology are: Compatibility, Big Data Dimensions, Self-Efficacy, Job Relevance, Complexity and Anxiety. Although the literature shows that the use of Big Data technology by accountants is at an earlier stage than it should be, the results of the research foreshadow the forthcoming change presenting the tendency of the majority of Greek accountants to accept the use of Big Data technology.

Key words: Big Data, Greek Accountants, TAM (Technology Acceptance Model), Intention of Use

Εισαγωγή

Στο εξελιγμένο, πλέον, τεχνολογικό περιβάλλον όλα καταγράφονται, μετριοούνται και εντοπίζονται ψηφιακά. Η ανάγκη για επεξεργασία και διαχείριση αυτών των δεδομένων οδήγησε στην ανάπτυξη και εξέλιξη της τεχνολογίας των Μαζικών Δεδομένων (Big Data) (Laney, 2001).

Τα Μαζικά Δεδομένα είναι σύνολα δεδομένων που έχουν συλλεχθεί και είναι τόσο μεγάλα σε όγκο και περίπλοκα που απαιτούν νέες τεχνολογίες για την επεξεργασία τους. Τα δεδομένα μπορεί να προέρχονται από πολλές διαφορετικές πηγές. Συχνά είναι δεδομένα του ίδιου τύπου, για παράδειγμα, δεδομένα GPS που προέρχονται από εκατομμύρια κινητά τηλέφωνα και χρησιμοποιούνται για τον περιορισμό της κυκλοφοριακής συμφόρησης (Dogo et al., 2021). Ωστόσο μπορούμε να έχουμε και συνδυασμούς τύπων, όπως στην περίπτωση συλλογής δεδομένων από μητρώα υγείας και εφαρμογές εξυπηρέτησης ασθενών. Η τεχνολογία επιτρέπει τη συλλογή αυτών των δεδομένων με υψηλές ταχύτητες, σε σχεδόν πραγματικό χρόνο, και την ανάλυση τους για την απόκτηση νέων πληροφοριών. (www.europarl.europa.eu¹)

Σκοπός της παρούσας Διπλωματικής εργασίας είναι, αρχικά να παρουσιάσει τα Μαζικά Δεδομένα, τη χρησιμότητα αυτών και τους τομείς στους οποίους εφαρμόζονται, ενώ στη συνέχεια να παρουσιάσει τη σχέση της χρήσης τεχνολογίας Μαζικών Δεδομένων με τον κλάδο της Λογιστικής, εξετάζοντας τη στάση των Ελλήνων λογιστών απέναντί της και την πρόθεση χρήσης της εν λόγω τεχνολογίας από αυτούς. Τα πρώτα κεφάλαια αφιερώνονται στην επεξήγηση και κατανόηση του ορισμού και των χαρακτηριστικών της τεχνολογίας των Μαζικών Δεδομένων. Μέσω της βιβλιογραφίας αναλύονται οι προκλήσεις που δημιουργούνται από τη χρήση εργαλείων Μαζικών Δεδομένων, καθώς και η εφαρμογή τους στην καθημερινότητα, στους τομείς της Χρηματοοικονομικής Λογιστικής και της Ελεγκτικής. Παρόλη την αύξηση στην έρευνα των Μαζικών Δεδομένων και την έξαρση των δημοσιεύσεων το τελευταίο διάστημα (Abbasi et al., 2016; Akter & Wamba, 2016; Frizzo-Barker et al., 2016), εντύπωση προκαλεί πως κανένας δεν εστιάζει ειδικά στη λογιστική και τη χρηματοοικονομική. Παραδόξως, αφού οι συγκεκριμένοι κλάδοι παραδοσιακά ασχολούνται με μεγάλους όγκους δεδομένων στις αναλύσεις τους, η πραγματική χρήση των Μαζικών Δεδομένων θα περίμενε κανείς να μη βρίσκεται σε τόσο πρώιμο στάδιο (Jeacle & Carter, 2011; O'Shea & Shah, 2014; Scott & Orlikowski, 2012; Shen & Chen, 2018).

¹ <https://www.europarl.europa.eu/news/en/headlines/priorities/artificial-intelligence-in-the-eu/20210211STO97614/big-data-definition-benefits-challenges-infographics>

Η παρούσα εργασία παρουσιάζει έντονο ενδιαφέρον αν παρατηρήσουμε την ελληνική πραγματικότητα, όσον αφορά τους λογιστές (είτε ελεύθερους επαγγελματίες είτε μισθωτούς), και την τεχνολογική εξέλιξη στο επάγγελμά τους. Εύκολα θα μπορούσε κάποιος να ισχυριστεί πως η διαδικασία της όποιας αλλαγής στα τεχνολογικά μέσα που χρησιμοποιούν οι λογιστές χαρακτηρίζεται από έντονη δυσκαμψία, μεγάλες καθυστερήσεις κι αναβολές. Υπάρχει γενικότερα η τάση για άρνηση από τους ίδιους τους λογιστές, πολλές φορές και δικαιολογημένη, ως προς τη νέα τάξη πραγμάτων. Χαρακτηριστικό κι επίκαιρο παράδειγμα αποτελεί η έναρξη της υποχρεωτικής διαβίβασης παραστατικών στην πλατφόρμα Mydata που σίγουρα αποτελεί εργαλείο μαζικών δεδομένων . Η πρώτη νομοθετική ρύθμιση όπου αναφέρεται η υποχρέωση της ηλεκτρονικής διαβίβασης πληροφοριών δημοσιεύτηκε τον Δεκέμβριο του 2019 με τον Ν. 4646/2019 (ΦΕΚ Α' 201/12-12-2019) (Προσθήκη νέου άρθρου στο Νόμο 4174/2013). Σήμερα, 2 χρόνια μετά, κι αφού οι αναβολές διαδέχθηκαν η μία την άλλη με πιο πρόσφατη την αναβολή έως 01/10/2021 (Taxheaven, 2021²), η λειτουργία της εν λόγω ρύθμισης τέθηκε σε εφαρμογή χωρίς ωστόσο να βεβαιώνονται πρόστιμα έως τέλος του έτους 2021, δίνοντας έτσι ακόμα μία άτυπη παράταση. Τέτοια γεγονότα από μόνα τους αυξάνουν τη σημασία διερεύνησης του συγκεκριμένου ζητήματος πρόθεσης χρήσης τεχνολογίας μαζικών δεδομένων από τον συγκεκριμένο επαγγελματικό κλάδο.

Για την επίτευξη του σκοπού της παρούσας εργασίας, χρησιμοποιήθηκαν τα αποτελέσματα που προέκυψαν από έρευνα με τη χρήση ερωτηματολογίου, η διάθεση του οποίου πραγματοποιήθηκε μόνο με ηλεκτρονικά μέσα, σε όλη την ελληνική επικράτεια. Στο Κεφάλαιο 3, εκτός από την παρουσίαση της μεθοδολογίας, παρουσιάζονται τα TAM (Technology Acceptance Model) ως ένα ισχυρό, στιβαρό και συνήθως εφαρμόσιμο μοντέλο για την πρόβλεψη και την εξήγηση της συμπεριφοράς των χρηστών και ως προς τη χρήση τεχνολογιών πληροφορικής (Agarwal & Prasad, 1999; Davis et al., 1989; Legris et al., 2003). Η δημιουργία του ερωτηματολογίου της παρούσας εργασίας, έγινε με τη χρήση ενός πρόσφατα δημοσιευμένου μοντέλου TAM του Okcu και των συνεργατών (2019). Η συμβολή της παρούσας εργασίας, σε σχέση με παρεμφερή έργα, είναι η εφαρμογή του συγκεκριμένου μοντέλου σε διαφορετική χώρα (στην περίπτωση μας Ελλάδα) και επαγγελματικό κλάδο (Λογιστές) και η διερεύνηση της αποτελεσματικότητάς του.

Με βάση τα παραπάνω, τα ερευνητικά ερωτήματα της παρούσας μελέτης διατυπώνονται ως εξής:

² <https://www.taxheaven.gr/news/55122/platforma-mydata-kai-fhm-oi-nees-exelixeis?output=printer>

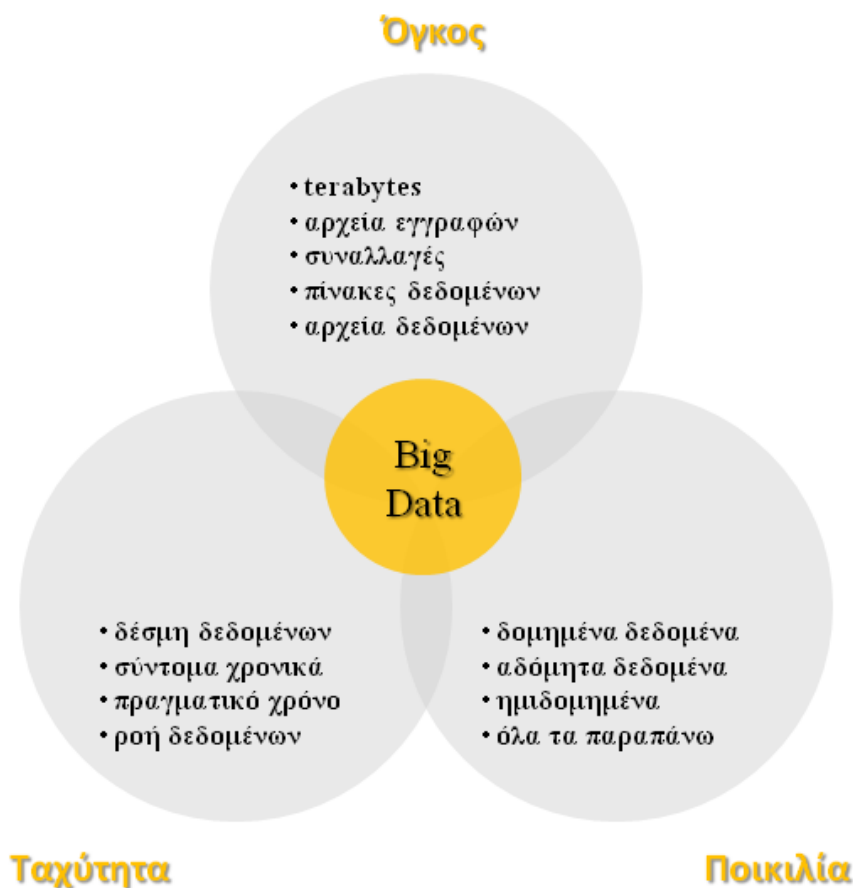
- Ποια είναι η επίδραση της Αντιλαμβανόμενης Χρησιμότητας και της Αντιλαμβανόμενης Ευκολίας Χρήσης, στην Πρόθεση Χρήσης;
- Ποια είναι η επίδραση των Διαστάσεων Μεγάλων Δεδομένων, της Αντιλαμβανόμενης Ευκολίας Χρήσης, της Συνάφειας Εργασίας και της Συμβατότητας, στην Αντιλαμβανόμενη Χρησιμότητα;
- Ποια είναι η επίδραση της Αυτό-αποτελεσματικότητας και του Άγχους, στην Αντιλαμβανόμενη Ευκολία Χρήσης

Στο Κεφάλαιο 4, συναντώνται τα αποτελέσματα της στατιστικής ανάλυσης. Πιο συγκεκριμένα, στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα των γραμμικών παλινδρομήσεων που διενεργήθηκαν με στόχο την απάντηση των ερευνητικών ερωτημάτων.

Στο Κεφάλαιο 5 παρουσιάζονται τα συμπεράσματα που προκύπτουν, προτάσεις για βελτίωση, καθώς και θέματα για περαιτέρω διερεύνηση.

1. Εισαγωγή στα Big Data

Οι περισσότεροι ορισμοί των Big Data εστιάζουν στο μέγεθος των δεδομένων ή στην αποθήκευση. Το μέγεθος έχει σημασία, αλλά υπάρχουν και άλλα σημαντικά χαρακτηριστικά αυτών, όπως η ποικιλία «*variety*» και η ταχύτητα «*velocity*» των δεδομένων (Diebold, 2012). Τα τρία V's των Big Data, όπως έχουν ονομαστεί στην αγγλική (volume, variety, and velocity), αποτελούν έναν ολοκληρωμένο ορισμό και καταρρίπτουν τον μύθο ότι τα Big Data αφορούν μόνο τον όγκο των δεδομένων (Emmanuel & Stanier, 2016). Επιπλέον, κάθε ένα από τα τρία V's έχει τις δικές του προεκτάσεις στα αναλυτικά στοιχεία (**Εικόνα 1**). Ωστόσο ο συγκεκριμένος κλάδος είναι δυναμικός και συνεχώς εξελίσσεται φθάνοντας στον ορισμό τεσσάρων και ακόλουθα των πέντε V's σήμερα (**βλέπε παράγραφο 1.1**)



Εικόνα 1, Αναλυτική παρουσίαση των τριών V's στα Big Data

Ο όγκος δεδομένων ως καθοριστικό χαρακτηριστικό των Big Data

Είναι προφανές ότι ο όγκος δεδομένων είναι το κύριο χαρακτηριστικό των Big Data. Έχοντας αυτό υπόψη, οι περισσότεροι άνθρωποι τα προσδιορίζουν σε terabytes ή μερικές φορές petabytes. Για παράδειγμα, ένας αριθμός χρηστών που πήραν συνέντευξη από τον

οργανισμό Transforming Data With Intelligence (TDWI) διαχειρίζονται 3 έως 10 terabyte (TB) δεδομένων για αναλυτικά στοιχεία. Ωστόσο άλλοι τρόποι ποσοτικοποίησης Big Data μπορεί να επιτευχθεί μετρώντας εγγραφές, συναλλαγές, πίνακες ή αρχεία. Ορισμένοι οργανισμοί θεωρούν πιο χρήσιμο να ποσοτικοποιήσουν τα Big Data από την άποψη του χρόνου. Για παράδειγμα, λόγω των περιορισμών του επταετούς καταστατικού στις ΗΠΑ, πολλές εταιρείες που εδρεύουν εκεί προτιμούν να διατηρούν επτά χρόνια δεδομένων διαθέσιμα για κινδύνους, συμμόρφωση και/ή νομική ανάλυση (Diebold, 2012; Sagiroglu & Sinanc, 2013).

Το εύρος των Big Data επηρεάζει επίσης την ποσοτικοποίησή τους. Για παράδειγμα, σε πολλούς οργανισμούς, τα δεδομένα που συλλέγονται για γενική αποθήκευση δεδομένων διαφέρουν από τα δεδομένα που συλλέγονται ειδικά για αναλυτικά στοιχεία. Διαφορετικές μορφές αναλυτικών στοιχείων μπορεί να έχουν διαφορετικά σύνολα δεδομένων. Ορισμένες πρακτικές ανάλυσης οδηγούν έναν επιχειρηματικό αναλυτή ή παρόμοιο χρήστη να δημιουργήσει *ad hoc* αναλυτικά σύνολα δεδομένων ανά αναλυτικό έργο. Στη συνέχεια υπάρχει ολόκληρη η επιχείρηση, η οποία συνολικά έχει τη δική της, ακόμη μεγαλύτερη εμβέλεια Big Data. Επιπλέον, κάθε στιγμή όλες αυτές οι ποσοτικοποιήσεις των Big Data μεταβάλλονται αυξητικά συνεχώς. Όλα αυτά τα καθιστούν έναν δυναμικό/κινούμενο στόχο που είναι δύσκολο να ποσοτικοποιηθεί (Tsai et al., 2015).

Η ποικιλομορφία δεδομένων ως χαρακτηριστικά προσδιορισμού των Big Data

Ένα από τα πράγματα που κάνουν τα Big Data να είναι πολύ μεγάλα είναι ότι προέρχονται από μεγαλύτερη ποικιλία πηγών από ποτέ. Πολλές από τις νεότερες πηγές εντοπίζονται στο διαδίκτυο, συμπεριλαμβανομένων αρχείων καταγραφής, ροών κλικ και κοινωνικών μέσων. Οργανισμοί που συλλέγουν δεδομένα Ιστού, για χρόνια εφαρμόζουν ένα γενικό είδος συσσώρευσης (Drosou et al., 2017). Έχουμε δει παρόμοια ανεκμετάλλευτα δεδομένα που συλλέγονται και αποθηκεύονται, όπως δεδομένα RFID από εφαρμογές αλυσίδας εφοδιασμού, δεδομένα κειμένου από εφαρμογές τηλεφωνικών κέντρων, ημιδομημένα δεδομένα από διάφορες διεργασίες από επιχείρηση σε επιχείρηση και γεωχωρικά δεδομένα στην εφοδιαστική. Αυτό που έχει αλλάξει είναι ότι μεγαλύτερος αριθμός χρηστών πλέον αναλύουν τα Big Data σε συνεχή ροή αντί να τα αποθηκεύουν (Tsai et al., 2015). Οι λίγοι οργανισμοί που έχουν αναλύσει αυτά τα δεδομένα το κάνουν τώρα σε πιο περίπλοκο και εξελιγμένο επίπεδο. Τα Big Data δεν είναι καινούργια, αλλά η αποτελεσματική αναλυτική μόχλευσή τους είναι.

Η πρόσφατη αξιοποίηση αυτών των πηγών για αναλυτικά στοιχεία σημαίνει ότι τα λεγόμενα δομημένα δεδομένα (τα οποία προηγουμένως είχαν μη αμφισβητούμενη ηγεμονία στα αναλυτικά στοιχεία) (Mohammed et al., 2016) ενώνονται πλέον με μη δομημένα δεδομένα (π.χ. κείμενο και ανθρώπινη γλώσσα) (Eberendu, 2016) και ημιδομημένα δεδομένα (π.χ. XML, RSS feeds) (Mohammed et al., 2016). Υπάρχουν επίσης δεδομένα που είναι δύσκολο να κατηγοριοποιηθούν, καθώς προέρχονται από ήχο, βίντεο και άλλες συσκευές. Επιπλέον, πολυδιάστατα δεδομένα μπορούν να αντληθούν από μια αποθήκη δεδομένων για να προσδώσουν ακόμη και ιστορικό πλαίσιο στα Big Data. Αυτός είναι ένας πολύ πιο εκλεκτικός συνδυασμός τύπων δεδομένων από ότι έχει δει ποτέ το επιστημονικό πεδίο της ανάλυσης δεδομένων. Έτσι, με τα Big Data η ποικιλία είναι εξίσου μεγάλη με τον όγκο. Επιπλέον, η ποικιλία και ο όγκος τείνουν να τροφοδοτούν ο ένας τον άλλον.

Η ταχύτητα δεδομένων ως χαρακτηριστικό προσδιορισμού των Big Data.

Τα Big Data μπορούν να περιγραφούν από την ταχύτητα, αν θεωρηθούν ως προς τη συχνότητα παραγωγής δεδομένων ή τη συχνότητα παράδοσης δεδομένων. Για παράδειγμα, ας σκεφτούμε τη ροή δεδομένων που προέρχονται από οποιοδήποτε είδος συσκευής ή αισθητήρα, ας πούμε ρομποτικές μηχανές κατασκευής, θερμομέτρα θερμοκρασίας ανίχνευσης, μικρόφωνα που ακούνε κίνηση σε ασφαλή περιοχή ή βιντεοκάμερες σάρωσης για συγκεκριμένο πρόσωπο σε πλήθος. Η συλλογή μεγάλων δεδομένων σε πραγματικό χρόνο δεν είναι νέα ιδέα, πολλές εταιρείες συλλέγουν δεδομένα ροής κλικ από ιστότοπους για πολλά χρόνια, χρησιμοποιώντας δεδομένα ροής για να κάνουν προτάσεις αγορών σε επισκέπτες του διαδικτύου (Zheng et al., 2015). Με τα δεδομένα από αισθητήρες και του διαδικτύου να ανανεώνονται ασταμάτητα σε πραγματικό χρόνο, οι όγκοι των δεδομένων αυξάνονται εκθετικά. Ακόμα πιο δύσκολο, τα αναλυτικά στοιχεία που ακολουθούν τη ροή δεδομένων πρέπει να κατανοηθούν σε πραγματικό χρόνο και ενδεχομένως να χρειάζεται ανάληψη δράσης.

Η αυγή της εποχής των Big Data

Τη τελευταία εικοσαετία, τα δεδομένα έχουν αυξηθεί σε μεγάλη κλίμακα σε διάφορους τομείς. Σύμφωνα με έκθεση της Εταιρίας Διεθνών Δεδομένων (International Data Corporation, IDC) μόνο το 2011, ο συνολικός όγκος δεδομένων που δημιουργήθηκε και αντιγράφηκε στον κόσμο ήταν 1,8ZB ($\approx 1021B$), ο οποίος αυξήθηκε σχεδόν εννέα φορές μέσα σε πέντε χρόνια (Gantz & Reinsel, 2011). Οι προβλέψεις λένε ότι αυτός ο αριθμός α διπλασιάζεται τουλάχιστον κάθε δύο χρόνια στο εγγύς μέλλον (Gantz & Reinsel, 2011).

Με την εκρηκτική αύξηση των παγκόσμιων δεδομένων, ο όρος Big Data χρησιμοποιείται κυρίως για την περιγραφή τεράστιων συνόλων δεδομένων. Σε σύγκριση με τα παραδοσιακά σύνολα δεδομένων, τα Big Data περιλαμβάνουν συνήθως μάζες μη δομημένων δεδομένων που χρειάζονται περισσότερη ανάλυση σε πραγματικό χρόνο. Επιπλέον, το Big Data προσφέρει επίσης νέες ευκαιρίες για την ανακάλυψη νέων τιμών, μας βοηθά να αποκτήσουμε μια εις βάθος κατανόηση των κρυφών τιμών και επίσης αντιμετωπίζει νέες προκλήσεις, π.χ. πώς να οργανώσουμε και να διαχειριστούμε αποτελεσματικά τέτοια σύνολα δεδομένων.

Πρόσφατα ο βιομηχανικός κλάδος έδειξε το ενδιαφέρον του για το υψηλό δυναμικό των Big Data και πολλές κυβερνητικές υπηρεσίες ανακοίνωσαν σημαντικά σχέδια για την επιτάχυνση της έρευνας και των εφαρμογών στα Big Data (White House, 2012). Επιπλέον, ζητήματα σχετικά με τα Big Data καλύπτονται συχνά σε δημόσια μέσα όπως ο *Economist* (Cukier, 2010; Labrinidis & Jagadish, 2012), οι *New York Times* (Lohr, 2012), και το *National Public Radio* των ΗΠΑ (Noguchi, 2011a, 2011b). Δύο κορυφαία επιστημονικά περιοδικά όπως το *Nature* και το *Science*, άνοιξαν επίσης ειδικές στήλες για να συζητούνται οι προκλήσεις και επιπτώσεις των Big Data (Frankel & Reid, 2008; 2011). Η εποχή των Big Data έχει έλθει χωρίς καμία αμφιβολία (Manyika, 2011).

Σήμερα, τα Big Data που σχετίζονται με την υπηρεσία των εταιρειών Διαδικτύου αναπτύσσονται ραγδαία. Για παράδειγμα, η Google επεξεργάζεται δεδομένα εκατοντάδων Petabyte (PB), το Facebook δημιουργεί δεδομένα καταγραφής άνω των 10 PB ανά μήνα, η Baidu, μια κινεζική εταιρεία, επεξεργάζεται δεδομένα δεκάδων PB και η Taobao, θυγατρική της Alibaba, δημιουργεί δεδομένα δεκάδων της Terabyte (TB) για online συναλλαγές ανά ημέρα (Padgavankar & Gupta, 2014). Ενώ η ποσότητα των μεγάλων συνόλων δεδομένων αυξάνεται δραστικά, δημιουργεί επίσης πολλά προκλητικά προβλήματα που απαιτούν άμεσες λύσεις:

- Οι τελευταίες εξελίξεις στην πληροφορική (information technology, IT) το καθιστούν ευκολότερο να δημιουργούνται δεδομένα. Για παράδειγμα, κατά μέσο όρο, 72 ώρες βίντεο μεταφορτώνονται στο YouTube κάθε λεπτό (Mayer-Schönberger & Cukier, 2013). Επομένως, είμαστε αντιμετώπι με την κύρια πρόκληση της συλλογής και ολοκλήρωσης τεράστιων δεδομένων από ευρέως κατανεμημένες πηγές
- Η ταχεία ανάπτυξη του «*Cloud Computing*» και του «*Internet of Things*» (IoT) προωθούν περαιτέρω την απότομη ανάπτυξη των δεδομένων. Το *Cloud Computing*

χρησιμοποιείται για παροχή προστασίας, πρόσβαση σε ιστότοπους και κανάλια για στοιχεία δεδομένων. Το IoT συναντάται σε αισθητήρες παγκοσμίως όπου συλλέγονται και μεταδίδονται δεδομένα για αποθήκευση και επεξεργασία στο cloud. Τέτοια δεδομένα τόσο σε ποσότητα όσο και σε αμοιβαίες σχέσεις θα ξεπεράσουν κατά πολύ τις ικανότητες των αρχιτεκτονικών πληροφορικής και της υποδομής των υπάρχουσών επιχειρήσεων, και η απαίτηση σε πραγματικό χρόνο θα τονίσει επίσης πολύ τη διαθέσιμη υπολογιστική ικανότητα. Τα ολοένα αυξανόμενα δεδομένα προκαλούν προβληματισμό στον τρόπο που θα αποθηκεύονται και θα διαχειρίζονται τόσο τεράστια ετερογενή σύνολα δεδομένων με υψηλές απαιτήσεις σε υποδομή υλικού και λογισμικού

- Λαμβάνοντας υπόψη την ετερογένεια, την επεκτασιμότητα, τον πραγματικό χρόνο, την πολυπλοκότητα και το απόρρητο των Big Data, θα πρέπει να συλλέγονται αποτελεσματικά τα σύνολα δεδομένων σε διαφορετικά επίπεδα κατά τη διάρκεια της ανάλυσης, της μοντελοποίησης, της οπτικοποίησης και της πρόβλεψης, ώστε να αποκαλύπτεται η εγγενής πληροφορία τους και να βελτιώνει τη λήψη αποφάσεων (J. Chen et al., 2013).

1. 1. Ορισμός και χαρακτηριστικά των Big Data

Τα Big Data επί της ουσίας μπορεί να θεωρηθεί ότι είναι μια αφηρημένη έννοια. Εκτός από τις μάζες δεδομένων, έχει επίσης κάποια άλλα χαρακτηριστικά, τα οποία καθορίζουν τη διαφορά μεταξύ αυτών και των «μαζικών δεδομένων» ή «πολύ μεγάλων δεδομένων». Προς το παρόν, αν και η σημασία των Big Data έχει αναγνωριστεί γενικά, οι άνθρωποι εξακολουθούν να έχουν διαφορετικές απόψεις σχετικά με τον ορισμό τους. Σε γενικές γραμμές, τα Big Data θα σημαίνουν σύνολα δεδομένων που δεν μπορούσαν να γίνουν αντιληπτά, να ληφθούν, να διαχειριστούν και να υποβληθούν σε επεξεργασία με παραδοσιακά εργαλεία πληροφορικής και λογισμικού εντός ανεκτού χρόνου. Λόγω διαφορετικών ανησυχιών, οι επιστημονικές και τεχνολογικές επιχειρήσεις, οι ερευνητές, οι αναλυτές δεδομένων και οι τεχνικοί έχουν διαφορετικούς ορισμούς για τα Big Data. Οι ακόλουθοι ορισμοί μπορεί να μας βοηθήσουν να κατανοήσουμε καλύτερα τις βαθιές κοινωνικές, οικονομικές και τεχνολογικές παραμέτρους των Big Data.

Το 2010, το ίδρυμα λογισμικού Apache Hadoop³ καθόρισε τα Big Data ως «σύνολα δεδομένων που δεν μπορούσαν να συλληφθούν, να διαχειριστούν και να υποστούν επεξεργασία από γενικούς υπολογιστές εντός αποδεκτού πεδίου». Με βάση αυτόν τον ορισμό, τον Μάιο του 2011, η McKinsey & Company, μια παγκόσμια εταιρεία συμβούλων ανακοίνωσε τα Big Data ως τα επόμενα σύνορα για την καινοτομία, τον ανταγωνισμό και την παραγωγικότητα (Manyika, 2011). Τα Big Data θα σημαίνουν τέτοια σύνολα δεδομένων που δεν θα μπορούσαν να αποκτηθούν, να αποθηκευτούν και να διαχειριστούν από κλασικό λογισμικό βάσης δεδομένων. Αυτός ο ορισμός περιλαμβάνει δύο έννοιες: Πρώτον, οι όγκοι των συνόλων δεδομένων που συμμορφώνονται με το πρότυπο των Big Data αλλάζουν και ενδέχεται να αυξηθούν με την πάροδο του χρόνου ή με την τεχνολογική πρόοδο. Δεύτερον, οι τόμοι συνόλων δεδομένων που συμμορφώνονται με το πρότυπο Big Data σε διαφορετικές εφαρμογές διαφέρουν μεταξύ τους. Προς το παρόν, τα Big Data κυμαίνονται γενικά από πολλά TB έως αρκετά PB (Manyika, 2011). Από τον ορισμό της McKinsey & Company, φαίνεται ότι ο όγκος ενός συνόλου δεδομένων δεν είναι το μόνο κριτήριο για τα Big Data. Η αυξανόμενη κλίμακα δεδομένων και η διαχείρισή τους που δεν μπορούν να αντιμετωπιστούν από τις παραδοσιακές τεχνολογίες βάσεων δεδομένων είναι τα επόμενα δύο βασικά χαρακτηριστικά.

Στην πραγματικότητα τα Big Data είχαν οριστεί ήδη από το 2001, όπου ο Doug Laney, αναλυτής της META (γνωστή πλέον ως Gartner) καθόρισε προκλήσεις και ευκαιρίες που προκύπτουν από αυξημένα δεδομένα με ένα μοντέλο των τριών V's, δηλαδή την αύξηση του

³ <https://hadoop.apache.org/>

όγκου, της ταχύτητας, και της ποικιλίας σε μια ερευνητική έκθεση (Laney, 2001). Αν και ένα τέτοιο μοντέλο δεν χρησιμοποιήθηκε αρχικά για τον ορισμό των Big Data, η Gartner και πολλές άλλες επιχειρήσεις, συμπεριλαμβανομένης της IBM (Zikopoulos & Eaton, 2011) και ορισμένα ερευνητικά τμήματα της Microsoft (Meijer, 2011) ακόμα χρησιμοποιούν το μοντέλο των τριών V's για να περιγράψουν τα Big Data και την εξέλιξή τους στα επόμενα δέκα χρόνια (Beyer, 2011). Στο μοντέλο των τριών V's, το *Volume* σημαίνει ότι, με τη δημιουργία και τη συλλογή μάζας δεδομένων, η κλίμακα δεδομένων γίνεται όλο και μεγαλύτερη. Για το *Velocity* σημαίνει ότι η επικαιρότητα των Big Data, συγκεκριμένα, η συλλογή και ανάλυση δεδομένων κ.λπ. πρέπει να πραγματοποιείται γρήγορα και έγκαιρα, ώστε να αξιοποιείται στο μέγιστο η εμπορική αξία των Big Data. Για το *Variety* δείχνει τους διάφορους τύπους δεδομένων, που περιλαμβάνουν ημι-δομημένα και μη δομημένα δεδομένα, όπως ήχο, βίντεο, ιστοσελίδα και κείμενο, καθώς και παραδοσιακά δομημένα δεδομένα.

Ωστόσο, άλλοι έχουν διαφορετικές απόψεις, συμπεριλαμβανομένης της εταιρίας International Data Corporation (IDC), ενός από τους πιο σημαντικούς ηγέτες στα Big Data και στους ερευνητικούς του τομείς. Το 2011, μια αναφορά της IDC καθόρισε τα Big Data ως «Οι τεχνολογίες Big Data περιγράφουν μια νέα γενιά τεχνολογιών και αρχιτεκτονικών, σχεδιασμένων να εξάγουν οικονομική αξία από πολύ μεγάλους όγκους και μια μεγάλη ποικιλία δεδομένων, επιτρέποντας τη λήψη, την ανακάλυψη και την υψηλή ταχύτητα ή ανάλυση» (Gantz & Reinsel, 2011). Με αυτόν τον ορισμό, τα χαρακτηριστικά των Big Data μπορούν να συνοψιστούν πλέον ως τέσσερα V's, δηλαδή τα κλασικά *Volume*, *Variety*, *Velocity* και πλέον με την εισαγωγή του *Value* (αξία). Αυτός ο ορισμός των τεσσάρων V's αναγνωρίστηκε ευρέως, καθώς επισημαίνει τη σημασία και την αναγκαιότητα των Big Data, δηλαδή την εξερεύνηση των τεράστιων κρυφών τιμών. Αυτός ο ορισμός υποδεικνύει το πιο κρίσιμο πρόβλημα στα Big Data, το οποίο είναι πώς να ανακαλυφθούν τιμές από σύνολα δεδομένων με τεράστια κλίμακα, διάφοροι τύποι και ταχεία παραγωγή. Όπως είπε ο Jay Parikh, Αναπληρωτής Διευθύνων Σύμβουλος του Facebook, «Θα μπορούσατε να διαθέσετε μια δέσμη δεδομένων μόνο εκτός από τα Big Data, εάν δεν χρησιμοποιείτε τα δεδομένα που συλλέγονται» (Mayer-Schönberger & Cukier, 2013). Η μετεξέλιξη των τεσσάρων V's δεν άργησε να έρθει και πλέον είναι πέντε V's όταν προστεθεί και η ακρίβεια (*Veracity*) (Anuradha, 2015), αλλάζοντας ουσιαστικά τον αρχικό ορισμό των Big Data (Laney, 2001).

Επιπλέον, το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας των ΗΠΑ (National Institute of Standards and Technology, NIST)⁴ ορίζει πως τα Big Data αποτελούν δεδομένα για

⁴ <https://www.nist.gov/>

τα οποία ο όγκος τους, η ταχύτητα απόκτησης ή η αναπαράσταση δεδομένων περιορίζει την ικανότητα χρήσης παραδοσιακών σχεσιακών μεθόδων. Αυτό δείχνει ότι πρέπει να αναπτυχθούν και να χρησιμοποιηθούν νέες αποτελεσματικές μέθοδοι ή τεχνολογίες για την ανάλυση και επεξεργασία των Big Data.

Έχουν γίνει σημαντικές συζητήσεις τόσο από τη βιομηχανία όσο και από τον ακαδημαϊκό χώρο σχετικά με τον ορισμό των Big Data (Grobelnik, 2012; Team, 2011). Εκτός από την ανάπτυξη ενός σωστού ορισμού, η έρευνα στα Big Data θα πρέπει επίσης να επικεντρωθεί στον τρόπο εξαγωγής της αξίας τους, στον τρόπο χρήσης δεδομένων και στον τρόπο μετατροπής «μιας δέσμης δεδομένων» σε Big Data.

Η αξία των Big Data

Η McKinsey & Company παρατήρησε πώς τα Big Data δημιούργησαν αξία μετά από εμπεριστατωμένη έρευνα σχετικά με την υγειονομική περίθαλψη των ΗΠΑ, τη δημόσια διοίκηση της ΕΕ, το λιανικό εμπόριο των ΗΠΑ, την παγκόσμια κατασκευή και τα παγκόσμια προσωπικά δεδομένα τοποθεσίας. (Manyika, 2011) Μέσα από έρευνα σχετικά με τους πέντε βασικούς κλάδους που αντιπροσωπεύουν την παγκόσμια οικονομία, η έκθεση McKinsey επεσήμανε ότι τα Big Data μπορούν να δώσουν πλήρη ρόλο στην οικονομική λειτουργία, να βελτιώσουν την παραγωγικότητα και την ανταγωνιστικότητα των επιχειρήσεων και του δημόσιου τομέα και να δημιουργήσουν τεράστια οφέλη για τους καταναλωτές. Σε πρόσφατη έρευνα (Manyika, 2011), η McKinsey συνοψίζει τις αξίες που θα μπορούσε να δημιουργήσουν τα Big Data: εάν αυτά μπορούσαν να χρησιμοποιηθούν δημιουργικά και αποτελεσματικά για τη βελτίωση της αποδοτικότητας και της ποιότητας, η πιθανή αξία της ιατρικής βιομηχανίας των ΗΠΑ που αποκτήθηκε μέσω δεδομένων μπορεί να ξεπεράσει τα 300 δισεκατομμύρια δολάρια ΗΠΑ, μειώνοντας έτσι τις δαπάνες για την υγειονομική περίθαλψη των ΗΠΑ σε ποσοστό άνω του 8%, ενώ οι λιανοπωλητές που χρησιμοποιούν πλήρως τα Big Data ενδέχεται να βελτιώσουν το κέρδος τους κατά περισσότερο από 60%. Τα Big Data μπορούν επίσης να χρησιμοποιηθούν για τη βελτίωση της αποτελεσματικότητας των κυβερνητικών ενεργειών, έτσι ώστε οι ανεπτυγμένες οικονομίες στην Ευρώπη να μπορούν να εξοικονομήσουν πάνω από 100 δισεκατομμύρια ευρώ (αποκλείοντας την επίδραση μιας μείωσης στις απάτες, των σφαλμάτων και της φορολογικής διαφοράς).

Κατά τη διάρκεια της πανδημίας γρίπης του 2009, η Google έλαβε έγκαιρες πληροφορίες αναλύοντας τα Big Data, τα οποία παρείχαν ακόμη πιο πολύτιμες πληροφορίες από αυτές που παρείχαν τα κέντρα πρόληψης ασθενειών. Σχεδόν όλες οι χώρες χρειάζονταν να

αποκτήσουν υπηρεσίες νοσοκομειακής ενημέρωσης συνδεδεμένες με τα κέντρα πρόληψης ασθενειών με άμεση ενημέρωση πάνω στο νέο είδος κρουσμάτων γρίπης. Ωστόσο, οι ασθενείς συνήθως δεν επισκέπτονταν αμέσως γιατρούς όταν μολύνονταν. Χρειάστηκε επίσης αρκετός χρόνος για την αποστολή των πληροφοριών από νοσοκομεία στα κέντρα πρόληψης ασθενειών, και για τα κέντρα πρόληψης ασθενειών να αναλύσουν και να συνοψίσουν τις πληροφορίες. Επομένως, όταν το κοινό γνώρισε την πανδημία του νέου τύπου γρίπης, η ασθένεια μπορεί να είχε ήδη εξαπλωθεί για μία έως δύο εβδομάδες. Η Google διαπίστωσε ότι κατά τη διάρκεια της εξάπλωσης της γρίπης του 2009, οι όροι αναζήτησης ήταν συχνά διαφορετικές από αυτές των συνηθισμένων χρόνων και οι συχνότητες χρήσης των καταχωρήσεων συσχετίστηκαν με την εξάπλωση της γρίπης τόσο σε χρόνο όσο και σε τοποθεσία. Η Google βρήκε 45 ομάδες καταχώρησης αναζήτησης που ήταν πολύ σχετικές με το ξέσπασμα της γρίπης και τις ενσωμάτωσε σε συγκεκριμένα μαθηματικά μοντέλα για να προβλέψει την εξάπλωση της γρίπης και ακόμη και να προβλέψει μέρη από όπου εξαπλώθηκε η γρίπη. Τα σχετικά ερευνητικά αποτελέσματα έχουν δημοσιευτεί στο Nature (Ginsberg et al., 2009).

Το 2008, η Microsoft αγόρασε τη Farecast, μια εταιρεία επιχειρηματικών δραστηριοτήτων επιστημονικής τεχνολογίας στις ΗΠΑ. Η οποία διέθετε ένα σύστημα προβλέψεων εισιτηρίων αεροπορικών εταιρειών που προέβλεπε τις τάσεις και τα αυξανόμενα / μειωμένα εύρη τιμών εισιτηρίων των αεροπορικών εταιρειών. Το σύστημα έχει ενσωματωθεί στη μηχανή αναζήτησης Bing της Microsoft. Μέχρι το 2012, το σύστημα έχει εξοικονομήσει σχεδόν 50 USD ανά εισιτήριο ανά επιβάτη, με την προβλεπόμενη ακρίβεια να φτάνει το 75%.

Επί του παρόντος, τα δεδομένα έχουν γίνει ένας σημαντικός παράγοντας παραγωγής που θα μπορούσε να είναι συγκρίσιμος με τα υλικά περιουσιακά στοιχεία και το ανθρώπινο κεφάλαιο. Καθώς αναπτύσσονται τα πολυμέσα, τα κοινωνικά μέσα και το IoT, οι επιχειρήσεις θα συλλέγουν περισσότερες πληροφορίες, οδηγώντας σε εκθετική αύξηση του όγκου δεδομένων. Τα Big Data θα έχουν τεράστια και αυξανόμενη δυνατότητα δημιουργίας αξιών για επιχειρήσεις και καταναλωτές.

Η ανάπτυξη των Big Data

Στα τέλη της δεκαετίας του 1970, προέκυψε η έννοια του «μηχανή βάσης δεδομένων», η οποία είναι μια τεχνολογία που χρησιμοποιείται ειδικά για την αποθήκευση και την ανάλυση δεδομένων. Με την αύξηση του όγκου των δεδομένων, η χωρητικότητα αποθήκευσης και επεξεργασίας ενός μοναδικού συστήματος κεντρικού υπολογιστή έγινε ανεπαρκής. Στη δεκαετία του 1980, οι άνθρωποι πρότειναν το «τίποτα κοινόχρηστο», ένα παράλληλο σύστημα

βάσης δεδομένων, για να καλύψουν τη ζήτηση του αυξανόμενου όγκου δεδομένων (DeWitt & Gray, 1992). Η αρχιτεκτονική του συστήματος «τίποτα κοινόχρηστο» δεν βασίζεται στη χρήση συμπλέγματος και κάθε μηχανή έχει τον δικό του επεξεργαστή, χώρο αποθήκευσης και δίσκο. Το σύστημα Teradata ήταν το πρώτο επιτυχημένο σύστημα παράλληλης εμπορικής βάσης δεδομένων. Τέτοιες βάσεις δεδομένων έγιναν πολύ δημοφιλής τελευταία. Η ημερομηνία 2 Ιουνίου 1986 θεωρείται ορόσημο καθώς η Teradata παρέδωσε το πρώτο παράλληλο σύστημα βάσης δεδομένων με χωρητικότητα αποθήκευσης 1TB στο Kmart για να βοηθήσει τη μεγάλη εταιρεία λιανικής στη Βόρεια Αμερική να επεκτείνει την αποθήκευση δεδομένων (Walter, 2009). Στα τέλη της δεκαετίας του 1990, τα πλεονεκτήματα της παράλληλης βάσης δεδομένων αναγνωρίστηκαν ευρέως στον τομέα των βάσεων δεδομένων.

