



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ**
UNIVERSITY OF PATRAS

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

Big Data και Ανάπτυξη Επιχειρηματικότητας

Πτυχιακή Εργασία

των

ΛΑΔΑΣ ΜΑΡΙΟΣ

ΜΑΥΡΟΥΛΙΑΣ ΚΑΝΕΛΛΟΣ ΠΑΝΑΓΙΩΤΗΣ

Επιβλέπων: Γιωτόπουλος Κωνσταντίνος, Αναπληρωτής Καθηγητής

ΠΑΤΡΑ, 2021

.....
Λαδάς Μάριος & Μαυρούλιας Κανέλλος Παναγιώτης
Απόφοιτοι του τμήματος Διοικητικής Επιστήμης & Τεχνολογίας, Πανεπιστήμιο Πατρών

© Copyright συγγραφής Ladas Marios & Mavroulias Kanellos Panagiotis , 2020.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας του Πανεπιστημίου Πατρών δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

Ευχαριστίες

Με την ολοκλήρωση της εν λόγω πτυχιακής εργασίας, θα θέλαμε να ευχαριστήσουμε τον καθηγητή μας Γιωτόπουλο Κωνσταντίνο, ο οποίος μας βοήθησε με τις καίριες κατευθύνσεις και τις εύστοχες υποδείξεις του, καθ' όλη τη διάρκεια εκπόνησης αυτής. Επίσης, θα θέλαμε να ευχαριστήσουμε ο καθένας από εμάς ξεχωριστά τις οικογένειες μας, για την ενθάρρυνση και την στήριξη τους κατά τη διάρκεια εκπόνησης της παρούσας πτυχιακής εργασίας. Δεν μπορούμε να παραλείψουμε να ευχαριστήσουμε και τις επιχειρήσεις για τις απαντήσεις αλλά και την καθοδήγηση τους στο ερωτηματολόγιο μας.

Περίληψη

Η παρούσα πτυχιακή εργασία εκπονήθηκε για την ολοκλήρωση της φοίτησης μας στο τμήμα Διοικητικής Επιστήμης και Τεχνολογίας της σχολής Οικονομικών Επιστημών και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Πατρών, πρώην τμήμα Διοίκησης Επιχειρήσεων του Τ.Ε.Ι Πατρών. Η εκκίνηση της πτυχιακής εργασίας τοποθετείται χρονικά τον Μάιο του 2020.

Το αντικείμενο της παρούσας πτυχιακής εργασίας αφορά στη μελέτη των Μεγάλων Δεδομένων και των τεχνολογιών τους καθώς και πως μπορούν να εφαρμοστούν στην σημερινή ελληνική επιχειρηματικότητα για την καλύτερη ανάπτυξη των επιχειρήσεων σε όλους τους τομείς της. Κρίθηκε ακόμη σκόπιμο, να διεξαχθεί μία έρευνα πάνω στα όσα μελετήσαμε στα πρώτα κεφάλαια και, αν και πως οι επιχειρήσεις χρησιμοποιούν κατάλληλα αυτήν την τεχνολογία, με δείγμα τις ελληνικές επιχειρήσεις και οργανώσεις, ώστε να αποτυπώσουμε τί ισχύει για την ελληνική οικονομία. Τέλος, παραθέσαμε τα αποτελέσματα που λάβαμε και προχωρήσαμε στην εξαγωγή των συμπερασμάτων από αυτά.

Abstract

The present thesis was done for the completion of our studies in the Department of Management Science and Technology of the School of Economics and Business Administration of the University of Patras, former Department of Business Administration of the Technological Educational Institute of Patras. The start of the thesis is scheduled for May 2020.

The subject of this thesis concerns the study of Big Data and their technologies as well as how they can be applied in today's Greek entrepreneurship for the better development of the companies in all their sectors. It was also considered appropriate to conduct a research on what we studied on the first chapters and, although how these companies use this technology properly, with a sample of Greek companies and organizations, in order to see what applies to the Greek economy. At the end, we presented the results we received and proceeded to draw conclusions from them.

Περιεχόμενα

Ευχαριστίες.....	3
Περίληψη	4
Abstract.....	5
1. Εισαγωγή	9
2. Μεγάλα Δεδομένα(Big Data).....	11
2.1 Τι είναι τα Big Data;	11
2.2 Χαρακτηριστικά Big Data.....	13
2.3 Τύποι των Big Data.....	15
2.4 Κύκλος ζωής των Big Data	16
2.5 Οφέλη των Big Data.....	21
2.6 Προκλήσεις των Big Data	21
2.7 Πηγές άντλησης των Big Data	23
2.8 Εφαρμογές των Big Data	25
2.9 Νομικά ζητήματα των Big Data.....	27
3. Τεχνολογίες των Big Data	29
3.1 Data Analytics	29
3.1.1 Apache Spark.....	30
3.1.2 R.....	31
3.2 Data Storage	31
3.2.1 Apache Hadoop	32
3.2.2 MongoDB.....	34
3.3 Data Mining.....	34
3.3.1 RapidMiner	35
3.3.2 Orange	36
3.4 Data visualization	37
3.4.1 Tableau	37
3.4.2 D3.js	40
3.5 Αναδυόμενες τεχνολογίες των Big Data.....	40

4. Τα Big Data στην επιχειρηματικότητα	47
4.1 Τι είναι η επιχειρηματικότητα;.....	47
4.2 Πως συμβάλουν τα Big Data στην επιχειρηματικότητα;	48
4.3 Ο ρόλος των Big Data στην καινοτομία	49
4.4 Τα Big Data σε κλάδους των επιχειρήσεων/ Τα Big Data στην λειτουργικότητα των επιχειρήσεων:	51
4.4.1 Big Data στα Logistics	51
4.4.2 Big Data στο Λιανικό Εμπόριο	51
4.4.3 Big Data στο Ηλεκτρονικό Εμπόριο	52
4.4.4 Big Data στον Τουρισμό	53
4.4.5 Big Data στην Φαρμακοβιομηχανία.....	53
4.4.6 Big Data στον Τραπεζικό Κλάδο	54
5. Μελέτες Περίπτωσης των Big Data στις επιχειρήσεις	56
5.1 Walmart.....	56
5.2 Google	57
5.3 Sprint	59
5.4 LinkedIn	61
6. Ερευνητικό Μέρος	63
6.1 Σκοπός της έρευνας.....	63
6.2 Επιλογή δείγματος της έρευνας.....	63
6.3 Επιλογή ερευνητικού εργαλείου	64
6.4 Επιλογή των ερωτήσεων της έρευνας	64
6.5 Αποτελέσματα της έρευνας.....	65
6.6 Συνδυαστικά αποτελέσματα της έρευνας.....	76
7. Ανακεφαλαίωση	81
ΒΙΒΛΙΟΓΡΑΦΙΑ	83
Παράρτημα	88

1. Εισαγωγή

Τα Μεγάλα Δεδομένα είναι αρκετά διαδεδομένα τα τελευταία χρόνια. Η ανάγκη να αποθηκεύουμε, αναπαράγουμε και να αναλύουμε τις πληροφορίες που κατέχουμε, σε συνδυασμό με τον μεγάλο όγκο των πληροφοριών που διαχειριζόμαστε, έφερε την ανάπτυξη τεχνικών και μέσων αποθήκευσης και ανάλυσης που διευκολύνουν τη διαχείριση τους. Ωστόσο, τα θεμέλια πάνω στα οποία στηρίχθηκε η εμφάνιση και εξέλιξή τους, τέθηκαν πολλά χρόνια πριν.

