



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΤΜΗΜΑ ΒΙΟΜΗΧΑΝΙΚΗΣ ΣΧΕΔΙΑΣΗΣ ΚΑΙ
ΠΑΡΑΓΩΓΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Η ΕΠΙΣΤΗΜΗ ΤΩΝ ΜΕΓΑΛΩΝ ΔΕΔΟΜΕΝΩΝ
ΜΠΡΟΥΝΟ ΛΑΚΑ 47188
14^ο ΕΞΑΜΗΝΟ

ΕΙΣΗΓΗΤΗΣ: κ. Δρόσος Χρήστος

ΑΘΗΝΑ 2023

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο/η κάτωθι υπογεγραμμένος Λάκα Μπρούνο του Σοκόλ, με αριθμό μητρώου 47188 φοιτητής της Σχολής Πανεπιστημίου Δυτικής Αττικής του Τμήματος Βιομηχανικής Σχεδίασης και Παραγωγής, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο/Η Δηλών/ούσα



Εξεταστική επιτροπή:

Δρόσος Χρήστος,

Παπουτσιδάκης Μιχάλης,

Συμεωνάκη Ελένη.

Περιεχόμενα

Κατάλογος Πινάκων	5
Κατάλογος Σχημάτων	6
Περίληψη	7
Abstract	9
Εισαγωγή	10
Κεφάλαιο 1 Τεχνολογίες Μεγάλων Δεδομένων	13
1.1 Ορισμοί των Μεγάλων Δεδομένων (Big Data).....	13
1.2 Χαρακτηριστικά μεγάλων δεδομένων	14
1.3 Τεχνολογίες Μεγάλων Δεδομένων	21
1.3.1. Επισκόπηση Τεχνολογιών Μεγάλων Δεδομένων	22
1.3.2 Σύγκριση Τεχνολογιών Μεγάλων Δεδομένων	23
1.4 Πλατφόρμες μεγάλων δεδομένων	26
1.4.1. Big Data Europe (BDE).....	26
1.4.2. Σύγκριση πλατφορμών μεγάλων δεδομένων.....	28
Κεφάλαιο 2 Εφαρμογές Μεγάλων Δεδομένων -Επιχειρηματικότητα	31
2.1 Εφαρμογές Μεγάλων Δεδομένων	31
2.1.1 Φροντίδα υγείας.....	32
2.1.2 Τραπεζικός κλάδος	32
2.1.3 Λιανεμποριο	33
2.1.4 Τηλεπικοινωνίες	33
2.2 Μεγάλα Δεδομένα στην επιχειρηματικότητα	34
2.2.1 Το εξελισσόμενο επιχειρηματικό τοπίο.....	36
2.2.2 Ευκαιρίες	36
2.2.4 Μεγάλα Δεδομένα και Ανοιχτή Καινοτομία	37

2.2.5 Η χρήση των μεγάλων δεδομένων για την κοινωνική καινοτομία.....	38
Κεφάλαιο 3 Ερευνητικό Μέρος.....	40
3.1 Σκοπός της έρευνας.....	40
3.2 Επιλογή δείγματος.....	40
3.3 Μεθοδολογία.....	41
3.4 Στατιστική ανάλυση.....	42
Κεφάλαιο 4 Αποτελέσματα – Συζήτηση.....	43
4.1 Δημογραφικά χαρακτηριστικά.....	43
Συμπεράσματα.....	55
Βιβλιογραφία.....	57
Παράρτημα.....	62

Κατάλογος Πινάκων

Πίνακας 1 Σύγκριση των δυνατοτήτων τεχνολογιών Μεγάλων Δεδομένων.....	26
Πίνακας 2 Ανασκόπηση των πλατφορμών Μεγάλων Δεδομένων	30
Πίνακας 3 Δημογραφικά χαρακτηριστικά: Φύλο	43
Πίνακας 4 Δημογραφικά χαρακτηριστικά: Ηλικία	43
Πίνακας 5 Δημογραφικά χαρακτηριστικά: Επίπεδο σπουδών	43
Πίνακας 6 Δημογραφικά χαρακτηριστικά: Τομέας δραστηριοποίησης της εταιρείας.....	44
Πίνακας 7 Αριθμός των εργαζομένων στην εταιρεία	45
Πίνακας 8 Έτη δραστηριοποίησης της επιχείρησης.....	45
Πίνακας 9 Μεγάλα δεδομένα και αποδοτικότητα στην επιχείρηση	50
Πίνακας 10 Μεγάλα δεδομένα και συμβολή στον τομέα του εσωτερικού ελέγχου στην επιχείρηση	54

Κατάλογος Σχημάτων

Σχήμα 1 Ανάπτυξη των χαρακτηριστικών V _s των Μεγάλων Δεδομένων.....	21
Σχήμα 2 Πηγή άντληση των μεγάλων δεδομένων	46
Σχήμα 3 Συλλογή των μεγάλων δεδομένων.....	47
Σχήμα 4 Κυριότερος λόγος χρήσης των μεγάλων δεδομένων.....	47
Σχήμα 5 Εργαλεία για τη συλλογή των μεγάλων δεδομένων	48

Περίληψη

Τα μεγάλα δεδομένα αναφέρονται στα τεράστια δεδομένα που παράγονται από άτομα, συναλλαγές και μηχανές. Συνήθως χαρακτηρίζεται από τέσσερις διαστάσεις γνωστές ως τέσσερα V: όγκος, ταχύτητα, ποικιλία και ακρίβεια. Η ανάλυση δεδομένων περιλαμβάνει τη διαδικασία εξέτασης, καθαρισμού, μετατροπής και μοντελοποίησης μεγάλων δεδομένων για την αποκάλυψη πολύτιμων πληροφοριών και προτύπων που υποστηρίζουν τη λήψη αποφάσεων. Στόχος αυτής της διατριβής ήταν η διερεύνηση και αξιολόγηση της προοπτικής των ελληνικών επιχειρήσεων σχετικά με την υιοθέτηση των Μεγάλων Δεδομένων. Η μελέτη στόχευσε στον προσδιορισμό της τρέχουσας χρήσης της τεχνολογίας των Μεγάλων Δεδομένων μεταξύ των ελληνικών επιχειρήσεων, καθώς και του ρόλου που διαδραματίζει εντός του οργανισμού. Επιπλέον, διερεύνησε τις μεθόδους που χρησιμοποιούν αυτές οι εταιρείες για τη συλλογή και ανάλυση δεδομένων, την έκταση της χρήσης Μεγάλων Δεδομένων και τα οφέλη που προκύπτουν από την εφαρμογή τους.

Η παρούσα διπλωματική εργασία χωρίζεται σε δύο κύρια μέρη. Το πρώτο μέρος εστιάζει στην εξέταση τεχνολογία των Μεγάλων Δεδομένων παρέχοντας διάφορους ορισμούς και χαρακτηριστικά των δεδομένων. Έγινε εκτενής περιγραφή της τεχνολογίας και των πλατφορμών των Μεγάλων Δεδομένων. Ιδιαίτερη αναφορά έγινε σε εφαρμογές των μεγάλων δεδομένων σε διάφορους τομείς και στην επίδρασή τους στην επιχειρηματικότητα.

Το δεύτερο μέρος συνεπάγεται την ποσοτική μελέτη με την ανάπτυξη ενός ερωτηματολογίου με στόχο τη διερεύνηση της επίδρασης των Μεγάλων Δεδομένων στη λειτουργικότητα και την ανάπτυξη των επιχειρήσεων. Επιλέχθηκαν 40 επιχειρήσεις που ανήκουν σε ένα ευρύ φάσμα βιομηχανιών που χρησιμοποιεί εκτενώς τα Μεγάλα Δεδομένα. Οι κύριες πηγές από τις οποίες οι εταιρείες συλλέγουν τα δεδομένα τους βρέθηκε ότι ήταν τα μηνύματα ηλεκτρονικού ταχυδρομείου, οι οικονομικές συναλλαγές και οι μηχανές αναζήτησης (π.χ. Google) και τα κύρια εργαλεία περιλαμβάνουν τα R, Apache Hadoop και Apache Spark. Η πλειονότητα των εταιρειών συμφώνησε ότι έχει συγκεκριμένη στρατηγική αξιοποίησης των Μεγάλων Δεδομένων, υποδεικνύοντας ότι συνέβαλαν σημαντικά στην επίτευξη των στόχων τους, αν και είναι δυνατή η περαιτέρω αξιοποίησή τους. Τα αποτελέσματα αυτής της έρευνας υπογραμμίζουν τη σημασία και τη δύναμη των Μεγάλων Δεδομένων στη σύγχρονη ελληνική επιχειρηματικότητα και παρέχουν

ένα χρήσιμο μοντέλο που μπορεί να λειτουργήσει ως ερεθίσματα για περαιτέρω ερευνητικές προσπάθειες.

Λέξεις-κλειδιά: Μεγάλα Δεδομένα, επιχειρηματικότητα, αποδοτικότητα, λειτουργικότητα, εσωτερικός έλεγχος, οικονομικός έλεγχος

Abstract

Big data in entrepreneurship

Big data refers to the massive amounts of data generated by people, transactions and machines. It is typically characterized by four dimensions known as the four Vs: volume, speed, variety, and precision. Data analytics involves the process of examining, cleaning, transforming and modeling big data to uncover valuable insights and patterns that support decision making. The aim of this thesis was to investigate and evaluate the perspective of Greek businesses regarding the adoption of Big Data. The study aimed to determine the current use of Big Data technology among Greek businesses, as well as the role it plays within the organization. In addition, it explored the methods these companies use to collect and analyze data, the extent of Big Data use, and the benefits derived from its implementation.

This thesis is divided into two main parts. The first part focuses on examining Big Data technology by providing various definitions and characteristics of data. Big Data technology and platforms were extensively described. Special reference was made to applications of big data in various sectors and their impact on entrepreneurship.

The second part involves the quantitative study with the development of a questionnaire aimed at investigating the impact of Big Data on the functionality and development of businesses. 40 companies belonging to a wide range of industries that make extensive use of Big Data were selected. The main sources from which companies collect their data were found to be emails, financial transactions and search engines (eg Google) and the main tools include R, Apache Hadoop and Apache Spark. The majority of companies agreed that it has a specific Big Data utilization strategy, indicating that it has contributed significantly to achieving their goals, although further utilization is possible. The results of this research highlight the importance and power of Big Data in modern Greek entrepreneurship and provide a useful model that can act as stimuli for further research efforts.

Keywords: Big Data, entrepreneurship, efficiency, functionality, internal control, financial control

Εισαγωγή

Τα δεδομένα έχουν αυξηθεί όσον αφορά τη μεγάλη κλίμακα σε διάφορους τομείς. Σύμφωνα με μια έκθεση του Παγκόσμιου Ινστιτούτου McKinsey, αναφέρεται στα Μεγάλα Δεδομένα ως σύνολα δεδομένων με μέγεθος που υπερβαίνει την ικανότητα των παραδοσιακών εργαλείων λογισμικού, βάσης δεδομένων για τη συλλογή, αποθήκευση, διαχείριση και ανάλυση αυτών των δεδομένων. Οι περισσότεροι από τους ερευνητές όρισαν τα μεγάλα δεδομένα ως «έναν όρο που περιγράφει μεγάλους όγκους δεδομένων υψηλής ταχύτητας, πολύπλοκων και μεταβλητών που απαιτούν προηγμένες τεχνικές και τεχνολογίες για να καταστεί δυνατή η σύλληψη, αποθήκευση, διανομή, διαχείριση και ανάλυση των πληροφοριών». Τα τελευταία χρόνια, οι περισσότεροι από τους κύριους παίκτες τεχνολογίας, συμπεριλαμβανομένων των IBM, EMC, Microsoft, Oracle, Google, Facebook, SAS και Amazon κ.λπ. έχουν ξεκινήσει τα σχέδιά τους για Big Data. Αναδεικνύοντας την IBM ως περίπτωση, από το 2005, η IBM έχει επενδύσει 16 δισεκατομμύρια δολάρια σε 30 ακίνητα που σχετίζονται με Big Data. Πολλοί από τους βασικούς παράγοντες Big Data, όπως το SAS Institute, η Gartner, η IBM, οι συμβουλευτικές εταιρείες McKinsey υπέδειξαν στα Big Data ως το επόμενο πλεονέκτημα για την καινοτομία, την παραγωγικότητα, τον ανταγωνισμό και την ποιότητα (da Silva et al., 2018; Jabeen et al., 2017; Oussous et al., 2018; Sagiroglu & Sinanc, 2013).

Το 2011, η McKinsey προέβλεψε ότι τα Μεγάλα Δεδομένα σε πέντε τομείς (δημόσιος τομέας, υγειονομική περίθαλψη, λιανικό εμπόριο, παραγωγή και δεδομένα προσωπικής τοποθεσίας) θα δημιουργήσουν περίπου 223 δισεκατομμύρια δολάρια. Το Παγκόσμιο Οικονομικό Φόρουμ ανέδειξε τα Big Data ως την πιο αξιοσημείωτη τεχνολογία του 2012. Από ακαδημαϊκή άποψη, τα Big Data ήταν επίσης υπό το φως. Το μεγαλύτερο μέρος της έρευνας διερευνήθηκε στις δυνατότητες του Big Data Analytics (BDA) για να καλύψει τις διαστάσεις της τεχνολογίας, της διαδικασίας, του ταλέντου και της διαχείρισης που επηρέασαν θετικά την απόδοση του οργανισμού. Οι ακαδημαϊκοί των Big Data έχουν εντοπίσει διάφορες δυνατότητες και πόρους Big Data ως πιθανή σταθερή βάση για τη βελτίωση της απόδοσης του οργανισμού. Η ικανότητα του οργανισμού να κερδίσει τα οφέλη από διάφορες μορφές μαζικών δεδομένων είναι απαραίτητη και η ετοιμότητα για επενδύσεις σε Big Data βρίσκεται πλέον στο επίκεντρο της προσοχής (Soares, 2012).

Πρόσφατα είναι σύνηθες για τους οργανισμούς να αντιμετωπίζουν πιέσεις και προκλήσεις για να κερδίσουν και να διατηρήσουν τη θέση τους στην αγορά ανταγωνιστικών πλεονεκτημάτων, εντοπίζοντας τρόπους μείωσης του κόστους, βελτίωσης της ποιότητας και μείωσης του χρόνου στην αγορά. Η νέα εποχή του μετασχηματισμού των Μεγάλων Δεδομένων απαιτεί τεχνολογίες επόμενης γενιάς, ενώ οι παραδοσιακές δυνατότητες δεν θα επιτύχουν τον σκοπό τους για τη διαχείριση των Μεγάλων Δεδομένων. Οι οργανισμοί θα πρέπει να το χειριστούν κατάλληλα όχι μόνο για ανταγωνιστικό πλεονέκτημα αλλά και για επιβίωση στη νέα ψηφιακή αγορά. Ο οργανισμός πρέπει να αναζητήσει τις ανάγκες και τις επιθυμίες των πελατών του. Με άλλα λόγια, ο οργανισμός θα πρέπει να μετακινήσει τη συλλογή και την ανάλυσή του δεδομένων από απλώς έναν προσανατολισμό προϊόντος ή υπηρεσίας σε μια πλατφόρμα προσανατολισμένη στο μέλλον. Επιπλέον, οι οργανισμοί θα πρέπει να είναι σε θέση να προσδιορίζουν τις κρίσιμες πηγές δεδομένων, τη δομή, τις απαιτούμενες δεξιότητες και την αρχιτεκτονική, να ορίζουν την υποκείμενη υποδομή διαδικασίας που υποστηρίζει την ανάλυση Big Data, να ορίζουν μια στρατηγική Big Data και να μετρούν τις τεχνολογίες και τις εφαρμογές που υποστηρίζουν τις ανάγκες του οργανισμού σχετικά με την προετοιμασία για το ταξίδι μεγάλων δεδομένων (da Silva et al., 2018).

Πριν από την επανάσταση των μεγάλων δεδομένων, οι εταιρείες δεν μπορούσαν να αποθηκεύουν όλα τα αρχεία τους για μεγάλες περιόδους ούτε να διαχειρίζονται αποτελεσματικά τεράστια σύνολα δεδομένων. Πράγματι, οι παραδοσιακές τεχνολογίες έχουν περιορισμένη χωρητικότητα αποθήκευσης, άκαμπτα εργαλεία διαχείρισης και είναι ακριβές. Δεν διαθέτουν επεκτασιμότητα, ευελιξία και απόδοση που απαιτούνται στο πλαίσιο των Μεγάλων Δεδομένων. Στην πραγματικότητα, η διαχείριση μεγάλων δεδομένων απαιτεί σημαντικούς πόρους, νέες μεθόδους και ισχυρές τεχνολογίες. Πιο συγκεκριμένα, τα Big Data απαιτούν καθαρισμό, επεξεργασία, ανάλυση, ασφάλεια και παροχή λεπτομερούς πρόσβασης σε τεράστια εξελισσόμενα σύνολα δεδομένων. Οι εταιρείες και οι βιομηχανίες γνωρίζουν περισσότερο ότι η ανάλυση δεδομένων γίνεται ολοένα και περισσότερο ένας ζωτικός παράγοντας για να είμαστε ανταγωνιστικοί, να ανακαλύψουμε νέες ιδέες και να εξατομικεύσουμε τις υπηρεσίες.

Λόγω της ενδιαφέρουσας αξίας που μπορεί να εξαχθεί από τα Big Data, πολλοί παράγοντες σε διαφορετικές χώρες έχουν ξεκινήσει σημαντικά έργα. Οι ΗΠΑ ήταν ένας από τους ηγέτες που έπιασαν την ευκαιρία για Big Data. Τον Μάρτιο του 2012, η κυβέρνηση Ομπάμα ξεκίνησε την

Πρωτοβουλία Έρευνας και Ανάπτυξης Μεγάλων Δεδομένων με προϋπολογισμό 200 εκατομμυρίων. Στην Ιαπωνία, η ανάπτυξη Big Data έγινε ένας σημαντικός άξονας της εθνικής τεχνολογικής στρατηγικής τον Ιούλιο του 2012 (Sagiroglu & Sinanc, 2013). Τα Ηνωμένα Έθνη εξέδωσαν μια έκθεση με τίτλο Big Data for Development: Opportunities and Challenges. Στόχος του είναι να σκιαγραφήσει τις κύριες ανησυχίες σχετικά με τις προκλήσεις των Big Data και να προωθήσει τον διάλογο σχετικά με το πώς τα Big Data μπορούν να υπηρετήσουν τη διεθνή ανάπτυξη.

Ως αποτέλεσμα των διαφορετικών έργων Big Data σε όλο τον κόσμο, πολλά μοντέλα Big Data, πλαίσια και νέες τεχνολογίες δημιουργήθηκαν για να παρέχουν μεγαλύτερη χωρητικότητα αποθήκευσης, παράλληλη επεξεργασία και ανάλυση σε πραγματικό χρόνο διαφορετικών ετερογενών πηγών. Επιπλέον, έχουν αναπτυχθεί νέες λύσεις για τη διασφάλιση του απορρήτου και της ασφάλειας των δεδομένων. Σε σύγκριση με τις παραδοσιακές τεχνολογίες, τέτοιες λύσεις προσφέρουν μεγαλύτερη ευελιξία, επεκτασιμότητα και απόδοση. Επιπλέον, το κόστος των περισσότερων λύσεων αποθήκευσης και επεξεργασίας υλικού μειώνεται συνεχώς λόγω της βιώσιμης τεχνολογικής προόδου (Berisha et al., 2022).

Για την εξαγωγή γνώσης από τα Big Data, έχουν σχεδιαστεί και προταθεί διάφορα μοντέλα, προγράμματα, λογισμικά, υλικό και τεχνολογίες. Προσπαθούν να εξασφαλίσουν πιο ακριβή και αξιόπιστα αποτελέσματα για εφαρμογές Big Data. Ωστόσο, σε ένα τέτοιο περιβάλλον, μπορεί να είναι χρονοβόρο και δύσκολο να επιλέξετε ανάμεσα σε πολλές τεχνολογίες. Στην πραγματικότητα, θα πρέπει να ληφθούν υπόψη πολλές παράμετροι: τεχνολογική συμβατότητα, πολυπλοκότητα ανάπτυξης, κόστος, αποτελεσματικότητα, απόδοση, αξιοπιστία, υποστήριξη και κίνδυνοι ασφάλειας. Υπάρχουν πολλές έρευνες Big Data στη βιβλιογραφία, αλλά οι περισσότερες από αυτές τείνουν να επικεντρώνονται σε αλγόριθμους και προσεγγίσεις που χρησιμοποιούνται για την επεξεργασία Big Data και όχι σε τεχνολογίες. Στην παρούσα εργασία, παρουσιάζεται μια έρευνα σχετικά με τις πρόσφατες τεχνολογίες που αναπτύχθηκαν για Big Data. Πραγματοποιείται κατηγοριοποίηση και σύγκριση με τη χρήση, τα πλεονεκτήματα, τα όρια και τις δυνατότητές τους, αλλά και ανάλογα με τα διαφορετικά επίπεδα τους, όπως Επίπεδο αποθήκευσης δεδομένων, Επίπεδο επεξεργασίας δεδομένων, επίπεδο ερωτήματος δεδομένων, επίπεδο πρόσβασης δεδομένων και επίπεδο διαχείρισης. Αυτό βοηθά στην καλύτερη κατανόηση των συνδέσεων μεταξύ των διαφόρων τεχνολογιών Big Data και της λειτουργίας τους.