Ωστόσο, η εφαρμογή τους έχουν ήδη επισημάνει και τις όποιες προκλήσεις των Big Data. Με την ανάπτυξη υπηρεσιών διαδικτύου, η αναζήτηση συγκεκριμένου περιεχομένου αυξήθηκε ραγδαία. Επομένως, οι εταιρείες μηχανών αναζήτησης έπρεπε να αντιμετωπίσουν τις προκλήσεις του χειρισμού τέτοιων Big Data. Η Google δημιούργησε το GFS (Ghemawat et al., 2003) και το MapReduce (Dean & Ghemawat, 2008) μοντέλα προγραμματισμού για την αντιμετώπιση των προκλήσεων που προκύπτουν από τη διαχείριση δεδομένων και την ανάλυση σε κλίμακα Διαδικτύου. Επιπλέον, τα περιεχόμενα που δημιουργούνται από χρήστες, αισθητήρες και άλλες πανταχού παρούσες πηγές δεδομένων παραμόρφωσαν επίσης τις συντριπτικές ροές δεδομένων, οι οποίες απαιτούσαν μια θεμελιώδη αλλαγή στην αρχιτεκτονική υπολογιστών και στον μηχανισμό επεξεργασίας δεδομένων μεγάλης κλίμακας. Τον Ιανουάριο του 2007, ο Jim Gray, πρωτοπόρος του λογισμικού βάσης δεδομένων, ονόμασε έναν τέτοιο μετασχηματισμό «Το τέταρτο παράδειγμα» (Hey et al., 2009). Σκέφτηκε επίσης ότι ο μόνος τρόπος για να αντιμετωπίσει αυτό το παράδειγμα ήταν να αναπτύξει μια νέα γενιά υπολογιστικών εργαλείων για τη διαχείριση, την οπτικοποίηση και την ανάλυση τεράστιων δεδομένων. Τον Ιούνιο του 2011, ακολούθησε ένα άλλο ορόσημο, η EMC/IDC δημοσίευσε μια ερευνητική έκθεση με τίτλο «Εξαγωγή αξιών από το χάος» (Gantz & Reinsel, 2011), που εισήγαγε την ιδέα και τις δυνατότητες των Big Data για πρώτη φορά. Αυτή η ερευνητική έκθεση πυροδότησε το μεγάλο ενδιαφέρον τόσο για τη βιομηχανία όσο και για τον ακαδημαϊκό χώρο για τα Big Data.

Τα τελευταία χρόνια, σχεδόν όλες οι μεγάλες εταιρείες, συμπεριλαμβανομένων των EMC, Oracle, IBM, Microsoft, Google, Amazon και Facebook κ.λπ. έχουν ξεκινήσει έρευνες στα Big Data. Λαμβάνοντας ως παράδειγμα την IBM, από το 2005, η IBM έχει επενδύσει 16 δισεκατομμύρια δολάρια ΗΠΑ σε 30 εξαγορές που σχετίζονται με τη Big Data. Στην ακαδημαϊκή κοινότητα, το Big Data ήταν επίσης στο προσκήνιο. Το 2008, η Nature δημοσίευσε

ένα ειδικό τεύχος Big Data. Το 2011, η Science ξεκίνησε επίσης ένα ειδικό τεύχος για τις βασικές τεχνολογίες της επεξεργασίας δεδομένων στα Big Data. Το 2012 το European Research Consortium for Informatics and Mathematics (ERCIM) News δημοσίευσε ένα ειδικό τεύχος για τα Big Data. Στις αρχές του 2012, μια έκθεση με τίτλο Big Data, Big Impact που παρουσιάστηκε στο φόρουμ Davos στην Ελβετία, ανακοίνωσε ότι το Big Data έχουν γίνει ένα νέο είδος οικονομικών περιουσιακών στοιχείων, όπως το νόμισμα ή ο χρυσός. Η Gartner, ένας διεθνής ερευνητικός οργανισμός, εξέδωσε Hype Cycles από το 2012 έως το 2013, το οποίο ταξινόμησε τον υπολογισμό Big Data, την κοινωνική ανάλυση και την αποθήκευση της ανάλυσης δεδομένων σε 48 αναδυόμενες τεχνολογίες που αξίζουν την περισσότερη προσοχή.

Πολλές εθνικές κυβερνήσεις, όπως των ΗΠΑ, έδωσαν επίσης μεγάλη προσοχή στα Big Data. Τον Μάρτιο του 2012, η κυβέρνηση Ομπάμα ανακοίνωσε μια επένδυση ύψους 200 εκατομμυρίων δολαρίων για να ξεκινήσει το «Big Data Research and Development Plan», το οποίο ήταν μια δεύτερη σημαντική πρωτοβουλία επιστημονικής και τεχνολογικής ανάπτυξης μετά την πρωτοβουλία «Information Highway» το 1993. Τον Ιούλιο του 2012, Το έργο «Vigorous ICT Japan» που εκδόθηκε από το Υπουργείο Εσωτερικών και Επικοινωνιών της Ιαπωνίας έδειξε ότι η ανάπτυξη των Big Data θα πρέπει να αποτελεί εθνική στρατηγική και οι τεχνολογίες εφαρμογών πρέπει να είναι το επίκεντρο. Τον Ιούλιο του 2012, τα Ηνωμένα Έθνη εξέδωσαν την έκθεση Big Data for Development, η οποία συνοψίζει τον τρόπο με τον οποίο οι κυβερνήσεις χρησιμοποίησαν τα Big Data για καλύτερη εξυπηρέτηση και προστασία των ανθρώπων τους.

1. 2. Προκλήσεις από τη χρήση των Big Data

Ο απότομα αυξανόμενος κατακλυσμός δεδομένων στην εποχή των Big Data δημιουργεί τεράστιες προκλήσεις στην απόκτηση, αποθήκευση, διαχείριση και ανάλυση δεδομένων. Τα παραδοσιακά συστήματα διαχείρισης και ανάλυσης δεδομένων βασίζονται σε συστήματα διαχειριστικών σχέσεων βάσεων δεδομένων (relational database management system, RDBMS). Ωστόσο, τέτοια RDBMS ισχύουν μόνο για δομημένα δεδομένα, εξαιρουμένων των ημι-δομημένων ή μη δομημένων δεδομένων. Επιπλέον, τα RDBMS χρησιμοποιούν όλο και πιο ακριβό υλικό. Είναι προφανές ότι τα παραδοσιακά RDBMS δεν μπορούν να διαχειριστούν τον τεράστιο όγκο και την ετερογένεια των Big Data. Η ερευνητική κοινότητα έχει προτείνει ορισμένες λύσεις από διαφορετικές οπτικές γωνίες. Για παράδειγμα, το cloud computing χρησιμοποιείται για την κάλυψη των απαιτήσεων σχετικά με την υποδομή για τα Big Data, π.χ. αποδοτικότητα κόστους, ελαστικότητα και ομαλή αναβάθμιση/υποβάθμιση. Για λύσεις μόνιμης αποθήκευσης και διαχείρισης μεγάλης κλίμακας διαταραγμένων συνόλων δεδομένων, καταναμημένων συστημάτων αρχείων (Howard et al., 1988) και NoSQL (Cattell, 2011) βάσεις δεδομένων είναι καλές επιλογές. Τέτοια πλαίσια προγραμματισμού αποδίδουν με μεγάλη επιτυχία την επεξεργασία ομαδοποιημένων εργασιών, ειδικά για την κατάταξη ιστοσελίδων. Επιπρόσθετα μπορούν να αναπτυχθούν διάφορες εφαρμογές Big Data με βάση αυτές τις καινοτόμες τεχνολογίες ή πλατφόρμες. Τέλος, δεν είναι ασήμαντο να αναπτυχθούν τα συστήματα ανάλυσης Big Data.

Συγκεκριμένα πολλές μελέτες (Chaudhuri et al., 2011; Labrinidis & Jagadish, 2012) αναλύουν τα κύρια εμπόδια στην ανάπτυξη εφαρμογών των Big Data. Οι βασικές προκλήσεις παρατίθενται ως εξής:

– *Απεικόνιση δεδομένων*: Πολλά σύνολα δεδομένων έχουν συγκεκριμένα επίπεδα ετερογένειας στον τύπο, τη δομή, τη σημασιολογία, την οργάνωση, την ευαισθησία και την προσβασιμότητα. Η αναπαράσταση δεδομένων στοχεύει να κάνει τα δεδομένα πιο ουσιαστικά για ανάλυση υπολογιστή και ερμηνεία χρήστη. Ωστόσο, μια ακατάλληλη απεικόνιση των δεδομένων θα μειώσει την αξία των αρχικών δεδομένων και μπορεί ακόμη και να εμποδίσει την αποτελεσματική ανάλυσή τους. Η αποτελεσματική αναπαράσταση δεδομένων αντικατοπτρίζει τη δομή, την τάξη και τον τύπο των δεδομένων, καθώς και τις ολοκληρωμένες τεχνολογίες, έτσι ώστε να είναι δυνατή η αποτελεσματική λειτουργία σε διαφορετικά σύνολα δεδομένων.

– *Μείωση πλεονασμού και συμπίεση δεδομένων*: Υπάρχει υψηλό επίπεδο πλεονασμού σε σύνολα δεδομένων. Κατά συνέπεια, η μείωση του πλεονασμού και η συμπίεση δεδομένων είναι αποτελεσματικά για τη μείωση του έμμεσου κόστους ολόκληρου του συστήματος με την

προϋπόθεση ότι δεν επηρεάζονται οι πιθανές τιμές των δεδομένων. Για παράδειγμα, τα περισσότερα δεδομένα που δημιουργούνται από δίκτυα αισθητήρων είναι εξαιρετικά περιττά, τα οποία μπορούν να φιλτραριστούν και να συμπιεστούν σε τάξεις μεγέθους.

– *Διαχείριση κύκλου ζωής δεδομένων*: Σε σύγκριση με τις σχετικά αργές εξελίξεις των συστημάτων αποθήκευσης, η διεισδυτική ανίχνευση και ο υπολογισμός δημιουργούν δεδομένα σε πρωτοφανείς ρυθμούς και κλίμακες. Πλέον αντιμετωπίζουμε πολλές πιεστικές προκλήσεις, μία από τις οποίες είναι ότι το τρέχον σύστημα αποθήκευσης δεν μπορεί να υποστηρίξει τόσο τεράστια δεδομένα. Γενικά, οι τιμές που κρύβονται στα Big Data εξαρτώνται από την ανανέωση των δεδομένων. Επομένως, πρέπει να αναπτυχθεί μια αρχή σπουδαιότητας δεδομένων που σχετίζεται με την αναλυτική τιμή για να αποφασιστεί ποια δεδομένα θα αποθηκευτούν και ποια δεδομένα θα απορριφθούν.

– *Αναλυτικός μηχανισμός*: Το αναλυτικό σύστημα των Big Data επεξεργάζεται μάζες ετερογενών δεδομένων εντός περιορισμένου χρόνου. Ωστόσο, τα παραδοσιακά RDBMS είναι αυστηρά σχεδιασμένα με έλλειψη κλίμακας και επεκτασιμότητας, τα οποία δεν μπορούσαν να ικανοποιήσουν τις απαιτήσεις απόδοσης. Τα RDBMS έχουν δείξει ότι προσφέρουν πλεονέκτημα έναντι της επεξεργασίας μη δομημένων δεδομένων και άρχισαν να γίνονται *mainstream* στην ανάλυση των Big Data. Παρόλα αυτά, εξακολουθούν να υπάρχουν ορισμένα προβλήματα στην απόδοσή τους και σε συγκεκριμένες εφαρμογές. Θα βρούμε μια συμβιβαστική λύση μεταξύ RDBMS και μη σχετικών βάσεων δεδομένων. Για παράδειγμα, ορισμένες επιχειρήσεις έχουν χρησιμοποιήσει μια μικτή αρχιτεκτονική βάσης δεδομένων που ενσωματώνει τα πλεονεκτήματα και των δύο τύπων βάσεων δεδομένων (π.χ. Facebook και Taobao). Ωστόσο απαιτείται περισσότερη έρευνα στη βάση δεδομένων στη μνήμη και τα δείγματα δεδομένων με βάση την κατά προσέγγιση ανάλυση.

– *Προσωπικά δεδομένα*: Οι περισσότεροι πάροχοι ή κάτοχοι υπηρεσιών Big Data προς το παρόν δεν δύναται να διατηρήσουν και να αναλύσουν αποτελεσματικά τόσο τεράστια σύνολα δεδομένων λόγω της περιορισμένης χωρητικότητάς τους. Πρέπει να βασίζονται σε επαγγελματίες ή εργαλεία για την ανάλυση τέτοιων δεδομένων, τα οποία αυξάνουν τους πιθανούς κινδύνους για την ασφάλεια των όποιων προσωπικών δεδομένων. Για παράδειγμα, το σύνολο δεδομένων συναλλαγών περιλαμβάνει ένα σύνολο δεδομένων λειτουργίας για την προώθηση βασικών επιχειρηματικών διαδικασιών. Τέτοια δεδομένα περιέχουν σε επίπεδα χαμηλότερης λεπτομέρειας και μερικές ευαίσθητες πληροφορίες όπως αριθμούς πιστωτικών καρτών. Επομένως, η ανάλυση των Big Data μπορεί να παραδοθεί σε τρίτους για επεξεργασία

μόνο όταν λαμβάνονται κατάλληλα προληπτικά μέτρα για την προστασία τέτοιων ευαίσθητων δεδομένων, για να διασφαλιστεί η ασφάλειά τους.

– *Διαχείριση ενέργειας:* Η κατανάλωση ενέργειας των υπολογιστικών συστημάτων *mainframe* έχει προσελκύσει μεγάλη προσοχή τόσο από την άποψη της οικονομίας όσο και από το περιβάλλον. Με την αύξηση του όγκου των δεδομένων και των αναλυτικών απαιτήσεων, η επεξεργασία, η αποθήκευση και η μετάδοση των Big Data θα καταναλώσει αναπόφευκτα όλο και περισσότερη ηλεκτρική ενέργεια. Επομένως, ο μηχανισμός ελέγχου και διαχείρισης κατανάλωσης ισχύος σε επίπεδο συστήματος θα δημιουργηθεί για τα Big Data, ενώ διασφαλίζεται η δυνατότητα επέκτασης και προσβασιμότητας.

– *Κλίμακες και επεκτασιμότητα:* Το αναλυτικό σύστημα των Big Data πρέπει να υποστηρίζει τα υπάρχοντα και τα μελλοντικά σύνολα δεδομένων. Ο αναλυτικός αλγόριθμος πρέπει να είναι σε θέση να επεξεργάζεται όλο και πιο επεκτεινόμενα και πιο περίπλοκα σύνολα δεδομένων.

– *Συνεργασία:* Η ανάλυση των Big Data χρειάζεται να υπάρχει διεπιστημονικότητα, εννοώντας την ύπαρξη εμπειρογνομώνων από διάφορους τομείς να συνεργαστούν για να αξιοποιήσουν τις δυνατότητές τους. Πρέπει να δημιουργηθεί μια ολοκληρωμένη αρχιτεκτονική δικτύου Big Data για να βοηθήσει τους επιστήμονες και τους μηχανικούς σε διάφορους τομείς να έχουν πρόσβαση σε διαφορετικά είδη δεδομένων και να αξιοποιήσουν πλήρως την εμπειρία τους, ώστε να συνεργαστούν για την ολοκλήρωση των αναλυτικών στόχων.

1. 3. Δημιουργία και προστιθέμενη αξία των Big Data

Όπως έχουμε προαναφέρει είναι αρκετές οι βασικές τεχνολογίες που σχετίζονται με τα Big Data, π.χ. το cloud computing, το IoT, τα κέντρα δεδομένων και το Hadoop. Επομένως η λειτουργική αλυσίδα των Big Data διαχωρίζεται σε τέσσερις φάσεις: 1) τη παραγωγή δεδομένων, 2) την απόκτηση δεδομένων, 3) την αποθήκευση δεδομένων και 4) την ανάλυση δεδομένων. Αρχής γενομένης της λήψης δεδομένων ως πρώτη ύλη, η παραγωγή δεδομένων και η απόκτηση δεδομένων είναι μια διαδικασία εκμετάλλευσης, η αποθήκευση δεδομένων είναι μια διαδικασία αποθήκευσης και η ανάλυση δεδομένων είναι μια διαδικασία παραγωγής που χρησιμοποιεί την πρώτη ύλη για να δημιουργήσει νέα αξία.

Η δημιουργία δεδομένων είναι το πρώτο βήμα των Big Data. Λαμβάνοντας υπόψη τα δεδομένα του Διαδικτύου ως παράδειγμα, δημιουργείται τεράστιος όγκος δεδομένων από την άποψη των καταχωρίσεων αναζήτησης, των προωθητικών μέσων του Διαδικτύου, των εγγραφών συνομιλίας και των μηνυμάτων *microblog*. Αυτά τα δεδομένα σχετίζονται στενά με την καθημερινή ζωή των ανθρώπων και έχουν παρόμοια χαρακτηριστικά υψηλής αξίας και χαμηλής πυκνότητας. Τέτοια δεδομένα στο Διαδίκτυο μπορεί να είναι χωρίς αξία από μόνα τους, αλλά μέσω της εκμετάλλευσης των συσσωρευμένων Big Data, μπορούν να εντοπιστούν χρήσιμες πληροφορίες, όπως συνήθειες και χόμπι των χρηστών, και είναι ακόμη δυνατό να προβλεφθούν οι συμπεριφορές και οι συναισθηματικές διαθέσεις των χρηστών.

Επιπλέον, δημιουργούνται μέσω διαμήκων και/ή κατανεμημένων πηγών δεδομένων, τα σύνολα δεδομένων που είναι πιο μεγάλης κλίμακας, πολύ διαφορετικά και πολύπλοκα. Τέτοιες πηγές δεδομένων περιλαμβάνουν αισθητήρες, βίντεο, ροές κλικ και/ή όλες τις άλλες διαθέσιμες πηγές δεδομένων. Προς το παρόν, οι κύριες πηγές των Big Data είναι οι πληροφορίες για τη λειτουργία και την εμπορία επιχειρήσεων, οι πληροφορίες λογιστικής και ανίχνευσης στο IoT, οι πληροφορίες αλληλεπίδρασης ανθρώπου και οι πληροφορίες θέσης στον κόσμο του Διαδικτύου και τα δεδομένα που παράγονται στην επιστημονική έρευνα κ.λπ. Οι πληροφορίες ξεπερνούν κατά πολύ το ικανότητες αρχιτεκτονικών πληροφορικής και υποδομών υπαρχουσών επιχειρήσεων, ενώ η απαίτηση σε πραγματικό χρόνο τονίζει επίσης πολύ την υπάρχουσα υπολογιστική ικανότητα.

1. 4. Απόκτηση δεδομένων Big Data

Η απόκτηση των Big Data περιλαμβάνει τη συλλογή δεδομένων, μετάδοση δεδομένων και προεπεξεργασία δεδομένων. Κατά τη λήψη Big Data, μόλις συλλέξουμε τα ανεπεξέργαστα δεδομένα, θα χρησιμοποιήσουμε έναν αποτελεσματικό μηχανισμό μετάδοσης για να τα στείλουμε σε ένα κατάλληλο σύστημα διαχείρισης/αποθήκευσης για την υποστήριξη διαφορετικών αναλυτικών εφαρμογών. Τα συλλεγόμενα σύνολα δεδομένων μπορεί μερικές φορές να περιλαμβάνουν πολύ περιττά ή άχρηστα δεδομένα, τα οποία αυξάνουν άσκοπα τον αποθηκευτικό χώρο και επηρεάζουν την επόμενη ανάλυση δεδομένων. Για παράδειγμα, η περίσσεια είναι πολύ συχνό φαινόμενο μεταξύ των συνόλων δεδομένων που συλλέγονται από αισθητήρες για παρακολούθηση του περιβάλλοντος. Η τεχνολογία συμπίεσης δεδομένων μπορεί να εφαρμοστεί για τη μείωση του πλεονασμού. Επομένως, οι εργασίες προεπεξεργασίας δεδομένων είναι απαραίτητες για την εξασφάλιση αποτελεσματικής αποθήκευσης και εκμετάλλευσης δεδομένων.

Η συλλογή δεδομένων είναι η χρήση ειδικών τεχνικών συλλογής δεδομένων για την απόκτηση πρωτογενών δεδομένων από ένα συγκεκριμένο περιβάλλον παραγωγής δεδομένων. Τέσσερις κοινές μέθοδοι συλλογής δεδομένων συναντώνται:

– *Log files*: Ως μια ευρέως χρησιμοποιούμενη Μεταβλητές που αφαιρέθηκαν συλλογής δεδομένων, τα αρχεία καταγραφής είναι αρχεία εγγραφής που δημιουργούνται αυτόματα από το σύστημα προέλευσης δεδομένων, έτσι ώστε να καταγράφονται δραστηριότητες σε καθορισμένες μορφές αρχείων για επακόλουθη ανάλυση. Τα αρχεία καταγραφής χρησιμοποιούνται συνήθως σε όλες σχεδόν τις ψηφιακές συσκευές. Για παράδειγμα, οι διακομιστές ιστού καταγράφουν στα αρχεία καταγραφής τον αριθμό των κλικ, τα ποσοστά κλικ, τις επισκέψεις και άλλες εγγραφές ιδιοκτησίας των χρηστών ιστού (Wahab et al., 2008). Για την καταγραφή δραστηριοτήτων χρηστών στους ιστότοπους, οι διακομιστές ιστού περιλαμβάνουν κυρίως τις ακόλουθες τρεις μορφές αρχείων καταγραφής: μορφή αρχείου δημόσιου αρχείου καταγραφής (NCSA), εκτεταμένη μορφή καταγραφής (W3C) και μορφή αρχείου καταγραφής IIS (Microsoft). Και οι τρεις τύποι αρχείων καταγραφής είναι σε μορφή κειμένου ASCII. Μπορούν μερικές φορές να χρησιμοποιηθούν βάσεις δεδομένων εκτός από τα αρχεία κειμένου για την αποθήκευση πληροφοριών καταγραφής για τη βελτίωση της αποτελεσματικότητας των ερωτημάτων του μαζικού καταστήματος καταγραφής (Joshi & Yesha, 2003; Nanopoulos et al., 2002). Υπάρχουν επίσης ορισμένα άλλα αρχεία καταγραφής που βασίζονται στη συλλογή δεδομένων, συμπεριλαμβανομένων δεικτών αποθεμάτων σε

χρηματοοικονομικές εφαρμογές και προσδιορισμού των καταστάσεων λειτουργίας στην παρακολούθηση του δικτύου και τη διαχείριση της κυκλοφορίας.

– *Sensing*: Οι αισθητήρες είναι συνηθισμένοι στην καθημερινή ζωή για τη μέτρηση φυσικών ποσοτήτων και τη μετατροπή φυσικών ποσοτήτων σε αναγνώσιμα ψηφιακά σήματα για μετέπειτα επεξεργασία (και αποθήκευση). Τα αισθητηριακά δεδομένα μπορούν να ταξινομηθούν ως ηχητικό κύμα, φωνή, δονήσεις, αυτοκίνητα, χημικά, ρεύμα, καιρός, πίεση, θερμοκρασία κ.λπ. Για εφαρμογές που μπορούν εύκολα να αναπτυχθούν και να διαχειριστούν, π.χ. σύστημα παρακολούθησης βίντεο (Chandramohan & Christensen, 2002), το ενσύρματο δίκτυο αισθητήρων είναι μια βολική λύση για την απόκτηση σχετικών πληροφοριών. Μερικές φορές η ακριβής θέση ενός συγκεκριμένου φαινομένου είναι άγνωστη και μερικές φορές το παρακολουθούμενο περιβάλλον δεν διαθέτει υποδομές ενέργειας ή επικοινωνίας. Στη συνέχεια, η ασύρματη επικοινωνία πρέπει να χρησιμοποιείται για να επιτρέψει τη μετάδοση δεδομένων μεταξύ κόμβων αισθητήρων με περιορισμένη ικανότητα ενέργειας και επικοινωνίας. Τα τελευταία χρόνια, οι WSN έχουν λάβει μεγάλο ενδιαφέρον και έχουν εφαρμοστεί σε πολλές εφαρμογές, όπως η περιβαλλοντική έρευνα (Barrenetxea et al., 2008; Selavo et al., 2007), παρακολούθηση της ποιότητας του νερού (Kim et al., 2008), πολιτική μηχανική (Ceriotti et al., 2009; Kim et al., 2007), και την παρακολούθηση των συνηθειών της άγριας ζωής (Tolle et al., 2005). Ένα WSN αποτελείται γενικά από μεγάλο αριθμό γεωγραφικά κατανομημένων κόμβων αισθητήρων, καθένας από τους οποίους είναι μια μικρο-συσκευή που τροφοδοτείται από μπαταρία. Τέτοιοι αισθητήρες αναπτύσσονται σε καθορισμένες θέσεις όπως απαιτείται από την εφαρμογή για τη συλλογή δεδομένων τηλεπισκόπησης. Μόλις αναπτυχθούν οι αισθητήρες, ο σταθμός βάσης θα στείλει πληροφορίες ελέγχου για διαμόρφωση / διαχείριση δικτύου ή συλλογή δεδομένων σε κόμβους αισθητήρων. Με βάση αυτές τις πληροφορίες ελέγχου, τα αισθητήρια δεδομένα συγκεντρώνονται σε διαφορετικούς κόμβους αισθητήρων και αποστέλλονται πίσω στο σταθμό βάσης για περαιτέρω επεξεργασία (Wang & Liu, 2010).

– *Μέθοδοι απόκτησης δεδομένων δικτύου*: Προς το παρόν, η απόκτηση δεδομένων δικτύου πραγματοποιείται χρησιμοποιώντας έναν συνδυασμό προγράμματος ανίχνευσης ιστού, συστήματος τμηματοποίησης λέξεων, συστήματος εργασιών και συστήματος ευρετηρίου κ.λπ. Το πρόγραμμα ανίχνευσης ιστού είναι ένα πρόγραμμα που χρησιμοποιείται από μηχανές αναζήτησης για τη λήψη και αποθήκευση ιστοσελίδων (Cho & Garcia-Molina, 2002). Σε γενικές γραμμές, το πρόγραμμα ανίχνευσης ιστού ξεκινά από τον ομοιόμορφο εντοπισμό πόρων (URL) μιας αρχικής ιστοσελίδας για πρόσβαση σε άλλες συνδεδεμένες ιστοσελίδες, κατά τη διάρκεια της οποίας αποθηκεύει και αλληλουχίζει όλες τις ανακτημένες διευθύνσεις URL. Το πρόγραμμα ανίχνευσης ιστού αποκτά μια διεύθυνση URL κατά σειρά προτεραιότητας

μέσω μιας ουράς διευθύνσεων URL και, στη συνέχεια, κατεβάζει ιστοσελίδες και προσδιορίζει όλες τις διευθύνσεις URL στις ιστοσελίδες που έχουν ληφθεί και εξάγει νέες διευθύνσεις URL που πρέπει να τοποθετηθούν στην ουρά. Αυτή η διαδικασία επαναλαμβάνεται έως ότου σταματήσει το πρόγραμμα ανίχνευσης ιστού. Η απόκτηση δεδομένων μέσω ενός προγράμματος ανίχνευσης ιστού εφαρμόζεται ευρέως σε εφαρμογές που βασίζονται σε ιστοσελίδες, όπως μηχανές αναζήτησης ή προσωρινή αποθήκευση ιστού. Οι παραδοσιακές τεχνολογίες εξαγωγής ιστοσελίδων διαθέτουν πολλές αποτελεσματικές λύσεις και έχει γίνει σημαντική έρευνα σε αυτόν τον τομέα. Καθώς εμφανίζονται πιο προηγμένες εφαρμογές ιστοσελίδων, προτείνονται ορισμένες στρατηγικές εξαγωγής (Choudhary et al., 2012) για να αντιμετωπιστούν οι πλούσιες εφαρμογές διαδικτύου.

Οι τρέχουσες τεχνολογίες απόκτησης δεδομένων δικτύου περιλαμβάνουν κυρίως την παραδοσιακή τεχνολογία δέσμωσης πακέτων που βασίζεται σε Libpcap, τεχνολογίες δέσμωσης πακέτων μηδενικών αντιγράφων, καθώς και κάποιο εξειδικευμένο λογισμικό παρακολούθησης δικτύου όπως το Wireshark, το SmartSniff και το WinNetCap.

– *Libpcap-βασισμένη τεχνολογία δέσμωσης πακέτων*: Το Libpcap (βιβλιοθήκη συλλογής πακέτων) είναι μια ευρέως χρησιμοποιούμενη βιβλιοθήκη λειτουργιών συλλογής πακέτων δεδομένων δικτύου. Είναι ένα γενικό εργαλείο που δεν εξαρτάται από κανένα συγκεκριμένο σύστημα και χρησιμοποιείται κυρίως για την καταγραφή δεδομένων στο επίπεδο σύνδεσης δεδομένων. Διαθέτει απλότητα, εύκολο στη χρήση και φορητότητα, αλλά έχει σχετικά χαμηλή απόδοση. Επομένως, σε περιβάλλον δικτύου υψηλής ταχύτητας, ενδέχεται να προκύψουν σημαντικές απώλειες πακέτων όταν χρησιμοποιείται το Libpcap.

– *Zero-copy τεχνολογία δέσμωσης πακέτων*: Το λεγόμενο μηδενικών αντιγράφων (ZC) σημαίνει ότι δεν υπάρχουν αντίγραφα μεταξύ εσωτερικής μνήμης κατά τη λήψη και αποστολή πακέτων σε έναν κόμβο. Κατά την αποστολή, τα πακέτα δεδομένων ξεκινούν απευθείας από το buffer του χρήστη των εφαρμογών, περνούν μέσω των διεπαφών δικτύου και φτάνουν σε ένα εξωτερικό δίκτυο. Κατά τη λήψη, οι διασυνδέσεις δικτύου στέλνουν απευθείας πακέτα δεδομένων στο buffer του χρήστη. Η βασική ιδέα του μηδενικού αντιγράφου είναι να μειωθούν οι χρόνοι αντιγραφής δεδομένων, να μειωθούν οι κλήσεις συστήματος και να μειωθεί ο φόρτος των επεξεργαστών ενώ τα ddatagrams μεταφέρονται από εξοπλισμό δικτύου σε χώρο προγράμματος χρήστη. Η τεχνολογία μηδενικού αντιγράφου χρησιμοποιεί πρώτα την τεχνολογία άμεσης πρόσβασης στη μνήμη (DMA) για να μεταδίδει απευθείας δίκτυα δεδομένων σε ένα χώρο διευθύνσεων που έχει εκχωρηθεί εκ των προτέρων από τον πυρήνα του συστήματος, ώστε να αποφευχθεί η συμμετοχή της CPU. Εν τω μεταξύ, χαρτογραφεί την

εσωτερική μνήμη των γραφημάτων δεδομένων στον πυρήνα του συστήματος με εκείνη του προγράμματος ανίχνευσης ή δημιουργεί μια περιοχή προσωρινής μνήμης στο χώρο χρήστη και τη χαρτογραφεί στον χώρο του πυρήνα. Στη συνέχεια, το πρόγραμμα ανίχνευσης αποκτά άμεση πρόσβαση στην εσωτερική μνήμη, ώστε να μειωθεί το αντίγραφο της εσωτερικής μνήμης από τον πυρήνα του συστήματος στον χώρο του χρήστη και να μειωθεί ο αριθμός των κλήσεων συστήματος.

– *Κινητός εξοπλισμός*: Σήμερα οι κινητές συσκευές χρησιμοποιούνται ευρύτερα. Καθώς οι λειτουργίες της κινητής συσκευής γίνονται όλο και πιο ισχυρές, διαθέτουν πιο περίπλοκα και πολλαπλά μέσα απόκτησης δεδομένων καθώς και περισσότερη ποικιλία δεδομένων. Οι φορητές συσκευές μπορούν να αποκτήσουν πληροφορίες γεωγραφικής θέσης μέσω συστημάτων εντοπισμού θέσης. απόκτηση πληροφοριών ήχου μέσω μικροφώνων, βίντεο, δισδιάστατους γραμμωτούς κώδικες και άλλες πληροφορίες πολυμέσων μέσω κάμερας. Με την πάροδο των ετών, οι ασύρματοι φορείς εκμετάλλευσης έχουν βελτιώσει το επίπεδο υπηρεσιών του διαδικτύου αποκτώντας και αναλύοντας τέτοιες πληροφορίες. Για παράδειγμα, το ίδιο το iPhone είναι ένα «κινητός κατάσκοπος». Μπορεί να συλλέξει ασύρματα δεδομένα και πληροφορίες γεωγραφικής θέσης και στη συνέχεια να στείλει αυτές τις πληροφορίες πίσω στην Apple Inc. για επεξεργασία, για τις οποίες ο χρήστης δεν γνωρίζει. Εκτός από την Apple, λειτουργικά συστήματα έξυπνων τηλεφώνων όπως το Android της Google και το Windows Phone της Microsoft μπορούν επίσης να συλλέγουν πληροφορίες με παρόμοιο τρόπο.

Εκτός από τις προαναφερθείσες τρεις μεθόδους συλλογής δεδομένων των κύριων πηγών δεδομένων, υπάρχουν πολλές άλλες μέθοδοι ή συστήματα συλλογής δεδομένων. Για παράδειγμα, σε επιστημονικά πειράματα, πολλά ειδικά εργαλεία μπορούν να χρησιμοποιηθούν για τη συλλογή πειραματικών δεδομένων, όπως μαγνητικά φασματομέτρα και ράδιο τηλεσκοπία. Μπορούμε να ταξινομήσουμε τις μεθόδους συλλογής δεδομένων από διαφορετικές προοπτικές. Από τη σκοπιά των πηγών δεδομένων, οι μέθοδοι συλλογής δεδομένων μπορούν να ταξινομηθούν σε δύο κατηγορίες: καταγραφή μεθόδων συλλογής μέσω πηγών δεδομένων και καταγραφή μεθόδων συλλογής μέσω άλλων βοηθητικών εργαλείων.

1. 5. Εφαρμογή των δεδομένων Big Data στην καθημερινότητα

1. 5. 1. Δεδομένα εταιριών

Το 2013, η IBM δημοσίευσε μια ανάλυση με τον τίτλο «Οι Εφαρμογές των Big Data στον Πραγματικό Κόσμο», η οποία δείχνει ότι τα εσωτερικά δεδομένα των επιχειρήσεων είναι οι κύριες πηγές των Big Data. Τα εσωτερικά δεδομένα των επιχειρήσεων συνίστανται κυρίως σε δεδομένα συναλλαγών στο διαδίκτυο και σε απευθείας σύνδεση των δεδομένων ανάλυσης, τα περισσότερα από τα οποία είναι ιστορικά στατικά δεδομένα και διαχειρίζονται από RDBMSs με δομημένο τρόπο. Επιπλέον, δεδομένα παραγωγής, δεδομένα αποθέματος, δεδομένα πωλήσεων και οικονομικά δεδομένα κ.λπ. , επίσης, αποτελούν εσωτερικά δεδομένα επιχειρήσεων, τα οποία στοχεύουν στη καταγραφή των ανεπίσημων και βάσει δεδομένων πληροφοριών στις επιχειρήσεις, ώστε να καταγράφονται όλες οι πτυχές δραστηριότητας αυτών με τη μορφή εσωτερικών δεδομένων.

Τις τελευταίες δεκαετίες, η πληροφορική και τα ψηφιακά δεδομένα συνέβαλαν πολύ στη βελτίωση της κερδοφορίας των επιχειρηματικών τμημάτων. Εκτιμάται ότι ο όγκος των επιχειρηματικών δεδομένων όλων των εταιριών στον κόσμο μπορεί να διπλασιαστεί κάθε 1,2 χρόνια (Manyika et al., 2011) κατά την οποία, ο κύκλος εργασιών των επιχειρήσεων μέσω του διαδικτύου ανά ημέρα θα φτάσουν τα 450 δισεκατομμύρια δολάρια ΗΠΑ (Gantz & Reinsel, 2010). Ο συνεχώς αυξανόμενος όγκος επιχειρηματικών δεδομένων απαιτεί πιο αποτελεσματική ανάλυση σε πραγματικό χρόνο, ώστε να αξιοποιηθεί πλήρως το δυναμικό της. Για παράδειγμα, η Amazon επεξεργάζεται εκατομμύρια τερματικές λειτουργίες και περισσότερα από 500.000 ερωτήματα από πωλητές τρίτων ημερησίως (Laney, 2001). Η Walmart επεξεργάζεται ένα εκατομμύριο συναλλαγές πελατών ανά ώρα και τέτοια δεδομένα συναλλαγών εισάγονται σε μια βάση δεδομένων με χωρητικότητα άνω των 2,5PB (Cukier, 2010). Η Akamai αναλύει 75 εκατομμύρια εκδηλώσεις την ημέρα για τις διαφημίσεις-στόχους της (Zikopoulos & Eaton, 2011).

1. 5. 2. Δεδομένα ΙοΤ

Όπως προαναφέρθηκε, το ΙοΤ είναι μια σημαντική πηγή των Big Data. Μεταξύ έξυπνων πόλεων που κατασκευάζονται με βάση το ΙοΤ, τα Big Data μπορεί να προέρχονται από τη βιομηχανία, τη γεωργία, την κυκλοφορία, τη μεταφορά, την ιατρική περίθαλψη, τα δημόσια τμήματα, τις οικογένειες κ.λπ.

Σύμφωνα με τις διαδικασίες απόκτησης και μετάδοσης δεδομένων στο IoT, η αρχιτεκτονική του δικτύου μπορεί να χωριστεί σε τρία επίπεδα: το επίπεδο ανίχνευσης, το επίπεδο δικτύου και το επίπεδο εφαρμογής. Το επίπεδο ανίχνευσης είναι υπεύθυνο για την απόκτηση δεδομένων και αποτελείται κυρίως από δίκτυα αισθητήρων. Το επίπεδο δικτύου είναι υπεύθυνο για τη μετάδοση και επεξεργασία πληροφοριών, όπου η στενή μετάδοση μπορεί να βασίζεται σε δίκτυα αισθητήρων και η απομακρυσμένη μετάδοση εξαρτάται από το διαδίκτυο. Τέλος, το επίπεδο εφαρμογής υποστηρίζει συγκεκριμένες εφαρμογές του IoT.

Σύμφωνα με τα χαρακτηριστικά του IoT, τα δεδομένα που παράγονται από αυτό έχουν τα ακόλουθα χαρακτηριστικά:

- *Δεδομένα μεγάλης κλίμακας:* Στο IoT, αναπτύσσονται κατανεμημένα μάζες εξοπλισμού απόκτησης δεδομένων, τα οποία ενδέχεται να αποκτήσουν απλά αριθμητικά δεδομένα, π.χ. τοποθεσία; ή πολύπλοκα δεδομένα πολυμέσων όπως βίντεο παρακολούθησης. Προκειμένου να ικανοποιηθούν οι απαιτήσεις ανάλυσης και επεξεργασίας, πρέπει να αποθηκεύονται όχι μόνο τα δεδομένα που έχουν αποκτηθεί επί του παρόντος, αλλά και τα ιστορικά δεδομένα εντός συγκεκριμένου χρονικού πλαισίου. Επομένως, τα δεδομένα που δημιουργούνται από το IoT χαρακτηρίζονται από μεγάλες κλίμακες.
- *Ετερογένεια:* Λόγω της ποικιλίας των συσκευών απόκτησης δεδομένων, τα ληφθέντα δεδομένα είναι επίσης διαφορετικά και τέτοια δεδομένα διαθέτουν ετερογένεια.
- *Ισχυρή συσχέτιση χρόνου και χώρου:* Στο IoT, κάθε συσκευή λήψης δεδομένων τοποθετείται σε μια συγκεκριμένη γεωγραφική θέση και κάθε κομμάτι δεδομένων έχει χρονική σήμανση. Ο συσχετισμός χρόνου και χώρου είναι μια σημαντική ιδιότητα δεδομένων από το IoT. Κατά την ανάλυση και επεξεργασία δεδομένων, ο χρόνος και ο χώρος είναι επίσης σημαντικές διαστάσεις για στατιστική ανάλυση.
- *Τα αποτελεσματικά δεδομένα αντιπροσωπεύουν μόνο ένα μικρό μέρος των Big Data:* Μπορεί να προκύψει μεγάλη ποσότητα θορύβου κατά την απόκτηση και τη μετάδοση δεδομένων στο IoT. Μεταξύ των συνόλων δεδομένων που αποκτήθηκαν από συσκευές απόκτησης, είναι πολύτιμη μόνο μια μικρή ποσότητα ανώμαλων δεδομένων. Για παράδειγμα, κατά την απόκτηση βίντεο κίνησης, τα λίγα πλαίσια βίντεο που καταγράφουν την παραβίαση των κανονισμών κυκλοφορίας και τα τροχαία ατυχήματα είναι πιο πολύτιμα από αυτά που καταγράφουν μόνο την κανονική ροή της κυκλοφορίας.