Συγκεκριμένα 18.000 χρόνια πριν, οι άνθρωποι χρησιμοποιούσαν κλαδιά ή οστά, στα οποία σηματοδοτούσαν εγκοπές και συγκρίνανε μεταξύ τους, για να κάνουν υπολογισμούς, επιτρέποντας τους να παρακολουθούν την εμπορική τους δραστηριότητα ή τις προμήθειες. Χιλιάδες χρόνια αργότερα, κατά το 2400 π.Χ. εφευρίσκεται το αριθμητήριο και χτίζεται η βιβλιοθήκη της Βαβυλώνας, η οποία αντιπροσωπεύει τις πρώτες μας προσπάθειες μαζικής αποθήκευσης δεδομένων. Το 300 π.Χ. δημιουργείται η βιβλιοθήκη της Αλεξάνδρειας, η οποία αποτέλεσε το μεγαλύτερο κέντρο αποθήκευσης δεδομένων του κόσμου.

Έπειτα από 18 αιώνες και συγκεκριμένα το 1663 στο Λονδίνο, ο John Graunt πραγματοποίησε το πρώτο καταγραμμένο πείραμα στην ανάλυση στατιστικών δεδομένων. Το 1928, ο Flitz Pflumer δημιούργησε μια μέθοδο μαγνητικής αποθήκευσης πληροφοριών, το οποίο ήταν η αρχή για την σημερινή τεχνολογία αποθήκευσης των ψηφιακών δεδομένων. Το 1965 η αμερικάνικη κυβέρνηση δημιουργεί το πρώτο κέντρο δεδομένων. Μετά από 5 χρόνια, ο μαθηματικός Edgar F. Codd της IBM παρουσιάζει το πλαίσιο του για μια "σχεσιακή βάση δεδομένων", το μοντέλο που χρησιμοποιούν πολλές σύγχρονες υπηρεσίες δεδομένων, σήμερα, για την αποθήκευση πληροφοριών σε ιεραρχική μορφή. Έτσι, φθάνουμε στο 1989 όπου γίνεται η πρώτη, πιθανότατα, χρήση του όρου Big Data από τον συγγραφέα Erik Larson σε άρθρο του στο Harpers Magazine, με τις επερχόμενες χρονιές να κρίνονται καθοριστικές για την εξέλιξη του όρου, αλλά και της ίδιας της τεχνολογίας.

Τα Μεγάλα Δεδομένα όπως προκύπτει, υπήρχαν στη ζωή του ανθρώπου εδώ και πάρα πολλά χρόνια. Αυτό που δημιουργεί πολλά ερωτήματα, βέβαια, είναι για ποιο λόγο έγιναν γνωστά τα τελευταία χρόνια. Απάντηση σε αυτό έρχεται να δώσει η ίδια η επιστήμη. Η ραγδαία ανάπτυξη της τεχνολογίας έκανε την καθημερινότητα των ανθρώπων, ολοένα και πιο περίπλοκη, κάτι που είχε μεγάλο αντίκτυπο και στην καθημερινή λειτουργία των επιχειρήσεων. Ο μεγάλος όγκος πληροφοριών που είχαν να διαχειριστούν οι εταιρίες, καθώς και ο αναπτυσσόμενος ανταγωνισμός που υπήρχε, τις οδήγησε να βρουν νέους τρόπους και τεχνολογίες για να διαχειριστούν και να αντιμετωπίσουν τις δυσκολίες αυτές. Όλο αυτό συνέβαλε στην ανακάλυψη και αξιοποίηση των Μεγάλων Δεδομένων. Το γεγονός ότι μπορούμε να εξάγουμε περισσότερη γνώση συνδέοντας πληροφορίες και εντοπίζοντας συσχετίσεις μπορεί να αναδείξει ενδιαφέροντα πράγματα για τη καθημερινότητά μας, αλλά και να

βοηθήσουν τις επιχειρήσεις να επιτύχουν έναν τρέχοντα ή και μελλοντικό σκοπό, με αποτέλεσμα να βελτιωθούν. Παρακάτω θα ανακαλύψουμε πολλές ενδιαφέρουσες δυνατότητες των Μεγάλων Δεδομένων, τεχνολογίες που συμβάλουν στην ανάπτυξη και βελτίωση των επιχειρήσεων, καθώς και πως όλα αυτά εφαρμόζονται στην σημερινή επιχειρηματικότητα.

2. Μεγάλα Δεδομένα(Big Data)

2.1 Τι είναι τα Big Data;

Τα τελευταία χρόνια, η ανάγκη για την διαχείριση των Μεγάλων Δεδομένων (ή Big Data) έχει αυξηθεί λόγω των μεγάλων αναγκών της παραγωγής. Η πρώτη εμφάνιση του όρου έγινε το 1997 από τους επιστήμονες της NASA, οι οποίοι ανέφεραν ότι ήταν αδύνατη η αναπαράσταση συνόλων δεδομένων γραφικά, καθώς ήταν τόσο μεγάλα που ήταν αδύνατο να αποθηκευτούν στην κύρια μνήμη του δίσκου. Αυτό ήταν και ένα από τα πρώτα προβλήματα Μεγάλων Δεδομένων. Η τεχνολογική εξέλιξη στον τομέα των τηλεπικοινωνιών και των ολοκληρωμένων κυκλωμάτων έχει δώσει την δυνατότητα να δημιουργηθούν μηχανισμοί παρακολούθησης των λειτουργιών ενός οργανισμού σε ανεπτυγμένο επίπεδο. Μέσω αυτής της ψηφιοποίησης των διαδικασιών παραγωγής, μεγάλοι οργανισμοί αλλά και εταιρίες μικρού μεγέθους έχουν καταφέρει να παράγουν τεράστιους όγκους δεδομένων με πολύ γρήγορο ρυθμό. Τα δεδομένα αυτά έχουν πολύτιμες και χρήσιμες πληροφορίες για τις επιχειρήσεις, καθώς μέσω της ανάλυσης τους μπορεί η επιχείρηση να βελτιστοποιήσει την παραγωγή της. Ωστόσο μπορεί να φέρει και προβλήματα, καθώς οι υπάρχουσες τεχνολογικές λύσεις για την διαχείριση δεδομένων δεν ανταποκρίνονται πλήρως στο όγκο αλλά και στην φύση τους.

Τα ψηφιακά δεδομένα βρίσκονται πλέον παντού: σε κάθε τομέα, οικονομία, οργανισμό καθώς και χρήστη της τεχνολογίας. Ηγέτες από κάθε τομέα, αλλά και καταναλωτές προϊόντων και υπηρεσιών αρχίζουν να έλκονται ολοένα και περισσότερο από τα Μεγάλα Δεδομένα λόγω των οφελών που αναμένεται να αποκομίσουν από την αξιοποίησή τους. Σήμερα, η δυνατότητα αποθήκευσης, συγκέντρωσης δεδομένων και η ανάλυση των αποτελεσμάτων τους έχει γίνει πολύ πιο προσιτή και εφικτή. Επίσης, τα μέσα εξόρυξης γνώσης από τα δεδομένα σημειώνουν σημαντική βελτίωση, καθώς τα διαθέσιμα λογισμικά για την εφαρμογή τεχνικών αυξανόμενης πολυπλοκότητας συνδυάζονται με την αυξανόμενη υπολογιστική ισχύ. Παράλληλα, ο αυξημένος αριθμός των ατόμων, συσκευών αλλά και αισθητήρων έχει εκτοξεύσει την δυνατότητα παραγωγής, επικοινωνίας και πρόσβασης δεδομένων. Το 2010, σχεδόν, το 60 τοις εκατό του παγκόσμιου πληθυσμού, χρησιμοποιούσε κινητά τηλέφωνα, και περίπου 12 τοις εκατό από αυτούς τους ανθρώπους είχαν smartphones, τα οποία έχουν μια αυξητική πορεία με ποσοστό περίπου 20 τοις εκατό το χρόνο. Επίσης, περισσότεροι από 30 εκατ. συνδεδεμένοι κόμβοι αισθητήρων βρίσκονται πλέον στους κλάδους μεταφορών, αυτοκινητοβιομηχανίας, καθώς και σε τομείς του λιανικού εμπορίου. Έτσι η δημιουργία πολλών τεχνολογικών καινοτομιών έχει οδηγήσει σε μια δραματική αύξηση των δεδομένων και της συλλογής τους. Για αυτό τον λόγο και τα Μεγάλα δεδομένα έχουν γίνει πρόσφατη περιοχή των στρατηγικών επενδύσεων για του IT οργανισμούς.