Κεφάλαιο 1 Τεχνολογίες Μεγάλων Δεδομένων

1.1 Ορισμοί των Μεγάλων Δεδομένων (Big Data)

Υπάρχουν πολλοί ορισμοί των Μεγάλων Δεδομένων. Δεν υπάρχουν ενιαίοι ενοποιημένοι ορισμοί που να μοιράζονται μεταξύ του ακαδημαϊκού κόσμου, των επιχειρήσεων, της βιομηχανίας, των μέσων ενημέρωσης και των διαφόρων ενδιαφερόμενων μερών. Η έλλειψη ενός συστηματικού ορισμού δημιουργεί μυστήριο γύρω από την έννοια των Μεγάλων Δεδομένων. Ο ορισμός των μεγάλων δεδομένων είναι συνήθως ένα άτομο, διαφορετικό από κλάδο σε κλάδο και ανάλογα με τα είδη των διαθέσιμων εργαλείων λογισμικού και τα μεγέθη των συνόλων δεδομένων είναι κοινά σε έναν συγκεκριμένο κλάδο. Έχουν γίνει σημαντικές συζητήσεις τόσο από τη βιομηχανία όσο και από τον ακαδημαϊκό κόσμο σχετικά με τον ορισμό των Μεγάλων Δεδομένων. Συνδέοντας την έννοια των Μεγάλων Δεδομένων με τις τρέχουσες βασισμένες ακαδημαϊκές μελέτες, η έννοια των Μεγάλων Δεδομένων μπορεί να γίνει πιο κατανοητή. Μια σαφής εικόνα της έννοιας των Μεγάλων Δεδομένων θα βελτιώσει την ευαισθητοποίηση σχετικά με το φαινόμενο των Μεγάλων Δεδομένων τόσο για τους επαγγελματίες όσο και για τους ακαδημαϊκούς, οδηγώντας σε ταχύτερα αναπτυσσόμενη και αποτελεσματικότερη δημιουργία αξίας από τα Μεγάλα Δεδομένα (Oussous et al., 2018).

Το 1997, επιστήμονες από τη NASA γράφουν μια εργασία με τον πρώτο ορισμό των Μεγάλων Δεδομένων. Αναφέρθηκαν στον όγκο δεδομένων ως πάροχος για μια συναρπαστική πρόκληση για τα συστήματα υπολογιστών που αυξάνουν τον απαιτητικό έως μεγάλο όγκο της κύριας μνήμης, του τοπικού δίσκου, καθώς και του απομακρυσμένου δίσκου. Η NASA το εντόπισε αυτό ως πρόβλημα Μεγάλων Δεδομένων που πρέπει να αποκτήσουν περισσότερους πόρους (Al-Sai & Abdullah, 2019). Το 2001 ο αναλυτής του META Group Doug Laney (τόρα Gartner) όρισε τις προκλήσεις και τις ευκαιρίες ανάπτυξης δεδομένων ως τρισδιάστατες (όγκος, ταχύτητα και ποικιλία). Το 2013, ο ορισμός των μεγάλων δεδομένων ενημερώθηκε από την Gartner, ο οποίος όρισε την έννοια των Μεγάλων Δεδομένων ως «πληροφοριακά στοιχεία μεγάλου όγκου, υψηλής ταχύτητας και/ή μεγάλης ποικιλίας που απαιτούν οικονομικά αποδοτικές καινοτόμες μορφές επεξεργασίας πληροφοριών για βελτιωμένη γνώση, λήψη αποφάσεων και διαδικασία βελτιστοποίηση». Η SAS ορίζει τα Μεγάλα Δεδομένα ως «δημοφιλή όρο που χρησιμοποιείται για να περιγράψει την εκθετική ανάπτυξη, τη διαθεσιμότητα και τη χρήση πληροφοριών, τόσο δομημένων όσο και μη δομημένων». Η International Business Machines Corporation (IBM)

πρόσθεσε έναν άλλο ορισμό για τα Μεγάλα Δεδομένα «Δεδομένα, που προέρχονται από παντού. αισθητήρες που χρησιμοποιούνται για τη συλλογή πληροφοριών για το κλίμα, αναρτήσεις σε ιστοτόπους μέσω κοινωνικής δικτύωσης, ψηφιακές φωτογραφίες και βίντεο, καταγραφή συναλλαγών αγορών και σήμα GPS κινητού τηλεφώνου για να αναφέρουμε μερικά», «Τα μεγάλα δεδομένα ορίζονται ως ένα μεγάλο σύνολο δεδομένων που είναι πολύ αδόμητα και ανοργάνωτα» , "Τα Μεγάλα Δεδομένα είναι μια μορφή δεδομένων που υπερβαίνει τις δυνατότητες επεξεργασίας της παραδοσιακής υποδομής ή μηχανών βάσεων δεδομένων" (Al Nuaimi et al., 2015).

Σύμφωνα με τους Al-Sai & Abdullah (2019), σήμερα υπάρχουν πολλοί διαφορετικοί ορισμοί και θεωρίες για το τι διαμορφώνει τα Μεγάλα Δεδομένα. Ο πιο συχνά αναφερόμενος ορισμός που δείχνει ότι τα Μεγάλα Δεδομένα υπερβαίνουν την εμβέλεια των κοινώς χρησιμοποιούμενων περιβαλλόντων υλικού και εργαλείων λογισμικού για τη λήψη, τη διαχείριση και την επεξεργασία τους εντός αποδεκτού χρόνου για τον χρήστη του. Η έννοια των μεγάλων δεδομένων έχει βελτιωθεί για να περιγράψει τις νέες και ισχυρές υπολογιστικές τεχνολογίες που έχουν αναπτυχθεί για την επεξεργασία του τεράστιου όγκου δεδομένων. Τα μεγάλα δεδομένα έχουν οριστεί με διαφορετικούς τρόπους, αλλά βασικά είναι μια νέα τεχνολογία που χαρακτηρίζεται κυρίως και προέρχεται από το Business Intelligence και το Business Analytics (BI&BA). Μπορεί να δημιουργήσει επιχειρηματική αξία μέσω των προγνωστικών αναλύσεων και των δυνατοτήτων υποστήριξης αποφάσεων, οι οποίες παρέχουν τη δυνατότητα αντιμετώπισης δεδομένων που δεν μπορούσαν να υποστούν επεξεργασία χρησιμοποιώντας παραδοσιακές προσεγγίσεις. Σύμφωνα με προηγούμενους ορισμούς των Μεγάλων Δεδομένων, ο πιο κοινός ορισμός των Μεγάλων Δεδομένων επινοήθηκε ως «ένας όρος που περιγράφει μεγάλους όγκους δεδομένων υψηλής ταχύτητας και πολύπλοκων μεταβλητών που απαιτούν προηγμένες τεχνικές και τεχνολογίες για να καταστεί δυνατή η σύλληψη, αποθήκευση, διανομή, διαχείριση, και ανάλυση των πληροφοριών».

1.2 Χαρακτηριστικά μεγάλων δεδομένων

Με βάση τους διάφορους ορισμούς μεγάλων δεδομένων, είναι προφανές ότι το μέγεθος είναι το κυρίαρχο χαρακτηριστικό παρά τη σημασία άλλων χαρακτηριστικών. Ο Laney (2001) πρότεινε τα τρία V ως διαστάσεις πρόκλησης στη διαχείριση δεδομένων και τα τρία V αποτελούν ένα κοινό πλαίσιο. Αυτές οι τρεις διαστάσεις δεν είναι ανεξάρτητες η μία από την άλλη. Αν αλλάξει η

μονοδιάσταση, αυξάνεται και η πιθανότητα αλλαγής μιας άλλης διάστασης (Gandomi & Haider, 2015).

Συχνά προστίθενται άλλες δύο διαστάσεις στα χαρακτηριστικά των μεγάλων δεδομένων, η ακρίβεια και η μεταβλητότητα. Τα πέντε V αντικατοπτρίζουν την αυξανόμενη δημοτικότητα των μεγάλων δεδομένων. Το πρώτο V είναι, όπως πάντα, ο όγκος, ο οποίος σχετίζεται με την ποσότητα των δεδομένων που παράγονται. Το δεύτερο V είναι για την ταχύτητα (επικαιρότητα μεγάλων δεδομένων), καθώς όλη η συλλογή και ανάλυση δεδομένων πρέπει να διεξάγεται έγκαιρα. Το τρίτο V αναφέρεται στην ποικιλία, καθώς τα μεγάλα δεδομένα έρχονται σε πολλές διαφορετικές μορφές και δομές, όπως δεδομένα ERP, email και tweets ή ήχος και βίντεο. Το τέταρτο V αναφέρεται στην «τεράστια τιμή αλλά πολύ χαμηλή πυκνότητα» των μεγάλων δεδομένων, προκαλώντας κρίσιμα προβλήματα όσον αφορά την εξαγωγή τιμών από σύνολα δεδομένων. Το πέμπτο V αναφέρεται στην ακρίβεια και αμφισβητεί την αξιοπιστία των μεγάλων δεδομένων όπου οι πηγές είναι εξωτερικές, όπως στις περισσότερες περιπτώσεις. Η ακρίβεια σχετίζεται με την αξιοπιστία, την ακρίβεια της πηγής δεδομένων και το πόσο κατάλληλα είναι τα δεδομένα για την προτεινόμενη χρήση (Elragal, 2014).

Η χρήση μεγάλων δεδομένων απαιτεί τη σωστή τεχνική αρχιτεκτονική, αναλυτικά στοιχεία και εργαλεία για να επιτραπεί η ανάδυση πληροφοριών από την κρυφή γνώση για τη δημιουργία αξίας για την επιχείρηση, και αυτά εξαρτώνται από την κλίμακα δεδομένων, τη διανομή, την ποικιλομορφία και την ταχύτητα. Τα μεγάλα δεδομένα χαρακτηρίζονται πιο εύκολα από τα τρία κύρια χαρακτηριστικά τους, ωστόσο: Όγκος δεδομένων (μέγεθος), Ταχύτητα (ρυθμός αλλαγής δεδομένων) και Ποικιλία (μορφές και τύποι δεδομένων καθώς και οι απαιτούμενοι τύποι ανάλυσης δεδομένων) (Al-Shiakhli, 2019).

Η ροή δεδομένων είναι η αιχμή των μεγάλων δεδομένων, καθώς μπορούν να συλλεχθούν σε πραγματικό χρόνο από πολλούς ιστότοπους. Η προσθήκη του τελικού V, της αλήθειας, έχει συζητηθεί από αρκετούς ερευνητές και οργανισμούς σε αυτό το πλαίσιο. Το Veracity εστιάζει στην ποιότητα των δεδομένων, η οποία μπορεί να είναι καλή, κακή ή απροσδιόριστη λόγω ασυνέπειας, ελλιπούς, ασάφειας, καθυστέρησης, παραπλάνησης ή προσεγγίσεων δεδομένων. Καθώς οι περισσότερες πηγές μεγάλων δεδομένων είναι εξωτερικές, και στερούνται διακυβέρνησης (Elragal, 2014).

Το σημαντικό για τους σύγχρονους οργανισμούς που αναζητούν ανταγωνιστικά πλεονεκτήματα είναι ο τρόπος διαχείρισης και εξαγωγής της αξίας από τα δεδομένα. Τα μεγάλα δεδομένα συνδυάζουν τεχνικές προκλήσεις με πολλαπλές ευκαιρίες, και έτσι η εξαγωγή επιχειρηματικής αξίας αντιπροσωπεύει ταυτόχρονα πρόκληση και ευκαιρία. Αυτό θέτει την επιχειρηματική προοπτική των μεγάλων δεδομένων δίπλα-δίπλα με τις τεχνικές πτυχές και το να δείχνει πώς τα μεγάλα δεδομένα προσθέτουν αξία στους στόχους του οργανισμού έχει γίνει μια κρίσιμη πτυχή της έρευνας σε αυτόν τον τομέα. Οι Manyika et al. (2011) διευκρίνισαν πώς τα μεγάλα δεδομένα μπορούν να δημιουργήσουν προστιθέμενη αξία για τους οργανισμούς:

- καθιστώντας τις πληροφορίες σαφείς και πιο συχνά εφαρμόσιμες.
- επιτρέποντας στους οργανισμούς να δημιουργούν και να αποθηκεύουν δεδομένα συναλλαγών σε ψηφιακή μορφή, διευκολύνοντας τη συλλογή ακριβέστερων πληροφοριών σχετικά με τα αποθέματα και τα προϊόντα.
- χρήση εξελιγμένων αναλυτικών δεδομένων μεγάλων δεδομένων για τη βελτίωση της ποιότητας λήψης αποφάσεων.
- αξιοποίηση μεγάλων δεδομένων για τη διαμόρφωση της επόμενης γενιάς προϊόντων και υπηρεσιών (Elragal, 2014).

Ο ποσοτικός προσδιορισμός μεγάλων δεδομένων μπορεί να γίνει όσον αφορά το μέγεθος αποθήκευσης, τον αριθμό των εγγράφων, τις συναλλαγές, τους πίνακες ή τα αρχεία. Τα μεγάλα δεδομένα προέρχονται από πολλές διαφορετικές πηγές που συλλέγονται για πολλούς σκοπούς, συμπεριλαμβανομένων των δεδομένων IoT, των αρχείων καταγραφής, των ροών κλικ και των μέσων κοινωνικής δικτύωσης. Για να χρησιμοποιηθούν όλες αυτές οι πηγές για αναλυτικά στοιχεία απαιτείται η ένωση μη δομημένων δεδομένων (όπως κείμενα σε φυσική γλώσσα) και ημιδομημένων δεδομένων (όπως επεκτάσιμη γλώσσα σήμανσης (XML), JSON ή ροές σύνοψης εμπλουτισμένου ιστότοπου (RSS)) σε ένα κοινό πλαίσιο δομημένων δεδομένων (Al-Shiakhli, 2019).

Στη συνέχεια περιγράφονται τα χαρακτηριστικά των Μεγάλων δεδομένων πιο αναλυτικά:

Τρία Vs

Αρκετοί ερευνητές έχουν ασχοληθεί με τα τρία κύρια V των χαρακτηριστικών μεγάλων δεδομένων και πόσο σημαντικά είναι στον κόσμο της ανάλυσης μεγάλων δεδομένων. Αρχικά, η διαφοροποίηση αφορούσε τρία κύρια χαρακτηριστικά που περιλαμβάνουν τον όγκο, την ταχύτητα και την ποικιλία.

Ο όγκος (volume) αναφέρεται στον μεγάλο όγκο δεδομένων (μέγεθος) που αυξάνεται συνεχώς για κάθε τομέα, είναι τεράστιος και παρέχει καλύτερη πρόβλεψη για το μέλλον.

Η ταχύτητα (velocity) αναφέρεται στο πόσο γρήγορα μπορούν να αναλυθούν τα δεδομένα προκειμένου να ληφθούν αποφάσεις (πρόσβαση), τα δεδομένα έρχονται όλο και περισσότερο σε μεγάλα κύματα και είναι σημαντικό να ληφθούν χρήσιμες αναλυτικές πληροφορίες σε πραγματικό χρόνο.

Τέλος, το Variety αναφέρεται στις εξαιρετικά ετερογενείς πηγές δεδομένων τόσο στο επίπεδο του σχήματος (δομημένο, αδόμητο και ημι-δομημένο) συμπεριλαμβανομένου κειμένου, αισθητήρα, ήχου, βίντεο, γραφήματος και άλλα. Ως εκ τούτου, υπάρχει μεγάλη ποικιλία μεταξύ των δεδομένων που συλλέγονται και αυτό παρέχει μια καλύτερη ανάλυση των μεγάλων δεδομένων.

Με άλλα λόγια, όταν εξετάζουμε ένα σύνολο δεδομένων, η παροχή αξιόπιστων πληροφοριών θα πρέπει να είναι αρκετά μεγάλη, εύκολη στην πρόσβαση και σε διαφορετικές δομές. Αυτό σημαίνει ότι μεγάλη ποικιλία δεδομένων πρέπει να είναι εύκολα προσβάσιμα για τη λήψη αποφάσεων σε πραγματικό χρόνο από τα σύνολα δεδομένων. Οι συγγραφείς που αναφέρονται παραπάνω δεν είναι οι μόνοι, υπάρχουν πολλοί άλλοι ακαδημαϊκοί και ερευνητές που ερεύνησαν τα τρία Vs των μεγάλων δεδομένων. Οι απόψεις είναι αρκετά παρόμοιες, αλλά τα χαρακτηριστικά εφαρμόστηκαν σε διαφορετικούς τομείς. Για παράδειγμα, η μελέτη των Dong & Srivastava (2013) είναι ένα από τα παραδείγματα που έγιναν στα χαρακτηριστικά των Μεγάλων Δεδομένων για να δοθεί μια σαφής κατανόηση, όπου οι συγγραφείς ανέλυσαν και εξήγησαν κάθε χαρακτηριστικό από τη δική τους οπτική γωνία.

Τέσσερα Vs

Τα αναλυτικά χαρακτηριστικά των Μεγάλων Δεδομένων έχουν αυξηθεί με την πάροδο των ετών λόγω της ραγδαίας τεχνολογικής αλλαγής όπως το Διαδίκτυο των πραγμάτων, η μηχανική

μάθηση, η τεχνητή νοημοσύνη, η ρομποτική, η τρισδιάστατη εκτύπωση, η βιοτεχνολογία, η νανοτεχνολογία, οι τεχνολογίες ανανεώσιμων πηγών ενέργειας, οι τεχνολογίες δορυφόρων και drone. Ως εκ τούτου, οι μελετητές έχουν αντιμετωπίσει αυτό το ζήτημα επανεξετάζοντας τα τρία χαρακτηριστικά και ένα τέταρτο χαρακτηριστικό που προστέθηκε πρόσφατα (Veracity).

Μία από τις θεμελιώδεις δυσκολίες για την οποία εισήχθη το τέταρτο χαρακτηριστικό, που δηλώθηκε από τους Berti-Equille & Ba (2016) είναι ότι τα δεδομένα μπορεί να είναι προκατειλημμένα, θορυβώδη, παρωχημένα, λανθασμένα, παραπλανητικά και ως εκ τούτου αναξιόπιστα. Το πρόβλημα επιδεινώνεται από αντικρουόμενα δεδομένα από διάφορες πηγές, τα οποία με τη σειρά τους οδηγούν στην αξιοπιστία των δεδομένων που πρέπει να αξιολογηθούν.

Ως εκ τούτου, το Veracity αναφέρεται στις διάφορες τιμές, με σημαντικές διαφορές στην κάλυψη, την ακρίβεια και την επικαιρότητα δεδομένων. Είναι σημαντικό να διασφαλιστεί ότι το σύνολο δεδομένων που εξετάζεται είναι επαρκώς ακριβές, καθώς πολύ σημαντικές επιχειρηματικές αποφάσεις λαμβάνονται με βάση αυτά τα σύνολα δεδομένων. Οι Kerper et al. (2014) αντιμετώπισε το απόρρητο, την αξιοπιστία και τη διαθεσιμότητα των δεδομένων τους, γεγονός που είχε ως αποτέλεσμα την καθιέρωση της ακρίβειας.

Πέντε Vs

Ενώ πολλοί ερευνητές πίστευαν σε τρία και τέσσερα Vs για τα χαρακτηριστικά των Μεγάλων Δεδομένων, ο Nguyen (2018) συζήτησε τα χαρακτηριστικά των δεδομένων στην εργασία του και εξέτασε τα πέντε μεγάλα Vs. Σύμφωνα με τον Nguyen, υπάρχουν πέντε Vs για μεγάλα δεδομένα όπου προσθέτει την Αξία (Value) στα υπάρχοντα τέσσερα Vs (Volume, Velocity, Variety και Veracity).

Η αξία περιγράφεται ως η ατομική ή οργανωτική ικανότητα μετατροπής μεγάλων δεδομένων σε πραγματικές αξίες, η οποία περιλαμβάνει την ικανότητα συλλογής και στη συνέχεια αξιοποίησης των δεδομένων για την επίτευξη συγκεκριμένων στόχων. Επιπλέον, ο Ghasemaghahi (2021) έχει θίξει τη σημασία του πέμπτου V του BD στην εργασία του, όπου η αξία αναφέρεται στην κοινωνική και οικονομική αξία (κυρίως νομισματική) που μπορεί να δημιουργήσουν τα μεγάλα δεδομένα. Ο τελικός στόχος της ανάλυσης δεδομένων είναι να παρέχει πληροφορίες και στοιχεία που θα βοηθήσουν τους μεγάλους οργανισμούς να λάβουν καλύτερες αποφάσεις που προσθέτουν

αξία στις εταιρείες τους και, στις περισσότερες περιπτώσεις, η αξία είναι το χρηματικό εισόδημα. Πολλοί αναλυτές και σύμβουλοι απασχολούνται από τις επιχειρήσεις για να αξιολογήσουν διάφορες μεταβλητές από διαφορετικά τμήματα για να κατανοήσουν, να ερευνήσουν και στη συνέχεια να λάβουν αποφάσεις που βελτιώνουν τις επιχειρήσεις τους. Για παράδειγμα, οι Debattista et al. (2015) συζήτησαν τις δυνατότητες των μεθόδων Συνδεδεμένων Δεδομένων για την αντιμετώπιση και των πέντε Vs. Ως εκ τούτου, η αξία είναι μια από τις διαστάσεις που είναι κρίσιμες για τις επιχειρηματικές αποφάσεις και για το ουσιαστικό συμφέρον του κόσμου των Μεγάλων Δεδομένων.