1. 5. 3. Βιο-ιατρικά δεδομένα

Καθώς μια σειρά τεχνολογιών βιομετρήσεων υψηλής απόδοσης αναπτύσσονται καινοτόμα στις αρχές του 21ου αιώνα, η πρωτοποριακή έρευνα στον τομέα της βιοϊατρικής εισέρχεται επίσης στην εποχή των Big Data. Κατασκευάζοντας έξυπνα, αποτελεσματικά και ακριβή αναλυτικά μοντέλα και θεωρητικά συστήματα για εφαρμογές βιοϊατρικής, μπορεί να αποκαλυφθεί ο βασικός μηχανισμός που διέπει τα σύνθετα βιολογικά φαινόμενα. Δεν μπορεί να καθοριστεί μόνο η μελλοντική ανάπτυξη της βιοϊατρικής, αλλά και οι πρωταγωνιστικοί ρόλοι μπορούν να αναληφθούν στην ανάπτυξη μιας σειράς σημαντικών στρατηγικών βιομηχανιών που σχετίζονται με την εθνική οικονομία, τη διαβίωση των ανθρώπων και την εθνική ασφάλεια, με σημαντικές εφαρμογές όπως η ιατρική φροντίδα, νέα έρευνα και ανάπτυξη φαρμάκων και παραγωγή διαγονιδιακών καλλιεργειών.

Η ολοκλήρωση του HGP (Human Genome Project) και η συνεχής ανάπτυξη της τεχνολογίας προσδιορισμού αλληλουχιών οδηγούν επίσης σε εκτεταμένες εφαρμογές των Big Data στον τομέα. Οι μάζες των δεδομένων που δημιουργούνται από την αλληλουχία γονιδίων περνούν από εξειδικευμένη ανάλυση σύμφωνα με διαφορετικές απαιτήσεις εφαρμογής, για να το συνδυάσουν με την κλινική διάγνωση γονιδίων και να παρέχουν πολύτιμες πληροφορίες για έγκαιρη διάγνωση και εξατομικευμένη θεραπεία των νόσων. Μια αλληλουχία ανθρώπινου γονιδίου μπορεί να δημιουργήσει 100-600 GB ακατέργαστα δεδομένα. Στην Κίνα εντός της National Genebank στο Shenzhen, υπάρχουν 1,3 εκατομμύρια δείγματα συμπεριλαμβανομένων 1,15 εκατομμυρίων ανθρώπινων δειγμάτων και 150.000 δείγματα ζώων, φυτών και μικροοργανισμών. Μέχρι το τέλος του 2013, ακόμη 10 εκατομμύρια ανιχνεύσιμα βιολογικά δείγματα αποθηκεύονταν και μέχρι το τέλος του 2015, ο αριθμός αυτός τριπλασιάστηκε. Είναι προβλέψιμο ότι με την ανάπτυξη τεχνολογιών βιοϊατρικής, οι αλληλουχίες γονιδίων θα γίνουν ταχύτερες και πιο βολικές, και καθιστώντας έτσι τα Big Data στη βιοϊατρική αυξανόμενα πέρα από κάθε αμφιβολία.

Επιπλέον, τα δεδομένα που παράγονται από την κλινική ιατρική περίθαλψη και την ιατρική έρευνα αυξάνονται επίσης γρήγορα. Για παράδειγμα, το Ιατρικό Κέντρο του Πανεπιστημίου του Πίτσμπουργκ (UPMC) έχει αποθηκεύσει 2TB τέτοια δεδομένα. Η Exprolys, μια αμερικανική εταιρεία, παρέχει πλατφόρμες για ανταλλαγή κλινικών δεδομένων, δεδομένων λειτουργίας και συντήρησης και οικονομικών δεδομένων. Προς το παρόν, περίπου 13 εκατομμύρια πληροφορίες έχουν συγκεντρωθεί σε κλίμακα των περίπου 60-70 TB. Η Practice Fusion, μια άλλη αμερικανική εταιρεία, διαχειρίζεται ηλεκτρονικά ιατρικά αρχεία περίπου 200.000 ασθενών.

Εκτός από τέτοιες μικρές και μεσαίες επιχειρήσεις, άλλες γνωστές εταιρείες πληροφορικής, όπως η Google, η Microsoft και η IBM έχουν επενδύσει εκτενώς στην έρευνα και την υπολογιστική ανάλυση μεθόδων που σχετίζονται με βιολογικά Big Data υψηλής απόδοσης, για μερίδια στην τεράστια αγορά ως γνωστό ως "Next Internet". Η IBM προέβλεπε, στο Συνέδριο Στρατηγικής του 2013, ότι με την απότομη αύξηση των ιατρικών εικόνων και των ηλεκτρονικών ιατρικών αρχείων, οι επαγγελματίες του ιατρικού τομέα θα μπορούν να χρησιμοποιήσουν τα Big Data για να εξαγάγουν χρήσιμες κλινικές πληροφορίες από μάζες δεδομένων για να αποκτήσουν ιατρικό ιστορικό και να προβλέψουν αποτελέσματα θεραπείας, βελτιώνοντας έτσι τη φροντίδα των ασθενών με μείωση του κόστους. Το 2015, ο μέσος όγκος δεδομένων κάθε νοσοκομείου εκτιμάτε ότι έφτασαν τα 167-665 TB.

Για παράδειγμα επιστήμονες (Mayer-Schönberger & Cukier, 2013) αναφέρουν ότι οι ερευνητές στο Ινστιτούτο Τεχνολογίας του Πανεπιστημίου του Οντάριο και στην IBM συνεργάζονται με διάφορα νοσοκομεία σε λογισμικό για να βοηθήσουν τους γιατρούς να λάβουν καλύτερες διαγνωστικές αποφάσεις όταν φροντίζουν πρόωρα μωρά. Το λογισμικό λαμβάνει και επεξεργάζεται δεδομένα ασθενών σε πραγματικό χρόνο, παρακολουθώντας 16 διαφορετικές ροές δεδομένων, όπως καρδιακό ρυθμό, ρυθμό αναπνοής, θερμοκρασία, αρτηριακή πίεση και επίπεδο οξυγόνου στο αίμα, τα οποία μαζί ανέρχονται σε περίπου 1.260 σημεία δεδομένων ανά δευτερόλεπτο. Οι παρατηρήσεις μπορεί να επιτρέψουν στους γιατρούς να κατανοήσουν τελικά τις θεμελιώδεις αιτίες, απλώς γνωρίζοντας ότι κάτι πιθανό να συμβεί είναι αναμφισβήτητα πιο σημαντικό από το να κατανοήσουμε ακριβώς γιατί. Δεν είναι δύσκολο να φανταστεί κανείς μια ανάλογη εφαρμογή ελέγχου στην οποία οι επαναδιατυπώσεις ή άλλες ανεπιθύμητες ενέργειες συσχετίζονται με δείκτες που απορρίπτονται από κάθε δημόσια εταιρεία που υποβάλλει πληροφορίες και διάφορες άλλες περιπτώσεις.

2. Η εφαρμογή των Big Data στα οικονομικά και τη λογιστική

2.1. Εφαρμογές στη χρηματοοικονομική

Η λογιστική ως ο τομέας της οικονομικής επιστήμης που είναι επιφορτισμένος με τη μέτρηση των οικονομικών μιας επιχείρησης περιορίζεται συνήθως σε «χρηματοοικονομικές αναφορές». Για παράδειγμα, οι εκτεταμένες απαιτήσεις για πιστοποίηση εσωτερικών ελέγχων στη Βουλή των Αντιπροσώπων των ΗΠΑ (2002) περιορίζονται σε αυτόν τον τομέα. Η χρηματοοικονομική αναφορά συνοψίζει μια πληθώρα επιχειρηματικών δραστηριοτήτων που τελικά καταλήγουν στην καταχώριση μιας συναλλαγής, συμπεριλαμβάνοντας τις όποιες εκτιμήσεις πάνω σε τεχνολογικά ή κόστους/οφέλους θέματα επιχειρηματικών μετρήσεων σε αυτό το επίπεδο. Με την αλλαγή της τεχνολογίας και την έλευση των συστημάτων διαχείρισης επιχειρησιακών πόρων (Enterprise resource planning, ERP), που συλλαμβάνουν δεδομένα και υποστηρίζουν πολλές από τις διαδικασίες, οι διαδοχικές αιτιώδεις διαδικασίες μπορούν να μετρηθούν αλλάζοντας δραστικά τη φύση της επιχειρηματικής μέτρησης (Romero et al., 2012). Αν και οι αλλαγές στις λογιστικές πρακτικές και τα πρότυπα ως απάντηση στα Big Data δεν έχουν ακόμη συμβεί, τα Big Data έχουν τη δυνατότητα να προκαλέσουν μια αλλαγή παραδείγματος που επιτρέπει την ανίχνευση και τη μέτρηση των οικονομικών δραστηριοτήτων νωρίτερα και βαθύτερα. Μια συμπαρουσίαση της χρονολογικής εξέλιξης των τεχνολογικών γεγονότων με την επίδρασή τους στη λογιστική μπορεί να βοηθήσει στην αναγνώριση αυτής:

- Πριν από την εποχή των υπολογιστών οι οργανισμοί διατηρούσαν μόνο συνοπτικά λογιστικά δεδομένα μέσω γραφημάτων λογαριασμών/λογιστικών βιβλίων. Οι πληροφορίες μεμονωμένων συναλλαγών φυσικά σε αυτές τις εποχές δεν ήταν εύκολα προσβάσιμες ή όποτε ήταν αυτό γινόταν με μεγάλη προσπάθεια του λογιστή. Καθότι οι πληροφορίες ήταν αρχικά προφορικές, στη συνέχεια γραπτές και στη συνέχεια οργανώθηκαν σε ειδικά χαρτογραφικά αρχεία.
- Οι εξελίξεις στους υπολογιστές κατέστησαν δυνατή την αποθήκευση και τη διατήρηση πλήρων πληροφοριών για όλες τις συναλλαγές, οδηγώντας έτσι στη πρώτη εμφάνιση ουσιαστικά των Big Data στη λογιστική.
- Η τεχνολογία επέτρεψε στις επιχειρήσεις να γίνουν πραγματικά παγκόσμιες, οδηγώντας σε μια δεύτερη επέκταση του όγκου, της ταχύτητας και της ποικιλίας των λογιστικών δεδομένων. Ωστόσο, οι περιφερειακές διαφορές δημιουργούν σοβαρές δυσκολίες στη συμβατότητα μεταξύ των δεδομένων που αποκτώνται. Η εγγραφή, η αποθήκευση και η

αναφορά των πληροφοριών σε επίπεδο συναλλαγής επιτρέπει την επίλυση πολλών από αυτές τις διαφορές.

- Τα συστήματα των επιχειρήσεων, μαζί με τις έννοιες των δεδομένων αποθήκης, με τα ολοκληρωμένα λογιστικά δεδομένα και με δεδομένα από μη χρηματοοικονομικά συστήματα είναι επόμενο να οδηγούν σε περαιτέρω αύξηση του όγκου και της ποικιλίας των λογιστικών δεδομένων. Ένα τέτοιο παράδειγμα αποτελούν τα δεδομένα υπηρεσιών μετά την πώληση και την παρακολούθηση συμπεριφοράς πελατών που συνδέονται με συγκεκριμένες λογιστικές συναλλαγές. Συνεπώς επιτρέπεται η δημιουργία εξισώσεων συνέχειας με χρονική υστέρηση (Kogan et al., 2014) που συνδέουν οικονομικές και μη χρηματοοικονομικές μεταβλητές.
- Η περαιτέρω επέκταση των λογιστικών δεδομένων περιλαμβάνει την προσθήκη νέων τύπων πεδίων δεδομένων σε συγκεκριμένα αρχεία λογιστικών συναλλαγών πέρα από αυτά που καταγράφονταν παραδοσιακά, πολλά από τα οποία λαμβάνονται μέσω αυτοματοποιημένων αισθητήρων.
- Η αυτόματη λήψη δεδομένων μέσω αισθητήρων, RFID και ροών δεδομένων GPS επιτρέπει τον ορισμό της αυξημένης συχνότητας λήψης δεδομένων και ένα μεγάλο σύνολο επεκτάσεων της διαδικασίας διαχείρισης, λογιστικής και διασφάλισης, καθώς και την προσθήκη πολλαπλών πηγών στοιχείων ελέγχου.

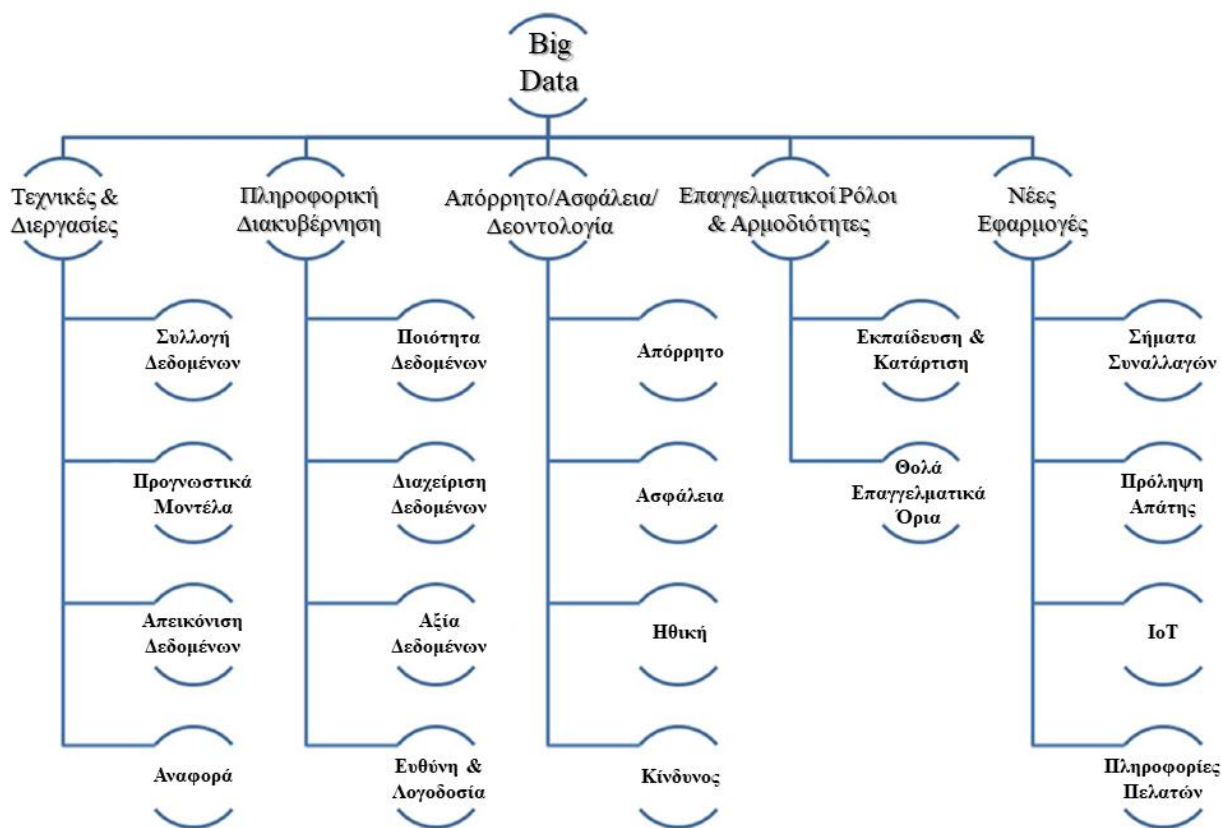
Η λογιστική μέτρηση έχει χάσει σταδιακά την πληροφοριακή της αξία με σημαντική μείωση της εξήγησης της αγοραίας αξίας από λογιστικές μεταβλητές (Ball & Brown, 1968; Lev & Zarowin, 1999). Αυτή η απώλεια της αξίας της πληροφορίας είναι ιδιαίτερα έντονη για τις νέες εταιρείες συσσώρευσης γνώσεων που έχουν μεγαλύτερη άυλη ένταση, οι οποίες διεκδικούν προοδευτικά αυξανόμενο μερίδιο της οικονομίας (Srivastava, 2014).

Μια οικονομία υποστηρίζεται από διαδικασίες σε πραγματικό χρόνο (Vasarhelyi & Greenstein, 2003) με τη δημόσια μέτρησή της να είναι τριμηνιαία ή ετήσια και πλούσια σε αναχρονιστικές συγκεντρώσεις (Krahel & Titera, 2015). Άυλα στοιχεία (Lev, 2000), FIFO, LIFO, μετρήσεις ιστορικού κόστους και εκτιμήσεις της ετήσιας απόσβεσης, δεν είναι ακριβείς μετρήσεις σε σύγκριση με αυτό που μπορεί να ληφθεί χρησιμοποιώντας την τρέχουσα τεχνολογία. Για παράδειγμα, οι νέοι λογιστές εκπλήσσονται όταν πρέπει να μάθουν διάφορες αναχρονιστικές λογιστικές μεθόδους απογραφής, από τη στιγμή που όλα στην εμπειρία τους προσδιορίζονται από RFID ή barcode. Δεδομένης της μετατόπισης από τις παραδοσιακές λογιστικές πρακτικές στις γνωστοποιήσεις αποτελεσμάτων σε μια βάση αναφοράς

πραγματικού χρόνου/εξαίρεσης θα οδηγούσε σε ένα πολύ διαφορετικό σύνολο επιχειρηματικών αναφορών (Gal, 2008), εκτός από μια πιο σχετική και γραφική αποκάλυψη υποστηρίζοντας την ανάλυση και την παροχή πινάκων ελέγχου διαχείρισης, ελεγκτή και ενδιαφερομένων.

Καθώς η απώλεια της λογιστικής πληροφορίας έχει προχωρήσει, οι εταιρείες έχουν επεκτείνει δραματικά τα παραδοσιακά τους αποθέματα δεδομένων, διατηρώντας έτσι μεγάλες βάσεις δεδομένων στα συστήματα ERP των οποίων μόνο ένα ελάχιστο μέρος σχετίζεται άμεσα με τις χρηματοοικονομικές αναφορές. Οι εσωτερικές αναφορές έχουν γίνει πολύ πλούσιες με τεράστιες ομάδες πληροφοριών που δεν έχουν άμεσο οικονομικό χαρακτήρα, όπως ανθρώπινοι πόροι, παραγωγή, μάρκετινγκ, αλυσίδα εφοδιασμού και έρευνα & ανάπτυξη. Παρόλο που αυτά τα «αποθηκευμένα δομημένα δεδομένα» σε ERP είναι μεγάλα, μπορούν να επισκιάζονται από εκτεταμένα σύνολα λιγότερο δομημένων δεδομένων.

Παρόλη την αύξηση στην έρευνα των Big Data και την εξάρση των δημοσιεύσεων το τελευταίο διάστημα (Abbasi et al., 2016; Akter & Wamba, 2016; Frizzo-Barker et al., 2016) όλως περιέργως κανένας δεν εστιάζει ειδικά στη λογιστική και τη χρηματοοικονομική. Παραδόξως, αφού οι συγκεκριμένοι κλάδοι παραδοσιακά ασχολούνται με μεγάλους όγκους δεδομένων στις αναλύσεις τους (**Σχήμα 2**), η πραγματική χρήση των Big Data θα περίμενε κανείς να μη βρίσκεται σε τόσο πρώιμο στάδιο (Jeacle & Carter, 2011; O'Shea & Shah, 2014; Scott & Orlikowski, 2012; Shen & Chen, 2018). Η έκθεση Turner (Turner et al., 2013) που παρήχθη από κοινού με τη Said School of Business και την IBM, σημειώνει ορισμένες εξελίξεις στη χρήση αναλυτικών στοιχείων στον τραπεζικό κλάδο. Η έκθεση σημειώνει ότι περισσότερες από τις μισές προσπάθειες των Big Data που πραγματοποιούνται από εταιρείες χρηματοοικονομικών υπηρεσιών επικεντρώνονται στην επίτευξη αποτελεσμάτων πελατών. Ωστόσο, ο τραπεζικός κλάδος υστερεί όσον αφορά τη Big Data δεδομένων και αναλυτικών στοιχείων για βασικές δυνατότητες ανάλυσης, ειδικά φυσική γλώσσα και επεξεργασία κειμένου, γεωχωρική, ροή και ανάλυση φωνής/βίντεο (Turner et al., 2013).



Σχήμα 2, Ταξινόμηση των Big Data

Η ανάλυση δεδομένων και τα Big Data θα αλλάξουν αναπόφευκτα τον ρόλο των λογιστών, αλλά αυτό δεν σημαίνει ότι οι λογιστές θα καταστούν άνευ αντικειμένου. Αντ' αυτού η επανάσταση των Big Data θα οδηγήσει σε αυτοματοποίηση των πιο συνηθισμένων και παρωχημένων λειτουργιών, επιτρέποντας στους λογιστές να επικεντρώσουν την προσοχή τους σε ευκαιρίες παροχής αξίας στους οργανισμούς και τους πελάτες τους. Οι Frey & Osborne (2017) κατηγοριοποιούν τη λογιστική και τον έλεγχο ως πιθανό να αυτοματοποιηθούν λόγω έλλειψης μη προσομοιώσιμων δεξιοτήτων, προωθώντας το επιχείρημα ότι οι λογιστές δεν έχουν την ικανότητα να σκέφτονται στρατηγικά και να αξιοποιούν τις επιχειρηματικές τους γνώσεις για να αυξήσουν την αξία που παρέχεται από τα Big Data analytics. Στην ουσία, υποστηρίζουν ότι οι δεξιότητες και οι γνώσεις των λογιστών είναι συμπληρωματικές των Big Data, όντας πολύτιμη για τη μεγιστοποίηση της αξίας μέσω της χρήσης των Big Data analytics σε ένα επιχειρηματικό περιβάλλον.

Αντί να δουν το ρόλο τους να μειώνεται, η εμπειρία των λογιστών στη συλλογή, συγχώνευση και ερμηνεία δεδομένων από πολλές πηγές είναι πιθανό να κάνει τους λογιστές ακόμη πιο πολύτιμους με την εμφάνιση των Big Data analytics. Αν και τα μη δομημένα

δεδομένα παρέχουν στις επιχειρήσεις πολύτιμες πληροφορίες, πρέπει να ερμηνεύονται στο πλαίσιο της εταιρείας. Το επιχειρηματικό πνεύμα των λογιστών, σε συνδυασμό με τις ισχυρές οικονομικές γνώσεις και τις δεξιότητες στρατηγικού σχεδιασμού, τους παρέχει τις απαραίτητες δυνατότητες για να το κάνουν αυτό - δεξιότητες που είναι επίσης ανθεκτικές στον αυτοματισμό. Παρόμοια με τη συμπλήρωση χρηματοοικονομικών μέτρων με μη χρηματοοικονομικά μέτρα στο BSC, η αύξηση δομημένων δεδομένων με μη δομημένα δεδομένα παρέχει στους διευθυντές μια σαφέστερη εικόνα.

Τα Big Data αντιπροσωπεύουν μια μεγάλη αλλαγή στον τρόπο με τον οποίο οι επιχειρήσεις θα δραστηριοποιούνται στο μέλλον. Πολλές θέσεις εργασίας και εργασίες θα καταστούν άνευ αντικειμένου κατά την εφαρμογή τους. Το αν τα Big Data αποτελούν απειλή ή ευκαιρία για το επάγγελμα του λογιστή εξαρτάται από τους λογιστές. Προκειμένου να διασφαλιστεί ότι οι λογιστές έχουν θέση στον κόσμο των Big Data, απαιτείται κατάλληλη εκπαίδευση και κατάρτιση σε όλα τα επίπεδα, από φοιτητές πανεπιστημίου έως τη συνεχιζόμενη εκπαίδευση εξασκούντων λογιστών. Επιπλέον, οι φορείς και τα ιδρύματα καθορισμού προτύπων που καθορίζουν τις βέλτιστες πρακτικές πρέπει να ενσωματώσουν τις αλλαγές που ενυπάρχουν στον κόσμο των Big Data στις οδηγίες και τις συστάσεις τους. Η λογιστική θα πρέπει να αλλάξει ως απάντηση στα Big Data (Richins et al., 2017).

Υπάρχουν βέβαια και οι ερευνητές που υποστηρίζουν ότι μια αλλαγή στα πρότυπα ώστε αυτά να επικεντρωθούν στα Big Data, τις διαδικασίες που τα δημιουργούν και την ανάλυσή τους, αντί της παρουσίας τους, θα προσθέσουν αξία και συνάφεια στο λογιστικό επάγγελμα, θα ενδυναμώσουν τους τελικούς χρήστες και θα βελτιώσουν την αποτελεσματικότητα των κεφαλαιαγορών (Kraheil & Titera, 2015). Τα λογιστικά και ελεγκτικά πρότυπα συμβάλλουν στη διαχείριση ενός συμβιβασμού μεταξύ της επιθυμίας των χρηστών για περισσότερες πληροφορίες και το κόστος προετοιμασίας και μετάδοσης αυτών. Παλαιότερα αναγκάστηκε να διαχειριστεί την έλλειψη πληροφοριών, οι επιχειρήσεις τώρα κοιτάν να εκμεταλλευτούν τους τεράστιους όγκους δεδομένων και μεταδεδομένων που γεμίζουν petabytes του χώρου των διακομιστών τους. Παρομοίως, δεδομένα από πολλές πηγές και σε πολλές μορφές, πολλές από αυτές που είναι άσχετοι, άφθονοι επενδυτές. Τα λογιστικά και ελεγκτικά πρότυπα δεν έχουν διατηρήσει το ρυθμό, δίνοντας έμφαση στην παρουσίαση, τη συγκέντρωση και τη δειγματοληψία.

2. 2. Εφαρμογές στην ελεγκτική

Τα ελεγκτικά τεκμήρια ορίζονται ως το σύνολο των πληροφοριών που συλλέγονται και αξιολογούνται από τους ελεγκτές όταν αποφασίζουν εάν οι οικονομικές καταστάσεις μιας εταιρείας δηλώνονται σύμφωνα με τις γενικά αποδεκτές λογιστικές αρχές (SAS No. 106, AICPA 2004). Στην πράξη οι εξωτερικοί ελεγκτές ασχολούνται με μεγάλες ποσότητες πληροφοριών που, παρόλο που είναι μικρότερες από το σύνολο των πιθανών ελεγκτικών τεκμηρίων, πληρούν την αποδεκτή επαγγελματική απαίτηση να είναι «επαρκής και κατάλληλες» (SAS No. 106, AICPA 2004). Τα ελεγκτικά αποδεικτικά στοιχεία μπορούν να ληφθούν μέσω της εξέτασης των υποκείμενων λογιστικών αρχείων καθώς και από άλλες επιβεβαιωτικές πηγές πληροφοριών όπως παρατηρήσεις, επιβεβαιώσεις από τρίτα μέρη και οποιαδήποτε άλλη πληροφορία που μπορεί να παρέχει εύλογη βάση για συμπεράσματα (Louwers et al., 2015).

Οι ελεγκτές γίνονται πιο ολιστικοί στις αξιολογήσεις κινδύνου ελέγχου, εξετάζοντας αποδεικτικά στοιχεία διαθέσιμα από διάφορες πηγές, προκειμένου να μειωθεί η πιθανότητα ουσιώδους ανακρίβειας και αποτυχίας ελέγχου (Bell et al., 2005). Αυτή η προσέγγιση διευκολύνθηκε από τις νέες τεχνολογίες που παρέχουν στους ελεγκτές μια ευρύτερη ποικιλία τόσο χρηματοοικονομικών όσο και μη χρηματοοικονομικών πληροφοριών, καθώς και βελτιωμένη απόδοση ελέγχου που προκύπτει από τη μηχανογράφηση και τον αυτοματοποιημένο έλεγχο (Trompeter & Wright, 2010).

Στην εποχή των Big Data, τεχνικές όπως η αναγνώριση προτύπων, η συλλογή δεδομένων και η επεξεργασία φυσικής γλώσσας έχουν βελτιώσει την προγνωστική ισχύ των ρουτινών ανάλυσης δεδομένων. Κατά συνέπεια, αναμένεται ότι οι αποφάσεις θα βασίζονται σε περισσότερα δεδομένα από ότι στην εμπειρία (Lohr, 2012). Δεδομένης αυτής της νέας πραγματικότητας αναμένεται ότι με τη χρήση των Big Data οι ελεγκτικές προσπάθειες και η συλλογή επαρκών και κατάλληλων ελεγκτικών τεκμηρίων να είναι ενισχυμένες. Τα Big Data μπορούν να θεωρηθούν ως αποδεικτικά στοιχεία ελέγχου από την προοπτική αποδεικτικών στοιχείων. Καθώς αποτελούν εν δυνάμει πολύτιμο συμπλήρωμα των παραδοσιακών στοιχείων ελέγχου λόγω των ιδιαίτερων χαρακτηριστικών τους. Χωρίς ωστόσο να παραλείψουμε ότι η χρήση των Big Data κρύβει προκλήσεις κατά τη διαδικασία ελέγχου.

Παρόλο που τα επαγγελματικά πρότυπα έχουν προσπαθήσει να κωδικοποιήσουν τις αποδεικτικές απαιτήσεις, η πειστικότητα των συλλεγόμενων αποδεικτικών στοιχείων είναι ζήτημα επαγγελματικής κρίσης (SAS No. 106, AICPA 2004). Σύμφωνα με τα επαγγελματικά

πρότυπα, τα ελεγκτικά τεκμήρια πρέπει να είναι «επαρκή και κατάλληλα», όπου «κατάλληλο» αναφέρεται στο «αξιόπιστο» και «σχετικό» (SAS No. 106, AICPA 2004). Τα Big Data συμβάλλουν στην απαίτηση «επάρκειας» λόγω του όγκου και της ποικιλίας των δεδομένων που παρέχονται σε πραγματικό χρόνο. Επειδή η επάρκεια εξαρτάται από τον κίνδυνο ανακρίβειας και την καταλληλότητα (δηλ. αξιοπιστία και συνάφεια) των ελεγκτικών τεκμηρίων που συλλέγονται (SAS No. 106, AICPA 2004), απαιτούνται όλο και περισσότερα στοιχεία από τα Big Data όταν έχουν χαμηλότερη αξιοπιστία και συνάφεια ή αντίστοιχα λιγότερα στοιχεία στην περίπτωση ύπαρξης υψηλής συνάφειας. Τα Big Data μπορούν να είναι αρκετά αξιόπιστα επειδή συχνά δημιουργούνται και αποκτώνται εξωγενώς από τους ελεγκτές απευθείας. Από την άλλη πλευρά, ο θόρυβος στα Big Data μπορεί να προκαλέσει υπερφόρτωση ψευδών θετικών, οδηγώντας σε χαμηλότερη αξιοπιστία. Ένα κρίσιμο βήμα για τη χρήση των Big Data είναι η εξάλειψη του θορύβου (αλήθεια) για τον στόχο ελέγχου. Όσον αφορά τη συνάφεια των Big Data, παρέχει μοναδικές και μερικές φορές πιο έγκαιρες αποδείξεις από τις παραδοσιακές πηγές. Για παράδειγμα, τα Big Data analytics είναι ιδιαίτερα χρήσιμο για στόχους ελέγχου που σχετίζονται με πελάτες (Russom, 2011).

Οι αποδεικτικές απαιτήσεις παρέχουν ένα πλαίσιο για την αξιολόγηση της χρησιμότητας των Big Data ως ελεγκτικών τεκμηρίων.

2. 2. 1. Εξέταση της επάρκειας των Big Data

Θεωρητικά, οι ελεγκτές θα πρέπει να έχουν πρόσβαση σε όλες τις απαραίτητες πληροφορίες μιας εταιρείας που υπόκειται έλεγχο (Dunn, 1996). Στην πραγματικότητα, το εύρος και η ποιότητα των συλλεγόμενων αποδεικτικών στοιχείων/δεδομένων προς έλεγχο επηρεάζονται από την τεχνολογία δηλαδή εάν και κατά πόσο τα αποδεικτικά στοιχεία είναι σε ηλεκτρονική μορφή, περιορισμοί κόστους/οφέλους και οι κοινωνικές αλληλεπιδράσεις με τους πελάτες (Bennett & Hatfield, 2013; Louwers et al., 2015). Τα σχετικά εξωτερικά Big Data θα μπορούσαν να αποτελέσουν συμπλήρωμα για τις εσωτερικές πληροφορίες ενός πελάτη που δεν είναι άμεσα διαθέσιμες στους ελεγκτές. Για παράδειγμα, κατά τον έλεγχο μιας κατασκευαστικής ανησυχίας, οι ελεγκτές μπορούν να ζητήσουν τις προβλέψεις πωλήσεων της διοίκησης, επειδή μπορούν να χρησιμοποιηθούν για την κατανόηση του όγκου παραγωγής και των επιπέδων αποθέματος (Louwers et al., 2015). Εάν οι προβλέψεις πωλήσεων των διαχειριστών δεν είναι διαθέσιμες ή είναι χαμηλής ποιότητας, τότε οι ελεγκτές θα μπορούσαν να χρησιμοποιήσουν την ανάλυση κειμένου για να αναλύσουν τα Big Data από άρθρα

ειδήσεων, φόρουμ συζήτησης προϊόντων και κοινωνικά δίκτυα για να κατανοήσουν καλύτερα τις τάσεις πωλήσεων του πελάτη.

Τα Big Data μπορούν να προσφέρουν υποστήριξη όταν τα παραδοσιακά αποδεικτικά στοιχεία ελέγχου είναι ανεπαρκή, όπως μπορεί να ισχύει σε περίπτωση απάτης. Η απόκτηση αποδεικτικών στοιχείων για απάτη είναι δύσκολη επειδή τα αυτούσια στοιχεία για δεδομένα που συνδέονται με κίνητρα και εξορθολογισμό σχετίζονται με τον τρόπο ζωής, τη συμπεριφορά και την ηθική ενός ατόμου (SAS No. 99, AICPA 2002), κανένα από τα οποία δεν είναι απαραίτητα παρατηρήσιμο. Η αξιολόγηση των μηνυμάτων ηλεκτρονικού ταχυδρομείου μπορεί να είναι ιδιαίτερα χρήσιμη για τον εντοπισμό των κινήτρων ενός ατόμου και τον πιθανό εξορθολογισμό, όπως η δυσαρέσκεια για μια εταιρεία. Ερευνητές (Holton, 2009) ήδη έχουν χρησιμοποιήσει την αυτοματοποιημένη συλλογή πληροφοριών από κείμενο για τον εντοπισμό μηνυμάτων ηλεκτρονικού ταχυδρομείου δυσαρεστημένων υπαλλήλων για περιπτώσεις εντοπισμού απάτης εκ μέρους τους.

Το κύριο όφελος που σχετίζεται με την επάρκεια είναι η αφθονία πληροφοριών που παρέχεται από διαφορετικές μορφές δεδομένων σε τεράστιες ποσότητες. Το πρωταρχικό κόστος είναι η προσπάθεια επεξεργασίας δεδομένων που είναι απαραίτητη για την επίτευξη συγκεκριμένου ισχυρισμού ελέγχου. Ευτυχώς, είναι διαθέσιμα προηγμένα αναλυτικά δεδομένα και αυτά τα εργαλεία ανάλυσης είναι πιο ισχυρά για μεγαλύτερα σύνολα δεδομένων και φιλικά προς μη δομημένα δεδομένα (Russom, 2011). Οι ελεγκτές μπορούν επίσης να δημιουργήσουν τις δικές τους αποθήκες δεδομένων για να επιτύχουν οικονομίες κλίμακας μεταξύ των πελατών για τη μείωση του κόστους επεξεργασίας δεδομένων.

2. 2. 2. Εξέταση αξιοπιστίας των Big Data

Ορισμένοι τύποι Big Data μπορούν να βοηθήσουν στην αξιολόγηση της αξιοπιστίας των παραδοσιακών στοιχείων ελέγχου. Για παράδειγμα, τα έγγραφα αποστολής χρησιμοποιούνται για την επαλήθευση των αποστολών, αλλά τα δεδομένα GPS παρέχουν μια πιο αξιόπιστη, ανθεκτική σε παραβίαση πηγή δεδομένων για επαλήθευση. Στον παραδοσιακό έλεγχο, τα έγγραφα ελέγχονται χειροκίνητα για την επαλήθευση επιχειρηματικών συναλλαγών. Στο περιβάλλον Big Data, οι ελεγκτές μπορούν να χρησιμοποιούν τεχνικές ανάλυσης κειμένου, όπως ομαδοποίηση, για την αυτόματη ανάλυση και σύνοψη εγγράφων (Dhillon & Modha, 2001). Αυτή η προσέγγιση είναι πιο αποτελεσματική και αποκαλυπτική από τη χειροκίνητη επιθεώρηση.

Η χρήση μη χρηματοοικονομικών πληροφοριών ως μέρος των αναλυτικών διαδικασιών μπορεί να αποτελέσει ανεξάρτητο σημείο αναφοράς για την αξιολόγηση των οικονομικών καταστάσεων (SAS No. 56, AICPA 1988). Σε αυτήν τη γραμμή, τα Big Data από εξωτερικές πηγές, όπως άρθρα ειδήσεων, αναφορές αναλυτών και κυβερνητικές εκθέσεις, θα μπορούσαν να προσφέρουν ανεξάρτητα σημεία αναφοράς για την αξιολόγηση τόσο των εσωτερικών όσο και των εξωτερικών τάσεων των χρηματοοικονομικών λογαριασμών. Ανάλυση του επιπέδου ικανοποίησης των καταναλωτών (Ittner & Larcker, 1998) ή του καθημερινού καιρού (Engle et al., 1986; Starr, 2000) μπορεί να μας βοηθήσει στην κατανόηση του επιπέδου των πωλήσεων. Για παράδειγμα, εάν η φήμη ενός συγκεκριμένου προϊόντος στα κοινωνικά δίκτυα είναι αρνητική, αλλά οι πωλήσεις του προϊόντος έχουν αυξηθεί, τότε οι ελεγκτές θα μπορούσαν να δουν αυτήν την ασυνέπεια ως «κόκκινη σημαία».

Μια σημαντική πηγή αξιοπιστίας των Big Data είναι ότι είναι εξαιρετικά δύσκολο να παραβιαστούν τα δεδομένα λόγω του τεράστιου μεγέθους τους, ειδικά όταν τα δεδομένα δημιουργούνται σε πραγματικό χρόνο από εξωτερικές πηγές. Το κύριο μέλημα είναι η ποιότητα των δεδομένων. Η αξιοπιστία μειώνεται όταν ο θόρυβος στα Big Data οδηγεί σε υπερφόρτωση ψευδών θετικών. Επίσης, Big Data από ιστότοπους κοινωνικών μέσων όπως το Twitter™ θα μπορούσε να ενέχουν μεροληπτικές τάσεις καθόσον οι χρήστες τους δεν αντιπροσωπεύουν ολόκληρο τον πληθυσμό των πελατών (Tufekci, 2013).

2. 2. 3. Εξέταση συνάφειας των Big Data

Οι ελεγκτές καλούνται να κατανοήσουν το ευρύ οικονομικό περιβάλλον στο οποίο λειτουργούν οι πελάτες τους και να αναλύσουν ανάλογα τους επιχειρηματικούς κινδύνους των πελατών τους (Louwers et al., 2015). Το περιεχόμενο των ειδησεογραφικών άρθρων είναι πιθανό να υποδεικνύει τα μελλοντικά κέρδη και τις προοπτικές των τιμών των μετοχών μιας εταιρείας (Tetlock, 2007; Tetlock et al., 2008), αλλά το παραδοσιακό ελεγκτικό έργο είναι συχνά εκ των υστέρων. Οι ελεγκτές μπορούν να αναλύσουν έγκαιρες αναφορές ειδήσεων για να αξιολογήσουν τις αλλαγές οικονομικής απόδοσης και τον επιχειρηματικό σχεδιασμό των πελατών τους.

Τα επαγγελματικά πρότυπα απαιτούν επίσης από τους ελεγκτές να αξιολογούν τους κινδύνους που σχετίζονται με αδυναμίες εσωτερικού ελέγχου και δόλιες δηλώσεις (SAS No. 107, AICPA 2007). Παλαιότερη βιβλιογραφία έχει προσφέρει διάφορους τρόπους για την αξιολόγηση των κινδύνων των πελατών (Johnstone, 2000) και οι γνωστοποιήσεις διαχείρισης

θα μπορούσαν να είναι ιδιαίτερα χρήσιμες για αυτήν την εργασία. Η SAS No. 99 (AICPA 2002) επισημαίνει ότι τα «υπερβολικά αισιόδοξα δελτία τύπου ή τα μηνύματα ετήσιας έκθεσης» είναι παράγοντες κινδύνου που σχετίζονται με πιθανή απάτη. Σε αυτή τη βάση, διάφοροι ερευνητές (Humpherys et al., 2011) μελέτησαν πιθανώς δόλιες τάσεις αναλύοντας ενότητες όπως η Συζήτηση και η Ανάλυση Διοίκησης. Ομοίως άλλοι (Larcker & Zakolyukina, 2012) κατέληξαν ότι η εκτελεστική χρήση παραπλανητικής γλώσσας σε κλήσεις συνδιάσκεψης μπορεί να βοηθήσει στον εντοπισμό οικονομικών αναφορών. Επομένως, η ανάλυση κειμένου των γνωστοποιήσεων διαχείρισης σχετίζεται με την εκτίμηση του κινδύνου απάτης διαχείρισης.