Ωστόσο, οι μεγάλοι όγκοι δεδομένων δεν παράγονται μόνο από τις επιχειρήσεις, αλλά και από την καθημερινή χρήση του κάθε ατόμου. Συγκεκριμένα κάθε άνθρωπος, σήμερα, έχει ένα ψηφιακό προφίλ στα κοινωνικά δίκτυα, από το οποίο παράγονται χιλιάδες χρήσιμες πληροφορίες-δεδομένα. Η Google εκτιμά ότι κάθε δύο μέρες το ψηφιακό υλικό που δημιουργείται από τους χρήστες είναι ίσο με το έντυπο υλικό που παρήγαγε η ανθρωπότητα από την αρχή της γραφής μέχρι το 2003. Μεγάλοι όγκοι δεδομένων παρουσιάζονται ακόμη και σε επιστημονικούς τομείς, όπως η ιατρική, η αστρονομία, η μετεωρολογία, αλλά και η βιολογία, χάρη στις νέες τεχνολογικές καινοτομίες που έχουν βοηθήσει κατά πολύ την ανθρωπότητα. Ο ρυθμός αύξησης των δεδομένων έχει τέτοια αυξητική πορεία που προβλέπεται πως τα επόμενα χρόνια θα υπάρξει μια ακόμη μεγαλύτερη “έκρηξη πληροφορίας” σαν ποτέ άλλοτε. Θα λέγαμε όμως, ότι δεν υπάρχει ένα όριο μεγέθους δεδομένων πάνω από το οποίο αποκαλούνται “Μεγάλα Δεδομένα”. Με τον συγκεκριμένο όρο υπολογίζεται πώς σήμερα γίνεται αναφορά σε όγκους δεδομένων που κυμαίνονται από μερικά terabytes έως δεκάδες ή και εκατοντάδες petabytes (1.024 terabytes) ή exabytes (1.024 petabytes) ή zetabytes(1.024 exabytes).

Ο όρος "Big Data" από μεγάλους οργανισμούς και επιστήμονες

Η **Gartner** (η μεγαλύτερη επιχείρηση στον κόσμο που ασχολείται με την τεχνολογική έρευνα και συμβουλευτική) δίνει τον εξής ορισμό για τον όρο των "Μεγάλων Δεδομένων": «Τα Μεγάλα Δεδομένα είναι στοιχεία υψηλού όγκου, υψηλής ταχύτητας και / ή υψηλής ποικιλίας που απαιτούν οικονομικά αποδοτικές και καινοτόμες μορφές επεξεργασίας πληροφοριών που επιτρέπουν την καλύτερη γνώση, τη λήψη αποφάσεων και την αυτοματοποίηση της διαδικασίας.»

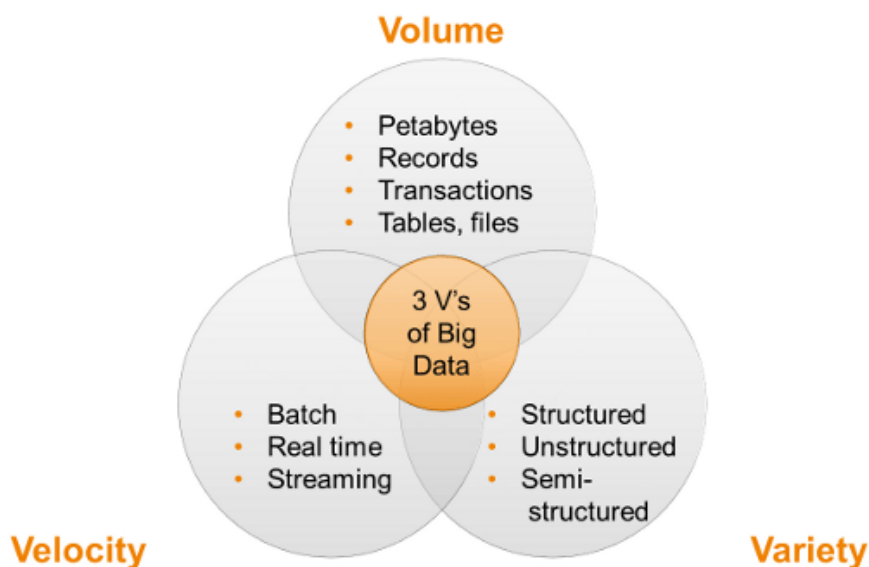
Η **SAS** πιστεύει πως ένας αποδοτικός ορισμός είναι: «Τα Μεγάλα Δεδομένα είναι ένας όρος που περιγράφει τον μεγάλο όγκο δεδομένων - τόσο δομημένων όσο και μη δομημένων - που κατακλύζουν μια επιχείρηση καθημερινά. Αλλά δεν είναι η ποσότητα των δεδομένων που είναι σημαντική. Είναι αυτό που κάνουν οι οργανισμοί με τα δεδομένα αυτά που έχει σημασία. Τα Μεγάλα Δεδομένα μπορούν να αναλυθούν για ιδέες που οδηγούν σε καλύτερες αποφάσεις και στρατηγικές επιχειρηματικές κινήσεις. »

Η **Oracle** επίσης θεωρεί με απλά λόγια και με βάση τον ορισμό που δίνει η Gartner(όπως αναφέραμε προηγουμένως) πως: «Τα Μεγάλα Δεδομένα είναι μεγαλύτερα, πιο σύνθετα σύνολα δεδομένων, ειδικά από νέες πηγές δεδομένων. Αυτά τα σύνολα δεδομένων είναι τόσο ογκώδη που το παραδοσιακό λογισμικό επεξεργασίας δεδομένων δεν μπορεί να τα διαχειριστεί. Αλλά αυτοί οι τεράστιοι όγκοι δεδομένων μπορούν να χρησιμοποιηθούν για την αντιμετώπιση επιχειρηματικών προβλημάτων που δεν θα μπορούσατε να αντιμετωπίσετε πριν.»

Δεκάδες είναι οι ορισμοί που δόθηκαν. Μελετώντας τους, μπορούμε να περιγράψουμε τον όρο ως συλλογές δεδομένων, τα οποία είναι δομημένα κατά ένα ποσοστό και αδόμητα στη πλειονότητά τους, και ο όγκος τους είναι τόσο μεγάλος που καθιστά πολύ δύσκολη την αποθήκευση, επεξεργασία και ανάλυσή τους με τη χρήση παραδοσιακών τεχνικών της επιστήμης της πληροφορικής. Ωστόσο, θα ήταν λάθος να δώσουμε έναν συγκεκριμένο ορισμό για τα Μεγάλα Δεδομένα, διότι κάθε επιχείρηση χρησιμοποιεί τα Big Data για διαφορετικούς σκοπούς, άρα είναι εμφανής η πολυδιάστατη φύση τους, η οποία με το πέρασμα του χρόνου, συνεχώς, εμπλουτίζεται.

2.2 Χαρακτηριστικά Big Data

Οι περισσότεροι ορισμοί των “Μεγάλων Δεδομένων” εστιάζουν στον όγκο των δεδομένων. Ωστόσο, εκτός από το μέγεθος των δεδομένων, υπάρχουν και άλλα εξίσου σημαντικά χαρακτηριστικά στα “Μεγάλα Δεδομένα”, μεταξύ των οποίων είναι η ποικιλία των δεδομένων και η ταχύτητα των δεδομένων. Ο συνδυασμός των Vs (Volume, Variety, Velocity) αποτελεί έναν πλήρη ορισμό για την έννοια των “Μεγάλων Δεδομένων”, στον οποίο δεν υπάρχουν περιορισμοί. Επιπλέον, κάθε ένα από αυτά τα τρία Vs έχει τις δικές του ιδιαιτερότητες ως προς την ανάλυση τους.(Σχήμα 2)



Σχήμα 1: The 3V's of Big Data
Πηγή: Pedro César Tebaldi Gomes (2014)

Τα τρία Vs

Volume: Αναφέρεται σε μεγάλες ποσότητες όλων των ειδών των δεδομένων από οποιαδήποτε πηγή δεδομένων, συμπεριλαμβανομένων επιχειρηματικών συναλλαγών, κοινωνικών δικτύων και πληροφορίες από δεδομένα που μεταδίδονται μεταξύ ψηφιακών συσκευών. Τα οφέλη από τη συλλογή και την ανάλυση αυτών των ποσοτήτων δημιουργούν μια σειρά προκλήσεων για την απόκτηση πολύτιμων γνώσεων για τους ανθρώπους και τις εταιρείες.