Επτά Vs

Μια άλλη έννοια που πρέπει να αντιμετωπιστεί για την ολοκλήρωση αυτής της έρευνας είναι οι συζητήσεις που γίνονται μεταξύ των ερευνητών για τα επτά Vs των μεγάλων δεδομένων. Υπάρχουν πολλοί μελετητές που συζητούν επτά V, και ένα από τα καλύτερα παραδείγματα είναι η εργασία των Khan et al. (2014). Αυτή η εργασία αποκάλυψε συνοπτικά τα επτά Vs. Τα επτά χαρακτηριστικά σύμφωνα με αυτά είναι: Όγκος, Ταχύτητα, Ποικιλία, Αυθεντικότητα, Αξία, Εγκυρότητα και Μεταβλητότητα. Τα δύο χαρακτηριστικά που προστέθηκαν είναι η εγκυρότητα (Validity) και η μεταβλητότητα (Variability) όπου:

Η εγκυρότητα είναι η ακρίβεια και η ορθότητα των δεδομένων σε σχέση με την προβλεπόμενη χρήση. Αν και ακούγεται τόσο παρόμοιο με την αλήθεια, είναι δύο διαφορετικές έννοιες. Ένα σύνολο δεδομένων μπορεί να μην έχει πρόβλημα με την ακρίβεια, αλλά μπορεί να μην είναι έγκυρο. Με άλλα λόγια, χωρίς να το επικυρώσουμε, δεν μπορούμε απλώς να πάρουμε ένα σύνολο δεδομένων και να το εμπιστευτούμε.

Η μεταβλητότητα είναι υπεύθυνη για γρήγορες και απροσδόκητες μεταβάσεις. Υπάρχουν πολλές επιχειρήσεις που έχουν παραδεχτεί ανοιχτά ότι δεν αποθηκεύουν παλαιότερα δεδομένα που δεν έχουν αξία. Οι διαδικτυακές εταιρείες, για παράδειγμα, μπορεί να μην θέλουν να διατηρήσουν το ιστορικό αγορών των παλαιότερων καταναλωτών, καθώς η εγγύηση μπορεί να λήξει. Είναι σημαντικό να διασφαλιστεί η αστάθεια ενός συνόλου δεδομένων για να επιτρέπεται η πλήρης αξιοπιστία του τελικού αποτελέσματος.

Δέκα Vs

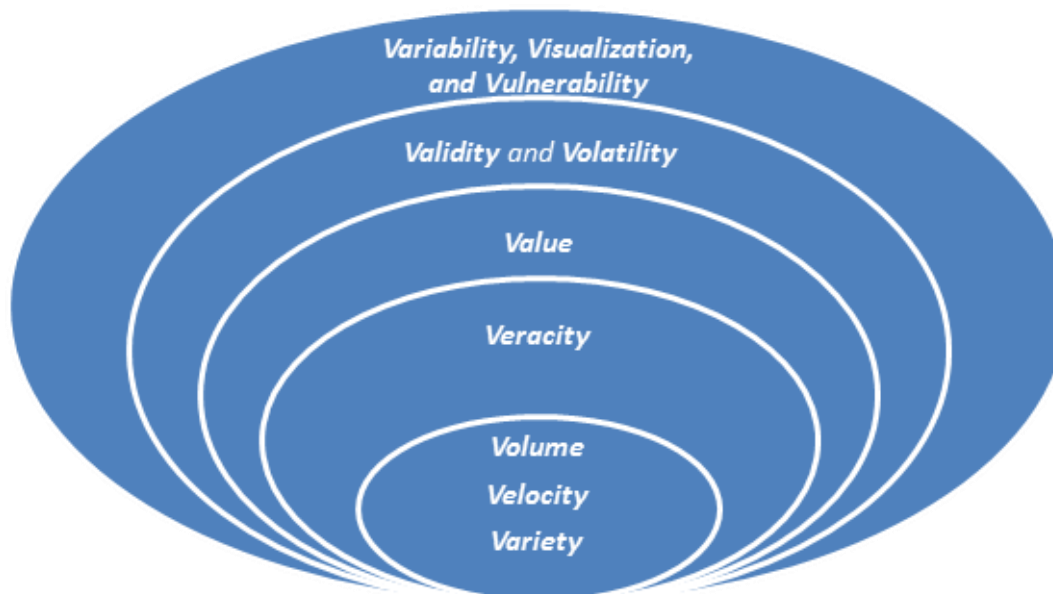
Είναι ενδιαφέρον ότι ο Ranjan (2019) είναι ένας από τους ερευνητές που συζήτησαν δέκα Vs του BD στην εργασία του. Τα 10 Vs είναι: Όγκος, Ταχύτητα, Ποικιλία, Αλήθεια, Τιμή, Εγκυρότητα, Μεταβλητότητα, Αστάθεια (Volatility), Οπτικοποίηση (Visualization) και Ευπάθεια (Vulnerability). Είναι χαρακτηριστικά BD για να διερευνηθεί εάν αυτές οι ιδιότητες κατέχουν το κλειδί για την επιτυχή υλοποίηση έργων Big Data. Ο Ranjan πρόσθεσε τη μεταβλητότητα, την οπτικοποίηση και την ευπάθεια σε αυτήν τη λίστα χαρακτηριστικών στις προηγούμενες μελέτες, όπου:

Η αστάθεια μπορεί να ακούγεται παρόμοια με το Variety, αλλά χρησιμοποιείται με την πάροδο του χρόνου για τον υπολογισμό της ακρίβειας των εισερχόμενων δεδομένων. Στο Big Data Analytics, η αστάθεια αναφέρεται στις ασυνέπειες στα αποτελέσματα. Λόγω του πλήθους των διαστάσεων δεδομένων που προκύπτουν από πολλούς διαφορετικούς τύπους και πηγές δεδομένων, τα μεγάλα δεδομένα μπορούν επίσης να είναι μεταβλητά. Ένα παράδειγμα αυτού είναι ότι οι αεροπορικές εταιρείες έχουν ένα μέτριο επίπεδο μεταβλητότητας, καθώς τα δεδομένα κοινωνικών πελατών μπορεί να είναι πολύ μεταβλητά, τα δεδομένα πτήσεων είναι αρκετά απλοποιημένα εκτός από περιπτώσεις όπως κακές καιρικές συνθήκες, προβλήματα κινητήρα κ.λπ. Για να προκύψουν σημαντικές αναλύσεις, ανωμαλία και θα πρέπει να εφαρμόζονται μέθοδοι ανίχνευσης ακραίων στοιχείων.

Η οπτικοποίηση περιλαμβάνει την παρουσίαση δεδομένων σχεδόν οποιουδήποτε τύπου σε γραφική μορφή που καθιστά εύκολη την κατανόηση και την ερμηνεία των δεδομένων. Η επιλογή λανθασμένου οπτικού βοηθήματος ή απλώς η προεπιλογή του πιο συνηθισμένου τύπου οπτικοποίησης δεδομένων μπορεί να προκαλέσει σύγχυση με τον θεατή ή να οδηγήσει σε εσφαλμένη ερμηνεία δεδομένων. Ένα από τα προβλήματα που επιβραδύνει την αναφορά και την ανάλυση είναι η κατανόηση των τύπων γραφημάτων που πρέπει να χρησιμοποιηθούν και γιατί.

Η ευπάθεια σχετίζεται με τα μέτρα ασφαλείας που πρέπει να ληφθούν ώστε τα δεδομένα που συλλέγονται να υποβάλλονται σε επεξεργασία σύμφωνα με τη νομοθεσία και τις επιθυμίες του πελάτη. Ο Ranjan (2019) έχει επίσης επεξηγήσει έξυπνα πώς κάθε κλάδος θα μπορούσε να είναι ευαίσθητος σε διαφορετικό επίπεδο.

Το Σχήμα 1 παρουσιάζει την ανάπτυξη στα V που αναφέρονται σε αυτή την ενότητα. Στη μελέτη που διεξήχθη από τους Husamaldin & Saeed (2019) για τη σύλληψη μιας ταξινόμησης συσχέτισης υπαρχουσών μεθόδων και σχετικών τεχνικών. Παρόλο που οι ερευνητές είχαν εισαγάγει πολλά νέα V, είναι, ωστόσο, δύσκολο για τους ερευνητές να συμφωνήσουν στα τρία κύρια χαρακτηριστικά, τα οποία είναι ο όγκος, η ταχύτητα και η ποικιλία.



Σχήμα 1 Ανάπτυξη των χαρακτηριστικών Vς των Μεγάλων Δεδομένων

1.3 Τεχνολογίες Μεγάλων Δεδομένων

Έχει προταθεί μια πληθώρα τεχνολογιών Μεγάλων Δεδομένων (Alexandron et al. 2014). Σε αυτήν την ενότητα, περιγράφονται εν συντομία ορισμένες από αυτές τις τεχνολογίες και παρέχεται μια σύγκριση μεταξύ τους σύμφωνα με ορισμένες κατηγορίες προβλημάτων.

1.3.1. Επισκόπηση Τεχνολογιών Μεγάλων Δεδομένων

Αυτή η ενότητα παρουσιάζει μια επισκόπηση των πιο ευρέως χρησιμοποιούμενων και πρόσφατα συζητημένων τεχνολογιών Μεγάλων Δεδομένων. Για το σκοπό αυτό, εξετάζονται τα κύρια χαρακτηριστικά των YARN/Hadoop, Spark, Flink και Hyracks/ASTERISK.

Σε πολλά σενάρια Μεγάλων Δεδομένων, το Apache Hadoop έχει γίνει το de facto πρότυπο δεδομένων και υπολογισμού για την κοινή χρήση και την πρόσβαση δεδομένων και υπολογιστικών πόρων (Vavilapalli et al., 2013). Το Hadoop είναι ένα επεκτάσιμο υπολογιστικό πλαίσιο ανοιχτού κώδικα που επιτρέπει την κατάτμηση των διαδικασιών υπολογισμού σε πολλούς κεντρικούς διακομιστές που δεν είναι απαραίτητα υπολογιστές υψηλής απόδοσης. Έχει δύο κύρια στοιχεία: μια μηχανή εκτέλεσης MapReduce και ένα κατανομημένο σύστημα αρχείων (DFS) που ονομάζεται HDFS – Hadoop Distributed FileSystem. Τα πλεονεκτήματα του Hadoop έγκεινται κυρίως στην υψηλή ευελιξία, την επεκτασιμότητα, το χαμηλό κόστος και την αξιοπιστία του για τη διαχείριση και την αποτελεσματική επεξεργασία μεγάλου όγκου δομημένων και μη δομημένων συνόλων δεδομένων, καθώς και στην παροχή χρονοδιαγραμμάτων εργασιών για εξισορρόπηση δεδομένων, πόρων και φορτίων εργασιών. Το Hadoop εξελίχθηκε σε YARN – Yet Another Resource Negotiator, του οποίου η αρχιτεκτονική αποσυνδέει το μοντέλο προγραμματισμού από τη διαχείριση πόρων και εκχωρεί πολλές λειτουργίες προγραμματισμού σε στοιχεία ανά εφαρμογή (Vavilapalli et al., 2013).

Το Apache Spark είναι μια ενοποιημένη μηχανή για την κατανομημένη επεξεργασία δεδομένων. Έχει ένα μοντέλο προγραμματισμού παρόμοιο με το MapReduce αλλά το επεκτείνει με μια αφαίρεση κοινής χρήσης δεδομένων που ονομάζεται Resilient Distributed Datasets ή RDDs (Zaharia et al., 2016). Χρησιμοποιώντας αυτήν την επέκταση, το Spark μπορεί να συλλάβει ένα ευρύ φάσμα φόρτων εργασίας επεξεργασίας που χρειάζονταν προηγουμένως ξεχωριστούς κινητήρες, όπως SQL, ροή, μηχανική εκμάθηση και επεξεργασία γραφημάτων. Το Spark σχεδιάστηκε επίσης για να ξεπεράσει τους περιορισμούς εισόδου/εξόδου του δίσκου και να βελτιώσει την απόδοση προηγούμενων συστημάτων. Το κύριο χαρακτηριστικό του Spark είναι η ικανότητά του να εκτελεί υπολογισμούς στη μνήμη. Επιτρέπει την αποθήκευση των δεδομένων στη μνήμη, εξαλείφοντας έτσι τον περιορισμό της επιβάρυνσης του δίσκου του YARN για επαναληπτικές εργασίες.

Το Apache Flink (Carbone et al., 2015) είναι ένα πλαίσιο επεξεργασίας ροής και παρτίδας ανοιχτού κώδικα για κατανεμημένες και υψηλής απόδοσης εφαρμογές που προέρχονται από έργο (Alexandrov et al., 2014). Βασίζεται στη φιλοσοφία ότι πολλές κατηγορίες εφαρμογών επεξεργασίας δεδομένων, συμπεριλαμβανομένων των αναλύσεων σε πραγματικό χρόνο, των συνεχών αγωγών δεδομένων, της ιστορικής επεξεργασίας δεδομένων και των επαναληπτικών αλγορίθμων μπορούν να εκφραστούν και να εκτελεστούν ως ροές δεδομένων με ανοχή σε σφάλματα. Το Flink μπορεί να εκτελεστεί ως εντελώς ανεξάρτητο πλαίσιο ή πάνω από το HDFS και το YARN. Αξιοποιεί την αποθήκευση στη μνήμη για τη βελτίωση της απόδοσης της εκτέλεσης του χρόνου εκτέλεσης. Οι κύριες καινοτομίες του Flink σε σύγκριση με προηγούμενες τεχνολογίες Μεγάλων Δεδομένων: ένας κατανεμημένος χρόνος εκτέλεσης ροής δεδομένων που εκμεταλλεύεται την εκτέλεση συνεχούς ροής για φόρτους εργασίας παρτίδας και ροής. Ακριβώς άπαξ κατάσταση συνέπεια μέσω ελαφρού σημείου ελέγχου, εγγενής επαναληπτική επεξεργασία, και μια εξελιγμένη σημασιολογία παραθύρων, που υποστηρίζει την ακατάλληλη επεξεργασία.

Το Hyracks/ASTERIX (Borkar et al., 2011) είναι μια παράλληλη πλατφόρμα λογισμικού που έχει σχεδιαστεί για να εκτελεί υπολογισμούς υψηλής έντασης δεδομένων σε μεγάλα συμπλέγματα κοινόχρηστου τύπου. Το Hyracks περιλαμβάνει μια συλλογή τελεστών που μπορούν να χρησιμοποιηθούν για τη συναρμολόγηση εργασιών επεξεργασίας δεδομένων χωρίς να χρειάζεται να γραφτεί ο χάρτης και η μείωση κώδικα. Επιπλέον, παρέχει επίσης ένα στρώμα συμβατό με το Yarn για την εκτέλεση υπαρχουσών εργασιών MapReduce. Το Hyracks παρουσιάζει ένα επεκτάσιμο σύστημα διαχείρισης πληροφοριών που υποστηρίζει την αποθήκευση, την αναζήτηση και την ανάλυση μεγάλων συλλογών ημιδομημένων ένθετων αντικειμένων δεδομένων. Το Hyracks παρέχει κέρδη απόδοσης σε σχέση με το MapReduce μέσω του πιο ευέλικτου μοντέλου χρήστη, ενώ είναι επίσης πιο αποτελεσματική εφαρμογή από τις εργασίες Hadoop for MapReduce για μια ποικιλία περιπτώσεων χρήσης υψηλής έντασης δεδομένων. Το Hyracks επιτυγχάνει επίσης κέρδη απόδοσης ανάκτησης σφαλμάτων σε σχέση με το Hadoop, προσφέροντας μια λιγότερο απαισιόδοξη προσέγγιση στον χειρισμό σφαλμάτων.

1.3.2 Σύγκριση Τεχνολογιών Μεγάλων Δεδομένων

Οι εταιρείες που χρησιμοποιούν τις τεχνολογίες Μεγάλων Δεδομένων συνήθως αντιμετωπίζουν προκλήσεις όπως: (i) συναλλαγές με την αποθήκευση ετερογενών πηγών όπως δομημένα, μη

δομημένα και ημιδομημένα δεδομένα. (ii) την ανάγκη ανακάλυψης γνώσης από μεγάλα και ετερογενή σύνολα δεδομένων όχι μόνο εφαρμόζοντας ερωτήματα SQL, αλλά και εκτελώντας πολύπλοκους αλγόριθμους μηχανικής μάθησης ή υπολογισμούς γραφημάτων. (iii) συνεχής λήψη ροών δεδομένων που πρέπει να υποβάλλονται σε συνεχή επεξεργασία κατά σειρά χιλιοστών του δευτερολέπτου για αναλύσεις (σχεδόν) σε πραγματικό χρόνο. Με βάση αυτές τις προκλήσεις, παρουσιάζεται μια περιεκτική συζήτηση σχετικά με το πώς αυτά τα πλαίσια είναι σε θέση ή όχι να παρέχουν υποστήριξη για ροή και επεξεργασία παρτίδας, γενική αποθήκευση και ανάλυση δεδομένων.

Επεξεργασία παρτίδων. Αυτό το είδος επεξεργασίας δεδομένων σχετίζεται στενά με τη μακροχρόνια εκτέλεση του υπολογισμού για μεγάλο όγκο δεδομένων, ταυτόχρονα, σε μια περίοδο. Συνήθως εκτελείται σε εργασίες ETL (Extract, Transform and Load), συγκέντρωση δεδομένων, εκπαίδευση και ενημέρωση μοντέλων μηχανικής εκμάθησης. Το Hadoop υιοθετήθηκε ευρέως στη μαζική επεξεργασία λόγω της εφαρμογής του MapReduce για τη διανομή της επεξεργασίας δεδομένων σε ένα υπολογιστικό σύμπλεγμα με πολλούς κόμβους. Το Hyracks εκτελεί επίσης ομαδική επεξεργασία δεδομένων. Ωστόσο, το Spark έχει γίνει ο κύριος υιοθετημένος κινητήρας για την επεξεργασία μεγάλων δεδομένων από διάφορες εταιρείες, καθώς προσφέρει γρήγορη δυνατότητα επεξεργασίας δεδομένων στη μνήμη, η οποία ξεπερνά τα γενικά έξοδα ανάγνωσης και εγγραφής του Hadoop.

Επεξεργασία ροής. Κατά την επεξεργασία ροής, τα δεδομένα επεξεργάζονται και τα αποτελέσματα παράγονται αυστηρά εντός συγκεκριμένων χρονικών περιορισμών (συνήα της τάξης των χιλιοστών του δευτερολέπτου και μερικές φορές μικροδευτερόλεπτων ανάλογα με την εφαρμογή και τις απαιτήσεις του χρήστη). Για παράδειγμα, το Spark Streaming λαμβάνει ζωντανές ροές δεδομένων εισόδου και διαιρεί τα δεδομένα σε μικρο-παρτίδες, οι οποίες επεξεργάζονται από τον κινητήρα Spark και χρησιμοποιούνται για τη δημιουργία της τελικής ροής αποτελεσμάτων σε παρτίδες. Το Micro-batching επιτρέπει το χειρισμό μιας ροής ως μια ακολουθία μικρών παρτίδων ή τμημάτων δεδομένων. Ωστόσο, μπορεί να εισαγάγει σημαντικά γενικά έξοδα με τη μορφή εργασιών προγραμματισμού. Από την άλλη πλευρά, το Flink μπορεί να προσφέρει όλα τα πλεονεκτήματα της προσωρινής αποθήκευσης χωρίς κανένα από τα γενικά έξοδα προγραμματισμού εργασιών (da Silva et al., 2018). Το Flink μπορεί επίσης να έχει καλή

απόδοση σε σενάρια πραγματικού ή σχεδόν πραγματικού χρόνου, όπου οι πληροφορίες από δεδομένα θα πρέπει να είναι διαθέσιμες σχεδόν την ίδια στιγμή της δημιουργίας δεδομένων.

Γενική αποθήκευση. Το HDFS μπορεί να αποθηκεύσει ένα διαφορετικό μείγμα δομημένων, αδόμητων και ημιδομημένων δεδομένων. Το Hyracks μπορεί να καταναλώσει τα δεδομένα από το HDFS και παρέχει επίσης την αποθήκευση δεδομένων AsterixDB για την απορρόφηση, αποθήκευση, ευρετηρίαση, αναζήτηση και ανάλυση μαζικών ποσοτήτων δεδομένων χρησιμοποιώντας ένα ευέλικτο μοντέλο δεδομένων (ADM). Το Spark υποστηρίζει την ενοποίηση με μια μεγάλη ποικιλία συστημάτων αρχείων, συμπεριλαμβανομένων των HDFS, MapR File System, Cassandra, Amazon S3 ή την υλοποίηση μιας προσαρμοσμένης λύσης. Το Flink επιτρέπει την ενοποίηση ετερογενών συνόλων δεδομένων, που κυμαίνονται από αυστηρά δομικά σχεσιακά δεδομένα, μη δομημένα δεδομένα κειμένου και ημιδομημένα δεδομένα. Λειτουργεί επίσης με HDFS και συνδέεται με διάφορα άλλα συστήματα αποθήκευσης δεδομένων. Το Flink και το Spark δεν παρέχουν μια κύρια λύση αποθήκευσης.