Η φύση του ηλεκτρονικού εμπορίου προσφέρει μια μοναδική ευκαιρία για χρήση τεχνικών ελέγχου βασισμένων σε Big Data. Υπήρξε μια σημαντική μετατόπιση στη βιομηχανία λιανικής σε πωλήσεις μέσω Διαδικτύου. Παρόλο που οι πωλήσεις ηλεκτρονικού εμπορίου αποτελούν μόνο το 6,6% των συνολικών λιανικών πωλήσεων στις ΗΠΑ το τρίτο τρίμηνο του 2014, αυξάνονται ταχύτερα (16,2% σε σχέση με το τρίτο τρίμηνο του 2013) σε σύγκριση με άλλους παραδοσιακούς τύπους λιανικών πωλήσεων.¹ Αυτή η τάση δείχνει ότι οι ελεγκτές θα αντιμετωπίζουν όλο και περισσότερο πελάτες που έχουν πολύ διαφορετικούς τύπους επιχειρηματικών διαδικασιών, προκαλώντας την ανάγκη συλλογής διαφορετικών μορφών ελεγκτικών τεκμηρίων. Για παράδειγμα, οι ελεγκτές μπορούν να συγκρίνουν τα δεδομένα κίνησης ιστότοπου ενός πελάτη με εκείνα των ανταγωνιστών με παρόμοιους πελάτες για την ίδια χρονική περίοδο. Οποιαδήποτε ασυνέπεια πρέπει να εντοπιστεί για ιδιαίτερη προσοχή, ακόμη και όταν το αρχείο πωλήσεων του πελάτη δεν δείχνει κανένα πρόβλημα.

Η προσέγγιση Big Data είναι σχετική επειδή παρέχει μοναδικά και μερικές φορές πιο έγκαιρα στοιχεία σε σύγκριση με την παραδοσιακή προσέγγιση ελέγχου. Το μεγαλύτερο κόστος είναι ότι τα στοιχεία που παράγονται από τα Big Data υποδηλώνουν κυρίως συσχέτιση, όχι αιτιώδη συνάφεια (Cao et al., 2015). Στο παραπάνω παράδειγμα, η παραπλανητική γλώσσα στις κλήσεις συνδιάσκεψης κερδών δεν προκαλεί οικονομική ανακρίβεια. Συνδέεται με την παραπλανητική συμπεριφορά των CEOs και CFOs, με τον τρόπο αυτό να σχετίζεται με οικονομική ανακρίβεια (Larcker & Zakolyukina, 2012).

2. 2. 4. Επίπεδα αποδεικτικών στοιχείων ελέγχου

Τα ελεγκτικά τεκμήρια μπορούν να κατηγοριοποιηθούν σε τρία επίπεδα: οικονομική δήλωση, ατομικός λογαριασμός και στόχος ελέγχου (Srivastava & Shafer, 1992). Η αποτελεσματικότητα του ελέγχου θα μπορούσε να βελτιωθεί σημαντικά εάν τα ελεγκτικά

τεκμήρια στο επίπεδο των οικονομικών καταστάσεων συμπληρώνονται από τα Big Data που σχετίζονται με την πιθανότητα ανακρίβειας, επειδή οι ελεγκτές τείνουν να συλλέγουν υπερβολικά αποδεικτικά στοιχεία σε λεπτομερή επίπεδα χωρίς να αναλύουν τα στοιχεία σε επίπεδο οικονομικής κατάστασης (Srinastava & Shafer, 1992). Το αποδεικτικό πλαίσιο και οι ουσιαστικές διαδικασίες θα είναι τόσο περιεκτικές όσο εάν εκτιμάται ότι ο κίνδυνος ανακρίβειας. Η ανάλυση κειμένου των γνωστοποιήσεων της εταιρείας και τα αντίγραφα κλήσεων συνδιάσκεψης βοηθούν τους ελεγκτές να εκτιμήσουν την πιθανότητα ανακρίβειας (Humpherys et al., 2011; Larcker & Zakolyukina, 2012). Σε επιμέρους επίπεδα λογαριασμού και αντικειμενικού ελέγχου, τα Big Data analytics θα μπορούσε να μειώσει τα ψευδώς θετικά και να εντοπίσει πιο σημαντικές ανωμαλίες για περαιτέρω έρευνα (Cao et al., 2015; Issa & Kogan, 2014). Τα Big Data από άρθρα ειδήσεων, φόρουμ συζήτησης προϊόντων και κοινωνικά δίκτυα είναι χρήσιμα για την αξιολόγηση των πωλήσεων (επίπεδο ατομικού λογαριασμού). Τα τσιπ αναγνώρισης ραδιοσυχνότητας (RFID) έχουν χρησιμοποιηθεί για την επαλήθευση της ύπαρξης αποθέματος (επίπεδο αντικειμενικού ελέγχου).

2. 2. 5. Συνεχής έλεγχος

Τα Big Data προέρχονται από παραδοσιακά συστήματα συναλλαγών, καθώς και από νέες πηγές όπως email, τηλεφωνικές κλήσεις, δραστηριότητες στο διαδίκτυο, μέσα κοινωνικής δικτύωσης, μέσα ενημέρωσης, ηχογραφήσεις αισθητήρων και βίντεο και ετικέτες RFID. Δεδομένου ότι μεγάλο μέρος αυτών των Big Data ενημερώνει και επηρεάζει εταιρικές αποφάσεις που είναι σημαντικές τόσο για εσωτερικούς όσο και για εξωτερικούς εταιρικούς ενδιαφερόμενους, οι ελεγκτές θα πρέπει να επεκτείνουν το τρέχον πεδίο εφαρμογής της ανάλυσης δεδομένων (Cao et al., 2015).

Λόγω του όγκου και της ταχύτητας, η εφαρμογή συνεχούς ελέγχου (continuous auditing, CA) έχει γίνει όλο και πιο σημαντική για την αυτοματοποίηση και την ανάλυση σε πραγματικό χρόνο των Big Data (Vasarhelyi et al., 2010). Ωστόσο, ο τεράστιος όγκος και η υψηλή ταχύτητα εισάγουν επίσης κενά μεταξύ της τρέχουσας κατάστασης των αναλυτικών στοιχείων ελέγχου και των απαιτήσεων του Big Data analytics σε ένα συνεχές πλαίσιο ελέγχου. Επιπλέον, η ποικιλία και η αβέβαιη αλήθεια παρουσιάζουν προκλήσεις πέρα από την ικανότητα των τρεχουσών μεθόδων CA.

Μια έρευνα του Ινστιτούτου Εσωτερικών Ελεγκτών (Institute of Internal Auditors, IIA) αναφέρει ότι «η ώθηση για χρηματοοικονομικές αναφορές σε πραγματικό χρόνο και η

προσπάθεια αυτοματοποίησης μη αυτόματων ελέγχων συλλογής πληροφοριακών πόρων ωθούν τους οργανισμούς να υιοθετήσουν το συνεχή έλεγχο» (Brannen, 2006). Ο συνεχής έλεγχος είναι «για να παρέχει διασφάλιση ταυτόχρονα με ή για σύντομο χρονικό διάστημα μετά, την εμφάνιση συμβάντων στα οποία βασίζεται το αντικείμενο» (CICA/AICPA 1999) προσδιορίζοντας αυτόματα εξαιρέσεις ή ανακρίβειες όπως ορίζονται από ορισμένα προκαθορισμένα κριτήρια στις ενσωματωμένες ενότητες ελέγχου (Groomer & Murthy, 2018) or a monitoring control layer (Vasarhelyi et al., 2018). Με την ώθηση των Big Data το CA απαιτείται για πρόσβαση και επεξεργασία πολλών πρόσθετων σχετικών πληροφοριών λόγω του τεράστιου όγκου δεδομένων και συναλλαγών. Οι ελεγκτές με αρμοδιότητα στην ανάλυση δεδομένων θα έχουν καλύτερες ευκαιρίες να διευρύνουν το εύρος και να αυξήσουν την κλίμακα του ελέγχου σε συχνότερη βάση μέσω συνεχούς παρακολούθησης δεδομένων (continuous data monitoring, CDM), συνεχούς παρακολούθησης ελέγχου (continuous control monitoring, CCM) και συνεχούς παρακολούθησης και αξιολόγησης κινδύνου (continuous risk monitoring and assessment, CRMA) (Vasarhelyi et al., 2010).

Ωστόσο, ένα από τα πιο σημαντικά ερωτήματα που ανακύπτουν στα τμήματα ελέγχου είναι ο τρόπος αποτελεσματικής παράδοσης αξίας από τα ολοένα αυξανόμενα αποθετήρια Big Data. Για παράδειγμα, μια ροή μεγάλων δεδομένων από πολλές πηγές δεν εξυπηρετεί κανένα πιθανό όφελος εάν δεν υπάρχει σύνδεση μεταξύ τους. Μια τέτοια σύνδεση βασίζεται σε μεγάλο βαθμό από εργαλεία CA και αναλυτικά δεδομένα για τον εντοπισμό ανωμαλιών και ανακρίβειών αυτόματα και έγκαιρα. Λαμβάνοντας υπόψη τα τέσσερα V του Big Data, οι ελεγκτές πρέπει να συνειδητοποιήσουν ότι υπάρχουν κενά μεταξύ των Big Data και των δυνατοτήτων των σύγχρονων συστημάτων CA, όπως φαίνεται στο Σχήμα 1. Συγκεκριμένα, τα αρχικά τρία V του τεράστιου όγκου, της υψηλής ταχύτητας και της τεράστιας ποικιλίας εισάγουν το κενά της συνοχής των δεδομένων, της αναγνώρισης δεδομένων και της συγκέντρωσης δεδομένων για τη σύνδεση βάσεων δεδομένων σε ένα σύστημα CA. Η αλήθεια ή ορθότητα των δεδομένων έχει προστεθεί στα αρχικά τρία V του Big Data, καθώς η ακεραιότητα των πληροφοριών γίνεται ζήτημα χωρίς ασφαλή αρχεία καταγραφής ή προέλευση δεδομένων. Εν τω μεταξύ, καθώς το Big Data γίνεται μια σημαντική πηγή ανάλυσης, το απόρρητό του είναι μια άλλη πηγή ανησυχίας τόσο για εταιρικά όσο και για συστήματα CA. Κάθε ένα από τα κενά δημιουργεί τις αντίστοιχες προκλήσεις στο Σχήμα 1, οι οποίες θα αναλυθούν παρακάτω.

Δεδομένου ότι τα κενά και οι προκλήσεις θα γίνουν κοινά για τη διαδικασία ελέγχου των Big Data, οι εφικτές λύσεις ζητούν μια ενημερωμένη εφαρμογή της CA. Τα κύρια στοιχεία

της αρχιτεκτονικής συνεχούς ελέγχου, όπως η παροχή δεδομένων, το φιλτράρισμα δεδομένων και το επίπεδο διαγνωστικών δεδομένων, πρέπει να προσαρμοστούν για να αντιμετωπίσουν τις προκλήσεις που παρουσιάζουν τα Big Data (Vasarhelyi et al., 2018).

Συνοχή δεδομένων

Η συνέπεια των δεδομένων είναι το πιο σημαντικό ζήτημα για το CA των Big Data και σχετίζεται με αλληλεξαρτώμενα δεδομένα μεταξύ εφαρμογών και σε έναν οργανισμό (Sheth & Rusinkiewicz, 1990). Δεδομένου ότι υπάρχει ένας αυξανόμενος αριθμός διαφορετικών πηγών δεδομένων στη σημερινή εποχή των Big Data, η συχνότητα των διενέξεων δεδομένων έχει αναπόφευκτα αυξηθεί. Η κύρια αιτία εδώ είναι ότι τα συστήματα Big Data που υποστηρίζουν βασικές επιχειρηματικές διαδικασίες αποτελούνται συνήθως από ένα συνονθύλευμα διαφορετικών συστημάτων, όπου τα δεδομένα μπορούν να αναπαραχθούν πλήρως ή εν μέρει, το ενημερωτικό περιεχόμενο μπορεί να επικαλύπτεται και μπορεί να αποθηκεύονται περισσότερα παράγωγα δεδομένα. Αυτή η κατάσταση δημιουργεί το σοβαρό κενό στη συνοχή των δεδομένων. Για παράδειγμα, τα δεδομένα μπορούν να χαρτογραφηθούν και να μειωθούν σε μια πλατφόρμα όπως του Hadoop. Επομένως, η νέα προσέγγιση του CA πρέπει να επαληθεύσει τη σχέση μεταξύ των πηγών δεδομένων και να ελέγξει τη συνέπεια των δεδομένων.

Υπάρχουν τουλάχιστον τρεις τύποι ασυνεπειών δεδομένων που πρέπει να διαχειρίζονται σε ένα σύστημα CA. Η πρώτη ασυνέπεια συμβαίνει σε μορφές δεδομένων, όπου δεδομένα από διαφορετικές πηγές ενδέχεται να διαφέρουν ως προς τη δομή. Σημαντικοί ερευνητές (Koudas, Saha, Srivastava, & Venkatasubramanian, 2009) ανέπτυξαν ένα μοντέλο για την επίλυση αυτού του τύπου ασυνέπειας, όπου διάφορες μορφές δεδομένων μπορούν να συσχετιστούν με βάση τις υποκείμενες εξαρτήσεις τους (δηλαδή, σχέσεις μεταξύ χαρακτηριστικών σε βάσεις δεδομένων). Για παράδειγμα, συμπεριλαμβάνοντας τη συναλλαγματική ισοτιμία στο μοντέλο εξάρτησης, το σύστημα CA μπορεί αυτόματα να συσχετίσει τις πωλήσεις 1 εκατομμυρίου δολαρίων ΗΠΑ που καταγράφονται από τα κεντρικά γραφεία στις ΗΠΑ με τις πωλήσεις 6,2 εκατομμυρίων RMB στη βάση δεδομένων θυγατρικών στην Κίνα. Το δεύτερο ζήτημα είναι ο συγχρονισμός δεδομένων, ο οποίος διασφαλίζει ότι τα δεδομένα είναι συγχρονισμένα σε ολόκληρο τον οργανισμό. Ένα παράδειγμα τέτοιου συγχρονισμού απεικονίστηκε από άλλους ερευνητές (Golab et al., 2009). Υποκείμενες διαδοχικές εξαρτήσεις (SD) που προέρχονται από φαινομενικά διαφορετικά δεδομένα

χρησιμοποιήθηκαν για την κατασκευή των σχέσεων και του συγχρονισμού με δεδομένα πραγματικού κόσμου. Η τρίτη και ίσως η πιο σοβαρή ασυνέπεια είναι αυτή της αντίφασης δεδομένων: τα δεδομένα από μια πηγή ενδέχεται να έρχονται σε αντίθεση με δεδομένα από άλλη πηγή. Οι Chiang και Miller (2008) παρείχε έναν αλγόριθμο βασισμένο σε εξαρτημένες λειτουργικές εξαρτήσεις για την αντιμετώπιση αυτής της μορφής ασυνέπειας. Επιπλέον, υπάρχουν πρόσθετες ερευνητικές μελέτες σχετικά με τις αξιολογήσεις της συνοχής των δεδομένων που μπορούν επίσης να χρησιμοποιηθούν για την ενημέρωση του συστήματος CA. Για παράδειγμα ο Fan και συνεργάτες (Fan & Geerts, 2010; Fan et al., 2012) σχεδίασαν ένα σύστημα ελέγχου συνέπειας δεδομένων σε ένα περιβάλλον αποθήκευσης όπου τα δεδομένα διαχωρίστηκαν, όπου άλλοι ερευνητές (Fuxman & Miller, 2007) χρησιμοποίησαν μια μέθοδο ερωτήματος δεδομένων για να επιδιορθώσουν απλά ασυνεπή δεδομένα.

Ακόμα κι αν έχουμε την εμπειρία να εφαρμόσουμε τις προαναφερθείσες τεχνικές στα τρία ζητήματα ασυνέπειας δεδομένων στην αποθήκευση δεδομένων, πώς να ενσωματώσουμε αυτές τις τεχνικές σε ένα σύστημα CA Big Data χωρίς να χάσει την αποτελεσματικότητά του αξίζει μελλοντική έρευνα. Ταυτόχρονα, για να αποφευχθεί ουσιαστικά το ζήτημα χαμηλής απόδοσης, ένα καθολικό πρότυπο μορφής δεδομένων ελέγχου για το CA μπορεί να είναι μια πιθανή λύση για την αντιμετώπιση του ζητήματος της ασυνέπειας της μορφής δεδομένων. Επιπλέον, χρειαζόμαστε περαιτέρω συζήτηση σχετικά με το πώς αυτό το καθολικό πρότυπο δεδομένων θα μπορούσε να περιλαμβάνει πρόσθετα χαρακτηριστικά δεδομένων, όπως η διαδοχική εξάρτηση δεδομένων και πηγή δεδομένων, σε μια προσπάθεια να ελαχιστοποιηθούν αυτά τα θέματα ασύγχρονης και αντίφασης.

Ακεραιότητα δεδομένων

Η ακεραιότητα των δεδομένων είναι ένας όρος ευρείας εμβέλειας. Στον τομέα του ελέγχου, η έλλειψη ακεραιότητας δεδομένων σχετίζεται συνήθως με παραβίαση ή αλλοίωση δεδομένων (Menezes et al., 2018) και ελλιπή δεδομένα (Motro, 1989). Στο περιβάλλον των Big Data, ο όγκος και οι τύποι δεδομένων είναι τόσο εκτεταμένοι που καθίσταται πιο δύσκολο να εντοπιστούν μεμονωμένα δεδομένα καθώς και σύνολα δεδομένων που έχουν τροποποιηθεί/διαγραφεί/κρυφτεί/καταστραφεί λόγω σφάλματος λειτουργίας, διαδικαστικού σφάλματος, παράνομης πρόσβασης, και/ή αποτυχίες μετάδοσης δικτύου. Αυτή η δυσκολία στον εντοπισμό ζητημάτων ακεραιότητας μπορεί να δημιουργήσει ένα φαινόμενο ντόμινο που αναγκάζει άλλα αξιόπιστα δεδομένα να χάσουν την αξία τους για σκοπούς ανάλυσης ελέγχου,

αυξάνοντας έτσι τον κίνδυνο ελέγχου σε ένα συνεχές περιβάλλον ελέγχου μεγάλων δεδομένων. Οι παραδοσιακές μέθοδοι επαλήθευσης της ακεραιότητας των δεδομένων, όπως η λογική, οι έλεγχοι επεξεργασίας και η σύγκριση με άλλες πηγές, ενδέχεται να μην είναι πρακτικές προς το παρόν με άμεση εφαρμογή ελέγχου στα Big Data. Πρέπει να αναπτυχθούν προσεγγίσεις για την επιδιόρθωση προβλημάτων ακεραιότητας δεδομένων που όχι μόνο βελτιώνουν τον πραγματικό χρόνο και τον αυτοματοποιημένο έλεγχο, αλλά βελτιώνουν επίσης την αποτελεσματικότητα των συνεχών ρουτινών παρακολούθησης της διοίκησης.

Υπάρχουν δύο κυρίαρχοι τύποι ζητημάτων όσον αφορά την ακεραιότητα στα Big Data που πρέπει να αντιμετωπιστούν. Το πρώτο είναι η τροποποίηση δεδομένων, όπου μπορεί να αναφέρεται στην ακούσια τροποποίηση που μπορεί να συμβεί με τέτοια συστήματα συγκεντρωτικών δεδομένων Big Data, όπως το Hadoop ή MapReduce. Επιπρόσθετα, με τον αυξημένο αριθμό ατόμων που ασχολούνται με την παραγωγή, επεξεργασία και μετασχηματισμό δεδομένων, αυξάνοντας τη πιθανότητα εκούσιας τροποποίησης δεδομένων. Οι τεχνικές εύρεσης μηνυμάτων παρέχουν έναν αποτελεσματικό τρόπο δοκιμής τροποποιημένων δεδομένων, ιδίως βασικών δεδομένων σύμφωνα με το Menezes και τους συνεργάτες του (Menezes et al., 2018) που έχουν συνοψίσει μια ποικιλία αλγορίθμων συλλογής μηνυμάτων που μπορούν να χρησιμοποιηθούν σε συνθήκες συνεχούς ελέγχου. Το δεύτερο ζήτημα ακεραιότητας είναι τα ελλιπή δεδομένα, τα οποία μπορούν να προκύψουν όταν ένας υπάλληλος εισάγει μερικές εγγραφές ημιτελών συναλλαγών ή εγγραφές με πεδία που λείπουν σε μια εταιρική βάση δεδομένων. Οι κλασικές μέθοδοι αντιμετώπισης της περίπτωσης ελλιπών δεδομένων είναι η διαγραφή κατά διαστήματα και η αντικατάστασή τους (Schafar & Graham, 2002). Επομένως, προτείνονται εποικοδομητικοί τρόποι όπως η εκτίμηση και ο καταλογισμός. Για παράδειγμα ερευνητές έχουν ήδη προσφέρει ένα σχήμα για την αναζήτηση πλήρων αποτελεσμάτων/ερωτημάτων από μια ελλιπή βάση δεδομένων (Motro, 1989). Ο αλγόριθμος βασίζεται στο πλαίσιο που η ακεραιότητα των δεδομένων συνάδει με την εγκυρότητα και πληρότητα αυτών. Επιπλέον οι Mayfield, Neville και Prabhakar (2010) πρότειναν ένα μοντέλο πιθανότητας με το όνομα ERACER για την εκτίμηση της τιμής και της επίδρασης των ελλιπών δεδομένων σε ένα σύστημα.

Παρόλο που οι παραπάνω τεχνικές επιδιόρθωσης ελλιπών δεδομένων μπορούν και είναι αποτελεσματικές σε περιπτώσεις, αυτό δεν είναι δυνατό κατά την επεξεργασία μεγάλου όγκου δεδομένων. Επομένως, είναι απαραίτητες νέες και πολύ αποδοτικές τεχνικές επισκευής δεδομένων. Ταυτόχρονα, απαιτούνται νέα αναλυτικά παραδείγματα για τον έλεγχο των Big

Data που περιέχουν ελλιπείς πληροφορίες, ώστε να αποφευχθεί η αναποτελεσματική διαδικασία επισκευής τους.

Αναγνώριση δεδομένων

Η αναγνώριση δεδομένων με τη σειρά της αναφέρεται σε εγγραφές που συνδέουν δύο ή περισσότερα ξεχωριστά καταγεγραμμένα στοιχεία για το ίδιο άτομο ή οντότητα (Newcombe et al., 1959). Όταν τα δεδομένα είναι δομημένα, η αναγνώρισή τους είναι εύκολη. Ωστόσο, στον έλεγχο των Big Data όπου πολλά από τα δεδομένα ενδέχεται να είναι σε μη δομημένη μορφή, η αναγνώριση γίνεται δύσκολη. Για παράδειγμα το ποσό εσόδων για μια δεδομένη πώληση μπορεί εύκολα να αναγνωριστεί από το σύστημα συνεχούς ελέγχου, αλλά μπορεί να είναι δύσκολο να συνδεθούν αυτόματα αυτές οι πληροφορίες με τους σχετικούς όρους και προϋποθέσεις πωλήσεων, οι οποίοι είναι σε μη δομημένη μορφή κειμένου.

Υπάρχουν πολλές καινοτόμες μέθοδοι για την αντιμετώπιση του ζητήματος αναγνώρισης δεδομένων και θα πρέπει να εξεταστεί η καταλληλότητά τους για τον συνεχή έλεγχο των Big Data. Οι κύριες μέθοδοι μπορούν γενικά να συνοψιστούν σε δύο τεχνικές: η μία βασίζεται σε εφαρμογή σημασιολογικών κανόνων και η άλλη βασίζεται σε μέτρα ομοιότητας. Όσον αφορά τη πρώτη, συχνά βασίζεται στην εμπειρία και τις γνώσεις που απαιτούνται για τον καθορισμό των κανόνων αναγνώρισης. Οι Lim, Srivastava, Prabhakar & Richardson (1996), καθώς και οι Hernandez, Stolfo (1998) πρότειναν απλές μεθόδους για τον προσδιορισμό της ισοδυναμίας των σημείων δεδομένων χρησιμοποιώντας χαρακτηριστικά αναγνώρισης κλειδιών που αποθηκεύονται ως παρουσίες. Από την πλευρά τους οι Arasu, Chaudhuri, Kaushik (2008) παρουσίασε ένα νέο πλαίσιο ελαφρύς μετασχηματισμένο για να αναγνωρίζει παραλλαγές και γενικές μορφές των δεδομένων. Άλλοι ερευνητές (Fan et al., 2009) συνήγαγαν εναλλακτικά χαρακτηριστικά των δεδομένων για την ανάπτυξη ομοιοτήτων και ταυτοποίησης, ενώ άλλοι (Getoor, 2010) εισήγαγαν την έννοια της αναγνώρισης γραφήματος για την επισήμανση των δεδομένων που είναι ασαφή. Όσον αφορά τα μέτρα ομοιότητας, υπάρχουν ασαφείς αντιστοιχίες με βάση την απόσταση κατά Levenshtein και άλλους παρόμοιους αλγόριθμους, οι οποίοι μπορούν να εφαρμοστούν σε ένα σύστημα CA. Ενδεικτικά όσο μεγαλύτερη είναι η απόσταση Levenshtein τόσο διαφορετικές είναι μεταξύ τους δύο οντότητες. Τέλος οι Ananthkrishna, Chaudhuri & Ganti (2002) πρότειναν μια άλλη εναλλακτική μέσω της μέτρησης ταυτόχρονα συχνοτήτων, ενδεικτικά των A και B με C. Για παράδειγμα, η ίδια διεύθυνση IP και ο χρόνος σύνδεσης/αποσύνδεσης μπορούν να

χρησιμοποιηθούν ως χαρακτηριστικά για τον προσδιορισμό των δραστηριοτήτων ενός μέλους του προσωπικού του οποίου το όνομα σύνδεσης είναι συγκεκριμένο για μία πηγή δεδομένων, ενώ ένα διαφορετικό όνομα χρήστη θα είναι για μια άλλη πηγή δεδομένων.

Εκτός από τις παραπάνω τεχνικές αναγνώρισης δεδομένων, οι προγραμματιστές έχουν προτείνει πολλές τεχνικές ταυτοποίησης για συγκεκριμένα σενάρια που καθιστούν απαραίτητη την αναθεώρηση και την κατηγοριοποίηση αυτών των τεχνικών, ώστε να προσφέρουν σε ένα σύστημα CA. Επιπλέον, οι περισσότερες από τις τρέχουσες τεχνικές παρέχουν απλώς πιθανολογικά μέτρα για την αναγνώριση δεδομένων, τα οποία με τη σειρά τους θα επηρεάσουν τον κίνδυνο ελέγχου του συστήματος CA.

Συγκέντρωση δεδομένων

Στο πλαίσιο των Big Data ο μεγάλος όγκος δεδομένων που ρέει στο σύστημα CA ενδέχεται να προκαλέσει υπερβολικές απαιτήσεις πόρων αποθήκευσης σε υπολογιστή. Επιπλέον, τα σύνολα δεδομένων με μεγάλες διακυμάνσεις μπορεί να οδηγήσουν σε ασταθή μοντέλα ανάλυσης ελέγχου που θα μπορούσαν είτε να προκαλέσουν πάρα πολλούς συναγερμούς είτε να βιώσουν ελαττούμενη ισχύ ανίχνευσης (Alleset al., & Vasarhelyi, 2018; Perols & Murthy, 2012). Έτσι, η συγκέντρωση δεδομένων είναι απαραίτητη για την κανονική λειτουργία του συνεχούς ελέγχου χρησιμοποιώντας Big Data και για να συνοψίσουμε και να απλοποιήσουμε ουσιαστικά τα Big Data που πιθανότατα προέρχονται από διαφορετικές πηγές.

Οι υπάρχουσες αρχιτεκτονικές συνεχούς ελέγχου εστιάζουν όχι μόνο στη συγκέντρωση πρωτογενών δεδομένων για τη συλλογή δεδομένων εξαιρέσεων, αλλά και στη συγκέντρωση και ανάλυση των ίδιων των δεδομένων εξαιρέσεων. Οι Kogan, Alles, Vasarhelyi, και Wu (2014) πρότειναν ότι η συγκέντρωση δεδομένων εξαίρεσης που βασίζονται σε ακατέργαστα Big Data θα μπορούσε να διευκολύνει τον προσδιορισμό γενικών προτύπων για μια χρονική περίοδο. Για παράδειγμα, ο αριθμός των συναλλαγών μπορεί να διαφέρει σημαντικά μεταξύ των μεμονωμένων συναλλαγών, καθώς και των χρόνων καθυστέρησης μεταξύ παραγγελίας, παράδοσης και πληρωμής. Συγκεντρώνοντας τις μεμονωμένες συναλλαγές, αυτή η διακύμανση μπορεί να μειωθεί σημαντικά, επιτρέποντας αποτελεσματικότερη ανίχνευση υλικών ανωμαλιών. Ο Alles και οι συνεργάτες (2018) πρότειναν ότι παρόλο που ένα σύστημα CA δημιουργεί αυτόματα συναγερμούς για κρίσιμες εξαιρέσεις, όπως μεμονωμένοι λογαριασμοί χωρίς κωδικούς πρόσβασης, αυτοί οι συναγερμοί θα πρέπει επίσης να συγκεντρωθούν για να εντοπίσουν αδυναμίες σε ορισμένες περιοχές ελέγχου (π.χ., διαχωρισμός καθηκόντων). Τέλος

οι Perols και Murthy (2012) πρότειναν ένα πολυεπίπεδο πλαίσιο που συγκεντρώνει συναγερμούς και όχι τα μη επεξεργασμένα δεδομένα. Με τις εξαιρέσεις που εντοπίζονται στο επίπεδο παρακολούθησης να ομαδοποιούνται στο επίπεδο συγκέντρωσης ανάλογα με τη συσχέτιση τους με συγκεκριμένους στόχους ελέγχου.

Ωστόσο, υπάρχει ανταλλαγή στη συσσωμάτωση δεδομένων. Καθώς, όσο πιο συγκεντρωτικά είναι αυτά τόσο πιο φυσιολογικές είναι οι αναλυτικές σχέσεις, αυξάνοντας έτσι τη στατιστική ισχύ στην ανάλυση. Η κανονικότητα από την άλλη έρχεται στην τιμή των χαμένων ανιχνεύσεων στο επίπεδο λεπτομέρειας. Έτσι, η επιλογή των επιπέδων συγκέντρωσης πρέπει να γίνεται κατά περίπτωση λαμβάνοντας υπόψη τα εγγενή χαρακτηριστικά και το επίπεδο κινδύνου των υποκείμενων δεδομένων συναλλαγών. Αυτές οι προκλήσεις που εισήγαγε η συγκέντρωση γίνονται πιο εμφανείς σε έναν συνεχή έλεγχο των Big Data.

Απόρρητο δεδομένων

Το απόρρητο δεδομένων σημαίνει ότι ορισμένα δεδομένα ή οι συσχετίσεις αυτών είναι ευαίσθητα και δεν μπορούν να κοινοποιηθούν σε τρίτους (Ciriani et al., 2009). Στην εποχή των Big Data τα δεδομένα μπορούν εύκολα να συσχετιστούν με άλλα δεδομένα. Μόλις διαρρεύσουν ορισμένα ευαίσθητα δεδομένα, μπορούν να μεταδοθούν με μεγάλη ταχύτητα και να συνδεθούν σε μεγάλο αριθμό σχετικών δεδομένων. Έτσι, το απόρρητο των Big Data καθίσταται ακόμη πιο επείγον και σημαντικό για τη διατήρηση της εικόνας της επωνυμίας και την εξασφάλιση ανταγωνιστικού πλεονεκτήματος.

Το CA είναι ελκυστικό για τον έλεγχο εμπιστευτικών Big Data, επειδή το λογισμικό αυτοματοποίησης προστατεύει τα δεδομένα από την καθημερινή ανθρώπινη παρατήρηση και χειρισμό. Ωστόσο, οποιαδήποτε ροή δεδομένων που επιδιώκει να προσδιορίσει μια τάση θα πρέπει να διατηρηθεί σε κάποια μορφή, τουλάχιστον έως ότου η τάση τεκμηριωθεί και τεκμηριωθεί σε ένα σύστημα CA δεδομένου ότι αυτή η διατήρηση δεδομένων μπορεί να ενέχει πιθανούς κινδύνους ασφάλειας και εμπιστευτικότητας (Alles et al., 2018). Για παράδειγμα, τα στοιχεία πληρωτέων λογαριασμών και απαιτήσεων ενδέχεται να χρειαστεί να αντιστοιχίσουν τιμολόγια ή παραγγελίες με αποδείξεις και θα διατηρηθούν σε μια βάση δεδομένων CA λόγω των απαιτήσεων διάρκειας χρονοσειρών για ανάλυση. Υπό αυτές τις συνθήκες, μία από τις εκτενώς εφαρμοζόμενες λύσεις για αυτό το πρόβλημα εμπιστευτικότητας Big Data είναι η κρυπτογράφηση δεδομένων από εταιρικό προσωπικό. Ωστόσο, αυτή η λύση κρυπτογράφησης οδηγεί στη συνέχεια στα ακόλουθα δύο ζητήματα σε ένα σύστημα CA: Το πρώτο ζήτημα

περιλαμβάνει την αναζήτηση δεδομένων. Ενώ ορισμένα δεδομένα, συμπεριλαμβανομένων ορισμένων ευαίσθητων οικονομικών δεδομένων, είναι συνήθως κρυπτογραφημένα, ένα σύστημα CA ενδέχεται να χρειαστεί να πραγματοποιήσει αναζήτηση και πρόσβαση σε αυτά τα δεδομένα. Σύμφωνα με τους Song, Wagner και Perrig (2000) για την αντιμετώπιση αυτού του ζητήματος τα συστήματα βασίζονται πλέον στην αναζήτηση κρυπτογραφημένων δεδομένων μέσω λέξεων-κλειδιών τα οποία μπορούν να εφαρμοστούν σε συστήματα CA. Ωστόσο, τα σχήματα απαιτούν από το άτομο που κρυπτογραφεί τα δεδομένα να δημιουργεί ταυτόχρονα τις αποδείξεις (π.χ. υπογραφές) λέξεων-κλειδιών, οι οποίες χρησιμοποιούνται από το σύστημα CA για να ελέγξουν εάν ορισμένες λέξεις-κλειδιά βρίσκονται στα κρυπτογραφημένα δεδομένα χωρίς να γνωρίζει το περιεχόμενο του υπόλοιπου κειμένου. Το δεύτερο ζήτημα περιλαμβάνει τον τρόπο ελέγχου των κρυπτογραφημένων δεδομένων. Υπάρχουν τεχνικές σχετικά με ορισμένους αλγόριθμους κρυπτογράφησης που μπορούν να χρησιμοποιηθούν από τους ελεγκτές, όπως αυτές που προτείνονται από τους Rivest, Adleman και Dertouzos (1978) αλλά και τον Gentry (2009). Για παράδειγμα, μόνο ορισμένοι ανώτεροι ελεγκτές θα έχουν πρόσβαση στο ευαίσθητο απλό κείμενο και θα το κρυπτογραφούν σε κρυπτογράφο, αφήνοντας τους κατώτερους ελεγκτές να εργάζονται στο κρυπτοκείμενο. Με αυτόν τον τρόπο, τα ευαίσθητα δεδομένα μπορεί να είναι γνωστά μόνο από λίγους εξουσιοδοτημένους ανώτερους ελεγκτές και η πιθανότητα διαρροής πληροφοριών κατά τη διάρκεια μιας διαδικασίας ελέγχου μπορεί να μειωθεί.

Ωστόσο, παραμένουν αρκετά ζητήματα όταν οι παραπάνω αλγόριθμοι αναζήτησης και κρυπτογράφησης εφαρμόζονται σε συστήματα CA: ο αλγόριθμος αναζήτησης μπορεί να χρησιμοποιηθεί μόνο για να κρίνει εάν μια λέξη-κλειδί είναι στο κείμενο κρυπτογράφησης και ο αλγόριθμος κρυπτογράφησης είναι αρκετά χρονοβόρος. Επομένως, απαιτείται συνεχής έρευνα για τη βελτιστοποίηση των αλγορίθμων κρυπτογράφησης. Η κύρια πρόκληση της τεχνικής ανάπτυξης είναι η προστασία της ιδιωτικής ζωής των δεδομένων, ενώ παράλληλα εγγυάται τη χρησιμότητα του ελέγχου, οπότε η καθιέρωση θεωρητικών αντισταθμίσεων μεταξύ ιδιωτικότητας και χρησιμότητας είναι ένας σημαντικός ανοιχτός χώρος για έρευνα.

3. Μεθοδολογία

3.1. Σχεδιασμός ερωτηματολογίου

Οι περιβαλλοντικές ιδιαιτερότητες (άτομα που συμμετέχουν στην έρευνα, βαθμός δυσκολίας εύρεσης των ατόμων, προηγούμενες μελέτες κ.λ.π.) συχνά καθορίζουν την επιλογή της μεθόδου συλλογής δεδομένων που θα χρησιμοποιηθεί (Karlan, 2004). Ως σπουδαιότερο εργαλείο συλλογής δεδομένων στη ποσοτική έρευνα μπορεί να θεωρηθεί το ερωτηματολόγιο (Allen et al., 2008). Τα μεγάλα δεδομένα αποτελούν από μόνα τους μια τεράστια τεχνική και ανθρώπινη πρόκληση όσον αφορά τη συλλογή, την ταξινόμηση, την αναγνώριση της χρήσιμης πληροφορίας και κατά συνεπακόλουθα την εκμετάλλευση αυτής υπέρ της λογιστικής. Αλλά για να φτάσουμε σε αυτό το σημείο πρέπει να αντιμετωπιστούν αυτές οι προκλήσεις παράλληλα με την όποια πολυπλοκότητα.

Πρωτοπόροι ερευνητές όπως ο Davis και οι συνεργάτες (Davis et al., 1989) επάνω στον τομέα της καλύτερης κατανόησης και ατομικής αποδοχής των τεχνολογιών της πληροφορίας, πρότειναν τα μοντέλα αποδοχής τεχνολογίας (technology acceptance model - TAM). Έτσι στη παρούσα έρευνα θα χρησιμοποιήσουμε τα TAM ως ένα ισχυρό, στιβαρό και συνήθως εφαρμόσιμο μοντέλο για την πρόβλεψη και την εξήγηση της συμπεριφοράς των χρηστών και της χρήσης πληροφορικής (Agarwal & Prasad, 1999; Davis et al., 1989; Legris et al., 2003). Στοιχεία που θα καθορίζουν, την αντιλαμβανόμενη ευκολία χρήσης, χρησιμότητα, στάση απέναντι στη χρήση, πρόθεση χρήσης και συμπεριφορά συστήματος. Η χρησιμότητα και η αντιλαμβανόμενη ευκολία χρήσης είναι οι σημαντικότεροι καθοριστικοί παράγοντες της πρόθεσης χρήσης και της χρήσης του συστήματος (Wu & Wang, 2005).

Επιπρόσθετα, το υιοθετούμενο TAM θα είναι προσαρμοσμένο και εμπλουτισμένο βάσει ενός πρόσφατα δημοσιευμένου μοντέλου του Okcu και των συνεργατών (Okcu et al., 2019) το οποίο επεκτάθηκε χρησιμοποιώντας παράγοντες όπως η συνάφεια εργασίας, διαστάσεις μαζικών δεδομένων, συμβατότητα, αυτο-αποτελεσματικότητα, πολυπλοκότητα και άγχος. Αυτή η προσαρμογή παρέχει πέντε νέα στοιχεία που έχουν άμεση επίδραση στην αντιλαμβανόμενη χρησιμότητα και την ευκολία χρήσης και τις έμμεσες επιπτώσεις τους στην πρόθεση χρήσης εργαλείων μαζικών δεδομένων. Υπάρχουν και άλλες μελέτες στη βιβλιογραφία που ασχολήθηκαν με τον αντίκτυπο ορισμένων παραγόντων στην αποδοχή εργαλείων εξόρυξης δεδομένων (data mining). Ενδεικτικά, στη μελέτη των Esteves και Curto (2013), η πραγματική υιοθέτηση μαζικών δεδομένων εξετάστηκε χρησιμοποιώντας την

αποσυντεθειμένη θεωρία της προγραμματισμένης συμπεριφοράς και του TAM, όπου οι συγκεκριμένοι ερευνητές συμπεριέλαβαν στις διαστάσεις των μαζικών δεδομένων, τη συμβατότητα, την αυτο-αποτελεσματικότητα και τις συνθήκες διευκόλυνσης στα μοντέλα τους. Επίσης αξιολόγησαν την πρόθεση υιοθέτησης μαζικών δεδομένων χρησιμοποιώντας την αντίληψη των κινδύνων και των οφελών. Στη μελέτη του Huang και των συνεργατών (Huang et al., 2012), η πρόθεση για χρήση εργαλείων εξόρυξης δεδομένων διερευνήθηκε χρησιμοποιώντας TAM. Αυτό το μοντέλο περιλάμβανε παράγοντες σε τέσσερις κατηγορίες: τη διάσταση προσανατολισμένη στην εργασία (συνάφεια εργασίας, ποιότητα παραγωγής, επίδειξη αποτελέσματος, χρόνος απόκρισης και μορφή), πεποιθήσεις ελέγχου (αυτο-αποτελεσματικότητα υπολογιστή και αντιλήψεις εξωτερικού ελέγχου), συναίσθημα (άγχος υπολογιστών) και κίνητρο χρήσης.

3. 2. Δειγματοληψία

Το δείγμα των συμμετεχόντων της παρούσας έρευνας είναι τυφλό, έγινε εξολοκλήρου με απομακρυσμένη πρόσβαση μετά από αποστολή σχετικού συνδέσμου στα μέλη του Οικονομικού Επιμελητηρίου Ελλάδος, όλων των Λογιστικών Συλλόγων της Επικράτειας, της Πανελληνίας Ομοσπονδίας Φοροτεχνικών Ελευθέρων Επαγγελματιών, διαφόρων λογιστικών ομάδων (taxheaven, e-forologia, κ.ά.) και μέσω κοινωνικών δικτύων. Όπως γίνεται κατανοητό το δείγμα βάσει του μητρώου των καταγεγραμμένων μελών των ανωτέρω συλλόγων θα μπορούσε δυνητικά να είναι πολύ μεγαλύτερο αλλά η συμμετοχή περιορίστηκε σε $N = 95$. Η προσπάθεια προώθησης του ερωτηματολογίου διήρκησε για διάστημα τεσσάρων εβδομάδων, όπου στις περισσότερες περιπτώσεις μετά την ηλεκτρονική αποστολή πραγματοποιούνταν και τηλεφωνική επικοινωνία για καλύτερη και πιο αποτελεσματική προώθηση. Όλοι οι συμμετέχοντες έλαβαν μέρος οικειοθελώς και υπήρξε ανωνυμία. Η δομή του πλήρες ερωτηματολογίου αναλύεται στην επόμενη παράγραφο, ενώ αυτούσιο παρατίθεται στο ΠΑΡΑΡΤΗΜΑ Α.