Velocity: Αναφέρεται στην ταχύτητα μεταφοράς δεδομένων. Με την συνεχή αύξηση των πηγών άντλησης δεδομένων, η ροή των δεδομένων, καθώς και το περιεχόμενο τους, αλλάζουν με μεγάλες ταχύτητες, πράγμα που κάνει δύσκολη την διαχείριση τους. Για αυτό τον λόγο, χρειάζονται νέοι αλγόριθμοι και τρόποι για την ομαλότερη επεξεργασία και ανάλυση τους.

Variety: Αναφέρεται σε μια ποικιλία από διαφορετικά είδη δεδομένων που συλλέγονται μέσω αισθητήρων, smartphones, social media και ιστοσελίδες, απεικονιζόμενα σε βίντεο, κείμενο, ήχο κ.α. Επίσης, τα δεδομένα αυτά αποτελούνται κυρίως από δομημένα, αριθμητικά δεδομένα (σε σχεσιακές βάσεις δεδομένων) έως και μη δομημένα αρχεία.

Ωστόσο, με το πέρασμα των χρόνων, προστέθηκαν επιπλέον χαρακτηριστικά γύρω από τα “Μεγάλα Δεδομένα”, μερικά από τα οποία είναι:

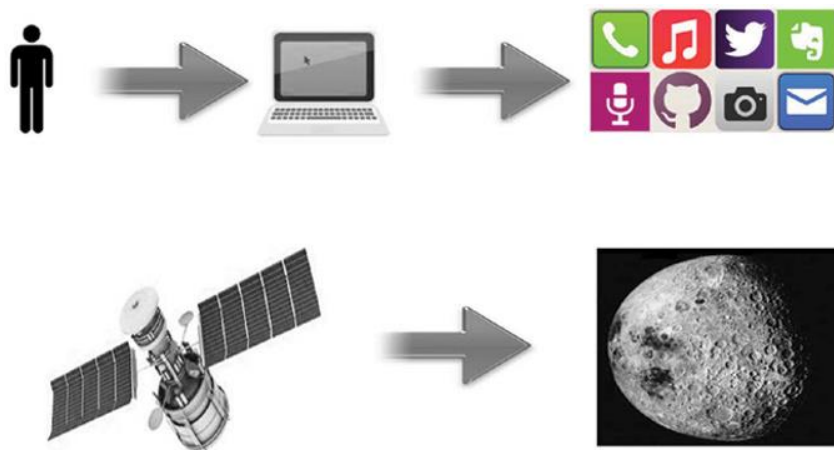
Veracity: Η ορθότητα των δεδομένων μπορεί να σώσει τους οργανισμούς από πολλά προβλήματα. Επειδή τα δεδομένα προέρχονται από πολλές διαφορετικές πηγές, η διασφάλιση της εγκυρότητας των δεδομένων από τις επιχειρήσεις είναι σημαντική για τις αναλύσεις τους και πιο συγκεκριμένα για την αυτοματοποιημένη λήψη αποφάσεων τους.

Variability: Αναφέρεται στην μεταβλητότητα των δεδομένων. Επειδή οι ροές δεδομένων είναι εξαιρετικά απρόβλεπτες και αλλάζουν συνεχώς σε μεγάλο βαθμό, είναι δύσκολο για τις εταιρίες να κατανοήσουν το νόημα και την σημασία των δεδομένων. Αν και δύσκολο, αυτό μπορεί να επιτευχθεί με την χρήση των κατάλληλων αλγορίθμων

Visualization: Η απεικόνιση των δεδομένων είναι ένα ιδιαίτερο κομμάτι, καθώς είναι σημαντική για την μετατροπή μεγάλου όγκου δεδομένων σε μορφή κατανοητή για τον αναγνώστη. Η απεικόνιση τους, αν και δεν είναι τεχνικά δύσκολη, είναι το πιο δύσκολο κομμάτι στα “μεγάλα δεδομένα”, διότι πίσω από κάθε γράφημα βρίσκεται μια σύνθετη ιστορία που πολλές φορές είναι δύσκολη να αποτυπωθεί, αλλά και εξαιρετικά σημαντική.

2.3 Τύποι των Big Data

Τα δεδομένα μπορεί να δημιουργούνται από μηχανή ή από τον άνθρωπο. Τα δεδομένα που παράγονται από τον άνθρωπο αναφέρονται σε δεδομένα που παράγονται ως αποτέλεσμα αλληλεπιδράσεων ανθρώπων με τις μηχανές. Τα ηλεκτρονικά ταχυδρομεία, τα έγγραφα, οι δημοσιεύσεις στο Facebook είναι μερικά από τα δεδομένα που δημιουργούνται από τον άνθρωπο. Τα δεδομένα που δημιουργούνται από τη μηχανή αναφέρονται σε δεδομένα που δημιουργούνται από εφαρμογές υπολογιστών ή συσκευές υλικού χωρίς την παρέμβαση του ανθρώπου. Τα δεδομένα από αισθητήρες, συστήματα προειδοποίησης καταστροφών, συστήματα πρόγνωσης καιρού και τα δορυφορικά δεδομένα είναι μερικά από τα δεδομένα που δημιουργούνται από τη μηχανή. Το σχήμα 3 αντιπροσωπεύει τα δεδομένα που παράγονται από έναν άνθρωπο σε διάφορα μέσα κοινωνικής δικτύωσης, τα ηλεκτρονικά μηνύματα που αποστέλλονται και τις εικόνες που έχει τραβήξει, και τα δεδομένα μηχανής που δημιουργούνται από τον δορυφόρο.



Σχήμα 2: Human- and Machine-generated data
Πηγή: Balamurugan Balusamy, Nandhini Abirami. R,
Seifedine Kadry, and Amir H. Gandomi (2021)

Τα δεδομένα που παράγονται από μηχανή και αυτά που παράγονται από τον άνθρωπο μπορούν να αναπαρασταθούν από τους ακόλουθους πρωτόγονους τύπους μεγάλων δεδομένων:

- Δομημένα δεδομένα
- Μη δομημένα δεδομένα
- Ημι-δομημένα δεδομένα

Δομημένα δεδομένα

Τα δεδομένα που μπορούν να αποθηκευτούν σε σχεσιακή βάση δεδομένων σε μορφή πίνακα με σειρές και στήλες ονομάζονται δομημένα δεδομένα. Τα δομημένα δεδομένα, που παράγονται συχνά από επιχειρήσεις, εμφανίζουν υψηλό βαθμό οργάνωσης και μπορούν εύκολα να υποστούν επεξεργασία χρησιμοποιώντας εργαλεία εξόρυξης δεδομένων. Παραδείγματα δομημένων δεδομένων αποτελούν τα στοιχεία εργαζομένων και οικονομικές συναλλαγές.

Μη δομημένα δεδομένα

Τα δεδομένα που είναι ακατέργαστα, μη οργανωμένα και δεν ταιριάζουν στα σχεσιακά συστήματα βάσεων δεδομένων ονομάζονται μη δομημένα δεδομένα. Σχεδόν το 80% των παραγόμενων δεδομένων δεν είναι δομημένα. Παραδείγματα μη δομημένων δεδομένων αποτελούν τα βίντεο, ο ήχος, οι εικόνες, τα μηνύματα ηλεκτρονικού ταχυδρομείου, τα αρχεία κειμένου και οι δημοσιεύσεις κοινωνικών μέσων. Τα μη δομημένα δεδομένα συνήθως βρίσκονται είτε σε αρχεία κειμένου είτε σε δυαδικά αρχεία. Τα δεδομένα που βρίσκονται σε δυαδικά αρχεία δεν έχουν καμία αναγνωρίσιμη εσωτερική δομή, όπως για παράδειγμα, ο ήχος, τα βίντεο και οι εικόνες. Τα δεδομένα που βρίσκονται σε αρχεία κειμένου είναι e-mail, δημοσιεύσεις κοινωνικών μέσων, αρχεία PDF και έγγραφα επεξεργασίας κειμένου.