Αναλύσεις δεδομένων. Το YARN/Hadoop υποστηρίζει ενεργά πολλά έργα ανώτατου επιπέδου για τη δημιουργία εργαλείων ανάπτυξης και τη διαχείριση της ροής δεδομένων και της επεξεργασίας του, όπως Giraph, Pig, Hive, Mahout και HBase. Το Spark υποστηρίζει επίσης ένα ευρύ φάσμα εφαρμογών, συμπεριλαμβανομένων των ETL, Machine Learning (MLib), Stream Processing (Spark Streaming) και Graph computation (GraphX). Η στοίβα του Flink προσφέρει βιβλιοθήκες με API υψηλού επιπέδου για διαφορετικές περιπτώσεις χρήσης: Επεξεργασία σύνθετων συμβάντων (CEP), Μηχανική μάθηση (FlinkML) και Ανάλυση γραφημάτων (Gelly). Η στοίβα λογισμικού του συστήματος Hyracks αποτελείται επίσης από διάφορες διεπαφές για αναλυτικά στοιχεία, όπως SQL (Hivesterix), XQuery (Apache VXQuery) και Graph (Pregelix). Ακόμη και τα Hyracks μπορούν να εκτελέσουν αποτελεσματικά πολύπλοκες κατανεμημένες λειτουργίες ροής δεδομένων και να εκφράσουν πλήρεις σχεσιακές άλγεβρες. Επίσης, εκθέτει API χαμηλού επιπέδου και απαιτεί από έναν ειδικό μηχανικής μάθησης (ML) να επαναδιατυπώσει τους αλγόριθμους του ως τελεστές ροής δεδομένων (Sparks et al., 2013).

Τέλος, ο Πίνακας 1 συνοψίζει τα παραπάνω και παρέχει μια σύντομη σύγκριση των δυνατοτήτων τεχνολογιών Μεγάλων Δεδομένων σύμφωνα με τις κατηγορίες προβλημάτων Μεγάλων Δεδομένων που αναλύονται.

Πίνακας 1 Σύγκριση των δυνατοτήτων τεχνολογιών Μεγάλων Δεδομένων

Χαρακτηριστικά/ Κατηγορίες	Hadoop	Spark	Flink	Hyracks
Τύπος επεξεργασίας	Ημιδομημένα δεδομένα	Μέτρια δομημένα δεδομένα	Δομημένα δεδομένα	Ημιδομημένα δεδομένα
Γενική αποθήκευση	HDFS	δεν υπάρχει κύριος χώρος αποθήκευσης	δεν υπάρχει κύριος χώρος αποθήκευσης	AsterixDB
Αναλύσεις δεδομένων	SQL, ML, Graph	ETL, ML, Graph	ML, CEP, Graph	SQL, XQuery, Graph

1.4 Πλατφόρμες μεγάλων δεδομένων

Μια πλατφόρμα Μεγάλων Δεδομένων είναι ένα οικοσύστημα υπηρεσιών και τεχνολογιών που χρειάζεται να εκτελεί αναλύσεις σε ογκώδη, πολύπλοκα και δυναμικά δεδομένα. Έτσι, η κλιμάκωση της πλατφόρμας υλικού καθίσταται επικείμενη και η επιλογή των σωστών τεχνολογιών υλικού/λογισμικού γίνεται μια κρίσιμη απόφαση εάν οι απαιτήσεις του χρήστη πρόκειται να ικανοποιηθούν σε εύλογο χρονικό διάστημα (Singh & Reddy, 2015). Πρόσφατα εμφανίστηκε ένα σύνολο πλατφορμών Μεγάλων Δεδομένων, συμπεριλαμβανομένου του Big Data Europe (BDE), Hortonworks και Cloudera (Jabeen et al., 2017).

1.4.1. Big Data Europe (BDE)

Η Πλατφόρμα BDE ανέπτυξε μια υπολογιστική υποδομή για το χειρισμό μεγάλων όγκων δεδομένων σε ποικίλες μορφές. Αντιμετωπίζει τις απαιτήσεις της απλούστευσης της χρήσης, της διευκόλυνσης της ανάπτυξης, της διαχείρισης της ετερογένειας και της βελτίωσης της επεκτασιμότητας και διευκολύνει την εκτέλεση και την ενσωμάτωση πλαισίων και εργαλείων

Μεγάλων Δεδομένων όπως το Hadoop, το Spark, το Flink και πολλά άλλα (Jabeen et al., 2017). Οι συγγραφείς αποφάσισαν να χρησιμοποιήσουν το Docker ως μεθοδολογία συσκευασίας και ανάπτυξης, καθώς και να διαχειριστούν αποτελεσματικά την ποικιλία των υποκείμενων πόρων υλικού παράλληλα με τις ποικίλες απαιτήσεις λογισμικού. Το BDE επιτρέπει την εκτέλεση μιας ποικιλίας εργασιών ροής μεγάλων δεδομένων, όπως η μετάδοση μηνυμάτων (μέσω Kafka, Flume), αποθήκευση (μέσω Hive, Cassandra), ανάλυση (μέσω Spark, Flink) ή δημοσίευση (μέσω GeoTriples). Επιπλέον, η πλατφόρμα είναι ανοιχτού κώδικα και εντελώς δωρεάν.

Το Hortonworks Data Platform (HDP) είναι μια σύγχρονη αρχιτεκτονική δεδομένων ανοιχτού κώδικα που προσφέρει άμεση αξία μειώνοντας το κόστος αποθήκευσης καθώς ενσωματώνει το Yarn στο κέντρο δεδομένων του και βελτιστοποιώντας το κόστος Enterprise Data Warehouse εκφορτώνοντας εργασίες υπολογιστών χαμηλής αξίας όπως το ETL. Το Yarn επιτρέπει στο HDP να ενσωματώνει όλες τις μηχανές επεξεργασίας δεδομένων σε όλη την κοινότητα και το εμπορικό οικοσύστημα για να παρέχει συνεπείς κοινές υπηρεσίες και πόρους σε όλη την πλατφόρμα. Το Ambari είναι μια διαισθητική διεπαφή ιστού και ένα ισχυρό API REST που κάνει τη διαχείριση του HDP πιο απλή, συνεπή και ασφαλή. Επιπλέον, το HDP είναι μια ολοκληρωμένη λύση που προσφέρει όχι μόνο επεξεργασία και διαχείριση δεδομένων, αλλά και τις επιχειρηματικές δυνατότητες που ανταποκρίνονται στις απαιτήσεις μιας επιχείρησης που καλύπτει την ασφάλεια, τη διακυβέρνηση και τις λειτουργίες (Oussous et al., 2018).

Η Cloudera ήταν η πρώτη εταιρεία που ανέπτυξε και διένειμε λογισμικό που βασίζεται στο Apache Hadoop και έχει κάνει την ανάλυση δεδομένων σε Μεγάλα Δεδομένα πιο βολική και προσβάσιμη σε οποιονδήποτε ενδιαφέρεται. Ενσωματώνει το Hadoop με περισσότερα από δώδεκα άλλα κρίσιμα έργα ανοιχτού κώδικα. Η Cloudera δημιούργησε ένα λειτουργικά προηγμένο σύστημα που βοηθά στην εκτέλεση ροών εργασίας Big Data από άκρο σε άκρο. Διαφορετικά έργα συνθέτουν το οικοσύστημα Cloudera για μια ποικιλία εργασιών Big Data: επεξεργασία ροής (μέσω Spark), μετάδοση μηνυμάτων (μέσω Kafka, Flume), αποθήκευση (μέσω Accumulo, Hive, Pig, HBase), ανάλυση (μέσω Flink, Impala), αναζήτηση (μέσω Cloudera Search) ή παροχή ενός επεκτάσιμου και παραγωγικού web GUI για τους χρήστες (μέσω HUE).

1.4.2. Σύγκριση πλατφορμών μεγάλων δεδομένων

Στη συνέχεια παρουσιάζεται μια σειρά από επαναλαμβανόμενα προβλήματα που αντιμετωπίζουν συνήθως οι οργανισμοί που μπορεί να γίνουν πιο περίπλοκα όταν ασχολούνται με Μεγάλα Δεδομένα: (i) ενσωμάτωση διαφορετικών πηγών Μεγάλων Δεδομένων και παρέχουν μια διαφανή άποψη στους χρήστες. (ii) διαχείριση και προστασία των δεδομένων του οργανισμού, προκειμένου να διασφαλίζει γενικά κατανοητά, σωστά, πλήρη και ασφαλή εταιρικά δεδομένα. (iii) παρακολούθηση των δεδομένων, των πόρων και των εφαρμογών για την ανασκόπηση και αξιολόγηση της υγείας και της απόδοσης ολόκληρου του συστήματος. Κάθε πρόκληση μπορεί να συνοψιστεί σε μία από τις ακόλουθες κατηγορίες προβλημάτων: Ενοποίηση δεδομένων, Διακυβέρνηση δεδομένων και Υπηρεσίες παρακολούθησης.

Ενοποίηση δεδομένων. Η ολοκλήρωση δεδομένων περιλαμβάνει το συνδυασμό δεδομένων από διαφορετικές πηγές και την παροχή στους χρήστες μιας ενοποιημένης εικόνας τους. Το HDP έχει συνεργαστεί με το Talend, μια ισχυρή και ευέλικτη λύση ανοιχτού κώδικα για την ενσωμάτωση Big Data που υποστηρίζει εγγενώς το Hadoop, συμπεριλαμβανομένων υποδοχών για HDFS, HBase, Pig, Sqoop και Hive χωρίς να χρειάζεται να γράψετε κώδικα. Το Talend υποστηρίζει επίσης το Cloudera Navigator. Μια άλλη εναλλακτική λύση για το HDP είναι το Oracle Data Integrator (ODI). Ένας χρήστης μπορεί να δημιουργήσει μια ροή από πηγές σε στόχους διαφορετικών τεχνολογιών, συμπεριλαμβανομένων σχεσιακών βάσεων δεδομένων, εφαρμογών, πίνακες XML, JSON, Hive, HBase, αρχεία HDFS και ούτω καθεξής. Η πλατφόρμα BDE προχωρά περισσότερο από το HDP και το Cloudera, συνθέτοντας μια πηγή Σημασιολογικών Δεδομένων – ένα αποθετήριο που παρέχεται για την επεξεργασία και ανάλυση των συνόλων δεδομένων στην αρχική τους μορφή – που ονομάζεται Ontario. Το Οντάριο δημιουργεί ένα Σημασιολογικό Επίπεδο πάνω από τη Πηγή Δεδομένων, το οποίο είναι υπεύθυνο για την αντιστοίχιση δεδομένων σε υπάρχοντα σημασιολογικά λεξιλόγια/οντολογίες. Μια επιτυχημένη διαδικασία χαρτογράφησης, που ονομάζεται Semantic Lifting, παρέχει μια άποψη για το σύνολο των δεδομένων (Sagiroglu & Sinanc, 2013). Με αυτόν τον τρόπο, τα δεδομένα μπορούν να εξαχθούν, να αναζητηθούν ή να αναλυθούν από τις ετερογενείς πηγές στη λίμνη σαν να ήταν σε μια ενιαία μορφή χρησιμοποιώντας μια γλώσσα ερωτημάτων υψηλού επιπέδου. Ένα άλλο σχετικό στοιχείο είναι το Semagrow, ένα σύστημα επεξεργασίας ερωτημάτων SPARQL που συνενώνει πολλαπλά απομακρυσμένα τελικά σημεία.

Διακυβέρνηση δεδομένων. Η Διακυβέρνηση Δεδομένων είναι ένα σύστημα δικαιωμάτων και λογαριασμών αποφάσεων για διαδικασίες που σχετίζονται με πληροφορίες, που εκτελείται σύμφωνα με συμφωνημένα μοντέλα που περιγράφουν ποιος μπορεί να προβεί σε ποιες ενέργειες με ποιες πληροφορίες και πότε, υπό ποιες συνθήκες, χρησιμοποιώντας ποιες μεθόδους (Dresner Advisory Services, 2017). Ο Soares (2012) επεκτείνει αυτόν τον ορισμό συμπεριλαμβάνοντας πολιτικές σχετικά με τη βελτιστοποίηση, το απόρρητο και τη δημιουργία εσόδων από Μεγάλα Δεδομένα. Η διαχείριση συστημάτων μεγάλων δεδομένων μπορεί να είναι πολύπλοκη. Η σταθερή διασφάλιση συνόλων δεδομένων σε πολλαπλά αποθετήρια μπορεί να είναι εξαιρετικά επιρρεπής σε σφάλματα. Το στοιχείο Cloudera Navigator Data Management είναι ένα πλήρως ενσωματωμένο εργαλείο διαχείρισης δεδομένων και ασφάλειας για το Hadoop που έχει σχεδιαστεί για να ανταποκρίνεται στις ανάγκες συμμόρφωσης, διακυβέρνησης δεδομένων και ελέγχου των παγκόσμιων επιχειρήσεων. Το HDP χρησιμοποιεί Apache Atlas και Apache Ranger, τα οποία συνδυάζουν την ταξινόμηση δεδομένων με την επιβολή της πολιτικής ασφαλείας. Το Apache Atlas δημιουργήθηκε ως μέρος της πρωτοβουλίας Hadoop Data Governance και προσφέρει τη δυνατότητα προβολής της σειράς πολλαπλών συστατικών, παρέχοντας μια πλήρη εικόνα της κίνησης των δεδομένων μέσω ορισμένων μηχανών ανάλυσης όπως οι Apache Storm, Kafka, Falcon και Hive. Το Apache Ranger παρέχει κεντρική διαχείριση ασφαλείας για το Hadoop. Με την ενσωμάτωση του Atlas και του Ranger, το HDP επιτρέπει στις εταιρείες να εφαρμόζουν δυναμικές πολιτικές πρόσβασης χρόνου εκτέλεσης που αποτρέπουν προληπτικά τις παραβιάσεις. Η BDE δεν εμβαθύνει πολύ στη διακυβέρνηση δεδομένων, καθώς δεν αντιμετωπίζει ζητήματα όπως το απόρρητο, η κοινή χρήση και τα δικαιώματα των δεδομένων.

Παρακολούθηση. Η παρακολούθηση είναι η διαδικασία της προληπτικής επανεξέτασης και αξιολόγησης όσων έχουν παρακολουθηθεί (ως δεδομένα, πόροι ή εφαρμογές). Το λογισμικό παρακολούθησης βοηθά στη μέτρηση και την παρακολούθηση των δεδομένων συνήθως χρησιμοποιώντας πίνακες εργαλείων, ειδοποιήσεις και αναφορές. Το Cloudera Manager παρέχει πολλές δυνατότητες για την παρακολούθηση της υγείας και της απόδοσης των στοιχείων των συμπλεγμάτων (κεντρικοί υπολογιστές, δαίμονες υπηρεσιών) καθώς και της απόδοσης και των απαιτήσεων πόρων των εργασιών που εκτελούνται σε συμπλέγματα. Το BDE κάνει διάκριση μεταξύ παρακολούθησης πόρων και παρακολούθησης κατάστασης. Το πρώτο επιτρέπει την παρακολούθηση της υγείας ενός διακομιστή ή ενός στοιχείου στην πλατφόρμα (χρήση CPU, χρήση μνήμης, δίκτυο I/O και χρήση δίσκου), ενώ το δεύτερο προσφέρει πληροφορίες για την

κατάσταση μιας συγκεκριμένης εφαρμογής (da Silva et al., 2018). Για την παρακολούθηση πόρων, τα εργαλεία Docker Stats, cAdvisor, Prometheus, InfluxDB και Grafana μπορούν να είναι χρήσιμα στην πλατφόρμα BDE. Για την παρακολούθηση της κατάστασης, το BDE υποστηρίζει ενσωματωμένη σύνδεση με docker και ELK. Ως μέρος του HDP, το Apache Ambari επιτρέπει τον σχεδιασμό, την εγκατάσταση και την ασφαλή διαμόρφωση συμπλεγμάτων υπολογιστών, διευκολύνοντας την παροχή συνεχούς συντήρησης και διαχείρισης συμπλέγματος.

Τέλος, μια περίληψη αυτής της ενότητας παρουσιάζεται στον Πίνακα 2 (da Silva et al., 2018), η οποία παρέχει μια σύντομη ανασκόπηση των πλατφορμών Μεγάλων Δεδομένων σύμφωνα με τις κατηγορίες προβλημάτων που αναλύονται.

Πίνακας 2 Ανασκόπηση των πλατφορμών Μεγάλων Δεδομένων

Πλατφόρμα/ Κατηγορίες	HDP	BDE	Cloudera
Ενοποίηση δεδομένων	Talend, ODI	Ontario, Semagrow	Talend
Διακυβέρνηση δεδομένων	Atlas, Ranger	Χωρίς υποστήριξη	Cloudera Navigator
Παρακολούθηση	Ambari	Prometheus, ELK stack	Cloudera Manager

Κεφάλαιο 2 Εφαρμογές Μεγάλων Δεδομένων -Επιχειρηματικότητα

2.1 Εφαρμογές Μεγάλων Δεδομένων

Μια πρόσφατη έρευνα που διεξήχθη το 2017 (Nashua, 2017) τόνισε την αυξανόμενη χρήση των αναλυτικών στοιχείων μεγάλων δεδομένων στις εταιρείες. Η μελέτη εξέτασε την υιοθέτηση μεγάλων δεδομένων από οργανισμούς σε σύγκριση με προηγούμενα έτη (2015, 2016 και 2017). Τα αποτελέσματα έδειξαν ότι μέχρι το 2017, πάνω από το 50 τοις εκατό των οργανισμών ενσωμάτωναν μεγάλα δεδομένα στις δραστηριότητές τους (Hugh J., 2019).

Τα μεγάλα δεδομένα έχουν αναδειχθεί ως σημαντική πηγή για τα επιχειρηματικά μοντέλα πολλών οργανισμών, τα οποία χρησιμεύουν ως αντανakλάσεις των υλοποιημένων στρατηγικών τους (Günther et al., 2017). Τα επιχειρηματικά μοντέλα αντιπροσωπεύουν την ικανότητα ενός οργανισμού να δημιουργεί και να συλλαμβάνει αξία. Κατά συνέπεια, οι οργανισμοί πρέπει να επανεξετάσουν τη χρήση των μεγάλων δεδομένων στο πλαίσιο των επιχειρηματικών μοντέλων τους. Αυτό περιλαμβάνει τη μόχλευση των αναλυτικών στοιχείων για την πρόσβαση σε νέες πηγές δεδομένων και τη χρήση τεχνικών που ενισχύουν την αποδοτικότητα και την αποτελεσματικότητα.

Οι Grover et al. (2018) υποστήριξαν ότι η επίτευξη στρατηγικής επιχειρηματικής αξίας από μεγάλα δεδομένα απαιτεί σημαντικές επενδύσεις σε υποδομές και αναλυτικές τεχνολογίες για να καταστεί δυνατή η εξειδικευμένη ανάλυση και η στρατηγική τοποθέτηση. Ως εκ τούτου, οι επιχειρήσεις πρέπει να έχουν πρόσβαση σε εργαλεία τελευταίας τεχνολογίας και να προσλαμβάνουν επαγγελματίες με γνώσεις δεδομένων που διαθέτουν βαθιά κατανόηση των σχετικών τεχνολογιών.

Ο Watson (2014) συνέγραψε μια εργασία για την ανάλυση μεγάλων δεδομένων, η οποία δημοσιεύτηκε από την Ένωση Πληροφοριακών Συστημάτων (CAIS). Αυτό το άρθρο συζήτησε τις προόδους στις τεχνολογίες, τις εφαρμογές και τον αντίκτυπο της ανάλυσης μεγάλων δεδομένων εκείνη την εποχή. Το 2019, ο ίδιος ερευνητής (Watson, 2019) τόνισε αρκετές σημαντικές πρόσφατες εξελίξεις στην ανάλυση μεγάλων δεδομένων, όπως:

- Συνεχής υιοθέτηση αναλυτικών στοιχείων μεγάλων δεδομένων.
- Αύξηση του αριθμού των εφαρμογών μεγάλων δεδομένων.
- Ανάπτυξη της τεχνολογίας του συστήματος Hadoop.

- Αξιοποίηση λιμνών δεδομένων.
- Αξιοποίηση προηγμένων μοντέλων ανάλυσης.
- Έμφαση στις αρχές αλγοριθμικής διαφάνειας.

Η ανάλυση μεγάλων δεδομένων έχει μεγάλες δυνατότητες για εφαρμογές όπως η πρόβλεψη ζήτησης, όπου η προηγούμενη εργασία έχει χρησιμοποιηθεί για την ταξινόμηση διαφόρων αναλυτικών τεχνικών (Hofmann & Rutschmann, 2018). Επιπλέον, η ανάλυση μεγάλων δεδομένων έχει βρει πρακτική χρήση σε διάφορους τομείς, εξυπηρετώντας μια σειρά σκοπών.