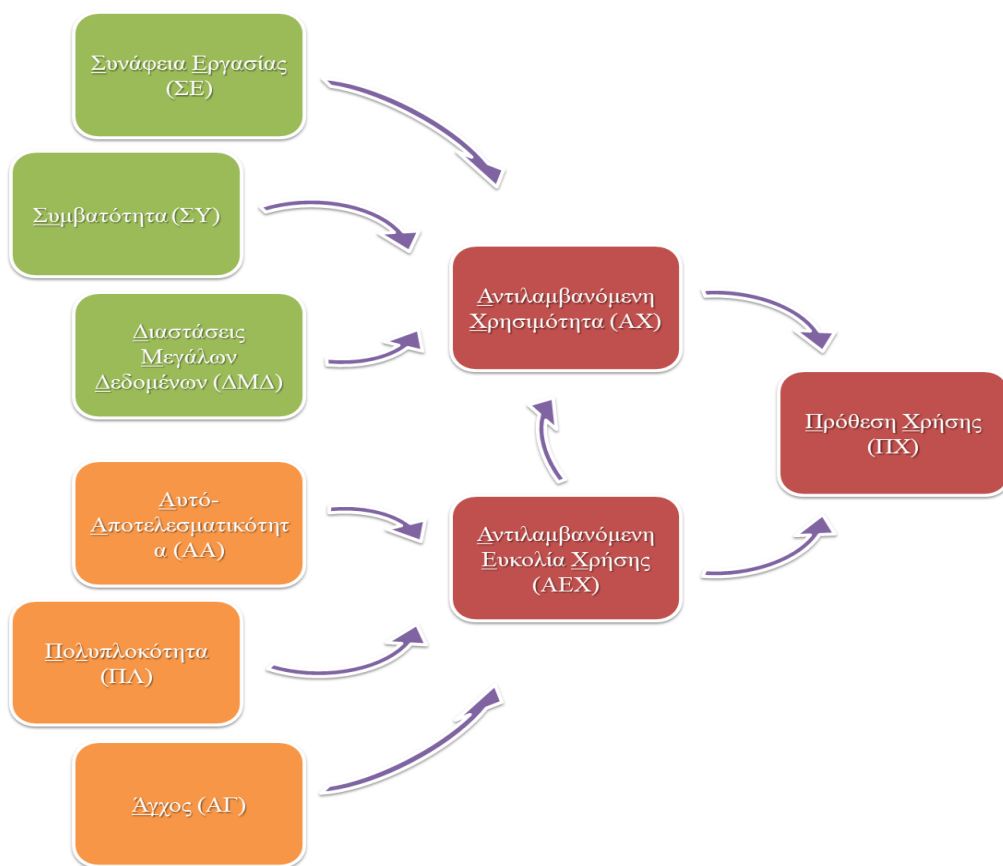
3. 3. Στατιστική ανάλυση

Στην παρούσα έρευνα η μέθοδος που χρησιμοποιήθηκε είναι αυτή της συλλογής δεδομένων μέσω ερωτηματολογίου (βλέπε ΠΑΡΑΡΤΗΜΑ Α). Ο πληθυσμός-στόχος ήταν λογιστές – φοροτέχνες και λοιποί υπάλληλοι λογιστηρίων. Το ερωτηματολόγιο αποτελείται από δύο κύρια μέρη. Το πρώτο μέρος περιλαμβάνει δημογραφικές ερωτήσεις που σχεδιάστηκαν για

να συλλέξουν πληροφορίες σχετικά με την ηλικιακή ομάδα, το φύλο, την εργασιακή εμπειρία και το επίπεδο εξειδίκευσης-μόρφωσης των συμμετεχόντων. Συνολικά 95 ερωτηματολόγια συμπληρώθηκαν. Η ανάλυση των δημογραφικών στοιχείων των συμμετεχόντων γίνεται στην ενότητα 4.

Το δεύτερο μέρος αποτελείται από ερωτήματα που αφορούν εννέα μεταβλητές που σκοπό έχουν να προσδιορίσουν - ποσοτικοποιήσουν την πρόθεση χρήσης (Agarwal & Karahanna, 2000), την αντιλαμβανόμενη χρησιμότητα (Davis, 1989), την αντιλαμβανόμενη ευκολία χρήσης (Davis, 1989), τη συνάφεια εργασίας (Davis et al., 1992), τη συμβατότητα (Moore & Benbasat, 1991), τις διαστάσεις μεγάλων δεδομένων (Sahoo & Yajnik, 2014), την αυτό-αποτελεσματικότητα (Venkatesh et al., 2003), το άγχος (Venkatesh et al., 2003) και την πολυπλοκότητα (Thompson et al, 1991). Οι παραπάνω εννέα μεταβλητές συνοψίζουν στο σύνολό τους το θεωρητικό μοντέλο της παρούσας μελέτης και δίνεται σχηματικά υπό τη μορφή δένδρογράμματος στην **Εικόνα 3.1**. Οι απαντήσεις στα ερωτήματα των εννέα μεταβλητών του δεύτερου μέρους του ερωτηματολογίου μετρούνται σε κλίμακα τύπου Likert πέντε σημείων (1-5, όπου 1: Διαφωνώ Απόλυτα και 5: Συμφωνώ Απόλυτα).

Λίγο αναλυτικότερα για το θεωρητικό υπόβαθρο των στοιχείων, η πρόθεση χρήσης (ΠΧ) σχετίζεται με τη συμπεριφορά και πως θα μπορούσε αυτή να επηρεάζεται από τα κίνητρα που είναι συστατικά της πρόθεσης συμπεριφοράς. Αυτά τα αποτελέσματα είναι δείκτες για το πώς τα άτομα είναι πρόθυμα να ενεργήσουν και πόση ενέργεια είναι διατεθειμένα να χρησιμοποιήσουν για να εκτελέσουν τη συμπεριφορά. Η συμπεριφορά καθορίζεται από τη συμπεριφορική πρόθεση, εφόσον ένα άτομο συμφωνεί να εκτελέσει μια συμπεριφορά (Ajzen, 1991). Επομένως, η πρόθεση χρήσης είναι καθοριστική για την πιθανότητα ενός ατόμου να εμπλακεί σε μια συγκεκριμένη συμπεριφορά (Ajzen, 1980). Οι άνθρωποι σκοπεύουν να χρησιμοποιήσουν μια δεδομένη τεχνολογία πριν τη χρησιμοποιήσουν. Ως εκ τούτου, ο άμεσος παράγοντας της πραγματικής χρήσης μπορεί να προσδιοριστεί ως η πρόθεση συμπεριφοράς για χρήση. Όσο περισσότερο ένα άτομο είναι πρόθυμο να χρησιμοποιήσει ένα σύστημα, τόσο μεγαλύτερη είναι η πιθανότητα χρήσης του (Ajzen & Madden, 1986).



Εικόνα 3.1, Παρουσίαση των εννέα σημαντικών στοιχείων του TAM σε μορφή δενδρογράμματος

Στην περίπτωση της αντιλαμβανόμενης χρησιμότητας (ΑΧ), αυτή ορίζεται ως «ο βαθμός στον οποίο ένα άτομο πιστεύει ότι η χρήση ενός συγκεκριμένου συστήματος θα ενισχύσει την απόδοσή του» (Davis et al., 1989; Saadé & Bahli, 2005). Στο αρχικό μοντέλο αποδοχής τεχνολογίας, η αντιληπτή χρησιμότητα είναι βασικός καθοριστικός παράγοντας της πρόθεσης χρήσης (Davis et al., 1989). Εκτεταμένη έρευνα σε διάφορους τομείς επιβεβαιώνει τον σημαντικό αντίκτυπο της αντιληπτής χρησιμότητας στη συμπεριφορά της πρόθεσης χρήσης (Agarwal & Prasad, 1999; Chang & Tung, 2008; Davis et al., 1989; Fu et al., 2006; Gallego et al., 2008; Gefen, 2003; Guriting & Ndubisi, 2006; Horst et al., 2007; Hu et al., 1999; Khalifa & Shen, 2008; Lee, 2006; Liao et al., 2007; Tung & Chang, 2008; Venkatesh, 2000; Venkatesh & Morris, 2000). Έτσι, εκτός από διάφορους τομείς εφαρμογής, η αντιληπτή χρησιμότητα των εργαλείων μεγάλων δεδομένων μπορεί να είναι κρίσιμη για την πρόθεση χρήσης.

Η τρίτη μεταβλητή, την οποία αποτελεί η αντιλαμβανόμενη ευκολία χρήσης (ΑΕΧ), θεωρητικά ορίζεται ως "ο βαθμός στον οποίο ένα άτομο πιστεύει ότι η χρήση ενός συγκεκριμένου συστήματος θα ήταν χωρίς προσπάθεια" (Davis et al., 1989). Η σχέση μεταξύ

της αντιληπτής ευκολίας χρήσης και της συμπεριφορικής πρόθεσης χρήσης επιβεβαιώνεται από TAM. Μετά την αντιληπτή χρησιμότητα, είναι ο δεύτερος βασικός καθοριστικός παράγοντας της πρόθεσης συμπεριφοράς να χρησιμοποιηθεί στο πρωτότυπο TAM (Davis et al., 1989). Οι άνθρωποι προτιμούν να μειώσουν την προσπάθεια που απαιτείται για τις πράξεις τους (Venkatesh, 2000). Εάν η χρήση εργαλείων μεγάλων δεδομένων απαιτεί υψηλό επίπεδο προσπάθειας, τότε οι εργαζόμενοι ενδέχεται να μην αρχίσουν να χρησιμοποιούν ή ακόμη και να εγκαταλείψουν τη χρήση του εργαλείου. Πολλές ερευνητικές μελέτες έχουν επαληθεύσει τις σημαντικές σχέσεις μεταξύ της αντιληπτής ευκολίας χρήσης και της πρόθεσης συμπεριφοράς για χρήση (Chang & Tung, 2008; Fusilier & Durlabhji, 2005; Gallego et al., 2008; Hong & Tam, 2006; Lee et al., 2006).

Εάν όλοι οι άλλοι παράγοντες διατηρούνται σταθεροί, τόσο πιο εύκολο είναι το σύστημα και τόσο υψηλότερη είναι η απόδοση εργασίας του χρήστη. Οι χρήστες ενδέχεται να μην αισθάνονται σίγουροι με ένα νέο σύστημα αφού δεν διαθέτουν τις απαραίτητες δεξιότητες και άνεση. Παρ'όλα αυτά, αφού αποκτήσουν μια ορισμένη ποσότητα γνώσης και εξοικειωθούν περισσότερο με αυτήν, η αντίληψή τους για την ευκολία χρήσης της μπορεί να μεταβληθεί (Hackbarth et al., 2003). Με άλλα λόγια, εάν άλλες πτυχές είναι ίσες, τότε όσο πιο εύκολο είναι να χρησιμοποιηθεί το σύστημα, τόσο πιο χρήσιμο μπορεί να είναι (Davis, 1993; Venkatesh & Davis, 2000). Εάν η χρήση μεγάλων εργαλείων δεδομένων είναι εύκολη, οι εργαζόμενοι δεν χρειάζεται να αφιερώσουν πολύ χρόνο για να μάθουν πώς να τα χρησιμοποιούν, γεγονός που μπορεί να επηρεάσει την απόδοση του χρήστη. Έρευνες σε διάφορους τομείς έχουν επιβεβαιώσει τις σημαντικές σχέσεις μεταξύ της αντιληπτής ευκολίας χρήσης και της αντιληπτής χρησιμότητας (Anandarajan et al., 2000; Calisir & Calisir, 2004; Chan, 2004; Fu et al., 2006; Gallego et al., 2008; Igbaria & Iivari, 1995; Kleijnen et al, 2004; Lee, 2006; Liker & Sindi, 1997; Lin, 2007; Oh et al., 2003; Tung et al., 2008; Venkatesh & Davis, 2000; Yi et al., 2006).

Όσον αφορά στη συνάφεια εργασίας (ΣΕ), οι χρήστες έχουν διαφορετικές αντιλήψεις για τα αποτελέσματα που αναμένουν να λάβουν από ένα νέο σύστημα λόγω της διαφορετικής φύσης των εργασιών τους. Επιπλέον, εκτίθενται σε εξωτερικές πληροφορίες που ενδέχεται να επηρεάσουν την επιλογή του εργαλείου που χρειάζονται. Η συνάφεια εργασίας μπορεί να οριστεί ως η αντίληψη ενός χρήστη σχετικά με το βαθμό εφαρμογής του συστήματος στη δουλειά του. Στην αναθεωρημένη έκδοση του TAM, Venkatesh and Davis (Venkatesh & Davis, 2000) προτείνουν ότι η συνάφεια με την εργασία είναι μια γνωστική εργαλειακή διαδικασία. Οι χρήστες μεγάλων εργαλείων δεδομένων χρησιμοποιούν μια νοητική αναπαράσταση για να

συσχετίσουν τους πρωταρχικούς εργασιακούς στόχους τους με τα αποτελέσματα της χρήσης αυτών των εργαλείων. Αυτή η γνωστική σχέση αποτελεί τη βάση για τη λήψη αποφάσεων σχετικά με το αν το εργαλείο είναι χρήσιμο. Αρκετές μελέτες επιβεβαιώνουν ότι η συνάφεια με την εργασία διαπιστώνεται ότι είναι ένας σημαντικός προγνωστικός παράγοντας της αντιληπτής χρησιμότητας (Chismar & Wiley-Patton, 2003; Liang et al., 2003; Venkatesh & Davis, 2000). Επιπλέον, έχουν επίσης μελετηθεί οι θετικές μετριαστικές επιδράσεις της συνάφειας εργασίας στην αντιληπτή χρησιμότητα (Bhattacharjee & Sanford, 2006; Kim, 2008).

Η συμβατότητα (ΣΥ) είναι αποδεκτή ως μία από τις πέντε κύριες θεωρήσεις που καθορίζουν την υιοθέτηση της καινοτομίας (Tornatzky & Klein, 1982). Ορίζεται ως "ο βαθμός στον οποίο μια καινοτομία εκλαμβάνεται ως συνεπής με τις υπάρχουσες αξίες, τις προηγούμενες εμπειρίες και τις ανάγκες των πιθανών υιοθετών". Σύμφωνα με τους Moore και Benbasat (Moore & Benbasat, 1991), υπάρχει σημαντική σχέση μεταξύ συμβατότητας και αντιληπτού πλεονεκτήματος, η οποία ισοδυναμεί με αντιληπτή χρησιμότητα στο πλαίσιο της TAM. Επιπλέον, οι Chau και Hu (2001) προτείνουν ότι η συμβατότητα είναι σημαντικός καθοριστικός παράγοντας της αντιληπτής χρησιμότητας αλλά όχι της αντιληπτής ευκολίας χρήσης. Δηλώνουν ότι η ασυμβατότητα ενός συστήματος μπορεί να απαιτεί μια μακρά διαδικασία εκμάθησης που οδηγεί σε πτώση της αντίληψης για ευκολία χρήσης. Προηγούμενες έρευνες σε διάφορους τομείς επιβεβαιώνουν τις σημαντικές σχέσεις μεταξύ συμβατότητας και αντιληπτής χρησιμότητας (Chang & Tung, 2008; Chen et al., 2004; Fu et al., 2006; Oh et al., 2003; Tung & Chang, 2008a, 2008b; Tung et al., 2008).

Οι διαστάσεις μεγάλων δεδομένων (ΔΜΔ) χαρακτηρίζονται ως μια νέα γενιά τεχνολογιών, που αποσκοπούν στην εξαγωγή οικονομικής αξίας από πολύ μεγάλους όγκους μεγάλης ποικιλίας δεδομένων επιτρέποντας τη λήψη και ανάλυση υψηλής ταχύτητας. Σύμφωνα με αυτήν την άποψη, οι πρώτες περιγραφές των μεγάλων δεδομένων χρησιμοποίησαν το μοντέλο «3V» για να περιγράψουν τις τρεις διαστάσεις του όγκου, της ταχύτητας και της ποικιλίας (βλέπε Κεφάλαιο 1). Πρόσφατη έρευνα έχει επαληθεύσει τη σχέση μεταξύ των διαστάσεων των μεγάλων δεδομένων και της αντιληπτής χρησιμότητας (Esteves & Curto, 2013; Shin, 2016).

Η αυτο-αποτελεσματικότητα (ΑΑ), που βασίζεται στην κοινωνική γνωστική θεωρία, είναι η διορατικότητα και η διαβεβαίωση των ατόμων σχετικά με την ικανότητά τους να επιτελούν και να διεξάγουν τις συγκεκριμένες δραστηριότητες που απαιτούνται για την επιτυχία σε συγκεκριμένους τύπους εργασιών (Bandura, 1986; Huffman et al., 2013). Έχει αντίκτυπο στις δραστηριότητες που προτιμούν να εκτελούν οι άνθρωποι, στην προσπάθεια που

είναι έτοιμοι να χρησιμοποιήσουν και στον χρόνο για τον οποίο θα συνεχίσουν αυτήν την προσπάθεια για να ξεπεράσουν τα εμπόδια (Bandura, 1982, 1986). Η θεωρία του Bandura αναφέρει ότι τα άτομα με υψηλή αυτο-αποτελεσματικότητα είναι πιο πιθανό να αξιολογήσουν τα δύσκολα καθήκοντα ως κάτι στο οποίο πρέπει να γίνουν εξειδικευμένοι και όχι κάτι που πρέπει να αποφεύγονται. Η σημασία της αντιληπτής ευκολίας χρήσης υποστηρίζεται από το Bandura's (1982) εκτεταμένη έρευνα για την αυτο-αποτελεσματικότητα (Davis et al., 1989). Ως εκ τούτου, αναμένεται ότι οι υπάλληλοι με υψηλή αυτο-αποτελεσματικότητα στη χρήση εργαλείων μεγάλων δεδομένων είναι πιο πιθανό να το δουν ως εύκολο και όχι περίπλοκο. Αρκετές μελέτες σε διάφορους τομείς εξέτασαν τη σχέση μεταξύ της αυτο-αποτελεσματικότητας και της αντιληπτής ευκολίας χρήσης (Alalwan et al., 2016; Igarria & Iivari, 1995; Mun & Hwang, 2003; Ozturk et al., 2016; Venkatesh, 2000).

Η πολυπλοκότητα (ΠΛ) ορίζεται ως "ο βαθμός στον οποίο μια καινοτομία εκλαμβάνεται ως σχετικά δύσκολη στην κατανόηση και στη χρήση"(Rogers & Shoemaker, 1971; Thompson et al., 1991). Η τεχνολογική πολυπλοκότητα αντιπροσωπεύει μια κατάσταση όπου μια νέα τεχνολογία είναι πιο δύσκολη για τους χρήστες από την προηγούμενη τεχνολογία που χρησιμοποιήθηκε για την ίδια εργασία, όπου υπάρχει αύξηση του αριθμού των εργασιών που πρέπει να εκτελέσει ο χρήστης στον ίδιο χρόνο. Παρόμοια με άλλα νέα τεχνολογικά εργαλεία, τα μεγάλα δεδομένα συχνά θεωρούνται πολύ πολύπλοκα και δύσκολα στην εφαρμογή. Όσο μεγαλύτερη είναι η πολυπλοκότητα, τόσο υψηλότερος ο διανοητικός φόρτος εργασίας και το άγχος (Sokol, 1994). Η πολυπλοκότητα των μεγάλων εργαλείων δεδομένων μπορεί να επηρεάσει αρνητικά την αντιληπτή ευκολία χρήσης. Στη βιβλιογραφία, αρκετές μελέτες συνδέουν την πολυπλοκότητα και την αντιληπτή ευκολία χρήσης (Rajan & Baral, 2015; Thompson et al., 1991; Venkatesh & Davis, 1996).

Συνοψίζοντας, το άγχος (ΑΓ) περιγράφεται ως «αγχωτικές ή συναισθηματικές αντιδράσεις όταν πρόκειται για την εκτέλεση μιας συμπεριφοράς» (Venkatesh et al., 2003). Σύμφωνα με τους Van Raaij και Schepers (2008), το άγχος είναι ένα προσωπικό χαρακτηριστικό και επηρεάζει αρνητικά την αποδοχή της τεχνολογίας. Επομένως, η δυσφορία ενός υπαλλήλου απέναντι σε μια τεχνολογία ή εργαλεία μεγάλων δεδομένων μπορεί να δημιουργήσει την αντίληψή τους ως πιο δύσκολη και πολύπλοκη. Ομοίως, οι εργαζόμενοι που είναι λιγότερο ανήσυχοι έχουν περισσότερες πιθανότητες να εργαστούν αποτελεσματικά με εργαλεία δεδομένων μεγάλης κλίμακας από τους εργαζόμενους που είναι περισσότερο ανήσυχοι. Αρκετές μελέτες στη βιβλιογραφία έχουν παρατηρήσει την αρνητική σημαντική

σχέση μεταξύ άγχους και αντιληπτής ευκολίας χρήσης (Calisir et al., 2014; Chatzoglou et al., 2009; Karaali et al., 2011; Van Raaij & Schepers, 2008; Venkatesh, 2000).

3. 4. Μέθοδοι στατιστικής ανάλυσης

Τα αποτελέσματα του ερωτηματολογίου εξήχθησαν από την πλατφόρμα του Google Documents σε μορφή λογιστικού φύλλου του προγράμματος MS-Excel (MS-Office XP) και στη συνέχεια όλα τα δεδομένα του μεταφέρθηκαν στο στατιστικό πακέτο SPSS ("IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp."), με το οποίο έγινε η επεξεργασία των αποτελεσμάτων και οι πίνακες.

Στους πίνακες σύνοψης αποτελεσμάτων γίνεται αναφορά αυτών, περιλαμβανομένων του μέσου όρου (M.O.), του τυπικού σφάλματος (Standard Error), του αριθμού πλήθους N και της μήτρας συσχέτισης ανά κατηγορία.

Η στατιστική ανάλυση των απαντήσεων που έχουν δοθεί από τους συμμετέχοντες έγινε σε δύο στάδια. Πρώτο στάδιο ήταν ο προσδιορισμός της εσωτερικής συνέπειας ή της αξιολόγηση αξιοπιστίας, στο μέτρο που είναι δυνατή, εντός του συνόλου των στοιχείων δοκιμής βάσει του συντελεστή α κατά Cronbach. Η αξιολόγηση κατά Cronbach's α έγινε συνολικά ανά μεταβλητή. Υπάρχουν πολλές διαφορετικές αναφορές σχετικά με τις αποδεκτές τιμές του α που κυμαίνονται από 0,7 έως 0,95 (Tavakol & Dennick, 2011). Στην παρούσα εργασία αποδεκτή τιμή για τον συντελεστή αξιοπιστίας είναι $\alpha > 0,7$.

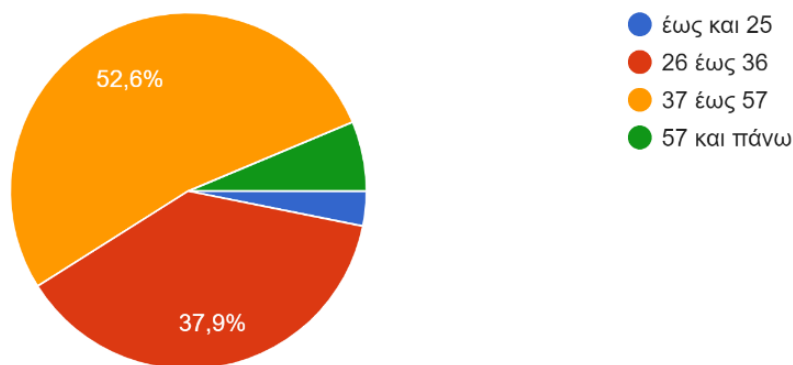
Στο δεύτερο στάδιο ακολούθησε η εφαρμογή της γραμμικής παλινδρόμησης (Linear Regression) ως προς την εξαγωγή στατιστικά σημαντικών συμπερασμάτων σε τέσσερα διαφορετικά μοντέλα που καταστρώθηκαν βάση ορθολογικής συσχέτισης στοιχείων όπως αυτά παρουσιάζονται και σχηματικά υπό μορφή δενδρογράμματος στην ενότητα 3.2 (βλέπε Εικόνα 3.1). Εδώ η στατιστική σημαντικότητα ορίστηκε στο 95% ($P = 0,05$). Στα τέσσερα μοντέλα πάντα ένα στοιχείο ορίζονταν ως εξαρτημένη μεταβλητή και πολλαπλές οι ανεξάρτητες με τη μέθοδο ορισμένη στην εισαγωγή. Επίσης επιλέχθηκαν η εξαγωγή στατιστικών δεδομένων ανά μοντέλο τόσο για τους συντελεστές (coefficients), την παρακολούθηση της μεταβολής του R^2 όσο και της ανάλυσης διακύμανσης ενός παράγοντα (One-way Analysis Of Variance - ANOVA). Κατά συνέπεια βάσει της τιμής ορισμού στατιστικής σημαντικότητας η εξαγωγή πιθανώς σημαντικών αλληλεπιδράσεων μεταξύ των επιμέρους εννέα στοιχείων του ερωτηματολογίου θα πρέπει να δίνει αντίστοιχες τιμές εντός των θεωρητικών τιμών των β , t , F και p για κάθε μοντέλο.

4. Αποτελέσματα έρευνας

Το πρώτο κομμάτι του ερωτηματολογίου στην παρούσα έρευνα αποτέλεσε η συλλογή γενικών δημογραφικών δεδομένων του δείγματος που συμμετείχε μέσω στοιχείων όπως η ηλικία, το φύλλο, η εργασιακή προϋπηρεσία και το μορφωτικό επίπεδο. Τα δημογραφικά στοιχεία που συλλέχθηκαν περιγράφουν σημαντικά ποιοτικά χαρακτηριστικά για την καλύτερη εξαγωγή συμπερασμάτων.

Σε ότι αφορά την ηλικιακή σύνθεση του δείγματος, αυτό δίνεται συνοπτικά στο **Διάγραμμα 4.1**. Εκ του συνόλου 95 συμμετεχόντων το 52,6% ανήκε στην ηλικιακή ομάδα των 37-57 ενώ η δεύτερη ηλικιακή ομάδα με ποσοστό 37,9% ήταν για τις ηλικίες 26-36. Κατά συνέπεια αθροιστικά οι κάτω των 25 και άνω των 57 αποτέλεσαν μόλις το 9,5%.

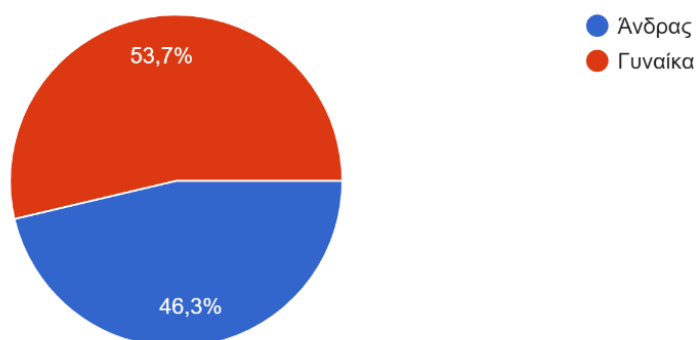
Ηλικία
95 απαντήσεις



Διάγραμμα 4.1, Συνοπτική παρουσίαση αποτελεσμάτων υπό μορφή πίτας για την κατανομή των ηλικιών των συμμετεχόντων της έρευνας

Αναφορικά με τη σύνθεση του φύλου των συμμετεχόντων, το 53,7% ήταν γυναίκες και το υπόλοιπο 46,3% ήταν άνδρες (**Διάγραμμα 4.2**). Όπως γίνεται διακριτό η αντιπροσωπευτικότητα των φύλων είναι ικανοποιητική με ελαφρά υπεροχή των γυναικών.

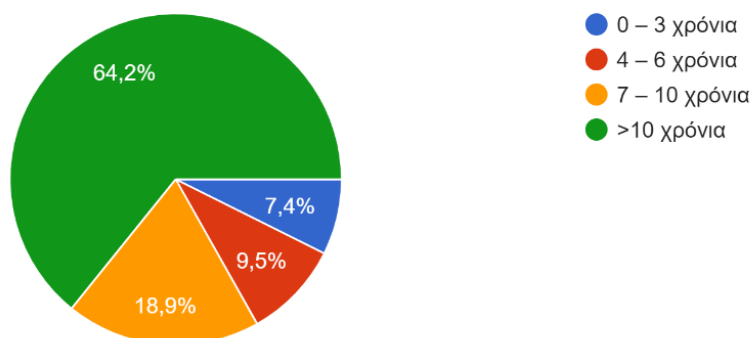
Φύλο
95 απαντήσεις



Διάγραμμα 4.2, Συνοπτική παρουσίαση αποτελεσμάτων υπό μορφής πίτας για την κατανομή ανά φύλο των συμμετεχόντων της έρευνας

Αντίστοιχα στο **Διάγραμμα 4.3** συγκεντρώνονται τα ποσοστά της εργασιακής εμπειρίας σε έτη των συμμετεχόντων, με την συντριπτική πλειονότητα (64,2%) να αποτελείται από άτομα προϋπηρεσίας μεγαλύτερης της μίας δεκαετίας. Προσδίδοντας ιδιαίτερη βαρύτητα στις απαντήσεις της έρευνας δεδομένης της πλούσιας εμπειρίας της πλειοψηφίας.

Εργασιακή Εμπειρία
95 απαντήσεις

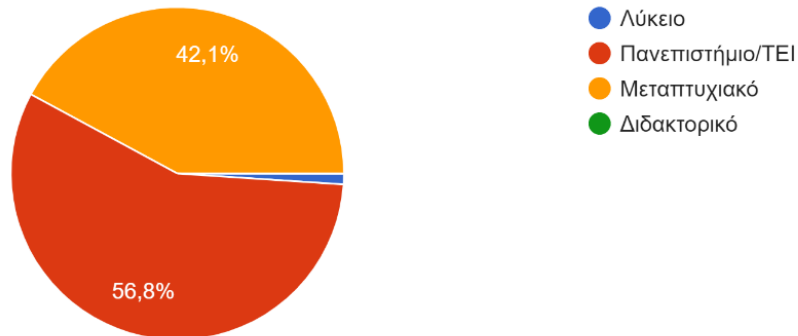


Διάγραμμα 4.3, Συνοπτική παρουσίαση αποτελεσμάτων υπό μορφής πίτας για την κατανομή της εργασιακής εμπειρίας των συμμετεχόντων της έρευνας

Τέλος, στο **Διάγραμμα 4.4** διαφαίνεται ξεκάθαρα από το συντριπτικό ποσοστό του 98,9% ότι οι συμμετέχοντες κατέχουν ανώτερη ή ανώτατη εξειδίκευση με το 50% σχεδόν αυτών να κατέχουν και μεταπτυχιακούς τίτλους. Ως εκ τούτου αποτελεί ακόμη ένα παράγοντα εγκυρότητας και ιδιαίτερης βαρύτητας των αποτελεσμάτων της παρούσας μελέτης.

Εκπαιδευτικό Επίπεδο

95 απαντήσεις



Διάγραμμα 4.4, Συνοπτική παρουσίαση αποτελεσμάτων υπό μορφής πίτας για την κατανομή του επιπέδου μόρφωσης των συμμετεχόντων της έρευνας

4. 1. Περιγραφική στατιστική και έλεγχος αξιοπιστίας

Η ανάλυση της αξιοπιστίας των απαντήσεων στο ερωτηματολόγιο αξιολογείται με τον προσδιορισμό του συντελεστή α κατά Cronbach. Με άλλα λόγια, η αξιοπιστία των μετρήσεων που έχουμε λάβει ανά στοιχείο αναφέρεται στον βαθμό κατά τον οποίο εμφανίζεται συνέπεια μεταξύ των εννοιών-ερωτήσεων και το α του Cronbach είναι ένας τρόπος μέτρησης της δύναμης αυτής της συνέπειας. Γενικά οι τιμές του α μπορούν να είναι από 0-1. Στην περίπτωση όπου όλα τα στοιχεία κλίμακας είναι εντελώς ανεξάρτητα το ένα από το άλλο (δηλαδή, δεν συσχετίζονται ή δεν μοιράζονται καμία συνδιακύμανση) η τιμή $\alpha = 0$. Αντίθετα, αν όλα τα στοιχεία μιας μελέτης έχουν υψηλές συνδιακυμάνσεις, τότε το α θα προσεγγίσει το 1 καθώς ο αριθμός των στοιχείων της κλίμακας πλησιάζει στο άπειρο. Με άλλα λόγια, όσο υψηλότερος είναι ο συντελεστής α , τόσο περισσότερο τα ερωτήματα έχουν κοινή συνδιακύμανση και πιθανώς μετρούν την ίδια βασική έννοια.

Το α του Cronbach υπολογίζεται συσχετίζοντας τη βαθμολογία για κάθε μία από τις εννέα μεταβλητές με τη συνολική βαθμολογία για κάθε παρατήρηση στους μεμονωμένους ερωτηθέντες της έρευνας, συγκρίνοντάς το με τη διακύμανση για όλες τις βαθμολογίες κάθε επιμέρους παρατήρησης.

Στην περίπτωση της πρόθεσης χρήσης (ΠΧ) των εργαλείων μεγάλων δεδομένων ο αριθμός των συμπεριλαμβανομένων αντικειμένων είναι 3. Όπως γίνεται διακριτό στον **Πίνακα 4.1** η τιμή του συντελεστή προσεγγίζει πολύ τη μονάδα και συγκεκριμένα είναι 0,96. Πράγμα το οποίο σημαίνει πως η συσχέτιση των ερωτήσεων 1-3 μεταξύ τους είναι πολύ ικανοποιητική.

Πίνακας 4.1. Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 1-3

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,960	0,960	3

Ωστόσο στο σημείο αυτό πρέπει να επισημανθεί ότι στην ερμηνεία του συντελεστή α μιας κλίμακας, μία υψηλή τιμή ($\alpha > 0,95$) μπορεί να είναι ταυτόχρονα συνάρτηση των συνδιακυμάνσεων μεταξύ των στοιχείων και του αριθμού των στοιχείων στην ανάλυση, οπότε ένας υψηλός συντελεστής δεν πρέπει να είναι από μόνος του το σήμα ενός «καλού» ή αξιόπιστου συνόλου αντικειμένων. Δεδομένου ότι συχνά μπορεί να αυξηθεί η τιμή α αλλά αυξάνοντας τον αριθμό των στοιχείων στην ανάλυση. Πράγμα το οποίο δεν παρατηρείται στην περίπτωσή μας.

Στον **Πίνακα 4.2** συνοψίζονται οι μέσοι όροι των απαντήσεων ανά αντικείμενο μαζί με το πλήθος της συμμετοχής και το στατιστικά προσδιοριζόμενο τυπικό σφάλμα ανά περίπτωση. Στο σύνολο ο μέσος όρος των απαντήσεων τείνει προς το 4 της κλίμακας απαντήσεων κατά Linkert, πράγμα το οποίο διακρίνεται και από τα γενικά στατιστικά της κάθε ερώτησης όπου προσεγγίζουν το 94% των απαντήσεων ανά περίπτωση να είναι μεταξύ 3-5. Συγκεκριμένα για την ερώτηση 1 είναι 93,7% των απαντήσεων στην κλίμακα 3-5 και για τις ερωτήσεις 2 & 3 είναι 93,6% στην ίδια κλίμακα αντίστοιχα.

Πίνακας 4.2. Στατιστικά στοιχεία ερωτήσεων 1-3

	M.O.	Τυπικό σφάλμα	N
E1: Σκοπεύω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων στο μέλλον	3,77	0,905	95

E2: Έχω πρόθεση να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων στο μέλλον	3,77	0,928	95
E3: Περιμένω να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων και στο μέλλον	3,79	0,955	95

Στο στοιχείο της αντιλαμβανόμενης χρησιμότητας (AX) των εργαλείων μεγάλων δεδομένων ο αριθμός των συμπεριλαμβανομένων αντικειμένων είναι 5. Ο υπολογισμός του συντελεστή δίνεται στον **Πίνακα 4.3** με τιμή που και εδώ προσεγγίζει πολύ τη μονάδα ($\alpha = 0,953$). Γεγονός που αναδεικνύει μια υψηλή συσχέτιση μεταξύ των ερωτήσεων 4-8.

Πίνακας 4.3. Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 4-8

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,953	0,953	5

Συνοπτικά στον **Πίνακα 4.4** παρουσιάζονται οι μέσοι όροι των απαντήσεων ανά αντικείμενο μαζί με το πλήθος της συμμετοχής και το στατιστικά προσδιοριζόμενο τυπικό σφάλμα ανά περίπτωση. Εδώ ο μέσος όρος των απαντήσεων ξεπερνά το 4 της κλίμακας απαντήσεων. Το ποσοστό των συμμετεχόντων που έδωσαν απάντηση μεταξύ 1 & 2 παρέμεινε κάτω του 7% στο σύνολο των ερωτήσεων, με την απάντηση 3 να στο ελάχιστο 15,8% για την ερώτηση 5 και μέγιστο το 20% για την ερώτηση 8. Οπότε γίνεται κατανοητό ότι η απάντηση της κλίμακας 4 & 5 δόθηκε πάνω από το 70% στο σύνολο των ερωτήσεων.

Πίνακας 4.4. Στατιστικά στοιχεία ερωτήσεων 4-8

	M.O.	Τυπικό σφάλμα	N
E4: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα μου επέτρεπε να διεκπεραιώνω διάφορες εργασίες γρηγορότερα	4,06	0,954	95
E5: Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου επέτρεπε να κάνω τη δουλειά μου με μεγαλύτερη ευκολία	4,05	0,938	95
E6: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα βελτίωνε την ποιότητά της	4,06	0,954	95
E7: Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου ήταν χρήσιμη	4,16	0,891	95
E8: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα ενίσχυε την αποδοτικότητά μου	4,05	0,949	95

Για την μεταβλητή της αντιλαμβανόμενης ευκολίας χρήσης (AEX) των εργαλείων μεγάλων δεδομένων (**Πίνακας 4.5**) περιέχει 5 αντικείμενα (ερωτήσεις 9-13) και παρατηρείται η πρώτη διαφορά εκ των υπολογιζόμενων συντελεστών α . Ο πρώτος συντελεστής έδωσε τιμή 0,922 ενώ στην περίπτωση των τυποποιημένων στοιχείων 0,925. Και στις δύο τιμές των συντελεστών αυτή προσεγγίζει πολύ τη μονάδα. Οι διαφορές μεταξύ των δύο συντελεστών είναι στη μέθοδο υπολογισμού δεδομένου ότι ο πρώτος α του Cronbach χρησιμοποιεί τις συνδιαλλαγές μεταξύ των αντικειμένων, ενώ το α που βασίζεται σε τυποποιημένα στοιχεία χρησιμοποιεί τους συσχετισμούς μεταξύ των αντικειμένων. Το τελευταίο άλφα βασίζεται στην υπόθεση ότι όλα τα στοιχεία έχουν ίσες αποκλίσεις, κάτι που συχνά είναι ψευδές στην πράξη και μπορεί να προκύψουν αποκλίσεις όπως στην περίπτωση του AEX.

Πίνακας 4.5, Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 9-13

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,922	0,925	5

Ο **Πίνακας 4.6** δείχνει όμοια και με την περίπτωση των στοιχείων ΠΧ, ΑΧ τον μέσο όρο των απαντήσεων στις ερωτήσεις 9-13 να τείνουν προς το 4. Στατιστικά αυτό καταγράφεται (**βλέπε ΠΑΡΑΡΤΗΜΑ Γ**) με σχεδόν το απόλυτο 50% των συμμετεχόντων ή/και περισσότερο κατά τις περιπτώσεις των ερωτήσεων 9, 10, 11, 13, να απαντούν με την κλίμακα 4.

Πίνακας 4.6, Στατιστικά στοιχεία ερωτήσεων 9-13

	M.O.	Τυπικό σφάλμα	N
E9: Θα μου ήταν εύκολο να γίνω επιδέξιος/α στη χρήση εργαλείων Μεγάλων Δεδομένων	3,85	0,838	95
E10: Θα μου ήταν εύκολη η χρήση εργαλείων Μεγάλων Δεδομένων	3,84	0,748	95
E11: Θα μου ήταν εύκολη η μάθηση χρήσης εργαλείων Μεγάλων Δεδομένων	3,83	0,781	95
E12: Η αλληλεπίδρασή μου με τα εργαλεία Μεγάλων Δεδομένων θα ήταν ξεκάθαρη και κατανοητή	3,63	0,923	95
E13: Θεωρώ πως θα υπάρχει ευελιξία στην αλληλεπίδραση με τα εργαλεία Μεγάλων Δεδομένων	3,63	0,935	95

Ακόλουθα στο στοιχείο της συνάφειας εργασίας (ΣΕ) των εργαλείων μεγάλων δεδομένων (**Πίνακας 4.7**) ο αριθμός των συμπεριλαμβανομένων αντικειμένων είναι 3 και η προκύπτουσα τιμή συντελεστή $\alpha = 0,916$. Εμφανίζοντας μεγάλο επίπεδο εμπιστοσύνης των ερωτήσεων 14-16.