Ημι-δομημένα δεδομένα

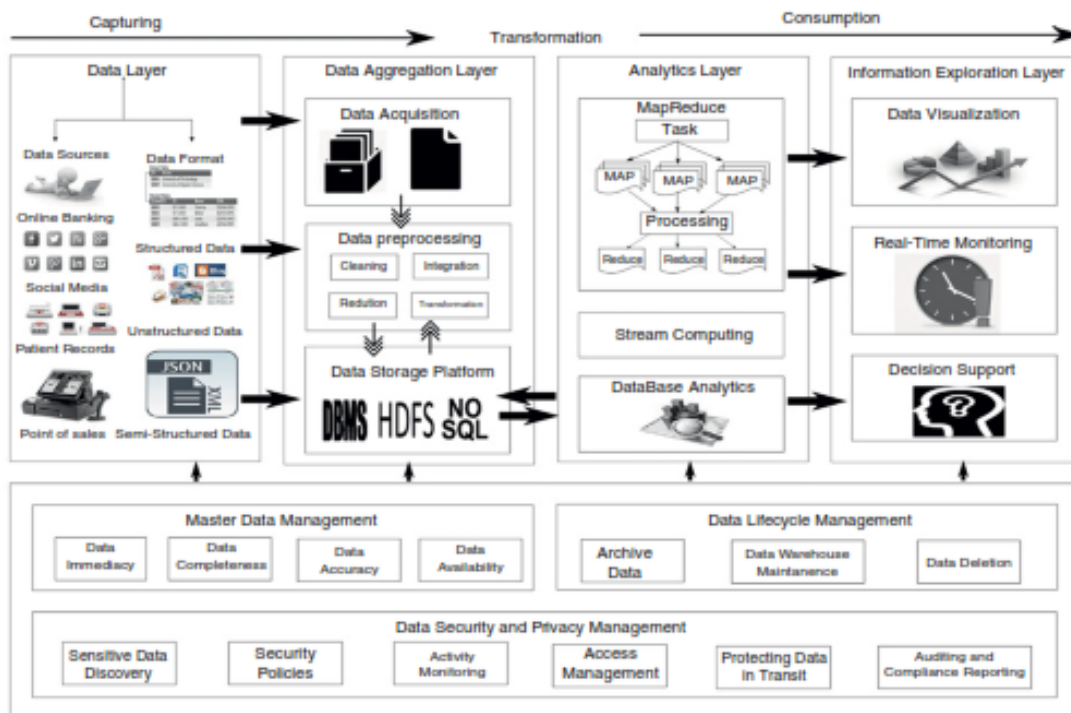
Τα ημι-δομημένα δεδομένα είναι εκείνα που έχουν δομή αλλά δεν ταιριάζουν στη σχεσιακή βάση δεδομένων. Τα ημι-δομημένα δεδομένα είναι οργανωμένα, γεγονός που καθιστά ευκολότερη την ανάλυση τους σε σύγκριση με τα μη δομημένα δεδομένα. Το JSON και το XML είναι παραδείγματα ημι-δομημένων δεδομένων.

2.4 Κύκλος ζωής των Big Data

Τα Big Data αποφέρουν μεγάλα οφέλη, ξεκινώντας από καινοτόμες επιχειρηματικές ιδέες έως και μη συμβατικούς τρόπους αντιμετώπισης ασθενειών, ξεπερνώντας έτσι πολλές προκλήσεις. Οι προκλήσεις προκύπτουν επειδή πολλά από τα δεδομένα συλλέγονται από την τεχνολογία σήμερα. Οι τεχνολογίες των Big Data είναι σε θέση να τις συλλάβουν και να τις αναλύσουν αποτελεσματικά. Η υποδομή των Big Data περιλαμβάνει νέα μοντέλα υπολογιστών με δυνατότητα επεξεργασίας τόσο καταναμημένων όσο και παράλληλων υπολογισμών με επεκτάσιμη αποθήκευση και απόδοση. Μερικά από τα στοιχεία των Big Data περιλαμβάνουν Hadoop (framework), HDFS (storage) και MapReduce (επεξεργασία).

Το σχήμα 3 απεικονίζει τον κύκλο ζωής των Big Data. Καταγράφονται δεδομένα που φτάνουν σε υψηλή ταχύτητα από πολλές πηγές με διαφορετικές μορφές δεδομένων. Τα καταγεγραμμένα δεδομένα αποθηκεύονται σε μια πλατφόρμα

αποθήκευσης, όπως HDFS και NoSQL και στη συνέχεια προεπεξεργάζονται για να είναι κατάλληλα για ανάλυση. Τα προεπεξεργασμένα δεδομένα που αποθηκεύονται στην πλατφόρμα αποθήκευσης μεταφέρονται στη συνέχεια στο επίπεδο των analytics, όπου τα δεδομένα υποβάλλονται σε επεξεργασία χρησιμοποιώντας εργαλεία Big Data όπως το MapReduce και το YARN και έπειτα πραγματοποιείται η ανάλυση των επεξεργασμένων δεδομένων για να αποκαλυφθεί η κρυφή γνώση από αυτά. Τα Analytics και η Μηχανική Μάθηση είναι σημαντικές έννοιες στον κύκλο ζωής των Big Data. Η ανάλυση κειμένου είναι ένας τύπος ανάλυσης που πραγματοποιείται σε μη δομημένα δεδομένα κειμένου. Με την ανάπτυξη των κοινωνικών μέσων και των συναλλαγών μέσω e-mail, η σημασία της ανάλυσης κειμένου έχει αυξηθεί. Η προγνωστική ανάλυση της συμπεριφοράς των καταναλωτών και η ανάλυση των συμφερόντων των καταναλωτών πραγματοποιούνται σε δεδομένα κειμένου που εξάγονται από διάφορες διαδικτυακές πηγές, όπως κοινωνικά μέσα, διαδικτυακούς ιστότοπους λιανικής πώλησης και πολλά άλλα. Η μηχανική μάθηση κατέστησε δυνατή την ανάλυση κειμένου. Τα αναλυθέντα δεδομένα απεικονίζονται οπτικά από εργαλεία οπτικοποίησης όπως το Tableau για να είναι εύκολα κατανοητό από τον τελικό χρήστη να λαμβάνει αποφάσεις.



Σχήμα 3: Big Data life cycle.

Πηγή: Balamurugan Balusamy, Nandhini Abirami. R, Seifedine Kadry, and Amir H. Gandomi (2021)

1. Δημιουργία των δεδομένων:

Η πρώτη φάση του κύκλου ζωής των Big Data είναι η δημιουργία δεδομένων. Η κλίμακα των δεδομένων που παράγονται από διαφοροποιημένες πηγές επεκτείνεται σταδιακά. Πηγές αυτού του μεγάλου όγκου δεδομένων αναφέρονται στην ενότητα 2.7.