Ακολουθούν οι εφαρμογές των Μεγάλων Δεδομένων σε βασικούς τομείς:

2.1.1 Φροντίδα υγείας

Σύμφωνα με έρευνα που διεξήχθη από τον Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh, & Byers (2011), τα μεγάλα δεδομένα έχουν τη δυνατότητα να μειώσουν τα απόβλητα και να βελτιώσουν την αποτελεσματικότητα στις κλινικές λειτουργίες, την έρευνα και ανάπτυξη και τη δημόσια υγεία. Αυτό μπορεί να επιτευχθεί μέσω της χρήσης στατιστικών εργαλείων και αλγορίθμων, μοντελοποίησης πρόβλεψης για ταχύτερη ανάπτυξη φαρμάκων και συσκευών, ανάλυση αρχείων ασθενειών για ενίσχυση της επιδημιολογίας, επιτάχυνση της ανάπτυξης εμβολίων και ταυτοποίηση σχετικών δεδομένων για την παροχή υπηρεσιών και την πρόληψη κρίσεων. Η ανάλυση μεγάλων δεδομένων στην υγειονομική περίθαλψη προσφέρει σημαντικά οφέλη όπως ταχύτερη ανίχνευση ασθενειών, βελτιωμένη αποτελεσματικότητα θεραπείας, έγκαιρη αναγνώριση της απάτης στον τομέα της υγειονομικής περίθαλψης και βελτιωμένη υγεία του πληθυσμού. Οι Zhong et al. (2016) τόνισε επίσης τον αντίκτυπο της ανάλυσης μεγάλων δεδομένων στην υγειονομική περίθαλψη, υπογραμμίζοντας την ικανότητά της να ενσωματώνεται στην καθημερινή ζωή και να παρέχει πληροφορίες για τις ασθένειες και τις εμπειρίες υγειονομικής περίθαλψης.

2.1.2 Τραπεζικός κλάδος

Η ανάλυση μεγάλων δεδομένων διαδραματίζει κρίσιμο ρόλο στον τραπεζικό κλάδο, επιτρέποντας στην ανάλυση των αρχείων καταγραφής πελατών και των αλληλεπιδράσεων για να αποκτήσει μια ολοκληρωμένη άποψη των πελατών και των λειτουργιών. Αυτό επιτρέπει στις τράπεζες να

κατηγοριοποιούν τους πελάτες με βάση τα προφίλ κινδύνου και τη χρήση πίστωσης, προσφέροντας εξατομικευμένα προϊόντα και υπηρεσίες. Η χρήση των αναλυτικών στοιχείων επεκτείνεται σε ολόκληρο τον κλάδο, συμπεριλαμβανομένων των τραπεζικών εργασιών λιανικής που παρακολουθούν και αντιστοιχούν τις συναλλαγές πελατών. Η υιοθέτηση της επιστήμης δεδομένων και της ανάλυσης βοηθά τις τράπεζες να αντιμετωπίσουν αποτελεσματικά τις επιχειρηματικές προκλήσεις επόμενης γενιάς, ενώ διαχειρίζονται αποτελεσματικά τους πόρους. Η ανάλυση μεγάλων δεδομένων βοηθά επίσης στην πρόβλεψη της συμπεριφοράς των πελατών, στην κατανόηση πελατών και εργαζομένων και στον μετριασμό των οικονομικών κινδύνων στον τομέα των χρηματοοικονομικών υπηρεσιών (Rana, 2019).

2.1.3 Λιανεμπόριο

Ο κλάδος του λιανικού εμπορίου αντιμετωπίζει σημαντικό αντίκτυπο από την ανάλυση μεγάλων δεδομένων, ενισχύοντας την εμπειρία του πελάτη και καταπολεμώντας την απάτη. Η πρόβλεψη της ζήτησης για προϊόντα δίνει τη δυνατότητα στους λιανοπωλητές να παρέχουν καλύτερες υπηρεσίες στους πελάτες και τα δεδομένα τιμολόγησης συμβάλλουν στη συγκέντρωση επιχειρηματικής ευφυΐας. Εργαλεία όπως το Hadoop Distributed File System (HDFS) διευκολύνουν την αποθήκευση, την επεξεργασία και την ανάλυση αυτών των δεδομένων, εξάγοντας πολύτιμες πληροφορίες. Η ανάλυση μεγάλων δεδομένων παρέχει στους λιανοπωλητές πληροφορίες για αποφάσεις αγοράς, δυνατότητες κατηγοριοποίησης πελατών και πληροφοριών κοινωνικών μέσων για να κατανοήσουν τις προτιμήσεις των πελατών και την ανάλυση συναισθήματος. Η κατηγοριοποίηση σε πραγματικό χρόνο και η στόχευση κοινωνικού μάρκετινγκ καθίστανται δυνατές μέσω προηγμένων τεχνικών μεγάλων δεδομένων, με αποτέλεσμα πιο ακριβή και στοχευμένη διαφήμιση. Υιοθετώντας τεχνικές ανάλυσης δεδομένων, οι έμποροι λιανικής μπορούν να αντιμετωπίσουν νέες προκλήσεις και να αξιοποιήσουν ευκαιρίες σε μια ταχέως μεταβαλλόμενη αγορά (Hofmann et al., 2018).

2.1.4 Τηλεπικοινωνίες

Η ανάλυση μεγάλων δεδομένων βελτιώνει την ποιότητα διαχείρισης στον τομέα των τηλεπικοινωνιών αξιοποιώντας την ανάλυση δεδομένων σε πραγματικό χρόνο και την παρακολούθηση των αρχείων καταγραφής μηχανών. Τα προγνωστικά αναλυτικά στοιχεία

μπορούν να εντοπίσουν ζητήματα απόδοσης και να παρέχουν έγκαιρες προειδοποιήσεις, μειώνοντας τη μεταβλητότητα και διασφαλίζοντας καλύτερη ποιότητα. Ωστόσο, η αποθήκευση και η επεξεργασία μεγάλων δεδομένων παρουσιάζει προκλήσεις για τον κλάδο των τηλεπικοινωνιών. Οι παραδοσιακές τεχνικές ανάλυσης είναι συχνά ακριβές, καθιστώντας τις τεχνικές μεγάλων δεδομένων όπως το Hadoop, το οποίο περιλαμβάνει ενότητες όπως το HDFS και το MapReduce, πολύτιμες για οικονομικά αποδοτική αποθήκευση και υπολογισμό. Η ανάλυση μεγάλων δεδομένων έχει τη δυνατότητα να χειρίζεται τις τεράστιες ποσότητες δομημένων και μη δομημένων δεδομένων που παράγονται από δίκτυα κινητής τηλεφωνίας, παρέχοντας πληροφορίες για τη βελτίωση της απόδοσης του δικτύου (He et al., 2011).

2.2 Μεγάλα Δεδομένα στην επιχειρηματικότητα

Η επιχειρηματικότητα είναι μια έννοια που περιλαμβάνει διάφορους κλάδους και δεν έχει έναν καθολικά αποδεκτό ορισμό. Η υπάρχουσα βιβλιογραφία επιβεβαιώνει την απουσία ενός μοναδικού ορισμού για τους επιχειρηματίες ή την επιχειρηματικότητα. Αντίθετα, μια σειρά από θεωρητικές προοπτικές σε ένα διεπιστημονικό πεδίο μελέτης ρίχνουν φως σε διαφορετικές πτυχές των επιχειρηματικών φαινομένων.

Η κύρια εστίαση στον ορισμό της επιχειρηματικότητας περιστρέφεται γύρω από τους κινδύνους και τις προκλήσεις που συνεπάγεται η δημιουργία και η παραγωγή αγαθών ή υπηρεσιών, τα οποία στη συνέχεια προσφέρονται προς πώληση σε συνεχώς μεταβαλλόμενες αγορές. Η καινοτομία διαδραματίζει επίσης κρίσιμο ρόλο στον ορισμό της επιχειρηματικότητας, όπως τόνισε ο Schumpeter, ο οποίος τόνισε ότι οι επιχειρηματίες έχουν θεμελιώδη ρόλο στον εντοπισμό και την ανταπόκριση στις οικονομικές ασυνέχειες. Επιπλέον, άλλα κοινά παρατηρούμενα χαρακτηριστικά των επιχειρηματιών, όπως η τάση ανάληψης ρίσκου και η ορμή για επιτεύγματα, συνδέονται συχνά με ατομικά χαρακτηριστικά προσωπικότητας και εμπειρίες ζωής.

Η έλλειψη συνέπειας στον ορισμό της επιχειρηματικότητας έχει θέσει προκλήσεις για την έρευνα σε αυτόν τον τομέα. Πολλοί ερευνητές έχουν επικεντρώσει τις προσπάθειές τους στην εξερεύνηση τεσσάρων βασικών παραδειγμάτων που συμβάλλουν στην ολοκληρωμένη κατανόηση της επιχειρηματικότητας.

Αναγνωρισμένη ως η κινητήρια δύναμη πίσω από την οικονομική ανάπτυξη, η επιχειρηματικότητα επηρεάζει σημαντικά την οικονομική ανάπτυξη, την ευημερία και την κοινωνική αλλαγή. Μελετητές, (Eakin & Luers, 2006) υποστηρίζουν ότι η προώθηση της επιχειρηματικότητας είναι ζωτικής σημασίας για τη διατήρηση μιας υγιούς οικονομίας, τη διατήρηση της ευημερίας και τη δημιουργία νέων ευκαιριών απασχόλησης. Είναι ευρέως αποδεκτό ότι οι επιχειρηματίες θεωρούν τους εαυτούς τους ως άτομα με ιδέες και όχι ως απλούς επιχειρηματίες. Όχι μόνο τρέφουν και αναπτύσσουν τις έννοιές τους αλλά επίσης διαταράσσουν το περιβάλλον οικοσύστημα ενώ το ανανεώνουν συνεχώς.

Οι κοινωνικές και περιβαλλοντικές διαστάσεις αντιμετωπίζονται ανοιχτά από την κοινωνική επιχειρηματικότητα και τη βιώσιμη επιχειρηματικότητα (Assembly, 2015). Αυτές οι προσεγγίσεις επικεντρώνονται στην επίτευξη κοινωνικών, οικονομικών και οικολογικών αποτελεσμάτων. Η επιδίωξη της βιώσιμης καινοτομίας απαιτεί την ανάπτυξη επαναλαμβανόμενων, βιώσιμων και μεταβιβάσιμων στρατηγικών που μπορούν να οδηγήσουν αυτή την καινοτομία προς τα εμπρός.

Οι Στόχοι Βιώσιμης Ανάπτυξης (ΣΒΑ) στοχεύουν άμεσα στη συμβολή στην εκπλήρωση των ανθρωπίνων δικαιωμάτων σε ευρεία κλίμακα. Αυτοί οι στόχοι περιλαμβάνουν τομείς όπως η εξάλειψη της φτώχειας, η ανακούφιση της πείνας, η υγειονομική περίθαλψη, η ισότητα των φύλων, η εκπαίδευση, η προστασία του περιβάλλοντος, η κοινωνική δικαιοσύνη και αμέτρητα άλλα κοινωνικά και οικονομικά προβλήματα που καθορίζουν τους πρωταρχικούς στόχους του κόσμου για πρόοδο προς μια βιώσιμη κοινωνία.

Η εμφάνιση των μεγάλων δεδομένων παρουσιάζει στους επιχειρηματίες ένα πλήθος προκλήσεων και ευκαιριών καθώς η τεχνολογία διαταράσσει τα υπάρχοντα συστήματα. Το κλειδί βρίσκεται στο πώς οι επιχειρηματίες μπορούν να αξιοποιήσουν αυτές τις ευκαιρίες και να μετατρέψουν υποσχόμενες επιχειρηματικές ιδέες σε εγχειρήματα που δημιουργούν αξία. Στο σημερινό ταχέως εξελισσόμενο τοπίο, οι τεχνολογικές εξελίξεις αναδιαμορφώνουν διάφορες πτυχές της ζωής μας. Οι επιχειρηματίες δεν έχουν την πολυτέλεια να αγνοήσουν τις δυνατότητες της τεχνητής νοημοσύνης, των μεγάλων δεδομένων, της ρομποτικής και του Διαδικτύου των Πραγμάτων (IoT). Αντίθετα, πρέπει να αξιοποιήσουν αυτές τις μεταμορφωτικές δυνάμεις. Η αγκαλιά της καινοτομίας στον ψηφιακό χώρο εξοπλίζει τους επιχειρηματίες με βασικές γνώσεις και εργαλεία για την πλοήγηση στη συνεχιζόμενη ψηφιακή επανάσταση (Abdelli et al., 2021).

2.2.1 Το εξελισσόμενο επιχειρηματικό τοπίο

Σύμφωνα με έκθεση του ΟΟΣΑ (2017), η «εγγενής διαφήμιση» αντιπροσωπεύει μια νέα γενιά επιχειρηματιών στην ψηφιακή εποχή. Τα πρόσφατα γεγονότα έχουν επηρεάσει σημαντικά το επιχειρηματικό οικοσύστημα, με πολλές νεοφυείς επιχειρήσεις να δημιουργούν την πρόταση αξίας τους στο διαδίκτυο και να αναδεικνύονται σε ηγέτες του κλάδου. Η δημιουργία μιας startup επιχείρησης στο Διαδίκτυο δεν ήταν ποτέ πιο προσιτή, με μειωμένο κόστος υποδομής και διαθεσιμότητα λύσεων λογισμικού με τη μορφή μοντέλων υπηρεσιών για τις επιχειρήσεις. Οι κοινωνίες αγκαλιάζουν όλο και περισσότερο την ψηφιακή επιχειρηματικότητα. Το κοινό του Διαδικτύου αντικαθιστά πλέον τις πατέντες ως μέσο καλύτερης προστασίας, καθώς επιτρέπουν στις νεοφυείς επιχειρήσεις να δημιουργήσουν πολύτιμα δίκτυα. Η κουλτούρα των startup υπογραμμίζει τη σημασία των επιχειρηματιών χάκερς και ενθαρρύνει την ανατρεπτική σκέψη. Τα πρώην εμπόδια εισόδου θεωρούνται πλέον ως ευκαιρίες και ο κόσμος των startup έχει υιοθετήσει όρους όπως το αναπτυξιακό hacking, αντικαθιστώντας τις παραδοσιακές πρακτικές μάρκετινγκ. Σε αυτό το νέο επιχειρηματικό πλαίσιο, η προτεραιότητα στον πελάτη είναι ζωτικής σημασίας. Οι νεοσύστατες εταιρείες που μπορούν πρώτα να συνεργαστούν με τους πελάτες και να δώσουν προτεραιότητα στις ανάγκες τους αποκτούν δεσπόζουσα θέση στην αγορά, δίνοντάς τους τη δυνατότητα να διαμορφώσουν τους όρους της ροής αξίας και να προσφέρουν αξία στους πελάτες.

2.2.2 Ευκαιρίες

Η επιχειρηματικότητα βρίσκεται αυτή τη στιγμή σε εξαιρετική φάση, καθώς η καθημερινή μας ζωή συνδέεται όλο και περισσότερο μέσω κινητών τηλεφώνων, υπολογιστών και δικτύων, διευκολύνοντας την πρόσβαση των επιχειρηματιών στους ανθρώπους. Η επιχειρηματικότητα έχει γίνει κοινή αξία πολλών ατόμων. Τα μεγάλα δεδομένα παρουσιάζουν μια σειρά από ευκαιρίες επιχειρηματικότητας και έχουν γίνει ένας σημαντικός τομέας μελέτης για την καινοτομία και την επιχειρηματικότητα. Ένα ζωντανό παράδειγμα του πώς η καινοτομία μπορεί να αιχμαλωτίσει τους ανθρώπους είναι η ιστορία του WhatsApp, μιας δημοφιλής εφαρμογή ανταλλαγής μηνυμάτων. Ο Brian Acton και ο Jan Koum, που εργάζονταν στο παρελθόν στο Yahoo!, συνέλαβαν την ιδέα για το WhatsApp το 2009. Αυτή η εφαρμογή απέκτησε γρήγορα μια μεγάλη βάση χρηστών παγκοσμίως λόγω της εύκολης προσβασιμότητας και των δυνατοτήτων συνομιλίας σε πραγματικό χρόνο. Οι Acton και Koum απέφεραν σημαντικά έσοδα χωρίς διαφημιστικές προσπάθειες.

Χρεώνανε τους χρήστες iPhone για την εγκατάσταση και χρήστες Android σε ετήσια βάση για αυτήν την εξαιρετικά επιτυχημένη εφαρμογή (Saeed & Husamaldin, 2021).

2.2.4 Μεγάλα Δεδομένα και Ανοιχτή Καινοτομία

Η ανοιχτή καινοτομία αναφέρεται στη συνεργασία μεταξύ επιχειρήσεων, εργολάβων, πελατών και διαδικτυακών χειριστών, που οδηγεί στη θεμελίωση της καινοτομίας. Βασίζεται στην ανταλλαγή γνώσεων για την ενίσχυση της εσωτερικής καινοτομίας και την επέκταση των αγορών για εξωτερική χρήση σχεδιασμού (Phills et al., 2014). Η ανοιχτή καινοτομία δίνει έμφαση σε ένα πιο προσιτό και κατανοημένο μοντέλο σχεδιασμού, προσφέροντας ποικίλες ευκαιρίες για τους επιχειρηματίες. Μπορούν να συμβάλουν στη διαδικασία δημιουργίας καθιερωμένων εταιρειών και να συμμετέχουν ως εταίροι σε κυρίαρχες πλατφόρμες. Υιοθετώντας ανοιχτές στρατηγικές καινοτομίας, οι επιχειρηματίες και οι επιχειρήσεις αλληλεπιδρούν με παραγωγούς, πελάτες, ανταγωνιστές και συμβούλους, δίνοντάς τους τη δυνατότητα να συλλέγουν δεδομένα από διάφορες πηγές.

Η ανοιχτή καινοτομία μπορεί να λάβει τρεις μορφές:

- Ανοιχτή καινοτομία μεταξύ εταιρειών: Τα τμήματα Έρευνας και Ανάπτυξης επιτρέπουν τη ροή ιδεών από εξωτερικές εταιρείες για την προώθηση της καινοτομίας.
- Συνεργασίες μεταξύ διαφορετικών επιχειρηματικών τομέων: Οι εταιρείες συνεργάζονται με παίκτες από πολλές αγορές ανοίγοντας τα επιχειρηματικά τους μοντέλα στην καινοτομία.
- Ανοικτή καινοτομία που περιλαμβάνει καταναλωτές: Χρησιμοποιώντας ένα μοντέλο διανομής peer-to-peer, το διαδίκτυο και τις κοινότητες χρηστών, οι καταναλωτές συνεργάζονται και συμβάλλουν στην καινοτομία. Αυτό είναι γνωστό ως οικονομία της συνεισφοράς, όπου άπειροι χρήστες του διαδικτύου παράγουν δεδομένα.

Η εποχή των μεγάλων δεδομένων και των ψηφιακών εξελίξεων έχει διευκολύνει την ανταλλαγή δεδομένων, ιδεών, υπηρεσιών και κερδών, παρουσιάζοντας πολλές ευκαιρίες για ανοιχτή καινοτομία. Η επιστήμη του πολίτη προέκυψε ως αποτέλεσμα του πολλαπλασιασμού των μεγάλων δεδομένων, δίνοντας τη δυνατότητα ακόμη και στους πολίτες να αναπτύξουν νέες ιδέες και προϊόντα.

2.2.5 Η χρήση των μεγάλων δεδομένων για την κοινωνική καινοτομία

Τα μεγάλα δεδομένα θεωρούνται πολύτιμο εργαλείο για την αντιμετώπιση σύνθετων κοινωνικών προβλημάτων μέσω της κοινωνικής καινοτομίας, όρος που έχει αποκτήσει ιδιαίτερη σημασία τα τελευταία χρόνια. Πιστεύεται ότι παρέχει πιο πρακτικές και αποτελεσματικές λύσεις σε σύγκριση με τις συμβατικές προσεγγίσεις, ωφελώντας το κοινωνικό σύνολο και όχι μόνο τα άτομα (Nicholls & Murdock, 2011). Η κοινωνική καινοτομία δημιουργεί νέες ιδέες και δομές, συμβάλλοντας στην κοινωνική ανασυγκρότηση, τη διαμόρφωση των προτύπων και την προώθηση της κοινοτικής ευημερίας, της αμεροληψίας και της δικαιοσύνης. Η αξία που δημιουργείται από την κοινωνική καινοτομία ωφελεί πρωτίστως την κοινωνία, καθώς οι επιχειρηματίες αναζητούν λύσεις σε κοινωνικά προβλήματα με έμφαση στην κοινωνική καινοτομία.