Πίνακας 4.7, Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 14-16

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,916	0,916	3

Στον **Πίνακα 4.8** συνοψίζονται οι μέσοι όροι των απαντήσεων στις ερωτήσεις 14-16. Όπως είναι διακριτό πέραν της ερώτησης 14 που προσεγγίζει το 4 οι ερωτήσεις 15 & 16 είναι οριακά ανάμεσα του 3-4. Οι συμμετέχοντες απάντησαν 1 & 2 σε αθροιστικά ποσοστά 7,4%, 15,8% και 14,8% αντίστοιχα. Με την πλειοψηφία να απαντά συγκρατημένα 3, 4, ενώ την απάντηση 5 έδωσε και στις τρεις ερωτήσεις ένα 20% και πάνω ομοιόμορφα.

Πίνακας 4.8, Στατιστικά στοιχεία ερωτήσεων 14-16

	M.O.	Τυπικό σφάλμα	N
E14: Στη δουλειά μου η χρήση εργαλείων Μεγάλων Δεδομένων είναι σημαντική	3,75	0,945	95
E15: Η δουλειά μου έχει σχέση με τη χρήση εργαλείων Μεγάλων Δεδομένων	3,54	1,080	95
E16: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απαραίτητη για τη διεκπεραίωση ορισμένων εργασιών που σχετίζονται με τη δουλειά μου	3,61	1,034	95

Στην περίπτωση του επόμενου στοιχείου, της συμβατότητας (ΣΥ), με τα εργαλεία μεγάλων δεδομένων (**Πίνακας 4.9**) περιέχονται 3 αντικείμενα (ερωτήσεις 17-19) και εμφανίζεται η δεύτερη περίπτωση ασυμφωνίας των δύο υπολογιζόμενων συντελεστών α . Όπου ο πρώτος συντελεστής έδωσε τιμή 0,857 ενώ στην περίπτωση των τυποποιημένων στοιχείων 0,861. Και στις δύο τιμές των συντελεστών αυτή προσεγγίζει πολύ τη μονάδα και άρα προσδίδει μεγάλα επίπεδα εμπιστοσύνης.

Πίνακας 4.9, Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 17-19

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,857	0,861	3

Στη μεταβλητή ΣΥ, οι μέσοι όροι ανά αντικείμενο (**Πίνακα 4.10**) εμφανίζουν ότι για τις ερωτήσεις 17 & 18 οι απαντήσεις έτειναν στο 3 και για την ερώτηση 19 ήταν μοιρασμένα μεταξύ 3-4. Πιο αναλυτικά το ποσοστό που απάντησαν στην κλίμακα 1 & 2 στην ερώτηση 17 είναι 20%, της ερώτησης 18 είναι 23,2% και της ερώτησης 19 είναι 13,7% αντίστοιχα.

Πίνακας 4.10, Στατιστικά στοιχεία ερωτήσεων 17-19

	M.O.	Τυπικό σφάλμα	N
E17: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι συμβατή με όλες τις πτυχές της δουλειάς μου	3,21	0,921	95
E18: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απόλυτα συμβατή με την τρέχουσα κατάστασή μου	3,19	1,003	95
E19: Πιστεύω πως η χρήση εργαλείων Μεγάλων Δεδομένων ταιριάζει αρκετά με τον τρόπο που μου αρέσει να εργάζομαι	3,65	1,060	95

Στην περίπτωση του στοιχείου που αφορά τις διαστάσεις των μεγάλων δεδομένων (ΔΜΔ) ο αριθμός των συμπεριλαμβανομένων αντικειμένων είναι επίσης 3 οπότε η τιμή του συντελεστή είναι 0,946 (**Πίνακας 4.11**) και είναι όμοια στις δύο εκφάνσεις του.

Πίνακας 4.11, Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 20-22

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,946	0,946	3

Στον **Πίνακα 4.12** συνοψίζονται οι μέσοι όροι των απαντήσεων ανά αντικείμενο μαζί με το πλήθος της συμμετοχής και το στατιστικά προσδιοριζόμενο τυπικό σφάλμα ανά περίπτωση. Στο σύνολο ο μέσος όρος των απαντήσεων τείνει προς το 4 της κλίμακας απαντήσεων κατά Linkert για τις ερωτήσεις 20-22. Μόλις σχεδόν το 40% ανά ερώτηση, με ελάχιστα πιο κάτω ποσοστιαία στην περίπτωση της ερώτησης 20, απάντησαν 4.

Πίνακας 4.12. Στατιστικά στοιχεία ερωτήσεων 20-22

	M.O.	Τυπικό σφάλμα	N
E20: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία τεράστιων όγκων δεδομένων, κάτι που δεν είναι δυνατό με τη χρήση παραδοσιακών συστημάτων	3,66	1,126	95
E21: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία διαφόρων τύπων δεδομένων, όπως αδόμητα, δομημένα και ημι-δομημένα	3,60	1,066	95
E22: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην ταχύτερη επεξεργασία σε σχέση με τις παραδοσιακές προσεγγίσεις	3,77	1,036	95

Στην περίπτωση της αυτό-αποτελεσματικότητας (AA) των μεγάλων δεδομένων ο αριθμός των συμπεριλαμβανομένων αντικειμένων είναι 4 και όπως γίνεται διακριτό στον **Πίνακα 4.13** η τιμή του συντελεστή εμφανίζεται με τιμή $\alpha = 0,883$.

Πίνακας 4.13. Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 23-26

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,883	0,883	4

Αντίθετα στον **Πίνακα 4.14** φαίνεται ότι για το στοιχείο AA ο μέσος όρος των απαντήσεων είναι μεταξύ 3-4 της κλίμακας απαντήσεων κατά Linkert. Το συγκεκριμένο στοιχείο παρουσιάζει λιγότερο από 10% προσθετικά ανά ερώτηση για τις απαντήσεις που κατέγραψαν 1 & 2 αλλά μοιράζονται στις υπόλοιπες κλίμακες 3-5 (**βλέπε ΠΑΡΑΡΤΗΜΑ Γ**).

Πίνακας 4.14. Στατιστικά στοιχεία ερωτήσεων 23-26

	M.O.	Τυπικό σφάλμα	N
E23: Εάν δεν υπήρχε κάποιος να με καθοδηγεί καθώς προχωρώ σε μια εργασία, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	3,46	0,873	95
E24: Εάν μπορούσα να καλέσω κάποιον για βοήθεια σε περίπτωση που κολλούσα, θα μπορούσα να ολοκληρώσω μια δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	3,60	0,804	95

E25: Εάν είχα πολύ χρόνο για να ολοκληρώσω μία εργασία για την οποία μου παρείχαν λογισμικό, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	3,64	0,849	95
E26: Εάν είχα για βοήθεια μόνο την ενσωματωμένη εγκατάσταση βοήθειας του προγράμματος, θα μπορούσα να ολοκληρώσω τη δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	3,67	0,818	95

Στην περίπτωση της πολυπλοκότητας (ΠΛ) των εργαλείων μεγάλων δεδομένων βασίστηκε σε 4 αντικείμενα (**Πίνακα 4.15**) με τιμή για το συντελεστή να είναι 0,883 και για τις δύο περιπτώσεις υπολογισμού του α .

Πίνακας 4.15, Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 27-30

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,883	0,883	4

Στην ΠΛ όπως διαφαίνεται και στον **Πίνακα 4.16** οι μέσοι όροι των απαντήσεων ανά αντικείμενο μαζί είναι κάτω του 3 με την ερώτηση 30 να προσεγγίζει το 2 ενώ τις υπόλοιπες να είναι ανάμεσα 2-3. Καθότι πιο συγκεκριμένα από τους 95 συμμετέχοντες πάνω από το 1/3 απάντησαν στην κλίμακα 1-2 στο σύνολο των ερωτήσεων και το υπόλοιπο ποσοστό να διαμοιράζεται στις υπόλοιπες κλίμακες.

Πίνακας 4.16, Στατιστικά στοιχεία ερωτήσεων 27-30

	M.O.	Τυπικό σφάλμα	N
E27: Η χρήση εργαλείων Μεγάλων Δεδομένων στερεί αρκετό χρόνο από τις συνηθισμένες υποχρεώσεις μου	2,75	0,934	95
E28: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι τόσο πολύπλοκη που μου είναι δύσκολο να καταλάβω τι γίνεται	2,61	0,992	95
E29: Η χρήση εργαλείων Μεγάλων Δεδομένων απαιτεί αρκετό χρόνο κάνοντας μηχανικές λειτουργίες (π.χ. εισαγωγή δεδομένων)	2,89	1,026	95
E30: Παίρνει τόσο πολύ χρόνο η μάθηση χρήσης του συστήματος, που δεν αξίζει η προσπάθεια	2,34	0,963	95

Τέλος, το άγχος (ΑΓ) συμπεριλήφθηκε ως παράγοντας επηρεασμού της χρήσης των εργαλείων μεγάλων δεδομένων. Τα αντικείμενα αριθμητικά είναι 4 και όπως φαίνεται στον **Πίνακα 4.17** οι τιμές του συντελεστή α εμφανίζουν μια μεταβολή κατά 0,001.

Πίνακας 4.25. Στατιστικά στοιχεία αξιοπιστίας ερωτήσεων 31-34

Cronbach's α	Cronbach's α (Βάσει τυποποιημένων στοιχείων)	N
0,890	0,891	4

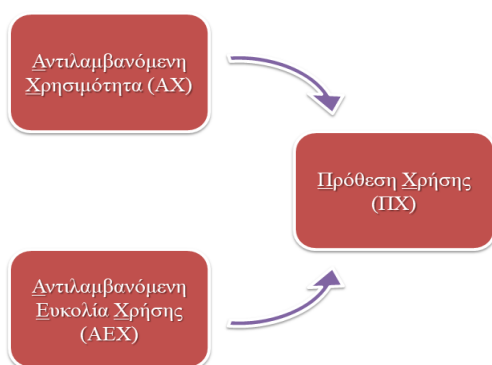
Στον **Πίνακα 4.18**, που συγκεντρώνονται τα στατιστικά των μέσων όρων του ΑΓ, όπως και στην περίπτωση του ΠΛ (**Πίνακα 4.16**) οι μέσοι όροι των απαντήσεων ανά αντικείμενο μαζί είναι κάτω του 3. Η ερώτηση 34 προσεγγίζει το 2 ενώ στις υπόλοιπες των περιπτώσεων πάλι είναι ανάμεσα των 2-3. Όπως αυτό μπορεί να φανεί και με μεγαλύτερη λεπτομέρεια στην παρουσίαση της κατανομής των ερωταπαντήσεων αναλυτικά στο ΠΑΡΑΡΤΗΜΑ Γ.

Πίνακας 4.18. Στατιστικά στοιχεία ερωτήσεων 31-34

	M.O.	Τυπικό σφάλμα	N
E31: Νιώθω ανησυχία σχετικά με τη χρήση εργαλείων Μεγάλων Δεδομένων	2,63	1,022	95
E32: Με τρομάζει η σκέψη πως θα μπορούσα να χάσω πολλές πληροφορίες, ενώ χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων, πατώντας ένα λάθος πλήκτρο	2,89	1,115	95
E33: Διστάζω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων γιατί φοβάμαι μην κάνω λάθη που δεν μπορώ να διορθώσω	2,59	1,047	95
E34: Θεωρώ τα εργαλεία Μεγάλων Δεδομένων εκφοβιστικά	2,22	1,023	95

4. 2. Αναλύσεις παλινδρόμησης

Στο παρόν κεφάλαιο λαμβάνει χώρα η ανάλυση με τον υπολογισμό της γραμμικής παλινδρόμησης σε συγκεκριμένες περιπτώσεις μοντέλων με σκοπό τον προσδιορισμό και επαλήθευση λογικών υποθέσεων του ερευνητικού μοντέλου (βλέπε ενότητες 3.2. & 3.3.). Πιο αναλυτικά θα προσδιορίζονται η τιμή F όπου υποδεικνύει εάν το συνολικό μοντέλο είναι στατιστικά σημαντικό στο να εξηγεί την ανεξάρτητη μεταβλητή και η τιμή του προσαρμοσμένου R^2 όπου υποδεικνύει το ποσοστό της συνολικής διακύμανσης της εξαρτημένης μεταβλητής όπως αυτή μεταφράζεται από τις αντίστοιχα χρησιμοποιούμενες ανεξάρτητες μεταβλητές στο μοντέλο παλινδρόμησης κάθε φορά.



Ως πρώτη περίπτωση μελέτης επιλέχθηκε η επίδραση στην Πρόθεση Χρήσης (PIX) από την Αντιλαμβανόμενη Χρησιμότητα (AX) και την Αντιλαμβανόμενη Ευκολία Χρήσης (AEX). Όπως φαίνεται στον **Πίνακα 4.20** το $F_{(2;92)} = 68,124$, τιμή που είναι αρκετά μεγαλύτερη του 1 και σε συνδυασμό με το πολύ χαμηλό p ($p < 0,001$), καταλήγουμε στο συμπέρασμα ότι το μοντέλο βελτιώνει σημαντικά την ερμηνεία της διασποράς της εξαρτημένης μεταβλητής σε σχέση με το βασικό μοντέλο. Στη συνέχεια παρατηρώντας το προσαρμοσμένο R^2 , επίσης στον **Πίνακα 4.20**, διαπιστώνουμε πως το μοντέλο εξηγεί το 58,8% της συνολικής διασποράς της εξαρτημένης μεταβλητής PIX. Τέλος, στον **Πίνακα 4.21** από τα αποτελέσματα της AX ($B = 0,638$) και AEX ($B = 0,272$) διαπιστώνουμε πως η AX έχει μεγαλύτερη επίδραση στην PIX σε σχέση με την AEX, αλλά και οι δύο επηρεάζουν θετικά την PIX.

Πίνακας 4.19, Μεταβλητές που εισήχθησαν/αφαιρέθηκαν^a κατά τη ανάλυση παλινδρόμησης

α/α	Εισαχθέντες μεταβλητές	Μεταβλητές που αφαιρέθηκαν	Μέθοδος
1	AEX, AX ^b	.	Εισαγωγή

a. Εξαρτημένη μεταβλητή: PIX

b. Όλες οι εισαχθέντες μεταβλητές.

Πίνακας 4.20, Περίληψη στατιστικών στοιχείων για το 1^ο μοντέλο παλινδρόμησης

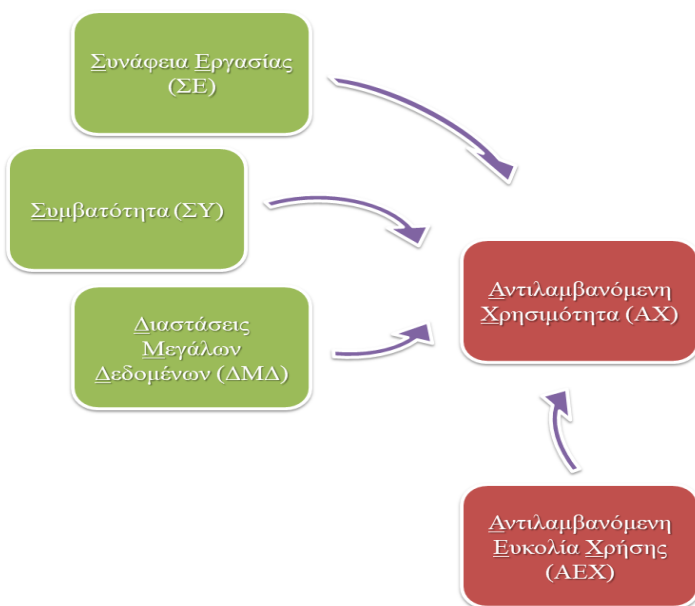
α/α	R	R ²	Προσαρμοσμένο R ²	Τυπικό σφάλμα εκτίμησης	Στατιστικά μεταβολής				
					R ² Μεταβλητό	F Μεταβλητό	df1	df2	p (F Μεταβλητό)
1	0,773 ^a	0,597	0,588	0,57399	0,597	68,124	2	92	0,000

a. Προγνωστικά: (Constant), AEX, AX

Πίνακας 4.21, Περίληψη στατιστικών συντελεστών^a για το 1^ο μοντέλο παλινδρόμησης

α/α		Μη τυποποιημένοι συντελεστές		Τυποποιημένοι συντελεστές	t	p	95,0% Διάστημα εμπιστοσύνης για το B	
		B	Τυπικό σφάλμα	β			Κάτω όριο	Άνω όριο
1	Σταθερά	0,153	0,329		0,466	0,642	-0,501	0,807
	AX	0,638	0,087	0,613	7,287	0,000	0,464	0,811
	AEX	0,272	0,102	0,225	2,674	0,009	0,070	0,474

a. Εξαρτημένη μεταβλητή: ΠΧ



Ως δεύτερη περίπτωση μελέτης επιλέχθηκε η επίδραση στην Αντιλαμβανόμενη Χρησιμότητα (ΑΧ) από τις μεταβλητές Διαστάσεις Μεγάλων Δεδομένων (ΔΜΔ), Αντιλαμβανόμενη Ευκολία Χρήσης (ΑΕΧ), Συνάφεια Εργασίας (ΣΕ) και Συμβατότητα (ΣΥ). Όπως προκύπτει από τον **Πίνακα 4.23**, $F_{(4;90)} = 40,529$, τιμή που είναι αρκετά μεγαλύτερη του 1 και σε συνδυασμό με το πολύ χαμηλό p ($p < 0,001$), καταλήγουμε και σε αυτή την περίπτωση στο συμπέρασμα ότι το μοντέλο βελτιώνει σημαντικά την ερμηνεία της διασποράς της

εξαρτημένης μεταβλητής σε σχέση με το βασικό μοντέλο. Το μοντέλο σε αυτή την περίπτωση εξηγεί το 62,7% της συνολικής διασποράς της εξαρτημένης μεταβλητής AX (προσαρμοσμένο $R^2 = 0,627$). Τέλος, στον Πίνακα 4.24 από τα αποτελέσματα των ΔΜΔ (B = 0,319), AEX (B = 0,212), ΣΕ (B = 0,104) και ΣΥ (B = 0,250) παρατηρούμε το μέγεθος επίδρασης, καθώς και την κατεύθυνση του αποτελέσματος που προκαλεί η κάθε μεταβλητή στην εξαρτημένη AX. Άξιο αναφοράς σε αυτή την περίπτωση, η τιμή του p για τη μεταβλητή ΣΕ ($p = 0,261$). Δηλαδή υπάρχει 26,1% πιθανότητα τα αποτελέσματα να έχουν εμφανιστεί με τυχαία κατανομή και μπορούμε να πούμε πως υπάρχει 73,9% πιθανότητα η μεταβλητή να έχει κάποιο αποτέλεσμα στο μοντέλο μας, τιμή που είναι εκτός του ορίου εμπιστοσύνης που έχει τεθεί ως όριο στη μελέτη μας (95%).

Πίνακας 4.22, Μεταβλητές που εισήχθηκαν/αφαιρέθηκαν^a κατά τη ανάλυση παλινδρόμησης

α/α	Εισαχθέντες μεταβλητές	Μεταβλητές που αφαιρέθηκαν	Μέθοδος
2	ΔΜΔ, AEX, ΣΕ, ΣΥ ^b	.	Εισαγωγή

a. Εξαρτημένη μεταβλητή: AX

b. Όλες οι εισαχθέντες μεταβλητές.

Πίνακας 4.23, Περίληψη στατιστικών στοιχείων για το 2^ο μοντέλο παλινδρόμησης

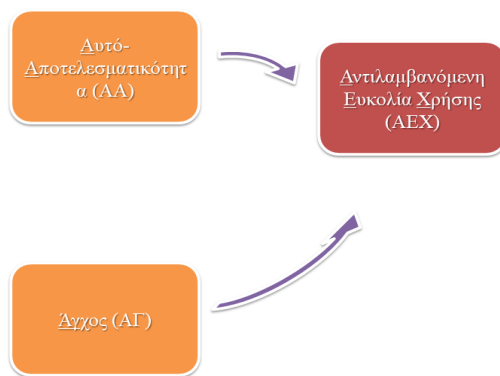
α/α	R	R ²	Προσαρμοσμένο R ²	Τυπικό σφάλμα εκτίμησης	Στατιστικά μεταβολής				
					R ² Μεταβλητό	F Μεταβλητό	df1	df2	P (F Μεταβλητό)
2	0,802 ^a	0,643	0,627	0,52520	0,643	40,529	4	90	0,000

a. Προγνωστικά: (Constant), ΔΜΔ, AEX, ΣΕ, ΣΥ

Πίνακας 4.24, Περίληψη στατιστικών συντελεστών^a για το 2^ο μοντέλο παλινδρόμησης

α/α		Μη τυποποιημένοι συντελεστές		Τυποποιημένοι συντελεστές	t	P	95,0% διάστημα εμπιστοσύνης για το B	
		B	Τυπικό σφάλμα	β			Κάτω όριο	Άνω όριο
2	Σταθερά	0,894	0,291		3,070	0,003	0,315	1,472
	AEX	0,212	0,096	0,183	2,210	0,030	0,021	0,403
	ΣΕ	0,104	0,092	0,114	1,131	0,261	-0,078	0,286
	ΣΥ	0,250	0,107	0,255	2,346	0,021	0,038	0,462
	ΔΜΔ	0,319	0,078	0,379	4,085	0,000	0,164	0,474

a. Εξαρτημένη μεταβλητή: AX



Σαν τρίτη περίπτωση μελέτης επιλέχθηκε η επίδραση στην Αντιλαμβανόμενη Ευκολία Χρήσης (AEX) από τις μεταβλητές Αυτό-αποτελεσματικότητα (AA) και Άγχος (ΑΓ). Όπως προκύπτει από τον **Πίνακα 4.26**, $F_{(2;92)} = 37,15$, τιμή που είναι αρκετά μεγαλύτερη του 1 και σε συνδυασμό με το πολύ χαμηλό p ($p < 0,001$), καταλήγουμε και σε αυτή την περίπτωση στο συμπέρασμα ότι το μοντέλο βελτιώνει σημαντικά την ερμηνεία της διασποράς της εξαρτημένης μεταβλητής σε σχέση με το βασικό μοντέλο. Το μοντέλο σε αυτή την περίπτωση εξηγεί το 43,5% της συνολικής διασποράς της εξαρτημένης μεταβλητής AEX (προσαρμοσμένο $R^2 = 0,435$). Τέλος, στον **Πίνακα 4.27** από τα αποτελέσματα των AA ($B = 0,602$), ΑΓ ($B = -0,212$), παρατηρούμε πως η AA έχει μεγάλη και θετική επίδραση στην εξαρτημένη AEX, σε αντίθεση με την ΑΓ που έχει μικρότερη και αρνητική επίδραση.

Πίνακας 4.25, Μεταβλητές που εισήχθηκαν/αφαιρέθηκαν^a κατά τη ανάλυση παλινδρόμησης

α/α	Εισαχθέντες μεταβλητές	Μεταβλητές που αφαιρέθηκαν	Μέθοδος
3	ΑΓ, AA ^b	.	Εισαγωγή

a. Εξαρτημένη μεταβλητή: AEX

b. Όλες οι εισαχθέντες μεταβλητές.

Πίνακας 4.26, Περίληψη στατιστικών στοιχείων για το 3^ο μοντέλο παλινδρόμησης

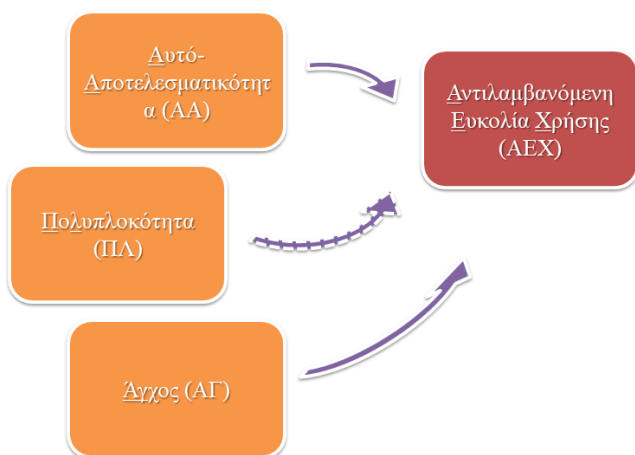
α/α	R	R ²	Προσαρμοσμένο R ²	Τυπικό σφάλμα εκτίμησης	Στατιστικά μεταβολής				
					R ² Μεταβλητό	F Μεταβλητό	df1	df2	P (F Μεταβλητό)
3	0,668 ^a	0,447	0,435	0,55655	0,447	37,150	2	92	0,000

a. Μεταβλητές: Σταθερά, ΑΓ, AA

Πίνακας 4.27. Περίληψη στατιστικών συντελεστών^a για το 3^ο μοντέλο παλινδρόμησης

α/α		Μη τυποποιημένοι συντελεστές		Τυποποιημένοι συντελεστές	t	P	95,0% Διάστημα εμπιστοσύνης για το Β	
		B	Τυπικό σφάλμα	β			Κάτω όριο	Άνω όριο
3	Σταθερά	2,144	0,353		6,068	0,000	1,442	2,845
	ΑΑ	0,602	0,080	0,585	7,488	0,000	0,442	0,761
	ΑΓ	-0,212	0,063	-0,262	-3,351	0,001	-0,338	-0,086

a. Εξαρτημένη μεταβλητή: ΑΕΧ



Τέλος, επαναυπολογίστηκε το τρίτο μοντέλο αλλά αυτή τη φορά με την προσθήκη της μεταβλητής της Πολυπλοκότητας (ΠΛ). Η σύνοψη των αποτελεσμάτων όπως αυτά παρατίθενται στον **Πίνακα 4.29** δεν διαφέρουν πολύ με αυτά της τρίτης περίπτωσης μελέτης, με $F_{(3;91)} = 24,664$ τιμή αρκετά μεγαλύτερη του 1 και σε συνδυασμό με το επίσης πολύ χαμηλό p ($p < 0,001$), καταλήγουμε και σε αυτή την περίπτωση στο συμπέρασμα ότι το μοντέλο βελτιώνει σημαντικά την ερμηνεία της διασποράς της εξαρτημένης μεταβλητής σε σχέση με το βασικό μοντέλο. Σε αυτή την περίπτωση το προσαρμοσμένο έχει σχεδόν όμοια τιμή $R^2 = 0,430$ κι εξηγεί το 43% της συνολικής διασποράς της εξαρτημένης μεταβλητής ΑΕΧ. Στον **Πίνακα 4.30** που συνοψίζονται τα αποτελέσματα των ΑΑ ($B = 0,610$), ΑΓ ($B = -0,181$) και ΠΛ ($B = -0,049$), προκύπτει πως η ΠΛ δεν έχει στατιστικά σημαντική συνεισφορά στην ΑΕΧ και η επίδρασή της είναι αρνητική όπως και της ΑΓ.

Πίνακας 4.28, Μεταβλητές που εισήχθηκαν/αφαιρέθηκαν^a κατά τη ανάλυση παλινδρόμησης

α/α	Εισαχθέντες μεταβλητές	Μεταβλητές που αφαιρέθηκαν	Μέθοδος
4	ΑΓ, ΑΑ, ΠΛ ^b	.	Εισαγωγή

- a. Εξαρτημένη μεταβλητή: AEX
 b. Όλες οι εισαχθέντες μεταβλητές.

Πίνακας 4.29, Περίληψη στατιστικών στοιχείων για το 4^ο μοντέλο παλινδρόμησης

Model	R	R ²	Προσαρμοσμένο R ²	Τυπικό σφάλμα εκτίμησης	Στατιστικά μεταβολής				
					R ² Μεταβλητό	F Μεταβλητό	df1	df2	P (F Μεταβλητό)
1	0,670 ^a	0,448	0,430	0,55876	0,448	24,664	3	91	0,000

- a. Προγνωστικά: (Constant), ΑΓ, ΑΑ, ΠΛ

Πίνακας 4.30, Περίληψη στατιστικών συντελεστών^a για το 4^ο μοντέλο παλινδρόμησης

Model		Μη τυποποιημένοι συντελεστές		Τυποποιημένοι συντελεστές	t	P	95,0% Διάστημα εμπιστοσύνης για το B	
		B	Τυπικό σφάλμα	β			Κάτω όριο	Άνω όριο
1	(Constant)	2,163	0,356		6,066	0,000	1,454	2,871
	ΑΑ	0,610	0,082	0,593	7,415	0,000	0,447	0,774
	ΠΛ	-0,049	0,094	-0,056	-0,525	0,601	-0,236	0,137
	ΑΓ	-0,181	0,087	-0,223	-2,081	0,040	-0,354	-0,008

- a. Εξαρτημένη μεταβλητή: AEX

5. Συμπεράσματα & συζήτηση

Η παρούσα έρευνα που πραγματοποιήθηκε με σκοπό να διερευνηθεί η πρόθεση χρήσης τεχνολογίας μαζικών δεδομένων από τους λογιστές – φοροτέχνες και υπαλλήλους λογιστηρίων της χώρας μας, βασίστηκε στο πρόσφατα δημοσιευμένο μοντέλο TAM του Okcu και των συνεργατών (2019). Κατά τον έλεγχο για την αξιοπιστία του μοντέλου προέκυψαν οι τιμές όπως αυτές φαίνονται στον πίνακα 5.1.

Πίνακας 5.1, Αξιοπιστία Cronbach's Alpha

<u>Μεταβλητές</u>	<u>Ερωτήσεις</u>	<u>Cronbach's Alpha</u>
ΠΧ	1 - 3	0,960
ΑΧ	4 - 8	0,953
ΑΕΧ	9 - 13	0,922
ΣΕ	14 - 16	0,916
ΣΥ	17 - 19	0,857
ΔΜΔ	20 - 22	0,946
ΑΑ	23 - 26	0,883
ΠΛ	27 - 30	0,883
ΑΓ	31 - 34	0,890

Σύμφωνα με τους Tavakol και Dennick (2011) κάθε τιμή Cronbach's alpha από 0,70 και πάνω θεωρείται αποδεκτή. Οι τιμές που προέκυψαν από το μοντέλο που χρησιμοποιήθηκε ήταν πολύ μεγαλύτερες (με μεγαλύτερη το 0,960 και μικρότερη το 0,857). Συνεπώς, το ερωτηματολόγιο που δημιουργήθηκε θεωρείται σε υψηλό βαθμό αξιόπιστο.

Η πρώτη εικόνα που προκύπτει μέσω των απαντήσεων που δόθηκαν παρουσιάζει μία ευνοϊκή στάση των ερωτηθέντων απέναντι στη χρήση και την πρόθεση χρήσης εργαλείων Μαζικών Δεδομένων. Παρατηρώντας τα στατιστικά ανά ερώτηση (βλέπε Παράρτημα Γ) είναι εμφανές πως το μεγαλύτερο ποσοστό και σχεδόν σε όλες τις ερωτήσεις απαντά επιλέγοντας 4 (που αντιστοιχεί σε «Συμφωνώ»). Στις ερωτήσεις των μεταβλητών Συνάφεια Εργασίας, Συμβατότητα, Διαστάσεις Μεγάλων Δεδομένων και Αυτό-αποτελεσματικότητα το μεγαλύτερο ποσοστό έχει δώσει απάντηση 3 και 4 (που αντιστοιχεί σε «Ούτε Συμφωνώ Ούτε Διαφωνώ» και σε «Συμφωνώ») και στις ερωτήσεις της μεταβλητής Αντιλαμβανόμενη Χρησιμότητα το μεγαλύτερο ποσοστό έχει δώσει απάντηση 5 (που αντιστοιχεί σε «Συμφωνώ Απόλυτα»). Η εικόνα αλλάζει στις ερωτήσεις των μεταβλητών Πολυπλοκότητα και Άγχος, με το μεγαλύτερο ποσοστό των απαντήσεων να συγκεντρώνεται στην απάντηση 2 (που αντιστοιχεί σε «Διαφωνώ»). Οι 2 αυτές μεταβλητές, όπως έχει ήδη αναφερθεί και στο κεφάλαιο 4, έχουν αρνητική σχέση με την Πρόθεση Χρήσης, επομένως οι συγκεκριμένες απαντήσεις επιβεβαιώνουν το συμπέρασμά μας για θετική στάση των συμμετεχόντων, απέναντι στη χρήση και την πρόθεση χρήσης εργαλείων Μαζικών Δεδομένων.

Ανατρέχοντας στα αποτελέσματα των γραμμικών παλινδρομήσεων, παρατηρούμε πως οι μεταβλητές Αντιλαμβανόμενη Χρησιμότητα, όπως προκύπτει από την πρώτη παλινδρόμηση, και Αυτό-αποτελεσματικότητα, όπως προκύπτει από την τρίτη και τέταρτη παλινδρόμηση, επηρεάζουν θετικά και σε μεγάλο βαθμό την Πρόθεση Χρήσης (63,8%, 60,2% και 61% αντίστοιχα). Οι μεταβλητές Άγχος και Πολυπλοκότητα, από την άλλη, έχουμε διαπιστώσει πως

έχουν αρνητική σχέση με την πρόθεση χρήσης. Αυτά τα αποτελέσματα θα μπορούσαν να αποτελέσουν τη βάση για τη διαμόρφωση νέων πολιτικών από μάνατζερς ή άλλους διαμορφωτές πολιτικής. Για παράδειγμα η απλούστευση της μορφής ή των διαδικασιών κατά τη χρήση ενός εργαλείου μαζικών δεδομένων, σε συνδυασμό με την εκπαίδευση σε αυτό ανά τακτά χρονικά διαστήματα, θα μπορούσε να οδηγήσει σε αύξηση της Αντιλαμβανόμενης Χρησιμότητας και της Αυτό-αποτελεσματικότητας, οι οποίες λόγω της θετικής επίδρασης θα προκαλούσαν αύξηση της Πρόθεσης Χρήσης. Όλα αυτά, όμως, θα προκαλούσαν και μείωση στην πολυπλοκότητα και το άγχος, οι οποίες λόγω της αρνητικής επίδρασης θα προκαλούσαν επιπλέον αύξηση της Πρόθεσης Χρήσης.

Η έρευνα, όπως έχει ήδη αναφερθεί, βασίστηκε στο δείγμα του πληθυσμού που απάντησε στο ερωτηματολόγιο, N=95. Αν αναλογιστεί κανείς το πλήθος των λογιστών της χώρας μας, το δείγμα που συγκεντρώθηκε μπορεί να θεωρηθεί πολύ μικρό. Εδώ προκύπτει κι ο μεγαλύτερος περιορισμός της έρευνας. Εξαιτίας της κατάστασης της Πανδημίας λόγω Covid-19, η διάδοση του ερωτηματολογίου έπρεπε να περιοριστεί μόνο σε ηλεκτρονικά μέσα κι απομακρυσμένα. Επιπρόσθετα, είναι γνωστό πως τους τελευταίους μήνες η κατάσταση της πανδημίας έχει προκαλέσει τέτοιο φόρτο εργασίας στους λογιστές, που ότι παρεκκλίνει από τις τρέχουσες κι επείγουσες εργασίες τους δύσκολα θα αποτελέσει αντικείμενο ενασχόλησης. Επομένως, ίσως να παρουσίαζε έντονο ενδιαφέρον η πραγματοποίηση της ίδιας έρευνας σε κάποια άλλη χρονική περίοδο.

Παρόμοια έρευνα δεν είχε πραγματοποιηθεί μέχρι σήμερα στην Ελλάδα και συγκεκριμένα για τον κλάδο των λογιστών, με αποτέλεσμα το γεγονός αυτό να προσδίδει αξία και κατ' επέκταση να ενισχύει τη χρησιμότητα της παρούσας έρευνας. Η τεχνολογική εξέλιξη είναι πλέον δεδομένη και ραγδαία και η πλειοψηφία των Ελλήνων λογιστών που συμμετείχαν δείχνουν να το κατανοούν, να αναμένουν τις επερχόμενες αλλαγές και να υιοθετούν τις ήδη υπάρχουσες. Αν μπορούσαμε να δώσουμε κάποια συμβουλή προς τους επαγγελματίες του κλάδου, αυτή θα ήταν να ακολουθήσουν το παράδειγμα των συναδέλφων τους που συμμετείχαν στην έρευνα, να προετοιμαστούν και να αγκαλιάσουν τις συνεχείς εξελίξεις διότι αυτές αποτελούν την πραγματικότητα, η οποία βρίσκεται ήδη εδώ.

6. Βιβλιογραφικές Αναφορές

- Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, 17(2), 3.
- Abd Wahab, M. H., Mohd, M. N., Hanafi, H. F., & Mohamad Mohsin, M. F. (2008). Data pre-processing on web server logs for generalized association rules mining algorithm. *World Academy of Science, Engineering and Technology*, 48, 190-197.
- Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS quarterly*, 665-694.
- Agarwal, R., & Prasad, J. (1999). Are individual differences germane to the acceptance of new information technologies? *Decision sciences*, 30(2), 361-391.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.
- Ajzen, I., & Madden, T. (1986). Prediction of goal-directed behavior from attitudinal and normative variables. *Journal of Experimental Social Psychology*, 22(5), 453-474.
- Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26(2), 173-194.
- Alalwan, A. A., Dwivedi, Y. K., Rana, N. P., & Williams, M. D. (2016). Consumer adoption of mobile banking in Jordan: Examining the role of usefulness, ease of use, perceived risk and self-efficacy. *Journal of enterprise information management*.
- Alles, M., Brennan, G., Kogan, A., & Vasarhelyi, M. A. (2018). Continuous Monitoring of Business Process Controls: A Pilot Implementation of a Continuous Auditing System at Siemens1 *Continuous Auditing*: Emerald Publishing Limited.
- Anandarajan, M., Igbaria, M., & Anakwe, U. P. (2000). Technology acceptance in the banking industry: A perspective from a less developed country. *Information Technology & People*.
- Ananthakrishna, R., Chaudhuri, S., & Ganti, V. (2002). *Eliminating fuzzy duplicates in data warehouses*. Paper presented at the VLDB'02: Proceedings of the 28th International Conference on Very Large Databases.
- Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science*, 48, 319-324.
- Arasu, A., Chaudhuri, S., & Kaushik, R. (2008). *Transformation-based framework for record matching*. Paper presented at the 2008 IEEE 24th International Conference on Data Engineering.
- Ajzen, I. (1980). Understanding attitudes and predicting social behavior. *Englewood Cliffs*.

- Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of accounting research*, 159-178.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American psychologist*, 37(2), 122.
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of social and clinical psychology*, 4(3), 359-373.
- Barrenetxea, G., Ingelrest, F., Schaefer, G., Vetterli, M., Couach, O., & Parlange, M. (2008). *Sensorscope: Out-of-the-box environmental monitoring*. Paper presented at the 2008 International Conference on Information Processing in Sensor Networks (ipsn 2008).
- Bell, T. B., Peecher, M. E., & Solomon, I. (2005). *The 21st century public company audit: conceptual elements of kpmg's global audit methodology*: KPMG international.
- Bennett, G. B., & Hatfield, R. C. (2013). The effect of the social mismatch between staff auditors and client management on the collection of audit evidence. *The Accounting Review*, 88(1), 31-50.
- Beyer, M. (2011). Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data. *Gartner*. Archived from the original on, 10.
- Bhattacharjee, A., & Sanford, C. (2006). Influence processes for information technology acceptance: An elaboration likelihood model. *MIS quarterly*, 805-825.
- Brannen, L. (2006). Upfront: Continuous auditing is ready for prime time. *Business Finance Magazine*(August 1).
- Calisir, F., Altin Gumussoy, C., Bayraktaroglu, A. E., & Karaali, D. (2014). Predicting the intention to use a web-based learning system: Perceived content quality, anxiety, perceived system quality, image, and the technology acceptance model. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 24(5), 515-531.
- Calisir, F., & Calisir, F. (2004). The relation of interface usability characteristics, perceived usefulness, and perceived ease of use to end-user satisfaction with enterprise resource planning (ERP) systems. *Computers in Human behavior*, 20(4), 505-515.
- Cao, M., Chychyla, R., & Stewart, T. (2015). Big Data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423-429.
- Cattell, R. (2011). Scalable SQL and NoSQL data stores. *Acm Sigmod Record*, 39(4), 12-27.
- Cerriotti, M., Mottola, L., Picco, G. P., Murphy, A. L., Guna, S., Corra, M., . . . Zanon, P. (2009). *Monitoring heritage buildings with wireless sensor networks: The Torre Aquila deployment*. Paper presented at the 2009 International Conference on Information Processing in Sensor Networks.

- Chan, S.-c. (2004). Understanding internet banking adoption and use behavior: A Hong Kong perspective. *Journal of Global Information Management (JGIM)*, 12(3), 21-43.
- Chandramohan, V., & Christensen, K. (2002). *A first look at wired sensor networks for video surveillance systems*. Paper presented at the 27th Annual IEEE Conference on Local Computer Networks, 2002. Proceedings. LCN 2002.
- Chang, S. C., & Tung, F. C. (2008). An empirical investigation of students' behavioural intentions to use the online learning course websites. *British Journal of Educational Technology*, 39(1), 71-83.
- Chatzoglou, P. D., Sarigiannidis, L., Vraimaki, E., & Diamantidis, A. (2009). Investigating Greek employees' intention to use web-based training. *Computers & education*, 53(3), 877-889.
- Chau, P. Y., & Hu, P. J. H. (2001). Information technology acceptance by individual professionals: A model comparison approach. *Decision sciences*, 32(4), 699-719.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of computer Science*, 7(2), 157-164.
- Chen, L.-d., Gillenson, M. L., & Sherrell, D. L. (2004). Consumer acceptance of virtual stores: a theoretical model and critical success factors for virtual stores. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 35(2), 8-31.
- Chiang, F., & Miller, R. J. (2008). Discovering data quality rules. *Proceedings of the VLDB Endowment*, 1(1), 1166-1177.
- Chismar, W. G., & Wiley-Patton, S. (2003). *Does the extended technology acceptance model apply to physicians*. Paper presented at the 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the.
- Cho, J., & Garcia-Molina, H. (2002). *Parallel crawlers*. Paper presented at the Proceedings of the 11th international conference on World Wide Web.
- Choudhary, S., Dincturk, M. E., Mirtaheri, S. M., Moosavi, A., Von Bochmann, G., Jourdan, G.-V., & Onut, I. V. (2012). *Crawling rich internet applications: the state of the art*. Paper presented at the Proceedings of the 2012 Conference of the Center for Advanced Studies on Collaborative Research.
- Ciriani, V., Di Vimercati, S. D. C., Foresti, S., Jajodia, S., Paraboschi, S., & Samarati, P. (2009). *Keep a few: Outsourcing data while maintaining confidentiality*. Paper presented at the European Symposium on Research in Computer Security.
- Cukier, K. (2010). *Data, data everywhere: A special report on managing information*: Economist Newspaper.

- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.
- Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International journal of man-machine studies*, 38(3), 475-487.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management science*, 35(8), 982-1003.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace 1. *Journal of applied social psychology*, 22(14), 1111-1132.
- Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.
- DeWitt, D., & Gray, J. (1992). Parallel database systems: the future of high performance database systems. *Communications of the ACM*, 35(6), 85-98.
- Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine learning*, 42(1-2), 143-175.
- Diebold, F. X. (2012). On the Origin (s) and Development of the Term 'Big Data'.
- Dogo, E. M., Makaba, T., Afolabi, O. J., & Ajibo, A. C. (2021). Combating Road Traffic Congestion with Big Data: A Bibliometric Review and Analysis of Scientific Research. In U. Z. A. Hamid & F. Al-Turjman (Eds.), *Towards Connected and Autonomous Vehicle Highways: Technical, Security and Social Challenges* (pp. 43-86). Cham: Springer International Publishing.
- Drosou, M., Jagadish, H. V., Pitoura, E., & Stoyanovich, J. (2017). Diversity in Big Data: A Review. *Big data*, 5(2), 73-84. doi: 10.1089/big.2016.0054
- Dunn, J. (1996). *Auditing: Theory and practice*: Prentice Hall.
- Eberendu, A. C. (2016). Unstructured Data: an overview of the data of Big Data. *International Journal of Computer Trends and Technology*, 38(1), 46-50.
- Emmanuel, I., & Stanier, C. (2016). *Defining big data*. Paper presented at the Proceedings of the International Conference on Big Data and Advanced Wireless Technologies.
- Engle, R. F., Granger, C. W., Rice, J., & Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association*, 81(394), 310-320.
- Esteves, J., & Curto, J. (2013). *A risk and benefits behavioral model to assess intentions to adopt big data*. Paper presented at the Proceedings of the 10th International Conference on Intellectual Capital, Knowledge Management and Organisational Learning: ICICKM 2013.

- Fan, W., & Geerts, F. (2010). *Capturing missing tuples and missing values*. Paper presented at the Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems.
- Fan, W., Geerts, F., & Wijsen, J. (2012). Determining the currency of data. *ACM Transactions on Database Systems (TODS)*, 37(4), 1-46.
- Fan, W., Jia, X., Li, J., & Ma, S. (2009). Reasoning about record matching rules. *Proceedings of the VLDB Endowment*, 2(1), 407-418.
- Frankel, F., & Reid, R. (2008). Big data: Distilling meaning from data. *Nature*, 455(7209), 30-30.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, 114, 254-280.
- Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., & Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management*, 36(3), 403-413.
- Fu, J.-R., Farn, C.-K., & Chao, W.-P. (2006). Acceptance of electronic tax filing: A study of taxpayer intentions. *Information & management*, 43(1), 109-126.
- Fusilier, M., & Durlabhji, S. (2005). An exploration of student internet use in India: the technology acceptance model and the theory of planned behaviour. *Campus-Wide Information Systems*.
- Fuxman, A., & Miller, R. J. (2007). First-order query rewriting for inconsistent databases. *Journal of Computer and System Sciences*, 73(4), 610-635.
- Gal, G. (2008). Query issues in continuous reporting systems. *Journal of Emerging Technologies in Accounting*, 5(1), 81-97.
- Gallego, M. D., Luna, P., & Bueno, S. (2008). User acceptance model of open source software. *Computers in Human behavior*, 24(5), 2199-2216.
- Gantz, J., & Reinsel, D. (2010). IDC–IVIEW.
- Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview*, 1142(2011), 1-12.
- Gefen, D. (2003). TAM or just plain habit: A look at experienced online shoppers. *Journal of Organizational and End User Computing (JOEUC)*, 15(3), 1-13.
- Gentry, C. (2009). *Fully homomorphic encryption using ideal lattices*. Paper presented at the Proceedings of the forty-first annual ACM symposium on Theory of computing.
- Getoor, L. (2010). *Graph identification*. Paper presented at the International Symposium on Intelligent Data Analysis.

- Ghemawat, S., Gobiuff, H., & Leung, S.-T. (2003). *The Google file system*. Paper presented at the Proceedings of the nineteenth ACM symposium on Operating systems principles.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012-1014.
- Golab, L., Karloff, H., Korn, F., Saha, A., & Srivastava, D. (2009). Sequential dependencies. *Proceedings of the VLDB Endowment*, *2*(1), 574-585.
- Grobelnik, M. (2012). Big data tutorial. *Kalamaki: Jožef Stefan Institute*.
- Groomer, S. M., & Murthy, U. S. (2018). Continuous Auditing of Database Applications: An Embedded Audit Module Approach1 *Continuous auditing*: Emerald Publishing Limited.
- Guriting, P., & Ndubisi, N. O. (2006). Borneo online banking: evaluating customer perceptions and behavioural intention. *Management research news*.
- Hackbarth, G., Grover, V., & Mun, Y. Y. (2003). Computer playfulness and anxiety: positive and negative mediators of the system experience effect on perceived ease of use. *Information & management*, *40*(3), 221-232.
- Hernández, M. A., & Stolfo, S. J. (1998). Real-world data is dirty: Data cleansing and the merge/purge problem. *Data mining and knowledge discovery*, *2*(1), 9-37.
- Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1): Microsoft research Redmond, WA.
- Holton, C. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support Systems*, *46*(4), 853-864.
- Hong, S.-J., & Tam, K. Y. (2006). Understanding the adoption of multipurpose information appliances: The case of mobile data services. *Information systems research*, *17*(2), 162-179.
- Horst, M., Kuttschreuter, M., & Gutteling, J. M. (2007). Perceived usefulness, personal experiences, risk perception and trust as determinants of adoption of e-government services in The Netherlands. *Computers in Human behavior*, *23*(4), 1838-1852.
- Howard, J. H., Kazar, M. L., Menees, S. G., Nichols, D. A., Satyanarayanan, M., Sidebotham, R. N., & West, M. J. (1988). Scale and performance in a distributed file system. *ACM Transactions on Computer Systems (TOCS)*, *6*(1), 51-81.
- Hu, P. J., Chau, P. Y., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of management information systems*, *16*(2), 91-112.

- Huang, T. C.-K., Liu, C.-C., & Chang, D.-C. (2012). An empirical investigation of factors influencing the adoption of data mining tools. *International Journal of Information Management*, 32(3), 257-270.
- Huffman, A. H., Whetten, J., & Huffman, W. H. (2013). Using technology in higher education: The influence of gender roles on technology self-efficacy. *Computers in Human behavior*, 29(4), 1779-1786.
- Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), 585-594.
- Igbaria, M., & Iivari, J. (1995). The effects of self-efficacy on computer usage. *Omega*, 23(6), 587-605.
- Issa, H., & Kogan, A. (2014). A predictive ordered logistic regression model as a tool for quality review of control risk assessments. *Journal of Information Systems*, 28(2), 209-229.
- Ittner, C. D., & Larcker, D. F. (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of accounting research*, 36, 1-35.
- Jeacle, I., & Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society*, 36(4-5), 293-309.
- Johnstone, K. M. (2000). Client-acceptance decisions: Simultaneous effects of client business risk, audit risk, auditor business risk, and risk adaptation. *Auditing: A Journal of Practice & Theory*, 19(1), 1-25.
- Joshi, K. P., Joshi, A., & Yesha, Y. (2003). On using a warehouse to analyze web logs. *Distributed and Parallel Databases*, 13(2), 161-180.
- Karaali, D., Gumussoy, C. A., & Calisir, F. (2011). Factors affecting the intention to use a web-based learning system among blue-collar workers in the automotive industry. *Computers in Human behavior*, 27(1), 343-354.
- Khalifa, M., & Shen, K. N. (2008). Explaining the adoption of transactional B2C mobile commerce. *Journal of enterprise information management*.
- Kim, S., Pakzad, S., Culler, D., Demmel, J., Fennes, G., Glaser, S., & Turon, M. (2007). *Health monitoring of civil infrastructures using wireless sensor networks*. Paper presented at the Proceedings of the 6th international conference on Information processing in sensor networks.
- Kim, S. H. (2008). Moderating effects of job relevance and experience on mobile wireless technology acceptance: Adoption of a smartphone by individuals. *Information & management*, 45(6), 387-393.
- Kim, Y., Schmid, T., Charbiwala, Z. M., Friedman, J., & Srivastava, M. B. (2008). *NAWMS: nonintrusive autonomous water monitoring system*. Paper presented at the Proceedings of the 6th ACM conference on Embedded network sensor systems.

- Kleijnen, M., Wetzels, M., & De Ruyter, K. (2004). Consumer acceptance of wireless finance. *Journal of financial services marketing*, 8(3), 206-217.
- Kogan, A., Alles, M. G., Vasarhelyi, M. A., & Wu, J. (2014). Design and evaluation of a continuous data level auditing system. *Auditing: A Journal of Practice & Theory*, 33(4), 221-245.
- Koudas, N., Saha, A., Srivastava, D., & Venkatasubramanian, S. (2009). *Metric functional dependencies*. Paper presented at the 2009 IEEE 25th International Conference on Data Engineering.
- Krahel, J., & Titera, W. (2015). How standards will/should change with Big Data? *Accounting Horizons*, 29(2), 409-422.
- Krahel, J. P., & Titera, W. R. (2015). Consequences of Big Data and formalization on accounting and auditing standards. *Accounting Horizons*, 29(2), 409-422.
- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032-2033.
- Laney, D. (2001). 3-d data management: Controlling data volume, velocity and variety," META Group, Research Note, February 2001.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.
- Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of accounting research*, 50(2), 495-540.
- Lee, H. H., Fiore, A. M., & Kim, J. (2006). The role of the technology acceptance model in explaining effects of image interactivity technology on consumer responses. *International Journal of Retail & Distribution Management*.
- Lee, Y. C. (2006). An empirical investigation into factors influencing the adoption of an e-learning system. *Online information review*.
- Legrís, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & management*, 40(3), 191-204.
- Lev, B. (2000). *Intangibles: Management, measurement, and reporting*: Brookings institution press.
- Lev, B., & Zarowin, P. (1999). The boundaries of financial reporting and how to extend them. *Journal of accounting research*, 37(2), 353-385.
- Liang, H., Xue, Y., & Byrd, T. A. (2003). PDA usage in healthcare professionals: testing an extended technology acceptance model. *International Journal of Mobile Communications*, 1(4), 372-389.

- Liao, C., Chen, J.-L., & Yen, D. C. (2007). Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: An integrated model. *Computers in Human behavior*, 23(6), 2804-2822.
- Liker, J. K., & Sindi, A. A. (1997). User acceptance of expert systems: a test of the theory of reasoned action. *Journal of Engineering and Technology management*, 14(2), 147-173.
- Lim, E.-P., Srivastava, J., Prabhakar, S., & Richardson, J. (1996). Entity identification in database integration. *Information Sciences*, 89(1-2), 1-38.
- Lin, H. F. (2007). The role of online and offline features in sustaining virtual communities: an empirical study. *Internet Research*.
- Lohr, S. (2012). The age of big data. *New York Times*, 11(2012).
- Louwers, T. J., Ramsay, R. J., Sinason, D. H., Strawser, J. R., & Thibodeau, J. C. (2015). *Auditing & assurance services*: McGraw-Hill Education.
- Manyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*: McKinsey Global Institute.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*: Houghton Mifflin Harcourt.
- Mayfield, C., Neville, J., & Prabhakar, S. (2010). *ERACER: a database approach for statistical inference and data cleaning*. Paper presented at the Proceedings of the 2010 ACM SIGMOD International Conference on Management of data.
- Meijer, E. (2011). The world according to LINQ. *Communications of the ACM*, 54(10), 45-51.
- Menezes, A. J., Van Oorschot, P. C., & Vanstone, S. A. (2018). *Handbook of applied cryptography*: CRC press.
- Mohammed, A. F., Humbe, V. T., & Chowhan, S. S. (2016). *A review of big data environment and its related technologies*. Paper presented at the 2016 International Conference on Information Communication and Embedded Systems (ICICES).
- Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information systems research*, 2(3), 192-222.
- Motro, A. (1989). Integrity= validity+ completeness. *ACM Transactions on Database Systems (TODS)*, 14(4), 480-502.

- Mun, Y. Y., & Hwang, Y. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International journal of human-computer studies*, 59(4), 431-449.
- Nanopoulos, A., Manolopoulos, Y., Zakrzewicz, M., & Morzy, T. (2002). *Indexing web access-logs for pattern queries*. Paper presented at the Proceedings of the 4th international workshop on Web information and data management.
- Newcombe, H. B., Kennedy, J. M., Axford, S. J., & James, A. P. (1959). Automatic linkage of vital records. *Science*, 130(3381), 954-959. doi: 10.1126/science.130.3381.954
- Noguchi, Y. (2011a). Following digital breadcrumbs to big data gold. *National Public Radio*, 3, 56-88.
- Noguchi, Y. (2011b). The search for analysts to make sense of big data. *National Public Radio*.
- O'Shea, V., & Shah, R. (2014). Big Data in capital markets: at the start of the journey. *Prepared for Thompson Reuters*.
- Oh, S., Ahn, J., & Kim, B. (2003). Adoption of broadband Internet in Korea: the role of experience in building attitudes. *Journal of Information Technology*, 18(4), 267-280.
- Okcu, S., Koksalmis, G. H., Basak, E., & Calisir, F. (2019). Factors affecting intention to use big data tools: an extended technology acceptance model *Industrial Engineering in the Big Data Era* (pp. 401-416): Springer.
- Ozturk, A. B., Bilgihan, A., Nusair, K., & Okumus, F. (2016). What keeps the mobile hotel booking users loyal? Investigating the roles of self-efficacy, compatibility, perceived ease of use, and perceived convenience. *International Journal of Information Management*, 36(6), 1350-1359.
- Padgavankar, M., & Gupta, S. (2014). Big data storage and challenges. *International Journal of Computer Science and Information Technologies*, 5(2), 2218-2223.
- Perols, J. L., & Murthy, U. S. (2012). Information fusion in continuous assurance. *Journal of Information Systems*, 26(2), 35-52.
- Rajan, C. A., & Baral, R. (2015). Adoption of ERP system: An empirical study of factors influencing the usage of ERP and its impact on end user. *IIMB Management Review*, 27(2), 105-117.
- Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big data analytics: Opportunity or threat for the accounting profession? *Journal of Information Systems*, 31(3), 63-79.
- Rivest, R. L., Adleman, L., & Dertouzos, M. L. (1978). On data banks and privacy homomorphisms. *Foundations of secure computation*, 4(11), 169-180.
- Rogers, E. M., & Shoemaker, F. F. (1971). *Communication of Innovations; A Cross-Cultural Approach*.

- Romero, S., Gal, G., Mock, T. J., & Vasarhelyi, M. A. (2012). A measurement theory perspective on business measurement. *Journal of Emerging Technologies in Accounting*, 9(1), 1-24.
- Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter, 19(4)*, 1-34.
- Saadé, R., & Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: an extension of the technology acceptance model. *Information & management*, 42(2), 317-327.
- Sagioglu, S., & Sinanc, D. (2013). *Big data: A review*. Paper presented at the 2013 international conference on collaboration technologies and systems (CTS).
- Sahoo, P., & Yajnik, N. (2014). Study of factors affecting customer behaviour using big data technology. *Computer Science & Information Technology*, 31, 39.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147.
- Scott, S. V., & Orlikowski, W. J. (2012). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society*, 37(1), 26-40.
- Selavo, L., Wood, A., Cao, Q., Sookoor, T., Liu, H., Srinivasan, A., . . . Young, D. (2007). *Luster: wireless sensor network for environmental research*. Paper presented at the Proceedings of the 5th international conference on Embedded networked sensor systems.
- Shen, D., & Chen, S.-H. (2018). Big data finance and financial markets *Big Data in Computational Social Science and Humanities* (pp. 235-248): Springer.
- Sheth, A., & Rusinkiewicz, M. (1990). *Management of interdependent data: Specifying dependency and consistency requirements*. Paper presented at the [1990] Proceedings. Workshop on the Management of Replicated Data.
- Shin, D.-H. (2016). Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy*, 40(9), 837-854.
- Sokol, M. B. (1994). Adaptation to difficult designs: Facilitating use of new technology. *Journal of Business and Psychology*, 8(3), 277-296.
- Song, D. X., Wagner, D., & Perrig, A. (2000). *Practical techniques for searches on encrypted data*. Paper presented at the Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000.
- Srivastava, A. (2014). Why have measures of earnings quality changed over time? *Journal of Accounting and Economics*, 57(2-3), 196-217.
- Srivastava, R. P., & Shafer, G. R. (1992). Belief-function formulas for audit risk. *Accounting Review*, 249-283.

- Starr, M. (2000). The effects of weather on retail sales.
- Tavakol M. & Dennick R.(2011), Making sense of Cronbach's alpha. *International Journal of medical education*
- Team, O. (2011). Big data now: current perspectives from O'Reilly Radar. *O'Reilly Media*.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3), 1139-1168.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), 1437-1467.
- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS quarterly*, 125-143.
- Tolle, G., Polastre, J., Szewczyk, R., Culler, D., Turner, N., Tu, K., . . . Gay, D. (2005). *A macroscope in the redwoods*. Paper presented at the Proceedings of the 3rd international conference on Embedded networked sensor systems.
- Tornatzky, L. G., & Klein, K. J. (1982). Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. *IEEE Transactions on engineering management*(1), 28-45.
- Trompeter, G., & Wright, A. (2010). The world has changed—Have analytical procedure practices? *Contemporary Accounting Research*, 27(2), 669-700.
- Tsai, C.-W., Lai, C.-F., Chao, H.-C., & Vasilakos, A. V. (2015). Big data analytics: a survey. *Journal of Big data*, 2(1), 1-32.
- Tufekci, Z. (2013). Big data: Pitfalls, methods and concepts for an emergent field. *Methods and Concepts for an Emergent Field (March 7, 2013)*.
- Tung, F.-C., & Chang, S.-C. (2008a). A new hybrid model for exploring the adoption of online nursing courses. *Nurse Educ Today*, 28(3), 293-300. doi: 10.1016/j.nedt.2007.06.003
- Tung, F.-C., & Chang, S.-C. (2008b). Nursing students' behavioral intention to use online courses: A questionnaire survey. *International journal of nursing studies*, 45(9), 1299-1309.
- Tung, F. C., Chang, S. C., & Chou, C. M. (2008). An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. *Int J Med Inform*, 77(5), 324-335. doi: 10.1016/j.ijmedinf.2007.06.006
- Turner, D., Schroeck, M., & Shockley, R. (2013). Analytics: The real-world use of big data in financial services. *IBM Global Business Services*, 27.
- Van Raaij, E. M., & Schepers, J. J. (2008). The acceptance and use of a virtual learning environment in China. *Computers & education*, 50(3), 838-852.

- Vasarhelyi, M., & Greenstein, M. (2003). Underlying principles of the electronization of business: A research agenda. *International Journal of Accounting Information Systems*, 4(1), 1-25.
- Vasarhelyi, M. A., Alles, M., & Williams, K. T. (2010). *Continuous assurance for the now economy*: Institute of Chartered Accountants in Australia Sydney, Australia.
- Vasarhelyi, M. A., Alles, M. G., & Kogan, A. (2018). Principles of Analytic Monitoring for Continuous Assurance¹. *Continuous Auditing: Theory and Application*, 191.
- Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4), 342-365.
- Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision sciences*, 27(3), 451-481.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 46(2), 186-204.
- Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS quarterly*, 115-139.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
- Walter, T. (2009). Teradata past, present, and future. *UCI ISG lecture series on scalable data management*, 1(1), 44-48.
- Wang, F., & Liu, J. (2010). Networked wireless sensor data collection: issues, challenges, and approaches. *IEEE Communications Surveys & Tutorials*, 13(4), 673-687.
- White House. (2012). Fact Sheet: Big Data Across the Federal Government. *Office of Science and Technology Policy*.
- Wu, J.-H., & Wang, S.-C. (2005). What drives mobile commerce?: An empirical evaluation of the revised technology acceptance model. *Information & management*, 42(5), 719-729.
- Yi, M. Y., Fiedler, K. D., & Park, J. S. (2006). Understanding the role of individual innovativeness in the acceptance of IT-based innovations: Comparative analyses of models and measures. *Decision sciences*, 37(3), 393-426.
- Zheng, Z., Wang, P., Liu, J., & Sun, S. (2015). Real-time big data processing framework: challenges and solutions. *Applied Mathematics & Information Sciences*, 9(6), 3169.
- Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*: McGraw-Hill Osborne Media.

ΠΑΡΑΡΤΗΜΑ Α - Ερωτηματολόγιο

ΜΕΡΟΣ Ι

Δημογραφικά Χαρακτηριστικά

Ηλικία	έως και 25 <input type="checkbox"/>	26 έως 36 <input type="checkbox"/>	37 έως 57 <input type="checkbox"/>	57 και πάνω <input type="checkbox"/>
Φύλο	Ανδρας <input type="checkbox"/>		Γυναίκα <input type="checkbox"/>	
Εργασιακή Εμπειρία	0 – 3 χρόνια <input type="checkbox"/>	4 – 6 χρόνια <input type="checkbox"/>	7 – 10 χρόνια <input type="checkbox"/>	>10 χρόνια <input type="checkbox"/>
Εκπαιδευτικό Επίπεδο	Λύκειο <input type="checkbox"/>	Πανεπιστήμιο/ΤΕΙ <input type="checkbox"/>	Μεταπτυχιακό <input type="checkbox"/>	Διδακτορικό <input type="checkbox"/>

ΜΕΡΟΣ ΙΙ

Παρακαλώ να απαντήσετε στις ακόλουθες ερωτήσεις επιλέγοντας από το 1 μέχρι το 5.

A/A	ΕΡΩΤΗΣΕΙΣ	<u>1= ΔΙΑΦΩΝΩ ΑΠΟΛΥΤΑ, 2= ΔΙΑΦΩΝΩ, 3= ΟΥΤΕ ΔΙΑΦΩΝΩ ΟΥΤΕ ΣΥΜΦΩΝΩ, 4= ΣΥΜΦΩΝΩ, 5= ΣΥΜΦΩΝΩ ΑΠΟΛΥΤΑ</u>				
	<u>ΠΡΟΘΕΣΗ ΧΡΗΣΗΣ</u>					
1	Σκοπεύω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων στο μέλλον	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
2	Έχω πρόθεση να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων στο μέλλον	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
3	Περιμένω να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων και στο μέλλον	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
	<u>ΑΝΤΙΛΑΜΒΑΝΟΜΕΝΗ ΧΡΗΣΙΜΟΤΗΤΑ</u>					
4	Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα μου επέτρεπε να διεκπεραιώνω διάφορες εργασίες γρηγορότερα	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
5	Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου επέτρεπε να κάνω τη δουλειά μου με μεγαλύτερη ευκολία	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
6	Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα βελτιώνε την ποιότητά της	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
7	Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου ήταν χρήσιμη	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

8	Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα ενίσχυε την αποδοτικότητά μου	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<u>ΑΝΤΙΑΑΜΒΑΝΟΜΕΝΗ ΕΥΚΟΛΙΑ ΧΡΗΣΗΣ</u>						
9	Θα μου ήταν εύκολο να γίνω επιδέξιος/α στη χρήση εργαλείων Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
10	Θα μου ήταν εύκολη η χρήση εργαλείων Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
11	Θα μου ήταν εύκολη η μάθηση χρήσης εργαλείων Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
12	Η αλληλεπίδρασή μου με τα εργαλεία Μεγάλων Δεδομένων θα ήταν ξεκάθαρη και κατανοητή	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
13	Θεωρώ πως θα υπάρχει ευελιξία στην αλληλεπίδραση με τα εργαλεία Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
A/A	ΕΡΩΤΗΣΕΙΣ	<u>1= ΔΙΑΦΩΝΩ ΑΠΟΛΥΤΑ, 2= ΔΙΑΦΩΝΩ, 3= ΟΥΤΕ ΔΙΑΦΩΝΩ ΟΥΤΕ ΣΥΜΦΩΝΩ, 4= ΣΥΜΦΩΝΩ, 5= ΣΥΜΦΩΝΩ ΑΠΟΛΥΤΑ</u>				
<u>ΣΥΝΑΦΕΙΑ ΕΡΓΑΣΙΑΣ</u>						
14	Στη δουλειά μου η χρήση εργαλείων Μεγάλων Δεδομένων είναι σημαντική	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
15	Η δουλειά μου έχει σχέση με τη χρήση εργαλείων Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
16	Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απαραίτητη για τη διεκπεραίωση ορισμένων εργασιών που σχετίζονται με τη δουλειά μου	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<u>ΣΥΜΒΑΤΟΤΗΤΑ</u>						
17	Η χρήση εργαλείων Μεγάλων Δεδομένων είναι συμβατή με όλες τις πτυχές της δουλειάς μου	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
18	Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απόλυτα συμβατή με την τρέχουσα κατάστασή μου	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
19	Πιστεύω πως η χρήση εργαλείων Μεγάλων Δεδομένων ταιριάζει αρκετά με τον τρόπο που μου αρέσει να εργάζομαι	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
<u>ΔΙΑΣΤΑΣΕΙΣ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ</u>						
20	Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία τεράστιων όγκων δεδομένων, κάτι που δεν είναι δυνατό με τη χρήση παραδοσιακών συστημάτων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
21	Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία διαφόρων τύπων δεδομένων, όπως αδόμητα, δομημένα και ημι-δομημένα	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

22	Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην ταχύτερη επεξεργασία σε σχέση με τις παραδοσιακές προσεγγίσεις	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
	ΑΥΤΟ-ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ					
23	Εάν δεν υπήρχε κάποιος να με καθοδηγεί καθώς προχωρώ σε μια εργασία, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
24	Εάν μπορούσα να καλέσω κάποιον για βοήθεια σε περίπτωση που κολλούσα, θα μπορούσα να ολοκληρώσω μια δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
25	Εάν είχα πολύ χρόνο για να ολοκληρώσω μία εργασία για την οποία μου παρείχαν λογισμικό, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
A/A	ΕΡΩΤΗΣΕΙΣ	<u>1= ΔΙΑΦΩΝΩ ΑΠΟΛΥΤΑ, 2= ΔΙΑΦΩΝΩ, 3= ΟΥΤΕ ΔΙΑΦΩΝΩ ΟΥΤΕ ΣΥΜΦΩΝΩ, 4= ΣΥΜΦΩΝΩ, 5= ΣΥΜΦΩΝΩ ΑΠΟΛΥΤΑ</u>				
26	Εάν είχα για βοήθεια μόνο την ενσωματωμένη εγκατάσταση βοήθειας του προγράμματος, θα μπορούσα να ολοκληρώσω τη δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
	ΠΟΛΥΠΛΟΚΟΤΗΤΑ					
27	Η χρήση εργαλείων Μεγάλων Δεδομένων στερεί αρκετό χρόνο από τις συνηθισμένες υποχρεώσεις μου	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
28	Η χρήση εργαλείων Μεγάλων Δεδομένων είναι τόσο πολύπλοκη που μου είναι δύσκολο να καταλάβω τι γίνεται	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
29	Η χρήση εργαλείων Μεγάλων Δεδομένων απαιτεί αρκετό χρόνο κάνοντας μηχανικές λειτουργίες (π.χ. εισαγωγή δεδομένων)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
30	Παίρνει τόσο πολύ χρόνο η μάθηση χρήσης του συστήματος, που δεν αξίζει η προσπάθεια	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
	ΑΓΧΟΣ					
31	Νιώθω ανησυχία σχετικά με τη χρήση εργαλείων Μεγάλων Δεδομένων	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
32	Με τρομάζει η σκέψη πως θα μπορούσα να χάσω πολλές πληροφορίες, ενώ χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων, πατώντας ένα λάθος πλήκτρο	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
33	Διστάζω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων γιατί φοβάμαι μην κάνω λάθη που δεν μπορώ να διορθώσω	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
34	Θεωρώ τα εργαλεία Μεγάλων Δεδομένων εκφοβιστικά	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

ΠΑΡΑΡΤΗΜΑ Β - Πίνακες του SPSS

Πίνακας Π.4.1, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Πρόθεσης Χρήσης (ΠΧ)

	E1: Σκοπεύω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων στο μέλλον	E2: Έχω πρόθεση να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων στο μέλλον	E3: Περιμένω να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων και στο μέλλον
E1: Σκοπεύω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων στο μέλλον	1,000	0,899	0,854
E2: Έχω πρόθεση να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων στο μέλλον	0,899	1,000	0,917
E3: Περιμένω να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων και στο μέλλον	0,854	0,917	1,000

Πίνακας Π.4.2, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Αντιλαμβανόμενη Χρησιμότητα (ΑΧ)

	E4: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα μου επέτρεπε να διεκπεραιώω διάφορες εργασίες	E5: Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου επέτρεπε να κάνω τη δουλειά μου με μεγαλύτερη ευκολία	E6: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα βελτιώνε την ποιότητά της	E7: Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου ήταν χρήσιμη	E8: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα ενίσχυε την αποδοτικότητά μου
E4: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα μου επέτρεπε να διεκπεραιώω διάφορες εργασίες γρηγορότερα	1,000	0,875	0,848	0,826	0,701
E5: Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου επέτρεπε να κάνω τη δουλειά μου με μεγαλύτερη ευκολία	0,875	1,000	0,852	0,792	0,749

E6: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα βελτιώνει την ποιότητά της	0,848	0,852	1,000	0,826	0,748
E7: Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου ήταν χρήσιμη	0,826	0,792	0,826	1,000	0,807
E8: Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα ενίσχυε την αποδοτικότητά μου	0,701	0,749	0,748	0,807	1,000

Πίνακας Π.4.3. Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Αντιλαμβανόμενη Ευκολίας Χρήσης (AEX)

	E9: Θα μου ήταν εύκολο να γίνω επιδέξιος/α στη χρήση εργαλείων Μεγάλων Δεδομένων	E10: Θα μου ήταν εύκολη η χρήση εργαλείων Μεγάλων Δεδομένων	E11: Θα μου ήταν εύκολη η μάθηση χρήσης εργαλείων Μεγάλων Δεδομένων	E12: Η αλληλεπίδρασή μου με τα εργαλεία Μεγάλων Δεδομένων θα ήταν ξεκάθαρη και κατανοητή	E13: Θεωρώ πως θα υπάρξει ευελιξία στην αλληλεπίδραση με τα εργαλεία Μεγάλων Δεδομένων
E9: Θα μου ήταν εύκολο να γίνω επιδέξιος/α στη χρήση εργαλείων Μεγάλων Δεδομένων	1,000	0,879	0,759	0,713	0,569
E10: Θα μου ήταν εύκολη η χρήση εργαλείων Μεγάλων Δεδομένων	0,879	1,000	0,810	0,670	0,509
E11: Θα μου ήταν εύκολη η μάθηση χρήσης εργαλείων Μεγάλων Δεδομένων	0,759	0,810	1,000	0,754	0,628
E12: Η αλληλεπίδρασή μου με τα εργαλεία Μεγάλων Δεδομένων θα ήταν ξεκάθαρη και κατανοητή	0,713	0,670	0,754	1,000	0,827
E13: Θεωρώ πως θα υπάρξει ευελιξία στην αλληλεπίδραση με τα εργαλεία Μεγάλων Δεδομένων	0,569	0,509	0,628	0,827	1,000

Πίνακας Π.4.4, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Συνάφεια Εργασίας (ΣΕ)

	E14: Στη δουλειά μου η χρήση εργαλείων Μεγάλων Δεδομένων είναι σημαντική	E15: Η δουλειά μου έχει σχέση με τη χρήση εργαλείων Μεγάλων Δεδομένων	E16: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απαραίτητη για τη διεκπεραίωση ορισμένων εργασιών που σχετίζονται με τη δουλειά μου
E14: Στη δουλειά μου η χρήση εργαλείων Μεγάλων Δεδομένων είναι σημαντική	1,000	0,791	0,736
E15: Η δουλειά μου έχει σχέση με τη χρήση εργαλείων Μεγάλων Δεδομένων	0,791	1,000	0,827
E16: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απαραίτητη για τη διεκπεραίωση ορισμένων εργασιών που σχετίζονται με τη δουλειά μου	0,736	0,827	1,000

Πίνακας Π.4.5, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Συμβατότητα (ΣΥ)

	E17: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι συμβατή με όλες τις πτυχές της δουλειάς μου	E18: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απόλυτα συμβατή με την τρέχουσα κατάστασή μου	E19: Πιστεύω πως η χρήση εργαλείων Μεγάλων Δεδομένων ταυριάζει αρκετά με τον τρόπο που μου αρέσει να εργάζομαι
E17: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι συμβατή με όλες τις πτυχές της δουλειάς μου	1,000	0,843	0,588
E18: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απόλυτα συμβατή με την τρέχουσα κατάστασή μου	0,843	1,000	0,593
E19: Πιστεύω πως η χρήση εργαλείων Μεγάλων Δεδομένων ταυριάζει αρκετά με τον τρόπο που μου αρέσει να εργάζομαι	0,588	0,593	1,000

Πίνακας Π.4.6, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Διαστάσεις Μεγάλων Δεδομένων (ΔΜΔ)

	E20: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία τεράστιων όγκων δεδομένων, κάτι που δεν είναι δυνατό με τη χρήση παραδοσιακών συστημάτων	E21: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία διαφόρων τύπων δεδομένων, όπως αδόμητα, δομημένα και ημι-δομημένα	E22: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην ταχύτερη επεξεργασία σε σχέση με τις παραδοσιακές προσεγγίσεις
E20: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία τεράστιων όγκων δεδομένων, κάτι που δεν είναι δυνατό με τη χρήση παραδοσιακών συστημάτων	1,000	0,879	0,826
E21: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία διαφόρων τύπων δεδομένων, όπως αδόμητα, δομημένα και ημι-δομημένα	0,879	1,000	0,859
E22: Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην ταχύτερη επεξεργασία σε σχέση με τις παραδοσιακές προσεγγίσεις	0,826	0,859	1,000

Πίνακας Π.4.7, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Αυτό-Αποτελεσματικότητα (ΑΑ)

	E23: Εάν δεν υπήρχε κάποιος να με καθοδηγεί καθώς προχωρώ σε μια εργασία, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	E24: Εάν μπορούσα να καλέσω κάποιον για βοήθεια σε περίπτωση που κολλούσα, θα μπορούσα να ολοκληρώσω μια δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	E25: Εάν είχα πολύ χρόνο για να ολοκληρώσω μία εργασία για την οποία μου παρείχαν λογισμικό, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων	E26: Εάν είχα για βοήθεια μόνο την ενσωματωμένη εγκατάσταση βοήθειας του προγράμματος, θα μπορούσα να ολοκληρώσω τη δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων
	1,000	0,661	0,685	0,557
	0,661	1,000	0,676	0,737
	0,685	0,676	1,000	0,611
	0,557	0,737	0,611	1,000

Πίνακας Π.4.8. Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Πολυπλοκότητα (ΠΛ)

	E27: Η χρήση εργαλείων Μεγάλων Δεδομένων στερεί αρκετό χρόνο από τις συνηθισμένες υποχρεώσεις μου	E28: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι τόσο πολύπλοκη που μου είναι δύσκολο να καταλάβω τι γίνεται	E29: Η χρήση εργαλείων Μεγάλων Δεδομένων απαιτεί αρκετό χρόνο κάνοντας μηχανικές λειτουργίες (π.χ. εισαγωγή δεδομένων)	E30: Παίρνει τόσο πολύ χρόνο η μάθηση χρήσης του συστήματος, που δεν αξίζει η προσπάθεια
E27: Η χρήση εργαλείων Μεγάλων Δεδομένων στερεί αρκετό χρόνο από τις συνηθισμένες υποχρεώσεις μου	1,000	0,673	0,627	0,604
E28: Η χρήση εργαλείων Μεγάλων Δεδομένων είναι τόσο πολύπλοκη που μου είναι δύσκολο να καταλάβω τι γίνεται	0,673	1,000	0,753	0,684
E29: Η χρήση εργαλείων Μεγάλων Δεδομένων απαιτεί αρκετό χρόνο κάνοντας μηχανικές λειτουργίες (π.χ. εισαγωγή δεδομένων)	0,627	0,753	1,000	0,574
E30: Παίρνει τόσο πολύ χρόνο η μάθηση χρήσης του συστήματος, που δεν αξίζει η προσπάθεια	0,604	0,684	0,574	1,000

Πίνακας Π.4.9, Μήτρα συσχέτισης των ερωτήσεων της κατηγορίας Πολυπλοκότητα (ΠΛ)

	E31: Νιώθω ανησυχία σχετικά με τη χρήση εργαλείων Μεγάλων Δεδομένων	E32: Με τρομάζει η σκέψη πως θα μπορούσα να χάσω πολλές πληροφορίες, ενώ χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων, πατώντας ένα λάθος πλήκτρο	E33: Διστάζω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων γιατί φοβάμαι μην κάνω λάθη που δεν μπορώ να διορθώσω	E34: Θεωρώ τα εργαλεία Μεγάλων Δεδομένων εκφοβιστικά
E31: Νιώθω ανησυχία σχετικά με τη χρήση εργαλείων Μεγάλων Δεδομένων	1,000	0,591	0,713	0,740
E32: Με τρομάζει η σκέψη πως θα μπορούσα να χάσω πολλές πληροφορίες, ενώ χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων, πατώντας ένα λάθος πλήκτρο	0,591	1,000	0,710	0,561
E33: Διστάζω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων γιατί φοβάμαι μην κάνω λάθη που δεν μπορώ να διορθώσω	0,713	0,710	1,000	0,712
E34: Θεωρώ τα εργαλεία Μεγάλων Δεδομένων εκφοβιστικά	0,740	0,561	0,712	1,000

Πίνακας Π.4.10, Περίληψη στατιστικών στοιχείων για την ανάλυση διακύμανσης απλής κατεύθυνσης (ANOVA^a) του 1^{ου} μοντέλου παλινδρόμησης

α/α		Άθροισμα τετραγώνων	df	Μέσο τετράγωνο	F	P
1	Παλινδρόμηση	44,889	2	22,445	68,124	0,000 ^b
	Υπόλοιπο	30,311	92	0,329		
	Σύνολο	75,200	94			

a. Εξαρτημένη μεταβλητή: ΠΧ

b. Προγνωστικά: (Constant), AEX, AX

Πίνακας Π.4.11, Περίληψη στατιστικών στοιχείων για την ανάλυση διακύμανσης απλής κατεύθυνσης (ANOVA^a) του 2^{ου} μοντέλου παλινδρόμησης

α/α		Άθροισμα τετραγώνων	df	Μέσο τετράγωνο	F	P
2	Παλινδρόμηση	44,718	4	11,180	40,529	0,000 ^b
	Υπόλοιπο	24,826	90	0,276		
	Σύνολο	69,544	94			

a. Εξαρτημένη μεταβλητή: AX

b. Προγνωστικά: (Constant), ΔΜΔ, AEX, ΣΕ, ΣΥ

Πίνακας Π.4.12, Περίληψη στατιστικών στοιχείων για την ανάλυση διακύμανσης απλής κατεύθυνσης (ANOVA^a) του 3^{ου} μοντέλου παλινδρόμησης

α/α		Άθροισμα τετραγώνων	df	Μέσο τετράγωνο	F	P
3	Παλινδρόμηση	23,014	2	11,507	37,150	0,000 ^b
	Υπόλοιπο	28,497	92	0,310		
	Σύνολο	51,512	94			

a. Εξαρτημένη μεταβλητή: AEX

b. Προγνωστικά: (Constant), AN, AA

Πίνακας Π.4.13, Περίληψη στατιστικών στοιχείων για την ανάλυση διακύμανσης απλής κατεύθυνσης (ANOVA^a) του 4^{ου} μοντέλου παλινδρόμησης

α/α		Άθροισμα τετραγώνων	df	Μέσο τετράγωνο	F	P
4	Παλινδρόμηση	23,101	3	7,700	24,664	0,000 ^b
	Υπόλοιπο	28,411	91	0,312		
	Σύνολο	51,512	94			

a. Εξαρτημένη μεταβλητή: AEX

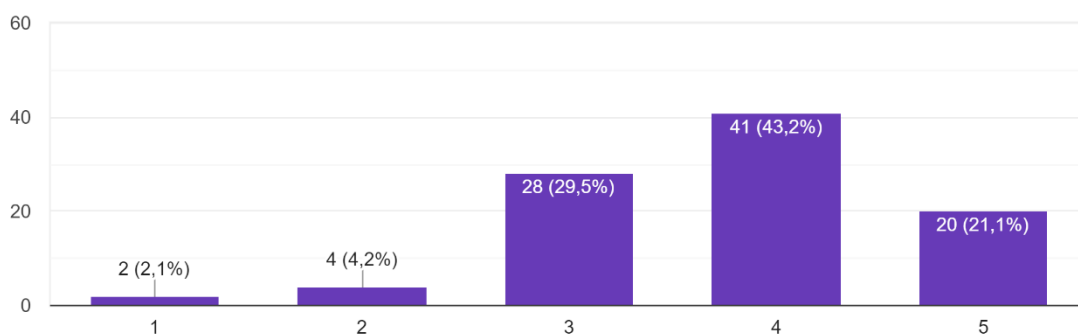
b. Προγνωστικά: (Constant), AN, AA, ΠΛ

ΠΑΡΑΡΤΗΜΑ Γ - Γενικά στατιστικά ανά ερώτηση

Ερωτήσεις 1-34 κατά αύξουσα σειρά και τα στατιστικά στοιχεία αυτών όπως εξήχθησαν απευθείας από την πλατφόρμα των Google Documents.