2. Συγκέντρωση των δεδομένων:

Η φάση της συγκέντρωσης των δεδομένων του κύκλου ζωής των Big Data περιλαμβάνει τη συλλογή των πρωτογενών δεδομένων, τη μετάδοση των δεδομένων στην πλατφόρμα αποθήκευσης και την προεπεξεργασία τους. Η απόκτηση δεδομένων στον κόσμο των Big Data σημαίνει την απόκτηση δεδομένων υψηλού όγκου που φθάνουν σε έναν συνεχώς αυξανόμενο ρυθμό. Τα ακατέργαστα δεδομένα που συλλέγονται με αυτόν τον τρόπο μεταδίδονται σε κατάλληλη υποδομή αποθήκευσης για την υποστήριξη της επεξεργασίας και διαφόρων αναλυτικών εφαρμογών. Η προεπεξεργασία περιλαμβάνει τον καθαρισμό δεδομένων, την ενσωμάτωση δεδομένων, τον μετασχηματισμό δεδομένων και τη μείωση δεδομένων για να καταστούν τα δεδομένα αξιόπιστα, χωρίς σφάλματα, συνεπή και ακριβή. Τα δεδομένα που συλλέγονται ενδέχεται να έχουν απολύσεις, οι οποίες καταλαμβάνουν το χώρο αποθήκευσης και αυξάνουν το κόστος αποθήκευσης και μπορούν να αντιμετωπιστούν μέσω της προεπεξεργασίας των δεδομένων. Επίσης, πολλά από τα δεδομένα που συλλέγονται ενδέχεται να μην σχετίζονται με τον στόχο ανάλυσης και, ως εκ τούτου, πρέπει να συμπειστούν κατά την προεπεξεργασία. Ως εκ τούτου, η αποτελεσματική προεπεξεργασία δεδομένων είναι απαραίτητη για να είναι οικονομικά αποδοτική, καθώς και για μια αποδοτική αποθήκευση δεδομένων. Τα προεπεξεργασμένα δεδομένα στη συνέχεια μεταδίδονται για διάφορους σκοπούς, όπως μοντελοποίηση δεδομένων και ανάλυση δεδομένων.

3. Προεπεξεργασία των δεδομένων:

Η προεπεξεργασία δεδομένων είναι μια σημαντική διαδικασία που πραγματοποιείται σε ανεπεξέργαστα δεδομένα για τη μετατροπή τους σε κατανοητή μορφή και παρέχει, έτσι, πρόσβαση σε συνεπή και ακριβή δεδομένα. Τα δεδομένα που παράγονται από πολλές πηγές είναι εσφαλμένα, ελλιπή και ασυνεπή λόγω του μαζικού όγκου και των ετερογενών πηγών τους, και δεν έχει νόημα να αποθηκεύονται άχρηστα και βρώμικα δεδομένα. Επιπλέον, ορισμένες αναλυτικές εφαρμογές έχουν καθοριστική προϋπόθεση για ποιοτικά δεδομένα. Ως εκ τούτου, για αποτελεσματική, αποδοτική και ακριβή ανάλυση δεδομένων, είναι απαραίτητη η συστηματική προεπεξεργασία δεδομένων. Η ποιότητα των δεδομένων προέλευσης επηρεάζεται από διάφορους παράγοντες. Για παράδειγμα, τα δεδομένα ενδέχεται να έχουν σφάλματα όπως ένα πεδίο μισθού που έχει αρνητική τιμή, το οποίο προκύπτει λόγω σφαλμάτων μετάδοσης ή τυπογραφικών λάθους ή σκόπιμης λανθασμένης εισαγωγής δεδομένων από χρήστες που δεν επιθυμούν να αποκαλύψουν τα προσωπικά τους στοιχεία. Τα σφάλματα αυτά συνεπάγονται ότι το πεδίο στερείται των

χαρακτηριστικών ενδιαφέροντος, τα οποία ενδέχεται να προέρχονται από μη εφαρμόσιμα σφάλματα πεδίου ή λογισμικού. Η ασυνέπεια στα δεδομένα αναφέρεται στις αποκλίσεις στα δεδομένα, όπως η ημερομηνία γέννησης και η ηλικία. Οι ασυνέπειες στα δεδομένα προκύπτουν όταν τα δεδομένα που συλλέγονται προέρχονται από διαφορετικές πηγές, λόγω ασυνεπειών στην ονομασία συμβάσεων μεταξύ διαφορετικών χωρών και ασυνεπειών στη μορφή εισαγωγής. Οι πηγές δεδομένων έχουν συχνά περιττά δεδομένα σε διαφορετικές μορφές, και ως εκ τούτου τα διπλά δεδομένα πρέπει επίσης να αφαιρεθούν κατά την προεπεξεργασία δεδομένων για να καταστούν τα δεδομένα ουσιαστικά και χωρίς σφάλματα. Υπάρχουν πολλά βήματα που εμπλέκονται στην προεπεξεργασία δεδομένων:

- Ενοποίηση δεδομένων
- Καθαρισμός δεδομένων
- Μείωση δεδομένων
- Μετασχηματισμός δεδομένων

Ενοποίηση δεδομένων

Η ενοποίηση δεδομένων περιλαμβάνει το συνδυασμό δεδομένων από διαφορετικές πηγές για να δώσει στους τελικούς χρήστες μια ενοποιημένη προβολή δεδομένων. Πολλές προκλήσεις υπάρχουν κατά την ενοποίηση των δεδομένων. Για παράδειγμα, κατά την εξαγωγή δεδομένων από το προφίλ ενός ατόμου, το όνομα και το επίθετο μπορεί να εναλλάσσονται σε μια συγκεκριμένη κουλτούρα, οπότε σε τέτοιες περιπτώσεις η ενοποίηση μπορεί να συμβεί εσφαλμένα. Οι απολύσεις δεδομένων συμβαίνουν συχνά κατά την ενοποίηση δεδομένων από πολλές πηγές. Πηγές όπως οργανώσεις, smartphones, προσωπικοί υπολογιστές, δορυφόροι και αισθητήρες δημιουργούν διαφορετικά δεδομένα όπως e-mail, στοιχεία υπαλλήλων, μηνύματα συνομιλίας WhatsApp, δημοσιεύσεις κοινωνικών μέσων, διαδικτυακές συναλλαγές, δορυφορικές εικόνες και δεδομένα αισθητήρων. Αυτοί οι διαφορετικοί τύποι δομημένων, μη δομημένων και ημι-δομημένων δεδομένων πρέπει να ενοποιηθούν και να παρουσιαστούν ως ενοποιημένα δεδομένα για τον καθαρισμό δεδομένων, τη μοντελοποίηση δεδομένων, την αποθήκευση δεδομένων και έπειτα για την εξαγωγή, μετατροπή και φόρτωση (ETL) των δεδομένων.

Καθαρισμός δεδομένων

Η διαδικασία καθαρισμού δεδομένων συμπληρώνει τις τιμές που λείπουν, διορθώνει τα σφάλματα και τις ασυνέπειες και αφαιρεί τον πλεονασμό στα δεδομένα για τη βελτίωση της ποιότητας των δεδομένων. Όσο μεγαλύτερη είναι η ετερογένεια των πηγών δεδομένων, τόσο υψηλότερος είναι ο βαθμός ακαθαρσίας. Κατά συνέπεια, ενδέχεται να περιλαμβάνονται περισσότερα βήματα καθαρισμού. Ο καθαρισμός δεδομένων περιλαμβάνει διάφορα βήματα, όπως εντοπισμό ή αναγνώριση του σφάλματος, διόρθωση του σφάλματος ή διαγραφή των λανθασμένων δεδομένων και

τεκμηρίωση του τύπου σφάλματος. Για την ανίχνευση του τύπου σφάλματος και της ασυνέπειας που υπάρχει στα δεδομένα, απαιτείται λεπτομερής ανάλυση των δεδομένων. Ο πλεονασμός δεδομένων είναι η επανάληψη δεδομένων, η οποία αυξάνει το κόστος αποθήκευσης και τα έξοδα μετάδοσης και μειώνει την ακρίβεια και την αξιοπιστία των δεδομένων. Οι διάφορες τεχνικές που εμπλέκονται στο χειρισμό του πλεονασμού δεδομένων είναι η ανίχνευση πλεονασμού και η συμπίεση δεδομένων. Οι τιμές που λείπουν μπορούν να συμπληρωθούν με μη αυτόματο τρόπο, αλλά είναι κουραστική, χρονοβόρα και δεν είναι κατάλληλη για τον τεράστιο όγκο δεδομένων. Μια καθολική σταθερά μπορεί να χρησιμοποιηθεί για να συμπληρώσει όλες τις τιμές που λείπουν, αλλά αυτή η μέθοδος δημιουργεί προβλήματα κατά την ενοποίηση των δεδομένων.