Οι κοινωνικοί επιχειρηματίες διαδραματίζουν κρίσιμο ρόλο στην εύρεση καινοτόμων τρόπων επίλυσης προβλημάτων και στην παροχή των τόσο απαραίτητων προϊόντων σε χώρες σε όλο τον κόσμο. Προσεγγίζουν τα κοινωνικά προβλήματα από μια διαφορετική οπτική γωνία, επαναπροσδιορίζοντάς τα και οδηγώντας την καινοτομία. Τα μεγάλα δεδομένα έχουν τη δυνατότητα να εξορθολογίσουν τις διαδικασίες λήψης αποφάσεων, επιτρέποντας στους επιχειρηματίες να αντιμετωπίσουν τα πιο προκλητικά κοινωνικά προβλήματα του κόσμου. Η ενασχόληση με την κοινωνική καινοτομία προσφέρει στους επιχειρηματίες μια ολοκληρωμένη προσέγγιση που αυξάνει τις πιθανότητές τους για επιτυχία στην εύρεση λύσεων. Η κατανόηση των συνθηκών, των προσεγγίσεων, των πολιτικών και των αρχών που οδηγούν σε επιτυχή κοινωνική αλλαγή και μακροπρόθεσμο αντίκτυπο είναι απαραίτητη (Westley et al., 2016).

Τα μεγάλα δεδομένα διευκολύνουν τη συνεργασία και τις συνεργασίες μεταξύ των ενδιαφερομένων που διαθέτουν τους απαραίτητους πόρους και τις δεξιότητες ανάλυσης δεδομένων. Η κοινωνική καινοτομία που υπερβαίνει τα κοινωνικά εμπόδια και υιοθετεί μια διεπιστημονική προσέγγιση μπορεί να προσεγγίσει αποτελεσματικά όλους τους ενδιαφερόμενους (Tsujimoto et al., 2018). Όλοι οι ενδιαφερόμενοι θα πρέπει να συμμετέχουν στη διαδικασία κοινωνικής καινοτομίας. Η παροχή προϊόντων ή υπηρεσιών απαιτεί ανάπτυξη διαχείρισης, νέας τεχνολογίας και καινοτομίας. Τα καλά οργανωμένα κοινωνικά δίκτυα, σχεδιασμένα μέσα σε ένα οργανωτικό πλαίσιο και με πολλαπλά επίπεδα, περιλαμβάνουν διαφορετικούς παράγοντες με διαφορετικά χαρακτηριστικά, πεποιθήσεις και ηθική (Calic & Ghasemaghaei, 2021).

Η αναγνώριση των αναγκών της κοινωνίας, η παροχή υπηρεσιών προς όφελός της και η διασφάλιση της διευκόλυνσης της βοήθειας στα άτομα και την κοινωνία απαιτεί τη χρήση μεγάλων δεδομένων με τη μορφή πληροφοριών που μπορούν να χρησιμοποιηθούν. Οι επιχειρηματίες μπορούν να αξιοποιήσουν διαφορετικές διαδικτυακές πλατφόρμες και να χρησιμοποιήσουν τεχνολογίες μέσω κοινωνικής δικτύωσης για τη γρήγορη συλλογή και χρήση δεδομένων για την αντιμετώπιση κοινωνικών προβλημάτων. Ωστόσο, οι κοινωνικοί επιχειρηματίες αντιμετωπίζουν προκλήσεις, όπως η δυσκολία συλλογής βασικών δεδομένων σε περιπτώσεις όπως η εμπορία ανθρώπων, όπου η αξιοπιστία είναι ζωτικής σημασίας για τα προγράμματα κατά της εμπορίας ανθρώπων. Η χρήση μεγάλων δεδομένων για κοινωνική αλλαγή, ιδιαίτερα για την αντιμετώπιση θεμάτων του κοινωνικού τομέα, είναι πιο περίπλοκη από τη χρήση τους σε επιχειρηματικά ή επιστημονικά πλαίσια. Οι υπεύθυνοι χάραξης πολιτικής και οι επιχειρηματίες μπορούν να παρέχουν λύσεις για κοινωνικά προβλήματα αξιοποιώντας τη δύναμη των μεγάλων δεδομένων, ειδικά όταν τα δεδομένα που συλλέγονται οργανώνονται και αναλύονται αποτελεσματικά (Calic & Ghasemaghaei, 2021).

Κεφάλαιο 3 Ερευνητικό Μέρος

3.1 Σκοπός της έρευνας

Στόχος της παρούσας έρευνας είναι η διερεύνηση και αξιολόγηση της προοπτικής των ελληνικών επιχειρήσεων σχετικά με την υιοθέτηση των Μεγάλων Δεδομένων. Η μελέτη στοχεύει στον προσδιορισμό της τρέχουσας χρήσης της τεχνολογίας των Μεγάλων Δεδομένων μεταξύ των ελληνικών επιχειρήσεων, καθώς και του ρόλου που διαδραματίζει εντός του οργανισμού. Επιπλέον, διερευνά τις μεθόδους που χρησιμοποιούν αυτές οι εταιρείες για τη συλλογή και ανάλυση δεδομένων, την έκταση της χρήσης Μεγάλων Δεδομένων και τα οφέλη που προκύπτουν από την εφαρμογή τους.

Μέσα από μια εκτενή βιβλιογραφική ανασκόπηση, οι πρωταρχικοί μας στόχοι είναι:

- Εντοπισμός των διαφορών που παρατηρούνται στη λήψη στρατηγικών αποφάσεων με την ενσωμάτωση της ανάλυσης μεγάλων δεδομένων.
- Προσδιορισμός των ευκαιριών για τον εξορθολογισμό των διαδικασιών ελέγχου, καθώς και των πιθανών απειλών που σχετίζονται με την ακατάλληλη χρήση και χειρισμό μεγάλων δεδομένων.
- Διερεύνηση των πλεονεκτημάτων των μεγάλων δεδομένων στην επιχείρηση και η ενσωμάτωσή τους με πρακτικές εσωτερικού και εξωτερικού ελέγχου για την παροχή ασφαλών και τεκμηριωμένων πληροφοριών για τη λήψη αποφάσεων.

Παραθέτοντας τα αποτελέσματα που προέκυψαν από αυτή την έρευνα, μπορεί να επιτευχθεί μια πιο ολοκληρωμένη κατανόηση της ελληνικής επιχειρηματικότητας σχετικά με τα Μεγάλα Δεδομένα.

3.2 Επιλογή δείγματος

Κατά τη διαδικασία επιλογής και συγκέντρωσης του δείγματος για την παρούσα έρευνα, ελήφθησαν υπόψη συγκεκριμένες παράμετροι για τη δημιουργία μιας ολοκληρωμένης λίστας επιχειρήσεων. Στόχος ήταν να διασφαλιστεί ότι το δείγμα ήταν όσο το δυνατόν πιο

αντιπροσωπευτικό, περιλαμβάνοντας περίπου 40 εταιρείες από διάφορους τομείς της βιομηχανίας και από διάφορες χρηματοοικονομικές κλίμακες.

Επιλέχθηκαν επιχειρήσεις που καλύπτουν μεγάλο εύρος σε διάφορους τομείς της τεχνολογίας Μεγάλων Δεδομένων, κυρίως από τους ακόλουθους κλάδους: α) Υγείας, β) Τραπεζών, γ) Τηλεπικοινωνιών, δ) Λιανικής, ε) Χρηματοοικονομικά, στ) Ενέργεια, κα.

3.3 Μεθοδολογία

Για την εκπλήρωση των στόχων της έρευνας που διεξήχθη, επιλέχθηκε μια ποσοτική προσέγγιση, χρησιμοποιώντας ένα ερωτηματολόγιο 23 ερωτήσεων ως κύριο εργαλείο συλλογής δεδομένων. Το ερωτηματολόγιο χωρίστηκε σε πέντε μέρη.

Το πρώτο μέρος περιλαμβάνει γενικές πληροφορίες σχετικά με τους ερωτηθέντες στην έρευνα, συμπεριλαμβανομένου του φύλου, της ηλικίας, του επιπέδου εκπαίδευσης και του πεδίου δραστηριότητας των αντίστοιχων επιχειρήσεων.

Το δεύτερο μέρος περιλαμβάνει πληροφορίες σχετικά με τον αριθμό των εργαζομένων και τη διάρκεια των εργασιών της επιχείρησης, προσφέροντας πληροφορίες για το μέγεθος και την εμπειρία των επιχειρήσεων που συμμετείχαν στην έρευνα.

Το τρίτο μέρος του ερωτηματολογίου αποτελούνταν από τέσσερις ερωτήσεις σχετικά με τον τρόπο με τον οποίο αποκτούν μεγάλα δεδομένα, το κυριότερο λόγο των δεδομένων που επεξεργάζονται περαιτέρω και τα εργαλεία που χρησιμοποιούνται για τη συλλογή και την ανάλυση των δεδομένων.

Το τέταρτο μέρος του ερωτηματολογίου αποτελούνταν από επτά ερωτήσεις σχετικά με την αποτελεσματικότητα και την αποδοτικότητα της επιχείρησης και τη σχέση της με την αξιοποίηση μεγάλων δεδομένων. Αυτές οι ερωτήσεις ακολουθούσαν μια κλίμακα Likert πέντε βαθμίδων, επιτρέποντας στους ερωτηθέντες να υποδείξουν το επίπεδο συμφωνίας ή διαφωνίας τους με μια σειρά προτάσεων που σχετίζονται με το ερευνητικό θέμα.

Το πέμπτο μέρος επικεντρώθηκε στη συνολική ποιότητα και συμβολή των υπηρεσιών των Μεγάλων Δεδομένων στον εσωτερικό έλεγχο της επιχείρησης, με τους ερωτηθέντες να βαθμολογούν το επίπεδο ικανοποίησής τους χρησιμοποιώντας μια κλίμακα Likert έξι ερωτήσεων.

3.4 Στατιστική ανάλυση

Η ανάλυση των δεδομένων διεξήχθη χρησιμοποιώντας IBM SPSS Statistics, έκδοση 23 (IBM Corporation, 2015). Για την επίτευξη των στόχων της μελέτης, πραγματοποιήθηκαν περιγραφικές αναλύσεις όπως συχνότητες, μέσος όρος και τυπική απόκλιση. Η στατιστική σημασία ορίστηκε ως $p < 0,05$.

Κεφάλαιο 4 Αποτελέσματα – Συζήτηση

4.1 Δημογραφικά χαρακτηριστικά

Σε ένα σύνολο 40 συμμετεχόντων, οι άνδρες ήταν το 62.5% του συνόλου (Πίνακας 3) και το ηλικιακό τους εύρος κυμαίνονταν μεταξύ 25-57 έτη του περισσότερους να ανήκουν στην κατηγορία 25-35 (47,5%), ακολουθούμενοι από εκείνους του ηλικιακού εύρους 36-46 (40%) (Πίνακας 4). Το επίπεδο της εκπαίδευσης που είχαν λάβει οι περισσότεροι συμμετέχοντες ήταν πανεπιστημιακή εκπαίδευση ΑΕΙ/ΑΤΕΙ (45%), ακολουθούμενοι από εκείνους που είχαν απολυτήριο λυκείου (32.5%) (Πίνακας 5). Επίσης, οι περισσότεροι συμμετέχοντες δήλωσαν ότι ο τομέας δραστηριοποίησης της επιχείρησής τους ήταν κάποια άλλη δραστηριότητα εκτός των αναφερόμενων (30%), ακολουθούμενοι από εκείνους που δραστηριοποιούνται στον τομέας της Βιομηχανίας (27.5%), των Χρηματοοικονομικών, (22.5%), των Τηλεπικοινωνιών (10%), ενώ μόλις το 5% και 2,5% του συνόλου δραστηριοποιείται στον Κατασκευαστικό κλάδο, στην Ενέργεια και στον Αγροτικό Τομέα, αντίστοιχα (Πίνακας 6).

Πίνακας 3 Δημογραφικά χαρακτηριστικά: Φύλο

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Άνδρας	25	62.5	62.5	62.5
Γυναίκα	15	37.5	37.5	100.0
Total	40	100.0	100.0	

Πίνακας 4 Δημογραφικά χαρακτηριστικά: Ηλικία

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 25-35	19	47.5	47.5	47.5
36-46	16	40.0	40.0	87.5
47-57	5	12.5	12.5	100.0
Total	40	100.0	100.0	

Πίνακας 5 Δημογραφικά χαρακτηριστικά: Επίπεδο σπουδών

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Απολυτήριο λυκείου	13	32.5	32.5	32.5
ΑΕΙ/ΑΤΕΙ	18	45.0	45.0	77.5
Μεταπτυχιακός τίτλος	7	17.5	17.5	95.0
Διδακτορικός τίτλος	2	5.0	5.0	100.0
Total	40	100.0	100.0	

Πίνακας 6 Δημογραφικά χαρακτηριστικά: Τομέας δραστηριοποίησης της εταιρείας

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Αγροτικός	1	2.5	2.5	2.5
Ενέργεια	1	2.5	2.5	5.0
Κατασκευαστικός	2	5.0	5.0	10.0
Τηλεπικοινωνίες	4	10.0	10.0	20.0
Χρηματοοικονομικά	9	22.5	22.5	42.5
Βιομηχανία	11	27.5	27.5	70.0
Άλλο	12	30.0	30.0	100.0
Total	40	100.0	100.0	

ΔΡΑΣΤΗΡΙΟΤΗΤΑ ΤΗΣ ΕΠΙΧΕΙΡΗΣΗΣ

Σχετικά με την επιχείρηση, όπως φαίνεται στον πίνακα 5, είναι προφανές ότι η πλειοψηφία των ερωτηθέντων, που αντιστοιχεί στο 42,5%, αντιπροσωπεύει εταιρείες με 11-50 εργαζόμενους, ακολουθούμενη από εκείνους που αντιπροσωπεύουν επιχειρήσεις έως 10 εργαζόμενους; (37,5%), ενώ μόνο 8 εταιρείες διαθέτουν περισσότερους από 51 εργαζόμενους (Πίνακας 7). Όσον αφορά την επιχειρηματική δραστηριότητα, η πλειονότητα των ερωτηθέντων απασχολούνται σε επιχειρήσεις που λειτουργούν εδώ και 5-10 χρόνια (57,5), ακολουθούμενοι από εκείνους που εργάζονται σε επιχειρήσεις με ιστορία 11-50 ετών (42,5%). Ένας θεμελιώδης στόχος της εργασίας

ήταν να συγκεντρωθούν απαντήσεις από υψηλά καταρτισμένο και έμπειρο προσωπικό με αποδεδειγμένο ιστορικό (Πίνακας 8).

Πίνακας 7 Αριθμός των εργαζομένων στην εταιρεία

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Έως 10 άτομα	15	37.5	37.5	37.5
11-50 άτομα	17	42.5	42.5	80.0
51-100 άτομα	3	7.5	7.5	87.5
101-200 άτομα	3	7.5	7.5	95.0
201 και πάνω άτομα	2	5.0	5.0	100.0
Total	40	100.0	100.0	

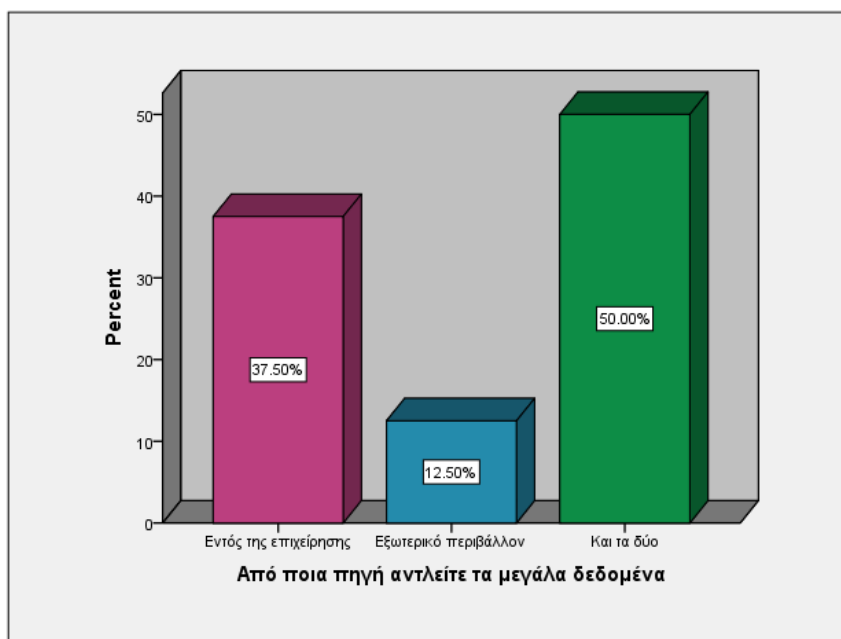
Πίνακας 8 Έτη δραστηριοποίησης της επιχείρησης

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 5-10 έτη	23	57.5	57.5	57.5
11-50 έτη	17	42.5	42.5	100.0
Total	40	100.0	100.0	

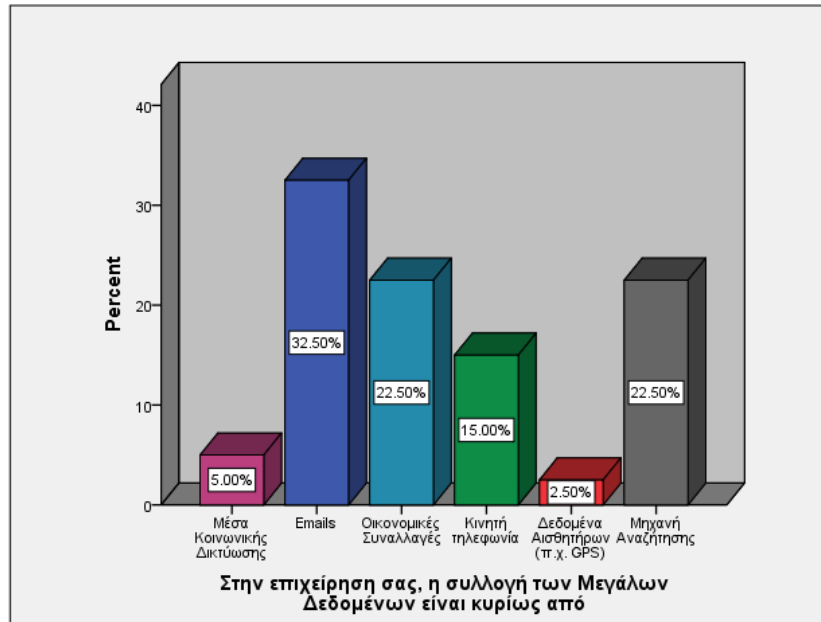
Πηγή άντλησης Μεγάλων Δεδομένων

Όσον αφορά την άντληση των μεγάλων δεδομένων, είναι προφανές ότι το 50% των ερωτηθέντων λαμβάνει τα δεδομένα τόσο από εσωτερικούς πόρους όσο και από το εξωτερικό περιβάλλον, ενώ μόνο το 12,5% τα αντλεί από το εξωτερικό περιβάλλον του (Σχήμα 2). Σχετικά με τη συλλογή των Μεγάλων Δεδομένων, οι περισσότερες ερωτηθέντες δήλωσαν ότι γίνεται μέσω Emails (32.5%), ακολουθούμενοι από εκείνους που δήλωσαν τις οικονομικές συναλλαγές και τη μηχανή αναζήτησης (22,5%) και την κινητή τηλεφωνία (15%), ως βασικές πηγές συλλογής Μεγάλων Δεδομένων (Σχήμα 3). Σχετικά με τον κύριο λόγο της χρήσης των Μεγάλων Δεδομένων, η πλειοψηφία των συμμετεχόντων θεώρησε την αύξηση των πωλήσεων (40%) ως πιο σημαντική για

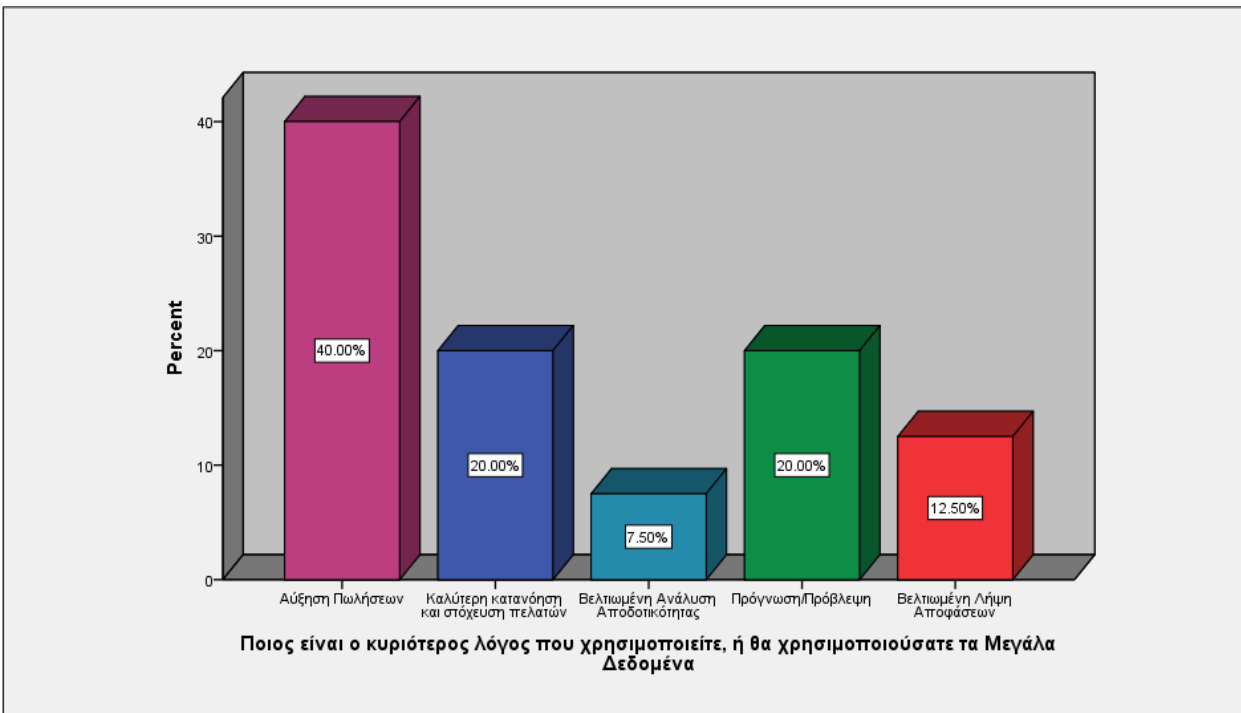
την επιχείρηση, ακολουθούμενη από την καλύτερη κατανόηση και στόχευση των πελατών (20%), την πρόγνωση/πρόβλεψη (20%), τη βελτιωμένη λήψη αποφάσεων (12,5%) και τη βελτιωμένη ανάλυση αποδοτικότητας (7,5%) (Σχήμα 4). Όσον αφορά τα εργαλεία για τη συλλογή των μεγάλων δεδομένων, οι περισσότεροι ερωτηθέντες χρησιμοποιούν το εργαλείο R (35%) στην επιχείρησή τους, ακολουθούμενο από το εργαλείο Apache Hadoop (32,5%) και Apache Sparke (22,5%), ενώ μόλις το 5% χρησιμοποιεί το MongoDB και το 2,5% το RapidMiner και Orange, αντίστοιχα (Σχήμα 5) .