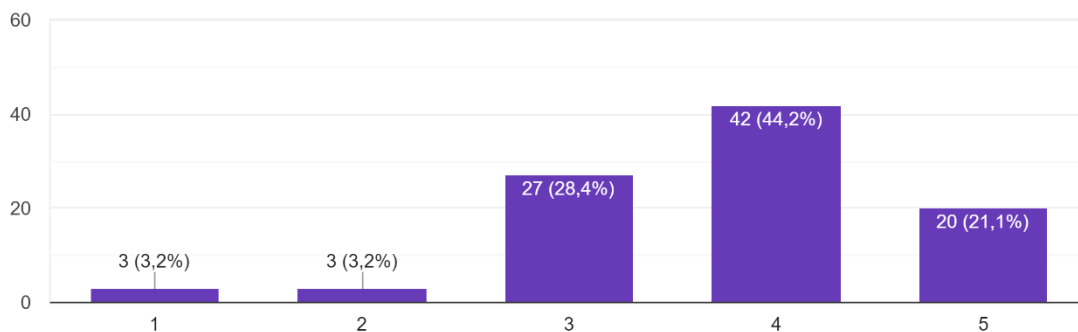
Σκοπεύω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων στο μέλλον

95 απαντήσεις

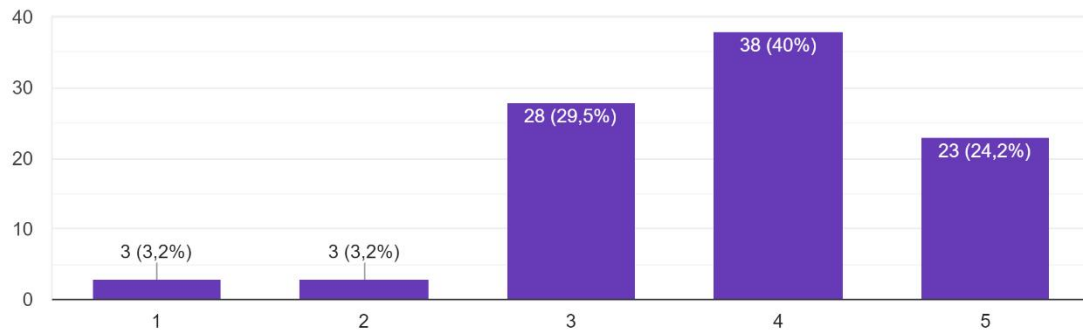


Έχω πρόθεση να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων στο μέλλον

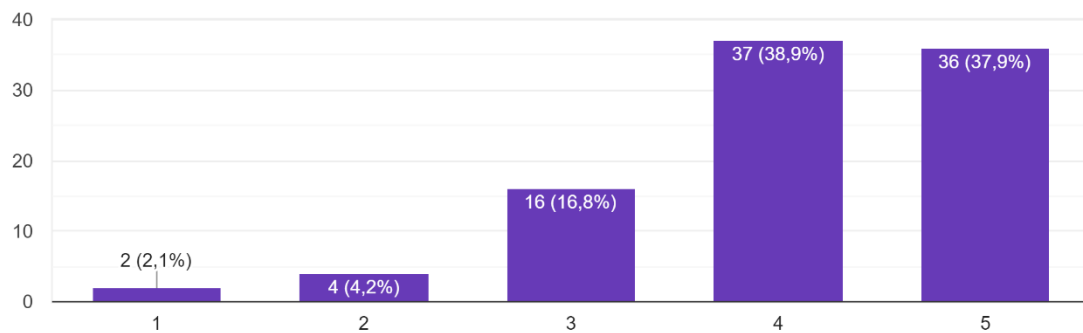
95 απαντήσεις



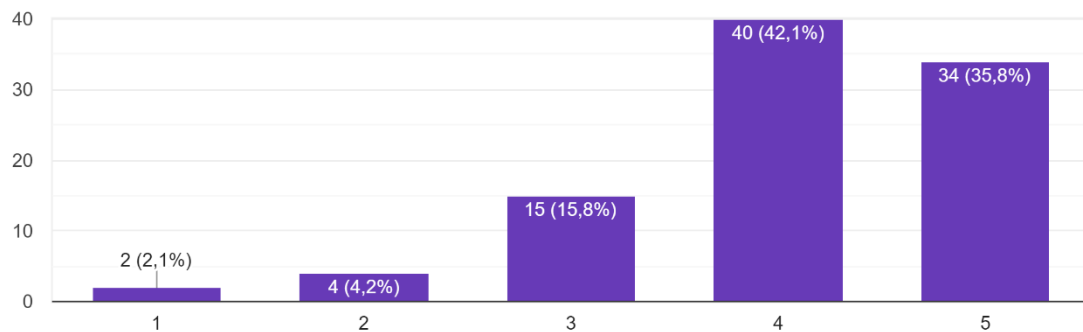
Περιμένω να συνεχίσω να χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων και στο μέλλον
95 απαντήσεις



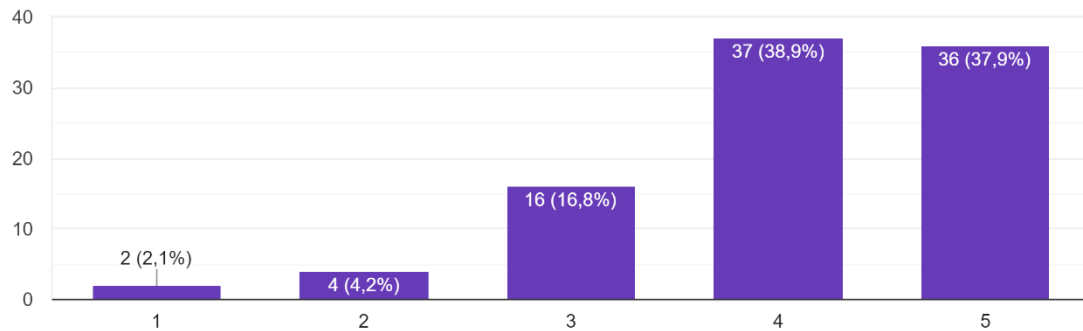
Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα μου επέτρεπε να διεκπεραιώνω
διάφορες εργασίες γρηγορότερα
95 απαντήσεις



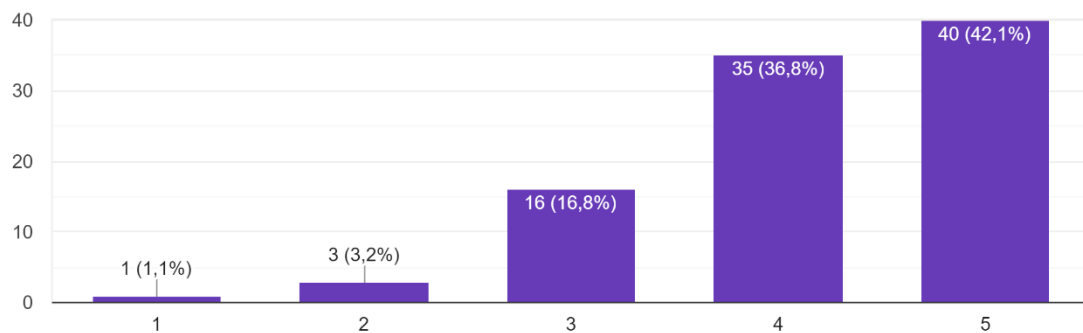
Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου επέτρεπε να κάνω τη δουλειά μου με
μεγαλύτερη ευκολία
95 απαντήσεις



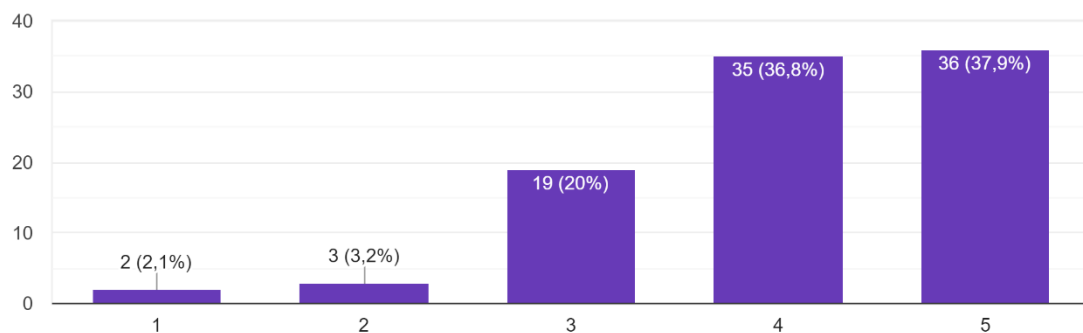
Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα βελτιώνει την ποιότητά της
95 απαντήσεις



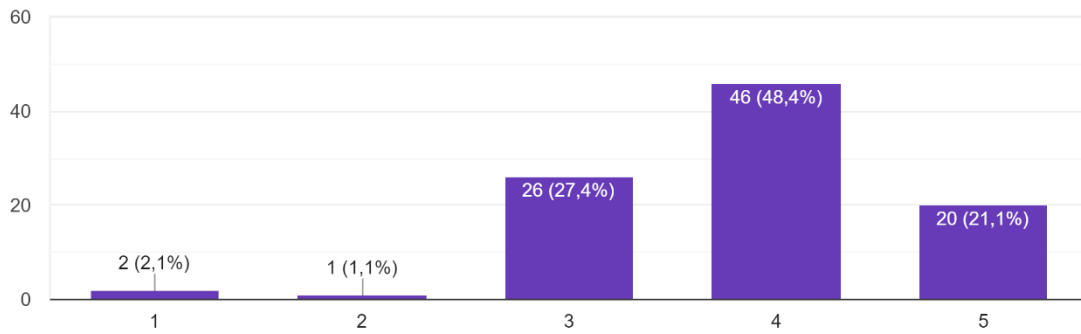
Η χρήση εργαλείων Μεγάλων Δεδομένων θα μου ήταν χρήσιμη
95 απαντήσεις



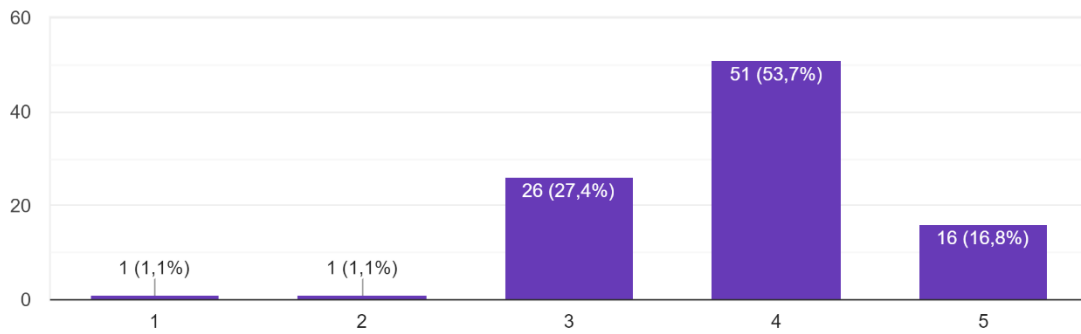
Η χρήση εργαλείων Μεγάλων Δεδομένων στη δουλειά μου, θα ενίσχυε την αποδοτικότητά μου
95 απαντήσεις



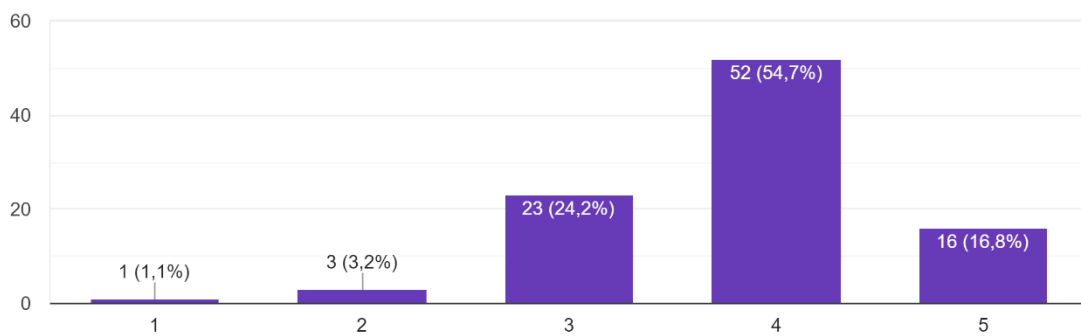
Θα μου ήταν εύκολο να γίνω επιδέξιος/α στη χρήση εργαλείων Μεγάλων Δεδομένων
95 απαντήσεις



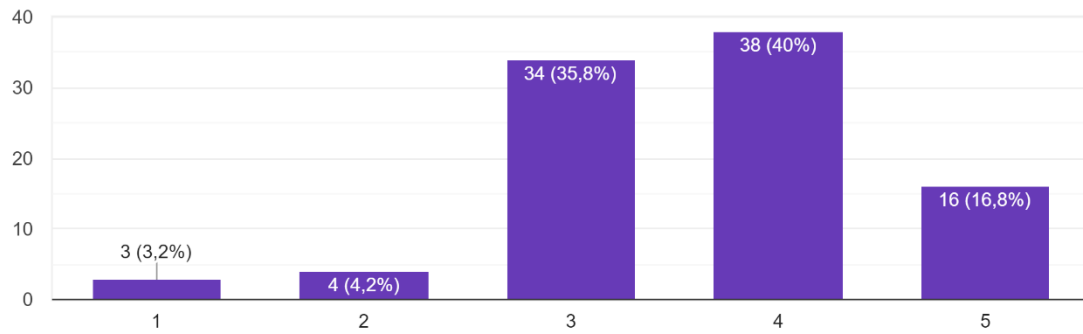
Θα μου ήταν εύκολη η χρήση εργαλείων Μεγάλων Δεδομένων
95 απαντήσεις



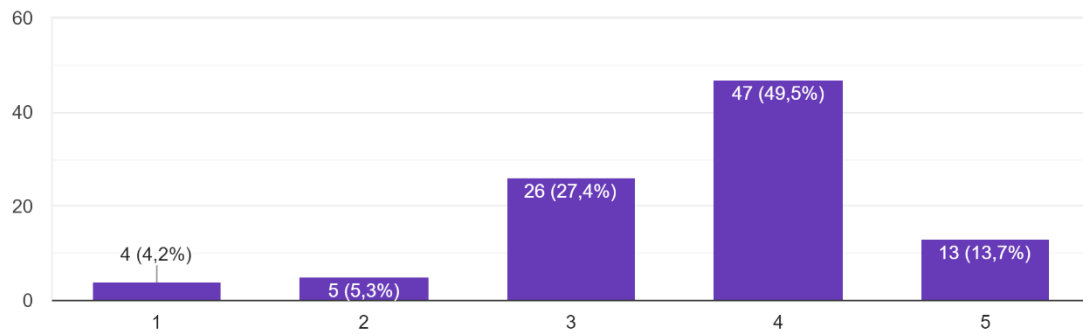
Θα μου ήταν εύκολη η μάθηση χρήσης εργαλείων Μεγάλων Δεδομένων
95 απαντήσεις



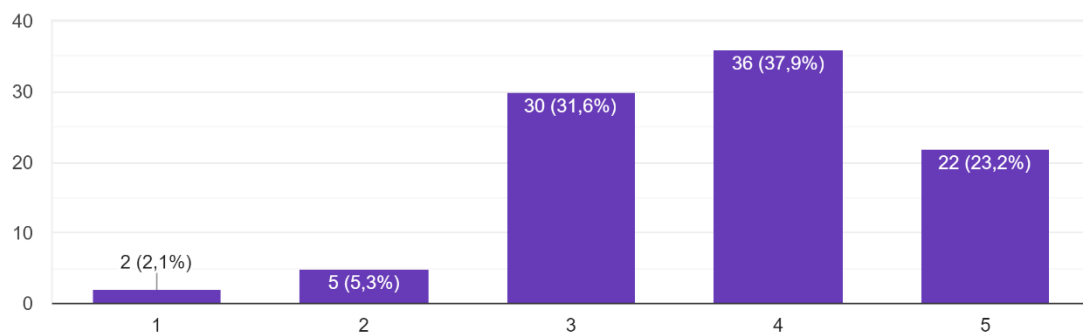
Η αλληλεπίδρασή μου με τα εργαλεία Μεγάλων Δεδομένων θα ήταν ξεκάθαρη και κατανοητή
95 απαντήσεις



Θεωρώ πως θα υπάρχει ευελιξία στην αλληλεπίδραση με τα εργαλεία Μεγάλων Δεδομένων
95 απαντήσεις

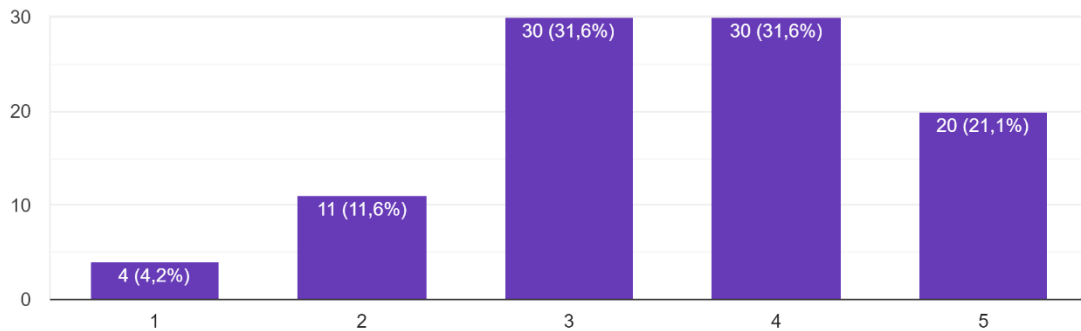


Στη δουλειά μου η χρήση εργαλείων Μεγάλων Δεδομένων είναι σημαντική
95 απαντήσεις



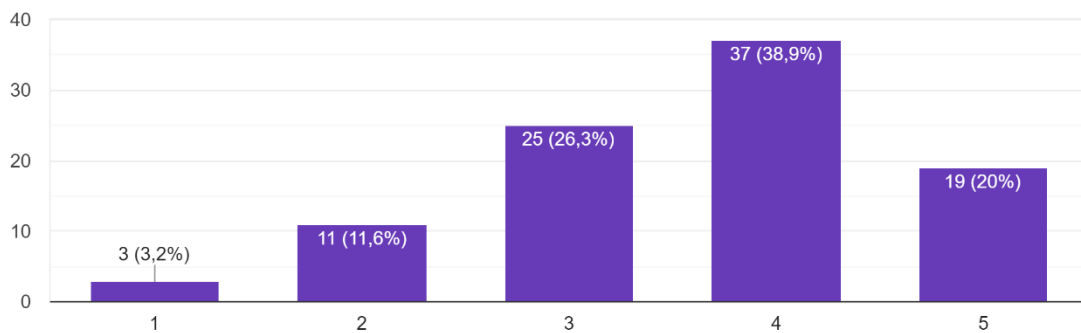
Η δουλειά μου έχει σχέση με τη χρήση εργαλείων Μεγάλων Δεδομένων

95 απαντήσεις



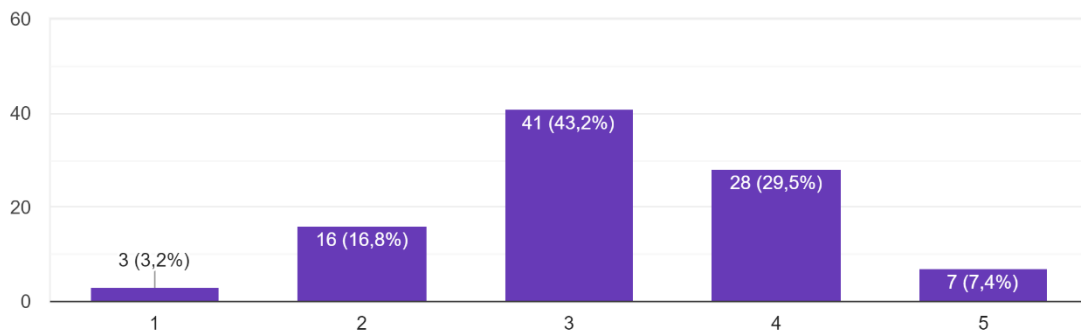
Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απαραίτητη για τη διεκπεραίωση ορισμένων εργασιών που σχετίζονται με τη δουλειά μου

95 απαντήσεις

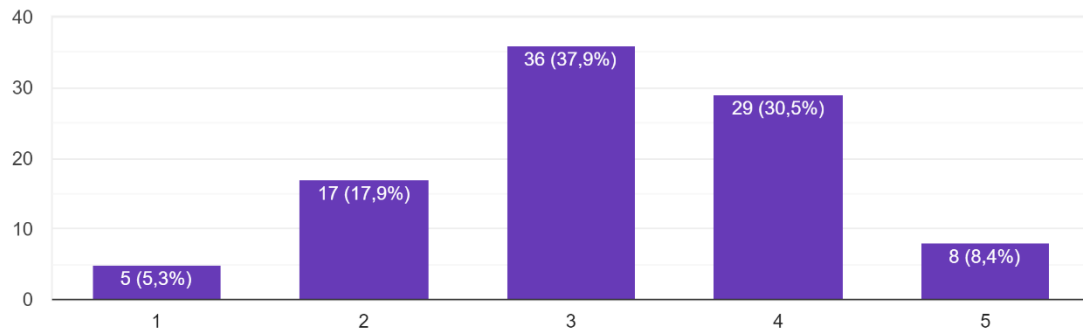


Η χρήση εργαλείων Μεγάλων Δεδομένων είναι συμβατή με όλες τις πτυχές της δουλειάς μου

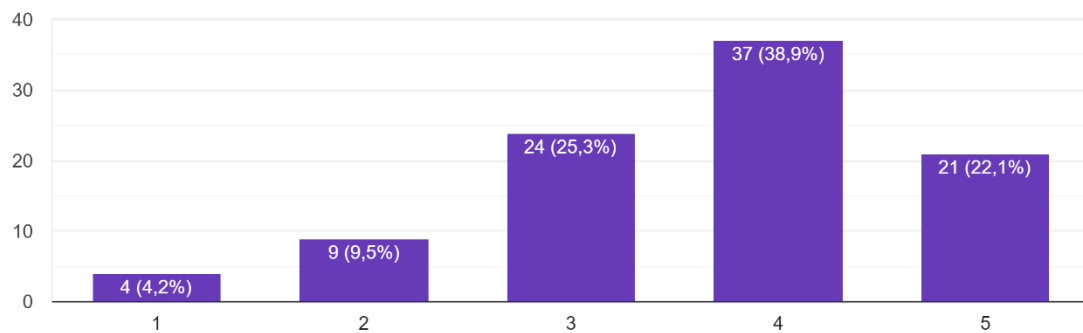
95 απαντήσεις



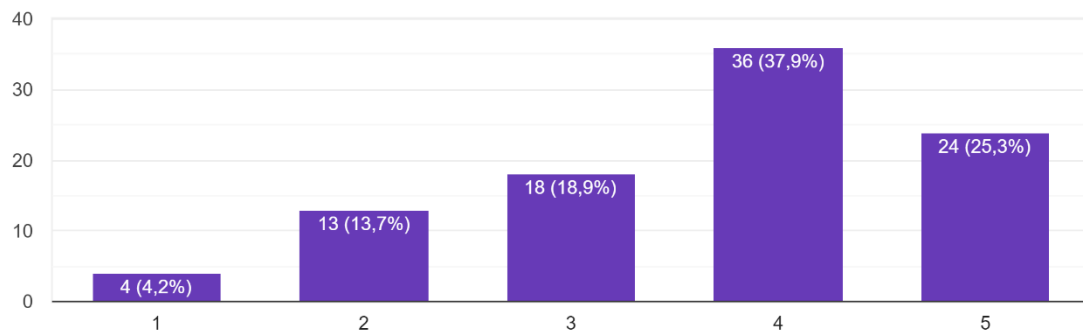
Η χρήση εργαλείων Μεγάλων Δεδομένων είναι απόλυτα συμβατή με την τρέχουσα κατάστασή μου
95 απαντήσεις



Πιστεύω πως η χρήση εργαλείων Μεγάλων Δεδομένων ταιριάζει αρκετά με τον τρόπο που μου
αρέσει να εργάζομαι
95 απαντήσεις

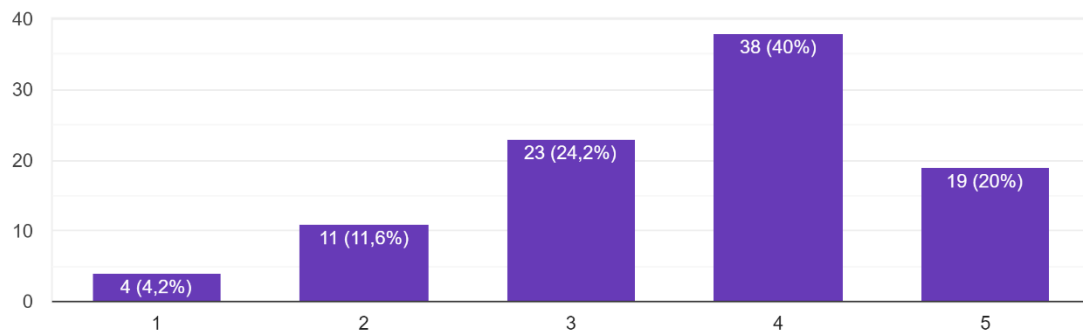


Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία τεράστιων όγκων
δεδομένων, κάτι που δεν είναι δυνατό με τη χρήση παραδοσιακών συστημάτων
95 απαντήσεις



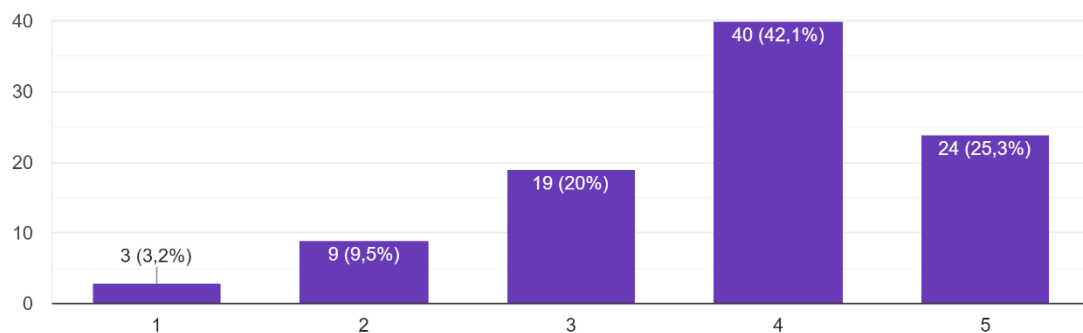
Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην επεξεργασία διαφόρων τύπων δεδομένων, όπως αδόμητα, δομημένα και ημι-δομημένα

95 απαντήσεις



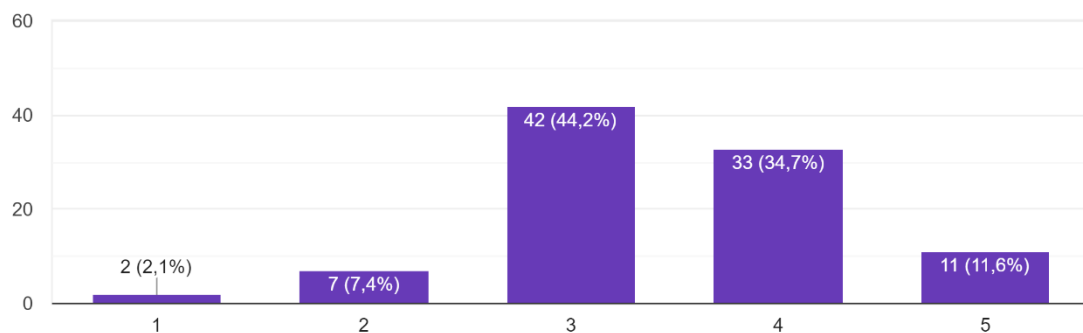
Χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων γιατί βοηθούν στην ταχύτερη επεξεργασία σε σχέση με τις παραδοσιακές προσεγγίσεις

95 απαντήσεις

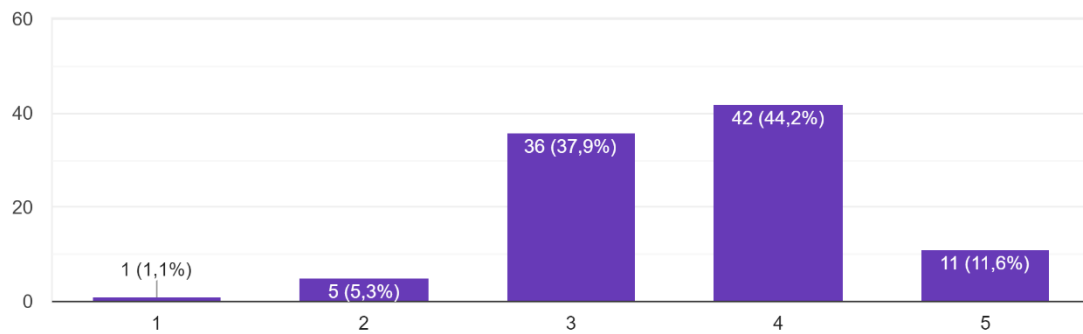


Εάν δεν υπήρχε κάποιος να με καθοδηγεί καθώς προχωρώ σε μια εργασία, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων

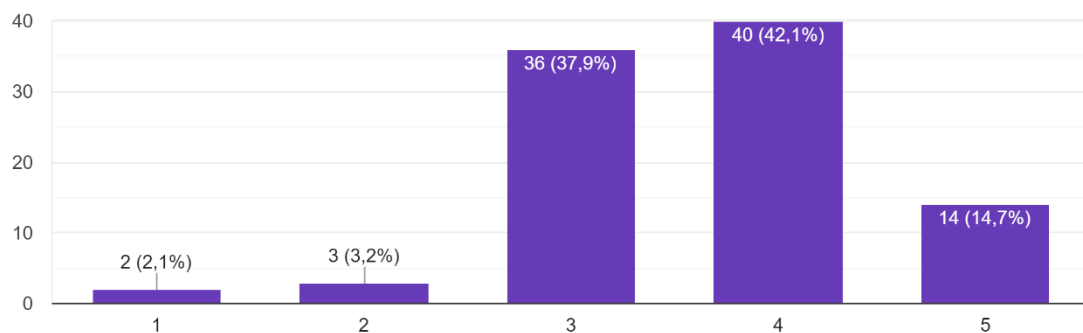
95 απαντήσεις



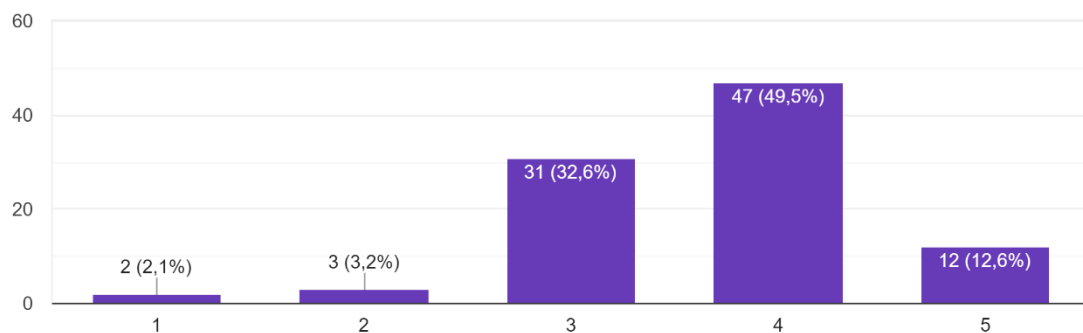
Εάν μπορούσα να καλέσω κάποιον για βοήθεια σε περίπτωση που κολλούσα, θα μπορούσα να ολοκληρώσω μια δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων
95 απαντήσεις



Εάν είχα πολύ χρόνο για να ολοκληρώσω μία εργασία για την οποία μου παρείχαν λογισμικό, θα μπορούσα να την ολοκληρώσω χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων
95 απαντήσεις

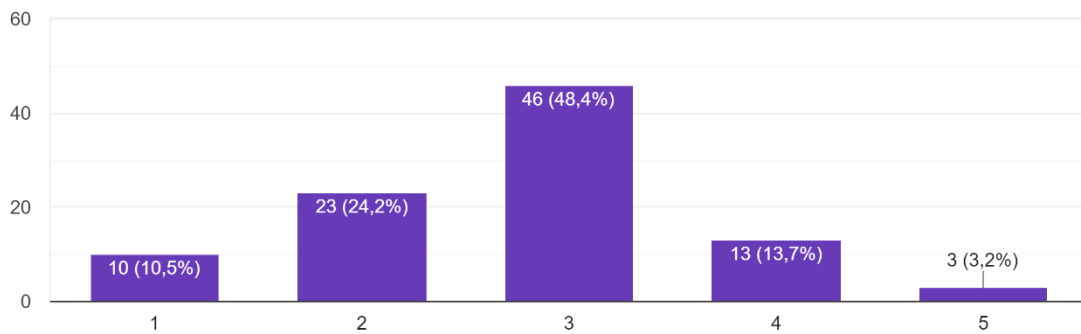


Εάν είχα για βοήθεια μόνο την ενσωματωμένη εγκατάσταση βοήθειας του προγράμματος, θα μπορούσα να ολοκληρώσω τη δουλειά χρησιμοποιώντας εργαλεία Μεγάλων Δεδομένων
95 απαντήσεις



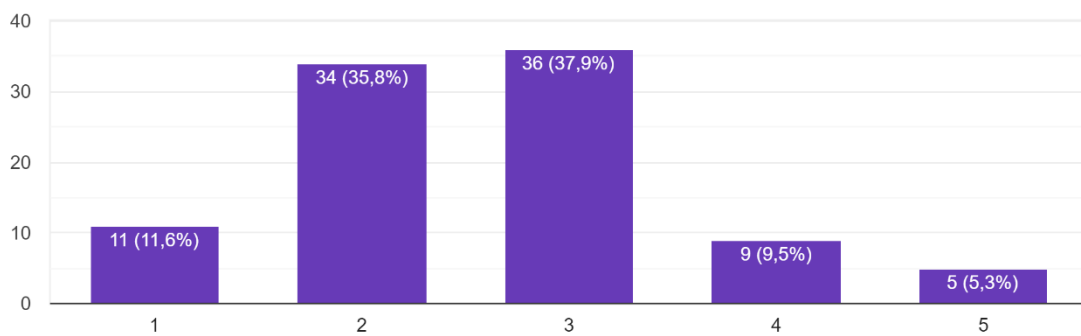
Η χρήση εργαλείων Μεγάλων Δεδομένων στερεί αρκετό χρόνο από τις συνηθισμένες υποχρεώσεις μου

95 απαντήσεις



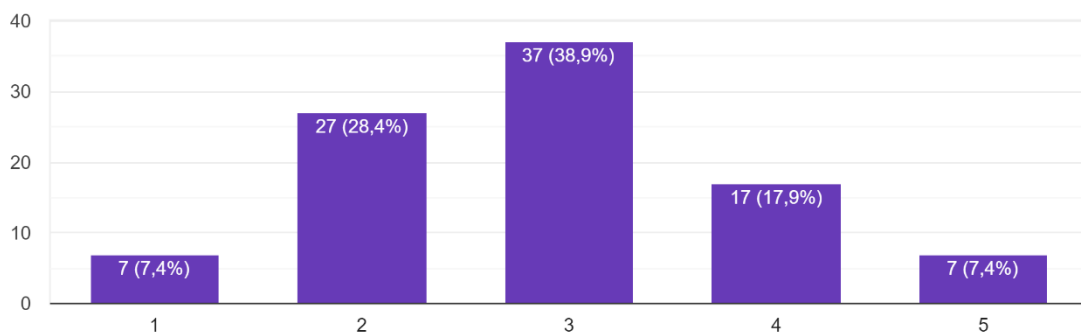
Η χρήση εργαλείων Μεγάλων Δεδομένων είναι τόσο πολύπλοκη που μου είναι δύσκολο να καταλάβω τι γίνεται

95 απαντήσεις

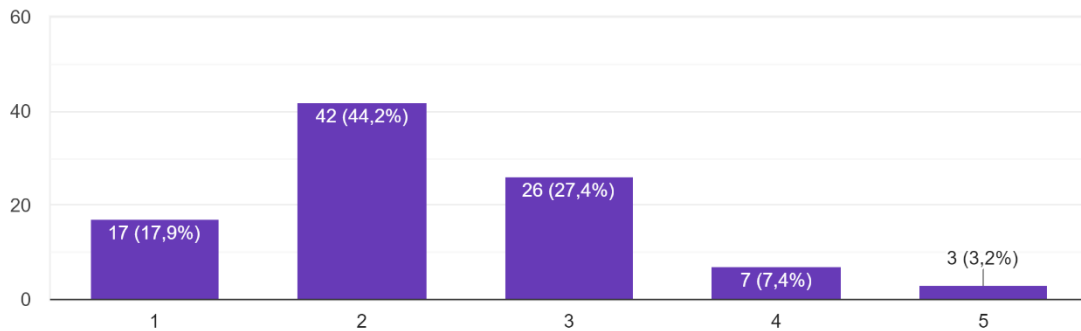


Η χρήση εργαλείων Μεγάλων Δεδομένων απαιτεί αρκετό χρόνο κάνοντας μηχανικές λειτουργίες (π.χ. εισαγωγή δεδομένων)

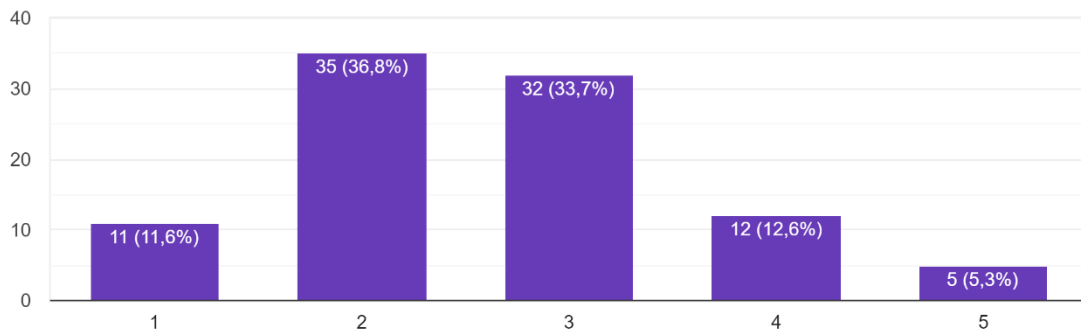
95 απαντήσεις



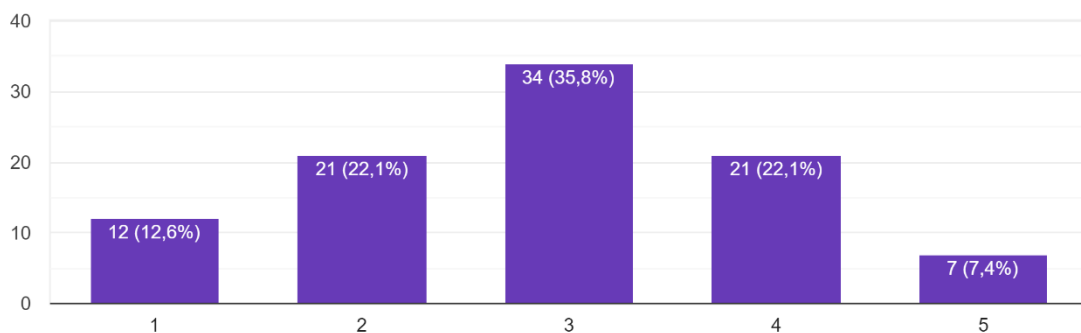
Παίρνει τόσο πολύ χρόνο η μάθηση χρήσης του συστήματος, που δεν αξίζει η προσπάθεια
95 απαντήσεις



Νιώθω ανησυχία σχετικά με τη χρήση εργαλείων Μεγάλων Δεδομένων
95 απαντήσεις

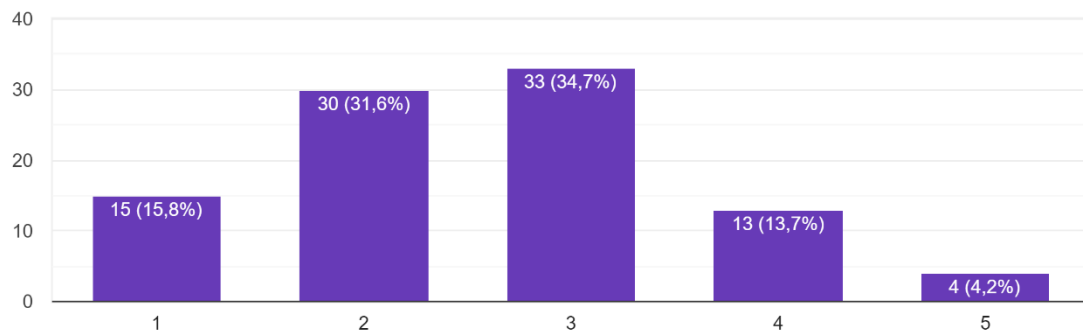


Με τρομάζει η σκέψη πως θα μπορούσα να χάσω πολλές πληροφορίες, ενώ χρησιμοποιώ εργαλεία Μεγάλων Δεδομένων, πατώντας ένα λάθος πλήκτρο
95 απαντήσεις



Διστάζω να χρησιμοποιήσω εργαλεία Μεγάλων Δεδομένων γιατί φοβάμαι μην κάνω λάθη που δεν μπορώ να διορθώσω

95 απαντήσεις



Θεωρώ τα εργαλεία Μεγάλων Δεδομένων εκφοβιστικά

95 απαντήσεις