Μείωση δεδομένων

Η επεξεργασία δεδομένων σε τεράστιο όγκο δεδομένων μπορεί να διαρκέσει πολύ, καθιστώντας την ανάλυση δεδομένων είτε ανέφικτη είτε μη πρακτική. Η μείωση δεδομένων είναι η έννοια της μείωσης του όγκου δεδομένων ή της μείωσης της διάστασης των δεδομένων, δηλαδή του αριθμού των χαρακτηριστικών. Οι τεχνικές μείωσης δεδομένων υιοθετούνται για την ανάλυση των δεδομένων σε μειωμένη μορφή χωρίς να χάνεται η ακεραιότητα των πραγματικών δεδομένων και παράλληλα να αποδίδονται ποιοτικά αποτελέσματα. Οι τεχνικές μείωσης δεδομένων περιλαμβάνουν συμπίεση δεδομένων, μείωση διαστάσεων και μείωση αριθμητικότητας. Οι τεχνικές συμπίεσης δεδομένων εφαρμόζονται για τη λήψη της συμπιεσμένης ή μειωμένης αναπαράστασης των πραγματικών δεδομένων. Η μείωση διαστάσεων είναι η μείωση ενός αριθμού χαρακτηριστικών, δηλαδή μια τεχνική που αφαιρεί τα άσχετα ή περιττά χαρακτηριστικά. Η μείωση αριθμητικότητας είναι μια τεχνική που υιοθετείται για τη μείωση του όγκου επιλέγοντας μικρότερα εναλλακτικά δεδομένα.

Μετασχηματισμός δεδομένων

Ο μετασχηματισμός δεδομένων αναφέρεται σε μετασχηματισμό ή ενοποίηση των δεδομένων σε κατάλληλη μορφή και μετατροπή τους σε λογικές και σημαντικές πληροφορίες για τη διαχείριση και ανάλυση δεδομένων. Η πραγματική πρόκληση στον μετασχηματισμό δεδομένων έρχεται στην εικόνα όταν τα πεδία σε ένα σύστημα δεν ταιριάζουν με τα πεδία σε άλλο σύστημα. Πριν από τον μετασχηματισμό δεδομένων, πραγματοποιείται καθαρισμός και χειρισμός των δεδομένων. Οι οργανισμοί συλλέγουν ένα τεράστιο ποσό δεδομένων και έτσι ο όγκος των δεδομένων αυξάνεται ραγδαία. Τα δεδομένα που συλλέγονται μετασχηματίζονται χρησιμοποιώντας εργαλεία ETL.

2.5 Οφέλη των Big Data

Ο όγκος των δεδομένων αυξάνεται με ταχύς ρυθμούς και οι επιχειρήσεις έχουν προχωρήσει σε αποθήκευση μεγάλων ποσοτήτων δεδομένων και αναλύσεων τους για την καλύτερη λήψη αποφάσεων και βελτιστοποίηση της στρατηγικής τους. Οι τεχνολογίες των Big Data βοηθούν σε αυτήν την αποθήκευση και ανάλυση των δεδομένων ταχύτερα.

Για τον λόγο αυτό, επιχειρήσεις από όλους τους τομείς έλκονται ολοένα και περισσότερο από τον τρόπο διαχείρισης και ανάλυσης των Μεγάλων Δεδομένων. Σημαντικό κριτήριο εδώ είναι οι επιχειρήσεις πρέπει να αντιδρούν γρηγορότερα στις αλλαγές, όχι μόνο στις παρούσες αλλά και στις μελλοντικές προκλήσεις. Έτσι, τα Μεγάλα Δεδομένα γίνονται όλο και πιο απαραίτητα για τους υπεύθυνους λήψης αποφάσεων για να λαμβάνουν καλύτερες αποφάσεις.

Τα Μεγάλα Δεδομένα μέσω της ανάλυσης δεδομένων από πολλές πηγές, όπως σαρωτές, κινητά τηλέφωνα, το διαδίκτυο και οι πλατφόρμες κοινωνικών μέσων προσφέρουν σημαντικά οφέλη στις επιχειρήσεις. Ως εκ τούτου, σύμφωνα με τον Frankel και Reid, (2008) ένα από τα σημαντικότερα οφέλη των Μεγάλων Δεδομένων είναι η ικανότητα τους να προβάλλουν τις σημαντικότερες και πιο χρήσιμες πληροφορίες για τις επιχειρήσεις.

Τα Μεγάλα Δεδομένα και οι αναλύσεις αυτών των στοιχείων βοηθούν τους διαχειριστές και τα στελέχη των επιχειρήσεων να αξιολογήσουν τις πληροφορίες που συνέλλεξαν και να τις χρησιμοποιήσουν προς όφελος τους, πετυχαίνοντας έτσι μείωση του κόστους και του χρόνου για την καλύτερη λήψη σημαντικών αποφάσεων καθώς και την δημιουργία νέων προϊόντων και υπηρεσιών που στοχεύουν στις ανάγκες των πελατών τους.

2.6 Προκλήσεις των Big Data

Οι ευκαιρίες συνυπάρχουν συχνά με τις προκλήσεις. Από την μία πλευρά, τα Big Data προσφέρουν ευκαιρίες σε πολλούς τομείς. Για την άλλη, οι προκλήσεις ασφάλειας και απορρήτου που προκαλούνται από τα Big data προσελκύουν επίσης το βλέμμα των ανθρώπων. Αυτές οι προκλήσεις διαρκούν καθ' όλη τη διάρκεια ζωής των Big Data, τα οποία μπορούν να χαρακτηριστούν ως συλλογές δεδομένων, καθώς και αποθήκευσης και διαχείρισης, μετάδοσης, ανάλυσης και καταστροφής δεδομένων. Ως εκ τούτου, ανεξάρτητα από τη συλλογή δεδομένων και την καταστροφή δεδομένων, οι προκλήσεις του Big data μπορούν να συνοψιστούν ως εξής:

1. Προκλήσεις απορρήτου και ασφάλειας:

Είναι το πιο σημαντικό ζήτημα με τα Big Data, καθώς είναι ένα πολύ ευαίσθητο θέμα με τεχνική αλλά και νομική σημασία. Ένα μεγάλο ζήτημα είναι οι προσωπικές πληροφορίες ενός ατόμου όταν συνδυάζονται με εξωτερικά μεγάλα σύνολα δεδομένων, οδηγούν στη σύναψη νέων γεγονότων για αυτό το άτομο και όπως είναι πιθανό αυτά τα είδη γεγονότων για το άτομο να είναι μυστικά και να μην θέλει ο κάτοχος αυτών των δεδομένων ή οποιοδήποτε άτομο να μάθει για αυτόν. Παράλληλα, οι πληροφορίες σχετικά με τους χρήστες συλλέγονται και χρησιμοποιούνται για να προσθέσουν αξία στην επιχείρηση του οργανισμού. Αυτό γίνεται δημιουργώντας ιδέες στη ζωή τους, τις οποίες δεν γνωρίζουν.

Μια άλλη σημαντική συνέπεια που προκύπτει είναι η κοινωνική διαστρωμάτωση όπου ένας αναλφάβητος θα επωφελούνταν από την προγνωστική ανάλυση Big Data και από την άλλη, οι μειονεκτούντες θα αναγνωρίζονται εύκολα και θα αντιμετωπίζονται χειρότερα.

Τα Big Data που χρησιμοποιούνται από την επιβολή του νόμου θα αυξήσουν τις πιθανότητες ορισμένων ατόμων να υποφέρουν από δυσμενείς συνέπειες χωρίς την ικανότητα να αντισταθούν ή ακόμη και να γνωρίζουν ότι υφίστανται διακρίσεις.