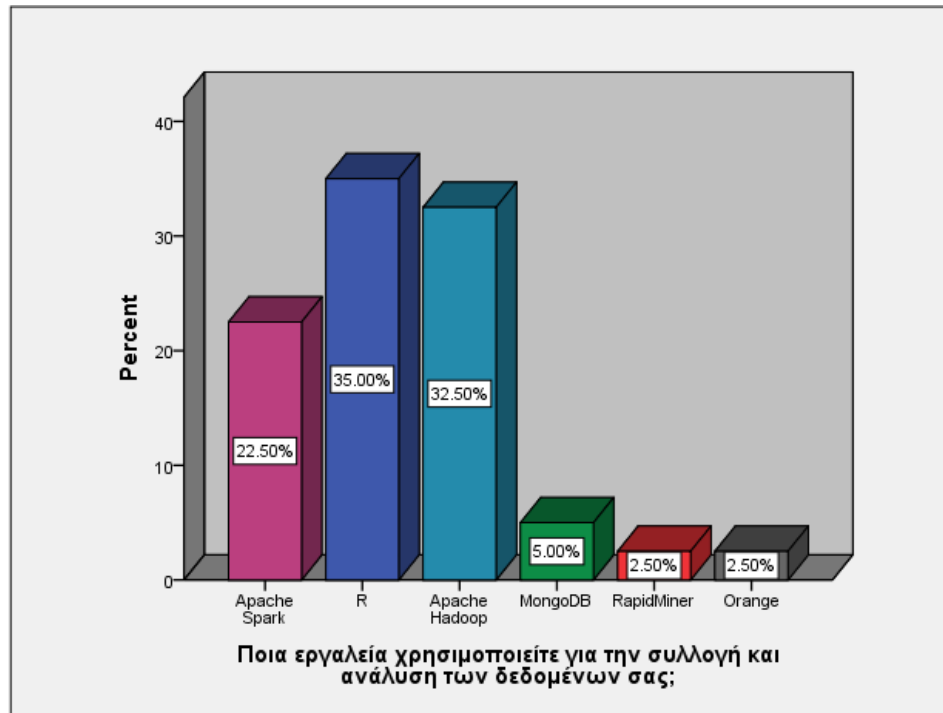
Σχήμα 2 Πηγή άντληση των μεγάλων δεδομένων



Σχήμα 3 Συλλογή των μεγάλων δεδομένων



Σχήμα 4 Κυριότερος λόγος χρήσης των μεγάλων δεδομένων



Σχήμα 5 Εργαλεία για τη συλλογή των μεγάλων δεδομένων

ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΑ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΤΗΝ ΑΠΟΔΟΤΙΚΟΤΗΤΑ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

Όσον αφορά το επόμενο μέρος των ερωτήσεων οι οποίες αφορούν την αποτελεσματικότητα των μεγάλων δεδομένων στην επιχείρηση, η πλειοψηφία των ερωτηθέντων (40%) δήλωσε ότι η εταιρεία τους ακολουθεί συγκεκριμένη στρατηγική ανάλυσης και επεξεργασίας των δεδομένων που λαμβάνονται. Με βάση αυτά τα ευρήματα, παρατηρείται ότι οι περισσότερες εταιρείες είναι εξοικειωμένες με την έννοια και τις μεθόδους ανάλυσης για την ανάκτηση μεγάλων δεδομένων (Πίνακας 9).

Όσον αφορά την ύπαρξη επιχειρηματικών ευκαιριών που προκύπτουν από την ανάλυση μεγάλων δεδομένων, το 57,5% των ερωτηθέντων αναγνωρίζει τέτοιες ευκαιρίες, ενώ το 17,5% ούτε συμφωνεί ούτε διαφωνεί. Αξίζει να τονιστεί ότι τα στελέχη επιχειρήσεων συχνά αμφισβητούν την απουσία εξειδικευμένων μεταπτυχιακών προγραμμάτων για την ανάπτυξη τεχνογνωσίας σε αυτή

την κρίσιμη πτυχή της σύγχρονης επιχείρησης. Σχετικά με το διαθέσιμο όγκο των μεγάλων δεδομένων, οι περισσότερες απαντήσεις (47,5%) καταλήγουν σε ένα συγκεκριμένο συμπέρασμα ότι ο όγκος δεδομένων είναι επαρκής για τη λήψη κατάλληλων αποφάσεων για την επιχείρηση. Όσον αφορά τον εντοπισμό οικονομικής απάτης μέσω μεγάλων δεδομένων, το 55% των ερωτηθέντων εκφράζει ουδετερότητα ως προς αυτή την άποψη, ενώ το 25% συμφωνεί ότι έχουν εντοπιστεί πολυάριθμες περιπτώσεις χρηματοοικονομικής απάτης χρησιμοποιώντας μεγάλα δεδομένα (Πίνακας 9).

Σχετικά με τη δυνατότητα πρόσβασης περισσότερων εργαζομένων σε μεγάλα δεδομένα, μια σημαντική πλειοψηφία (45%) των εργαζομένων συμφωνεί ότι οι επιχειρήσεις θα πρέπει να παρέχουν μεγαλύτερη πρόσβαση για να επιτρέψουν σε περισσότερους υπαλλήλους να συμμετέχουν στην επεξεργασία και ανάλυση μεγάλων δεδομένων, ενώ ένα εξίσου υψηλό ποσοστό (42,5%) των ερωτηθέντων ούτε συμφωνεί ούτε διαφωνεί με αυτήν την άποψη

Όσον αφορά την εμπιστοσύνη των πελατών στα προσωπικά τους δεδομένα, το 42,5% των ερωτηθέντων εκφράζει ουδετερότητα σε αυτό το θέμα, ενώ το 32,5% συμφωνεί ότι οι πελάτες εμπιστεύονται τα προσωπικά τους δεδομένα. 7 άτομα (17,5%) πιστεύουν ότι οι πελάτες δεν εμπιστεύονται τα προσωπικά τους δεδομένα.

Όπως φαίνεται στον πίνακα 13, το 52,5% των ερωτηθέντων συμφωνεί ότι πρέπει να υπάρχει σαφήνεια σχετικά με τις πτυχές της πνευματικής ιδιοκτησίας που σχετίζονται με τα μεγάλα δεδομένα. Αντίθετα, το 20% είναι ουδέτερο για το εάν τέτοιες διευκρινίσεις είναι απαραίτητες (Πίνακας 9).

Πίνακας 9 Μεγάλα δεδομένα και αποδοτικότητα στην επιχείρηση

Συχνότητα (Ποσοστό%)	1	2	3	4	5	Μέση τιμή	Τυπική απόκλιση
Η επιχείρησή σας έχει συγκεκριμένη στρατηγική για την ανάλυση των μεγάλων δεδομένων.	11(27,5)	16(40)	13(32,5)	-	-	2,05	0,78
Πιστεύετε ότι υπάρχουν νέες ευκαιρίες που βλέπετε για τον οργανισμό σας ως αποτέλεσμα της διαθεσιμότητας αυξημένου όγκου δεδομένων.	10(25)	25(57,5)	7(17,5)			1,93	0,66
Πιστεύετε ότι ο διαθέσιμος όγκος εσωτερικών και εξωτερικών μεγάλων δεδομένων επαρκεί για να λάβετε τις κατάλληλες αποφάσεις.	9(22,5)	19(47,5)	11(27,5)	1(2,5)		2,10	0,80
Έχετε εντοπίσει περιπτώσεις οικονομικής απάτης με την χρήση των μεγάλων δεδομένων που σε άλλες περιπτώσεις δεν ήταν διακριτές	4(10)	10(25)	22(55)	1(10)		2,65	0,80
Συμφωνείτε ότι οι επιχειρήσεις πρέπει να δώσουν τη δυνατότητα σε περισσότερους υπαλλήλους να έχουν πρόσβαση σε μεγάλα δεδομένα ώστε να μπορούν να λαμβάνουν επιχειρηματικές αποφάσεις με επίγνωση.	4(10)	18(45)	17(42,5)	1(2,5)		2,38	0,71
Οι πελάτες σας, σας εμπιστεύονται τα προσωπικά τους δεδομένα	3(7,5)	13(32,5)	17(42,5)	7(17,5)		2,70	0,85
Θεωρείτε απαραίτητη την αποσαφήνιση των πτυχών που αφορούν την πνευματική ιδιοκτησία της καινοτομίας που	11(27,5)	21(52,5)	8(20)			1,93	0,69

**βασίζεται στα δεδομένα (π.χ. όσον αφορά τεχνολογίες όπως η
εξόρυξη κειμένου και δεδομένων).**

ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΑ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΤΗ ΣΥΜΒΟΛΗ ΣΤΟΝ ΤΟΜΕΑ ΤΟΥ ΕΣΩΤΕΡΙΚΟΥ ΕΛΕΓΧΟΥ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

Όσον αφορά το ερώτημα εάν η εταιρεία επηρεάζεται από ευρωπαϊκούς και ενδοεπιχειρησιακούς ρυθμιστικούς περιορισμούς, τα αποτελέσματα παραμένουν ασαφή καθώς η πλειονότητα των ερωτηθέντων (45%) απέφυγε να δώσει σαφή θετική ή αρνητική απάντηση. Ωστόσο, 13 άτομα πιστεύουν ότι η εταιρεία στην οποία εργάζονται επηρεάζεται από διάφορους κανονισμούς, ενώ 9 άτομα πιστεύουν ότι το ρυθμιστικό πλαίσιο δεν επηρεάζει τον τρόπο με τον οποίο η εταιρεία διαχειρίζεται τα μεγάλα δεδομένα (Πίνακας 10).

Οι απαντήσεις που ελήφθησαν για τον οικονομικό έλεγχο της εταιρίας δείχνουν ότι η πλειονότητα, συγκεκριμένα το 47,5%, πιστεύει ότι η εταιρεία μπορεί να επωφεληθεί από τη χρήση μεγάλων δεδομένων για την άσκηση κατάλληλου ελέγχου. Αντίθετα, μόνο το 17,5% των ερωτηθέντων ήταν ουδέτερο σχετικά με το επιχειρηματικό όφελος από την ανάλυση και τη χρήση μεγάλων δεδομένων.

Οι απαντήσεις που ελήφθησαν στη χρήση μεγάλων δεδομένων για τον εντοπισμό οικονομικών λαθών αποκαλύπτουν ότι το μεγαλύτερο ποσοστό, συγκεκριμένα το 52.5%, πιστεύει ότι η εταιρεία μπορεί να χρησιμοποιήσει μεγάλα δεδομένα για να εντοπίσει οικονομικά λάθη. Σύμφωνα με τον πίνακα 23, διαπιστώνεται η πλειοψηφία των συμμετεχόντων υποστηρίζει ότι τα μεγάλα δεδομένα μπορούν να βοηθήσουν στην ανάλυση των λογιστικών καταστάσεων. Οι Barnaghi et al. (2013) δηλώνουν ότι τα μεγάλα δεδομένα θα μπορούσαν να έχουν σημαντικό αντίκτυπο στο μέλλον της χρηματοοικονομικής πληροφόρησης, ιδιαίτερα σε τομείς όπως η αναφορά περιουσιακών στοιχείων εκτός ισολογισμού και η λογιστική εύλογης αξίας. Οι Coyne et al. (2018) αναλύουν τον ρόλο των λογιστών στην εποχή των μεγάλων δεδομένων και προτείνουν ότι οι λογιστές θα πρέπει να διαδραματίσουν κρίσιμο ρόλο στη διαχείριση πληροφοριών που προέρχονται από μεγάλα δεδομένα, καθώς διαθέτουν την ικανότητα να εντοπίζουν τις ανάγκες πληροφοριών και ελέγχου των εσωτερικών και εξωτερικών υπευθύνων λήψης αποφάσεων (Πίνακας 10).

Με βάση τον πίνακα 26, είναι εμφανές ότι η πλειοψηφία και συγκεκριμένα το 55% υποστηρίζει τα πλεονεκτήματα που μπορεί να αποκομίσει η εταιρεία από τη χρήση μεγάλων δεδομένων για εσωτερικό έλεγχο. Αντίθετα, το 35% πιστεύει ότι ο εσωτερικός έλεγχος δεν μπορεί ούτε μπορεί να βελτιωθεί ούτε να επιδεινωθεί μέσω της χρήσης μεγάλων δεδομένων. Οι Watson et al. (2014) προτείνουν ότι οι τεχνολογικές εξελίξεις, συμπεριλαμβανομένης της χρήσης μεγάλων δεδομένων,

μπορούν να χρησιμοποιηθούν για την ενίσχυση του ελέγχου, την αύξηση της αποτελεσματικότητας και την παροχή μεγαλύτερης ασφάλειας.

Με βάση τον πίνακα 24, είναι προφανές ότι η πλειοψηφία, συγκεκριμένα το 50%, πιστεύει στη θετική συμβολή των μεγάλων δεδομένων στη διαμόρφωση στόχων και πολιτικής και στη λειτουργικότητα της επιχείρησης στα επόμενα 5 έτη (Πίνακας 10).

Πίνακας 10 Μεγάλα δεδομένα και συμβολή στον τομέα του εσωτερικού ελέγχου στην επιχείρηση

Συχνότητα (Ποσοστό%)	1	2	3	4	5	Μέση τιμή	Τυπική απόκλιση
Η επιχείρησή σας επηρεάζεται από ευρωπαϊκούς κανονιστικούς περιορισμούς (π.χ. GDPR) ή και ενδοεπιχειρησιακούς κανονιστικούς περιορισμούς στις απαιτήσεις εντοπισμού δεδομένων	2(5)	7(17,5)	18(45)	10(25)	3(7,5)	3,12	0,96
Πιστεύετε ότι η χρήση των μεγάλων δεδομένων μπορεί να βοηθήσει στο σωστό οικονομικό έλεγχο της εταιρίας			7(17,5)	19(47,5)	14(35)	4,18	0,71
Θεωρείτε ότι μπορούν να εντοπιστούν οικονομικά λάθη με τη χρήση των μεγάλων δεδομένων			9(22,5)	21(52,5)	10(25)	4,02	0,69
Θεωρείτε ότι η εξόρυξη & ανάλυση των μεγάλων δεδομένων μπορούσε να βοηθήσει τις επιχειρήσεις ως προς την ανάλυση των λογιστικών καταστάσεων			7(17,5)	25(62,5)	8(20)	4,02	0,62
Η επιχείρησή σας χρησιμοποιεί τα μεγάλα δεδομένα για επεξεργασία και προς όφελος του Εσωτερικού ελέγχου			14(35)	22(55)	4(10)	3,75	0,63
Σε 5 χρόνια, πιστεύετε ότι θα είναι η χρήση των Μεγάλων Δεδομένων σημαντική στη λειτουργικότητα της επιχείρησή σας			6(15)	14(35)	20(50)	4,35	0,74

Συμπεράσματα

Οι συνεχείς εξελίξεις στην τεχνολογία, ιδιαίτερα στη διαχείριση μεγάλου όγκου δεδομένων, έχουν μετατρέψει τα Μεγάλα Δεδομένα σε σημαντική τεχνολογική πρόκληση για πολλές εταιρείες και οργανισμούς. Στόχος της παρούσας διπλωματικής εργασίας ήταν η διερεύνηση της τεχνολογίας των Μεγάλων Δεδομένων, των εργαλείων που χρησιμοποιούνται για την ανάλυση και αποθήκευση σημαντικών ποσοτήτων δεδομένων, καθώς και των ρόλων και των ευκαιριών που παρουσιάζει για τις ελληνικές επιχειρήσεις.

Η έρευνα που διεξήχθη σε αυτή την εργασία έδωσε σημαντικά ευρήματα σχετικά με τα Μεγάλα Δεδομένα. Πρώτον, θα πρέπει να σημειωθεί ότι η πλειονότητα των εταιρειών που συμμετείχαν στην έρευνα χρησιμοποιεί Μεγάλα Δεδομένα, με πολλές από αυτές να διαθέτουν σημαντική εμπειρία σε αυτήν την τεχνολογία. Κατά την εξέταση του κλάδου και του μεγέθους των επιχειρήσεων που συμμετείχαν στην έρευνα, παρατηρήθηκε ότι ένα ευρύ φάσμα βιομηχανιών χρησιμοποιεί εκτενώς τα Μεγάλα Δεδομένα, υποδεικνύοντας την ευρεία επίδραση αυτής της τεχνολογίας σε διάφορους τομείς. Ομοίως, όσον αφορά το μέγεθος της εταιρείας, το υψηλότερο ποσοστό απαντήσεων προήλθε από εταιρείες με 11-50 υπαλλήλους και εταιρείες με έως και 10 υπαλλήλους. Αυτό υποδηλώνει ότι η τεχνολογία Big Data υιοθετείται πιο συχνά από μεγαλύτερες εταιρείες με ισχυρότερους οικονομικούς πόρους και τεχνογνωσία σε σύγκριση με μικρότερες επιχειρήσεις.

Ένα αξιοσημείωτο εύρημα ήταν ότι τα Μεγάλα Δεδομένα αντλούνται από το εσωτερικό της επιχείρησης αλλά και από το εξωτερικό της περιβάλλον. Επιπλέον, οι περισσότερες εταιρείες είτε χρησιμοποιούν επί του παρόντος είτε σχεδιάζουν να χρησιμοποιήσουν τα Μεγάλα Δεδομένα στο μέλλον για την αύξηση των πωλήσεων, την καλύτερη κατανόηση και στόχευση πελατών, την πρόβλεψη/πρόγνωση καθώς και για τη βελτίωση των διαδικασιών λήψης αποφάσεων.

Οι κύριες πηγές από τις οποίες οι εταιρείες συλλέγουν τα δεδομένα τους είναι τα μηνύματα ηλεκτρονικού ταχυδρομείου, οι οικονομικές συναλλαγές και οι μηχανές αναζήτησης (π.χ. Google). Τα κύρια εργαλεία που χρησιμοποιούνται από αυτές τις εταιρείες για τη συλλογή και ανάλυση των δεδομένων τους περιλαμβάνουν τα R, Apache Hadoop και Apache Spark. Η πλειονότητα των εταιρειών συμφώνησε ότι έχει συγκεκριμένη στρατηγική αξιοποίησης των Μεγάλων Δεδομένων, υποδεικνύοντας ότι συνέβαλαν σημαντικά στην επίτευξη των στόχων τους,

αν και είναι δυνατή η περαιτέρω αξιοποίησή τους. Τέλος, οι περισσότερες εταιρείες που συμμετείχαν στην έρευνα και έχουν στρατηγική Μεγάλων Δεδομένων δήλωσαν ότι ενισχύουν σημαντικά το ανταγωνιστικό τους πλεονέκτημα, τον οικονομικό έλεγχο και τον εντοπισμό οικονομικών λαθών μέσω της εφαρμογής τέτοιων στρατηγικών. Αυτά όλα καταδεικνύουν την αποτελεσματικότητα των Μεγάλων Δεδομένων στις ελληνικές επιχειρήσεις, παρά τις δύσκολες συνθήκες της ελληνικής οικονομίας.

Τα αποτελέσματα αυτής της έρευνας υπογραμμίζουν τη σημασία και τη δύναμη των Μεγάλων Δεδομένων στη σύγχρονη ελληνική επιχειρηματικότητα. Οι περισσότερες εταιρείες έχουν δηλώσει ότι σκοπεύουν να επενδύσουν σε αυτήν την τεχνολογία στο εγγύς μέλλον, με πολλούς να πιστεύουν ότι η χρήση της θα είναι κρίσιμη για τις επιχειρηματικές τους δραστηριότητες τα επόμενα πέντε χρόνια. Αυτό δείχνει ότι τα Μεγάλα Δεδομένα θα διαδραματίσουν σημαντικό ρόλο για τις επιχειρήσεις που στοχεύουν να αναπτυχθούν, να ευδοκιμήσουν και να καινοτομήσουν.

Αν και τα Μεγάλα Δεδομένα ήταν πάντα παρόντα στη ζωή μας, η ανακάλυψή τους και η εξερεύνηση τους έγιναν τα τελευταία χρόνια λόγω των τεχνολογικών προόδων. Η εμφάνισή τους έχει επιφέρει ριζικές αλλαγές τόσο στην κοινωνική όσο και στην επιχειρηματική λειτουργία. Τα δεδομένα που κάποτε θεωρούνταν μη διαχειρίσιμα αναγνωρίζονται πλέον ως εξαιρετικά πολύτιμα για την εξαγωγή ουσιαστικών πληροφοριών. Η συνεχής εξέλιξη της τεχνολογίας σε συνδυασμό με την επιστήμη των Μεγάλων Δεδομένων είχε βαθύ αντίκτυπο στις σημερινές επιχειρήσεις, δίνοντας τη δυνατότητα σε εταιρείες όλων των μεγεθών να λαμβάνουν βελτιωμένες αποφάσεις και να έχουν πρόσβαση σε πληροφορίες σε πραγματικό χρόνο σε όλες τις πτυχές των εργασιών τους. Είναι προφανές ότι τα Μεγάλα Δεδομένα προσφέρουν ένα ευρύ φάσμα δυνατοτήτων και εξαιρετικές προοπτικές για το μέλλον, δίνοντας τη δυνατότητα στις επιχειρήσεις να εξελίσσονται συνεχώς και να μεγιστοποιούν τα πιθανά οφέλη τους.

Βιβλιογραφία

- Abdelli, M. E. A., Youssef, W. A. B., Özgöker, U., & Slimene, I. B. (2021). *Big Data for Entrepreneurship and Sustainable Development*. CRC Press.
- Al-Sai, Z. A., & Abdullah, R. (2019). Big data impacts and challenges: a review. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 250–255.
- Al-Shiakhli, S. (2019). *Big data analytics: a literature review perspective*.
- Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 25. <https://doi.org/10.1186/s13174-015-0041-5>
- Alexandrov, A., Bergmann, R., Ewen, S., Freytag, J.-C., Hueske, F., Heise, A., Kao, O., Leich, M., Leser, U., Markl, V., Naumann, F., Peters, M., Rheinländer, A., Sax, M. J., Schelter, S., Höger, M., Tzoumas, K., & Warneke, D. (2014). The Stratosphere platform for big data analytics. *The VLDB Journal*, 23(6), 939–964. <https://doi.org/10.1007/s00778-014-0357-y>
- Assembly, G. (2015). *Sustainable Development Goals. SDGs Transform Our World, 2030*.
- Barnaghi, P., Sheth, A., & Henson, C. (2013). From Data to Actionable Knowledge: Big Data Challenges in the Web of Things [Guest Editors' Introduction]. *IEEE Intelligent Systems*, 28(6), 6–11. <https://doi.org/10.1109/MIS.2013.142>
- Berisha, B., Mëziu, E., & Shabani, I. (2022). Big data analytics in Cloud computing: an overview. *Journal of Cloud Computing*, 11(1), 24. <https://doi.org/10.1186/s13677-022-00301-w>
- Berti-Equille, L., & Ba, M. L. (2016). Veracity of Big Data. *Journal of Data and Information Quality*, 7(3), 1–3. <https://doi.org/10.1145/2935753>
- Borkar, V., Carey, M., Grover, R., Onose, N., & Vernica, R. (2011). Hyracks: A flexible and extensible foundation for data-intensive computing. In *ICDE* (pp. 1151–1162).
- Calic, G., & Ghasemaghaei, M. (2021). Big data for social benefits: Innovation as a mediator of the relationship between big data and corporate social performance. *Journal of Business Research*, 131, 391–401. <https://doi.org/10.1016/j.jbusres.2020.11.003>

- Carbone, P., Katsifodimos, A., Ewen, S., Markl, V., Haridi, S., & Tzoumas, K. (2015). Apache flink: Stream and batch processing in a single engine. *Bulletin of the IEEE Computer Society TCDE*, 36(4).
- Coyne, E. M., Coyne, J. G., & Walker, K. B. (2018). Big Data information governance by accountants. *International Journal of Accounting & Information Management*, 26(1), 153–170. <https://doi.org/10.1108/IJAIM-01-2017-0006>
- da Silva, T. L. C., Magalhães, R. P., Brilhante, I. R., de Macêdo, J. A., Araújo, D., Rego, P. A., & Neto, A. V. L. (2018). Big Data Analytics Technologies and Platforms: A Brief Review. *LADaS@ VLDB*, 25–32.
- Debattista, J., Lange, C., Scerri, S., & Auer, S. (2015). Linked “Big” Data: Towards a Manifold Increase in Big Data Value and Veracity. *2015 IEEE/ACM 2nd International Symposium on Big Data Computing (BDC)*, 92–98. <https://doi.org/10.1109/BDC.2015.34>
- Dong, X. L., & Srivastava, D. (2013). Big data integration. *2013 IEEE 29th International Conference on Data Engineering (ICDE)*, 1245–1248. <https://doi.org/10.1109/ICDE.2013.6544914>
- Dresner Advisory Services. (2017). *Big Data Analytics Market Study*.
- Eakin, H., & Luers, A. L. (2006). Assessing the vulnerability of social-environmental systems. *Annu. Rev. Environ. Resour.*, 31, 365–394.
- Elragal, A. (2014). *ERP and big data: the inept couple. s.l., Elsevier*.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Ghasemaghahi, M. (2021). Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics. *International Journal of Information Management*, 57, 102055. <https://doi.org/10.1016/j.ijinfomgt.2019.102055>
- Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating Strategic Business Value

- from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, 35(2), 388–423. <https://doi.org/10.1080/07421222.2018.1451951>
- Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3), 191–209. <https://doi.org/10.1016/j.jsis.2017.07.003>
- He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X., & Xu, Z. (2011). RCFile: A fast and space-efficient data placement structure in MapReduce-based warehouse systems. *2011 IEEE 27th International Conference on Data Engineering*, 1199–1208. <https://doi.org/10.1109/ICDE.2011.5767933>
- Hofmann, E., & Rutschmann, E. (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management*, 29(2), 739–766. <https://doi.org/10.1108/IJLM-04-2017-0088>
- Hugh J., W. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems*, 364–379. <https://doi.org/10.17705/1CAIS.04421>
- Husamaldin, L., & Saeed, N. (2019). Big Data Analytics Correlation Taxonomy. *Information*, 11(1), 17. <https://doi.org/10.3390/info11010017>
- Jabeen, H., Archer, P., Scerri, S., Verstedden, A., Ermilov, I., Mouchakis, G., Lehmann, J., & Auer, S. (2017). Big data europe. In *EDBT/ICDT Workshops*.
- Kepner, J., Gadepally, V., Michaleas, P., Schear, N., Varia, M., Yerukhimovich, A., & Cunningham, R. K. (2014). Computing on masked data: a high performance method for improving big data veracity. *2014 IEEE High Performance Extreme Computing Conference (HPEC)*, 1–6. <https://doi.org/10.1109/HPEC.2014.7040946>
- Khan, M. A., Uddin, M. F., & Gupta, N. (2014). Seven V's of Big Data understanding Big Data to extract value. *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*, 1–5. <https://doi.org/10.1109/ASEEZone1.2014.6820689>
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*.

- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*.
- 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.
- Nashua, N. (2017). *Big Data Analytics Market Study, s.l.: Dresner Advisory Services*.
- Nguyen, T. L. (2018). A Framework for Five Big V's of Big Data and Organizational Culture in Firms. *2018 IEEE International Conference on Big Data (Big Data)*, 5411–5413. <https://doi.org/10.1109/BigData.2018.8622377>
- Nicholls, A., & Murdock, A. (2011). *Social Innovation: Blurring Boundaries to Reconfigure Markets*.
- Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences*, 30(4), 431–448. <https://doi.org/10.1016/j.jksuci.2017.06.001>
- Phills, J. A., Deiglmeier, K., & Miller, D. T. (2014). Rediscovering Social Innovation. *Stanford Social Innovation Review*, 6(4), 34–53.
- Rana, S. (2019). Moving in the Realm of Big Data: Using Analytics in Management Research and Practices. *FIIB Business Review*, 8(1), 7–8. <https://doi.org/10.1177/2319714519839802>
- RANJAN, J. (2019). The 10 Vs of Big Data framework in the Context of 5 Industry Verticals. *PRODUCTIVITY*, 59(4), 324–342. <https://doi.org/10.32381/PROD.2019.59.04.2>
- Saeed, N., & Husamaldin, L. (2021). Big data characteristics (V's) in industry. *Iraqi Journal of Industrial Research*, 8(1), 1–9.
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47.
- Singh, D., & Reddy, C. K. (2015). A survey on platforms for big data analytics. *Journal of Big Data*, 2(1), 8. <https://doi.org/10.1186/s40537-014-0008-6>
- Soares, S. (2012). *Big data governance: an emerging imperative*. Mc Press.

- Sparks, E. R., Talwalkar, A., Smith, V., Kottalam, J., Pan, X., Gonzalez, J., Franklin, M. J., Jordan, M. I., & Kraska, T. (2013). *Mli: An api for distributed machine learning*. In *ICDM*.
- Tsujimoto, M., Kajikawa, Y., Tomita, J., & Matsumoto, Y. (2018). A review of the ecosystem concept — Towards coherent ecosystem design. *Technological Forecasting and Social Change*, *136*, 49–58. <https://doi.org/10.1016/j.techfore.2017.06.032>
- Vavilapalli, V. K., Murthy, A. C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., & Seth, S. (2013). Apache hadoop yarn: Yet another resource negotiator. *Proceedings of the 4th Symposium SOCC*.
- Watson, H. J. (2014). Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. *Communications of the Association for Information Systems*, *34*. <https://doi.org/10.17705/1CAIS.03465>
- Westley, F. R., McGowan, K. A., Antadze, N., Blacklock, J., & Tjornbo, O. (2016). How game changers catalyzed, disrupted, and incentivized social innovation: three historical cases of nature conservation, assimilation, and women's rights. *Ecology and Society*, *21*(4), art13. <https://doi.org/10.5751/ES-08811-210413>
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M. J., Ghodsi, A., Gonzalez, J., Shenker, S., & Stoica, I. (2016). Apache Spark. *Communications of the ACM*, *59*(11), 56–65. <https://doi.org/10.1145/2934664>
- Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*, *101*, 572–591. <https://doi.org/10.1016/j.cie.2016.07.013>

Παράρτημα

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Έρευνα σχετικά με την επίδραση των μεγάλων δεδομένων στην επιχειρηματικότητα

Αξιότιμοι κύριοι/κυρίες,

η έρευνα που διεξάγω πραγματοποιείται στα πλαίσια της μελέτης της επίδραση των μεγάλων δεδομένων στην επιχειρηματικότητα και την εξέλιξή της εκάστοτε επιχείρησης και σας προσκαλώ να συμμετέχετε σε αυτή.

Σκοπός της παρούσας έρευνας είναι να εξετάσει τις απόψεις των ατόμων με επιχειρήσεις και εταιρείες σχετικά με τη συνεισφορά και την επίδραση των μεγάλων δεδομένων στην επιχειρηματικότητα και την εξέλιξή της εκάστοτε επιχείρησης.

Παρακαλείσθε να συμπληρώσετε τις ερωτήσεις του ερωτηματολογίου που θα βρείτε συνημμένες με το έντυπο αυτό. Η διαδικασία θα διαρκέσει περίπου 15 λεπτά. Η έρευνα πραγματοποιείται απολύτως ανώνυμα και δεν πρόκειται να χρησιμοποιηθεί καμία προσωπική πληροφορία για άλλους σκοπούς, ενώ παράλληλα η διαδικασία αυτή έχει μηδενικό κίνδυνο για εσάς. Οι απαντήσεις σας θα βοηθήσουν στην παροχή χρήσιμων δεδομένων σχετικά με τη μελέτη της επίδρασης των μεγάλων δεδομένων στην επιχειρηματικότητα. Η συμμετοχή σας είναι εθελοντική και σε καμία περίπτωση δε θα επηρεάσει την επαγγελματική σας πορεία. Οι συμμετέχοντες είναι σε θέση να αποσυρθούν από τη μελέτη ανά πάσα στιγμή χωρίς να απαιτείται να δηλώσουν κάποιο λόγο.

Για οποιαδήποτε ερώτηση σχετικά με την έρευνα μπορείτε να επικοινωνήσετε μαζί μου μέσω email: bl.bigdata23@gmail.com. Η συμπλήρωση της έρευνας προϋποθέτει τη συναίνεση της συμμετοχής σας.

Η βοήθεια σας σε αυτή την έρευνα θα εκτιμηθεί ιδιαίτερα.

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Σας ενημερώνουμε ότι τα ερωτηματολόγια είναι ανώνυμα.

Παρακαλούμε πείτε μας λίγα λόγια για εσάς

ΔΗΜΟΓΡΑΦΙΚΑ ΣΤΟΙΧΕΙΑ

Οι ερωτήσεις αποσκοπούν στη συλλογή πληροφοριών σχετικά με τα δημογραφικά στοιχεία.

1. Φύλο

- 1) Άνδρας
- 2) Γυναίκα

2 Σε ποια ηλικιακή κατανομή βρίσκεστε;

- 1) Έως και 24
- 2) 25-35
- 3) 36-46
- 4) 47-57
- 5) 58 και άνω

3 Ποιο είναι το επίπεδο των σπουδών σας;

- 1) Απολυτήριο λυκείου
- 2) ΑΕΙ/ΑΤΕΙ
- 3) Μεταπτυχιακός τίτλος
- 4) Διδακτορικός τίτλος

4 Ποιος είναι ο Τομέας δραστηριοποίησης της εταιρείας;

- 1) Αγροτικός
- 2) Ενέργεια
- 3) Κατασκευαστικός
- 4) Τηλεπικοινωνίες
- 5) Χρηματοοικονομικά
- 6) Βιομηχανία
- 7) Άλλο

ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

Οι ερωτήσεις αποσκοπούν στη συλλογή πληροφοριών σχετικά με την εταιρεία.

5 Ποιος είναι ο αριθμός των εργαζομένων στην εταιρεία;

- 1) Έως 10 άτομα
- 2) 11-50 άτομα
- 3) 51-100 άτομα
- 4) 101-200 άτομα
- 5) 201 και πάνω άτομα

6 Πόσα έτη δραστηριοποιείται η επιχείρηση;

- 1) 5-10 έτη
- 2) 11-50 έτη
- 3) 51 και πάνω έτη

ΚΥΡΙΟ ΜΕΡΟΣ

Α' ΜΕΡΟΣ – ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΑ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

7 Από ποια πηγή αντλείτε τα μεγάλα δεδομένα;

- 1) Εντός της επιχείρησης
- 2) Εξωτερικό περιβάλλον
- 3) Και τα δύο

8 Στην επιχείρησή σας, η συλλογή των Μεγάλων Δεδομένων είναι κυρίως από

- 1) Μέσα Κοινωνικής Δικτύωσης
- 2) Emails
- 3) Οικονομικές Συναλλαγές
- 4) Κινητή τηλεφωνία
- 5) Δεδομένα Αισθητήρων(π.χ. GPS)
- 6) Μηχανή Αναζήτησης

9 Ποιος είναι ο κυριότερος λόγος που χρησιμοποιείτε, ή θα χρησιμοποιούσατε τα Μεγάλα Δεδομένα;

- 1) Αύξηση Πωλήσεων

- 2) Βελτιωμένη Εφοδιαστική Αλυσίδα
- 3) Καλύτερη κατανόηση και στόχευση πελατών
- 4) Βελτιωμένη Ανάλυση Αποδοτικότητας
- 5) Πρόγνωση/Πρόβλεψη
- 6) Βελτιωμένη Λήψη Αποφάσεων

10 Ποια εργαλεία χρησιμοποιείτε για την συλλογή και ανάλυση των δεδομένων σας;

- 1) Apache Spark
- 2) R
- 3) Apache Hadoop
- 4) MongoDB
- 5) RapidMiner
- 6) Orange

Β' ΜΕΡΟΣ – ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΑ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΤΗΝ ΑΠΟΔΟΤΙΚΟΤΗΤΑ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

11 Η επιχείρησή σας έχει συγκεκριμένη στρατηγική για την ανάλυση των μεγάλων δεδομένων;

Συμφωνώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφωνώ Απόλυτα
1	2	3	4	5

12 Πιστεύετε ότι υπάρχουν νέες ευκαιρίες που βλέπετε για τον οργανισμό σας ως αποτέλεσμα της διαθεσιμότητας αυξημένου όγκου δεδομένων;

Συμφωνώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφωνώ Απόλυτα
1	2	3	4	5

13 Πιστεύετε ότι ο διαθέσιμος όγκος εσωτερικών και εξωτερικών μεγάλων δεδομένων επαρκεί για να λάβετε τις κατάλληλες αποφάσεις;

Συμφωνώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφωνώ Απόλυτα
1	2	3	4	5

14 Έχετε εντοπίσει περιπτώσεις οικονομικής απάτης με την χρήση των μεγάλων δεδομένων που σε άλλες περιπτώσεις δεν ήταν διακριτές;

Συμφων ώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφων ώ Απόλυτα
1	2	3	4	5

15 Συμφωνείτε ότι οι επιχειρήσεις πρέπει να δώσουν τη δυνατότητα σε περισσότερους υπαλλήλους να έχουν πρόσβαση σε μεγάλα δεδομένα ώστε να μπορούν να λαμβάνουν επιχειρηματικές αποφάσεις με επίγνωση;

Συμφων ώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφων ώ Απόλυτα
1	2	3	4	5

16 Οι πελάτες σας, σας εμπιστεύονται τα προσωπικά τους δεδομένα;

Συμφων ώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφων ώ Απόλυτα
1	2	3	4	5

17 Θεωρείτε απαραίτητη την αποσαφήνιση των πτυχών που αφορούν την πνευματική ιδιοκτησία της καινοτομίας που βασίζεται στα δεδομένα (π.χ. όσον αφορά τεχνολογίες όπως η εξόρυξη κειμένου και δεδομένων);

Συμφων ώ Απόλυτα	Συμφωνώ	Είμαι ουδέτερος	Διαφωνώ	Διαφων ώ Απόλυτα
1	2	3	4	5

Γ' ΜΕΡΟΣ – ΕΡΩΤΗΣΕΙΣ ΣΧΕΤΙΚΑ ΜΕ ΤΑ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΚΑΙ ΤΗ ΣΥΜΒΟΛΗ ΣΤΟΝ ΤΟΜΕΑ ΤΟΥ ΕΣΩΤΕΡΙΚΟΥ ΕΛΕΓΧΟΥ ΣΤΗΝ ΕΠΙΧΕΙΡΗΣΗ

18 Η επιχείρησή σας επηρεάζεται από ευρωπαϊκούς κανονιστικούς περιορισμούς (π.χ. GDPR) ή και ενδοεπιχειρησιακούς κανονιστικούς περιορισμούς στις απαιτήσεις εντοπισμού δεδομένων;

Καθόλου	Λίγο	Ούτε ναι ούτε όχι	Αρκετά	Πολύ
1	2	3	4	5

19 Πιστεύετε ότι η χρήση των μεγάλων δεδομένων μπορεί να βοηθήσει στο σωστό οικονομικό έλεγχο της εταιρίας;

Καθόλου	Λίγο	Ούτε ναι ούτε όχι	Αρκετά	Πολύ
1	2	3	4	5

20 Θεωρείτε ότι μπορούν να εντοπιστούν οικονομικά λάθη με τη χρήση των μεγάλων δεδομένων;

Καθόλου	Λίγο	Ούτε ναι ούτε όχι	Αρκετά	Πολύ
1	2	3	4	5

21 Θεωρείτε ότι η εξόρυξη & ανάλυση των μεγάλων δεδομένων μπορούσε να βοηθήσει τις επιχειρήσεις ως προς την ανάλυση των λογιστικών καταστάσεων;

Καθόλου	Λίγο	Ούτε ναι ούτε όχι	Αρκετά	Πολύ
1	2	3	4	5

22 Η επιχείρησή σας χρησιμοποιεί τα μεγάλα δεδομένα για επεξεργασία και προς όφελος του Εσωτερικού ελέγχου;

Καθόλου	Λίγο	Ούτε ναι ούτε όχι	Αρκετά	Πολύ
1	2	3	4	5

23 Σε 5 χρόνια, πιστεύετε ότι θα είναι η χρήση των Μεγάλων Δεδομένων σημαντική στη λειτουργικότητα της επιχείρησή σας;

Καθόλου	Λίγο	Ούτε ναι ούτε όχι	Αρκετά	Πολύ
1	2	3	4	5

Σας ευχαριστώ πολύ για τη συμμετοχή σας!!