2. Προκλήσεις στην πρόσβαση δεδομένων και στον διαμοιρασμό πληροφοριών:

Εάν τα δεδομένα πρόκειται να χρησιμοποιηθούν για τη λήψη ακριβών αποφάσεων εγκαίρως, καθίσταται απαραίτητο να είναι διαθέσιμα με ακριβή, πλήρη και έγκαιρο τρόπο. Αυτό καθιστά τη διαδικασία διαχείρισης δεδομένων και διακυβέρνησης λίγο περίπλοκη, προσθέτοντας την ανάγκη να καταστούν τα δεδομένα ανοιχτά και να είναι διαθέσιμα σε κυβερνητικούς φορείς με τυποποιημένο τρόπο με τυποποιημένα API, μεταδεδωμένα και μορφές, οδηγώντας έτσι σε καλύτερη λήψη αποφάσεων, στην επιχειρηματική ευφυΐα και σε βελτιώσεις στην παραγωγικότητα.

Η αναμονή ανταλλαγής δεδομένων μεταξύ εταιρειών είναι περίεργη λόγω της ανάγκης των επιχειρήσεων να υπερέχουν από άλλες. Η κοινή χρήση δεδομένων για τους πελάτες και τις δραστηριότητές τους απειλεί την κουλτούρα της μυστικότητας και της ανταγωνιστικότητας.

3. Προκλήσεις αποθήκευσης και επεξεργασίας:

Ο διαθέσιμος χώρος αποθήκευσης δεν επαρκεί για την αποθήκευση του μεγάλου όγκου δεδομένων που παράγονται από σχεδόν όλα: Οι ιστότοποι κοινωνικών μέσων είναι οι ίδιοι μεγάλοι συντελεστές μαζί με τις συσκευές αισθητήρων κ.λπ.

Λόγω των αυστηρών απαιτήσεων των Big Data σε δίκτυα, μέσα αποθήκευσης και διακομιστές, η εξωτερική ανάθεση δεδομένων στο cloud μπορεί να φαίνεται μια επιλογή. Ωστόσο, η μεταφόρτωση αυτού του μεγάλου όγκου δεδομένων στο cloud δεν επιλύει το πρόβλημα. Και αυτό επειδή οι πληροφορίες για τα Big Data απαιτούν τη συλλογή όλων των δεδομένων και στη συνέχεια τη σύνδεσή τους με κάποιο μέσο για την εξαγωγή σημαντικών πληροφοριών. Terabytes δεδομένων απαιτούν μεγάλο χρονικό διάστημα για να μεταφορτωθούν στο cloud και επιπλέον αυτά τα δεδομένα

αλλάζουν τόσο γρήγορα, γεγονός που θα δυσκολέψει τη μεταφόρτωση αυτών των δεδομένων σε πραγματικό χρόνο.

Η μεταφορά δεδομένων από το σημείο αποθήκευσης στο σημείο επεξεργασίας μπορεί να αποφευχθεί με δύο τρόπους. Το ένα είναι η επεξεργασία μόνο στον χώρο αποθήκευσης και τα αποτελέσματα μπορούν να μεταφερθούν, ή να μεταφερθούν μόνο αυτά τα δεδομένα που είναι σημαντικά. Αλλά και οι δύο αυτές μέθοδοι απαιτούν τη διατήρηση της ακεραιότητας και της προέλευσης των δεδομένων.

Η επεξεργασία τόσο μεγάλου όγκου δεδομένων απαιτεί επίσης μεγάλο χρόνο. Για να βρεθούν κατάλληλα στοιχεία ολόκληρο το σύνολο δεδομένων πρέπει να σαρωθεί κάτι που είναι κάπως αδύνατο. Έτσι, η δημιουργία ευρετηρίων ακριβώς στην αρχή ενώ γίνεται η συλλογή και αποθήκευση των δεδομένων είναι μια καλή πρακτική και μειώνει σημαντικά το χρόνο επεξεργασίας.

4. Προκλήσεις ανάλυσης:

Τα Big Data φέρνουν μαζί τους μερικές τεράστιες αναλυτικές προκλήσεις. Ο τύπος ανάλυσης που πρέπει να γίνει σε αυτόν τον τεράστιο όγκο δεδομένων που μπορεί να είναι μη δομημένος, ημι-δομημένος ή δομημένος απαιτεί μεγάλο αριθμό δεξιοτήτων εκ των προτέρων. Επιπλέον, ο τύπος ανάλυσης που πρέπει να γίνει στα δεδομένα εξαρτάται σε μεγάλο βαθμό από τα αποτελέσματα που πρέπει να ληφθούν, δηλαδή τη λήψη αποφάσεων. Αυτό μπορεί να γίνει χρησιμοποιώντας μία από τις δύο τεχνικές: είτε ενσωματώνοντας τεράστιους όγκους δεδομένων στην ανάλυση είτε προσδιορίζοντας εκ των προτέρων ποια δεδομένα είναι σχετικά.

5. Προκλήσεις δεξιοτήτων:

Δεδομένου ότι το Big Data βρίσκονται στα ξεκινήματα τους και αποτελούν μια αναδυόμενη τεχνολογία, πρέπει να προσελκύσουν οργανισμούς και νέα άτομα με διαφορετικά νέα σετ δεξιοτήτων. Αυτές οι δεξιότητες δεν πρέπει να περιορίζονται μόνο σε τεχνικές αλλά και να επεκτείνονται σε ερευνητικές, αναλυτικές, ερμηνευτικές και δημιουργικές. Αυτές οι δεξιότητες πρέπει να αναπτυχθούν ατομικά και ως εκ τούτου απαιτεί την κατάρτιση προγραμμάτων εκπαίδευσης από τους οργανισμούς. Επιπλέον, τα Πανεπιστήμια πρέπει να εισαγάγουν πρόγραμμα σπουδών για τα Big Data για να παράγουν εξειδικευμένους υπαλλήλους σε αυτήν την τεχνολογία.

2.7 Πηγές άντλησης των Big Data

Η ύπαρξη πολλών πηγών, οδηγεί σε μία αύξηση των δεδομένων που χρησιμοποιούνται, σήμερα, για την εφαρμογή και την ανάλυση των Big Data. Ορισμένες από αυτές τις πηγές αποτελούν νέες πηγές δεδομένων, ενώ άλλες είναι μία παραλλαγή στην ανάλυση των ήδη υπάρχοντων δεδομένων που έχουν δημιουργηθεί.

Με τις εταιρίες να στρέφονται στην δημιουργία ψηφιακών αναπαραστάσεων των υπάρχοντων δεδομένων και στην απόκτηση όλων των νέων, ο ρυθμός αύξησης των

δεδομένων τα τελευταία χρόνια έχει φτάσει σχεδόν το άπειρο, λόγω των περισσότερων επιχειρήσεων να έχουν ξεκινήσει από το μηδέν.

Πολλές βιομηχανίες εμπίπτουν στην “ομπρέλα” της δημιουργίας νέων δεδομένων και στην ψηφιοποίηση υπαρχόντων, με τις περισσότερες να αποτελούν κατάλληλες πηγές για τους πόρους των Big Data. Αυτές οι βιομηχανίες περιλαμβάνουν τα ακόλουθα:

- **Μεταφορές, εφοδιαστική, λιανικό εμπόριο, επιχειρήσεις κοινής ωφέλειας και τηλεπικοινωνίες:** Τα δεδομένα αισθητήρων παράγονται σε μία επιταχυνόμενη ταχύτητα από GPS transceivers, αναγνώστες ετικετών RFID (αναγνώριση ραδιοσυχνότητων), έξυπνους μετρητές και κινητά τηλέφωνα (αρχεία δεδομένων κλήσεων). Αυτά τα δεδομένα χρησιμοποιούνται για βελτιστοποίηση λειτουργιών και χρησιμοποίηση εργαλείων Business Intelligence για την αναγνώριση άμεσων ευκαιριών επιχειρηματικότητας.
- **Τομέας Υγείας:** Η βιομηχανία υγειονομικής περίθαλψης οδηγείται γρήγορα σε ηλεκτρονικά ιατρικά αρχεία και εικόνες, τα οποία θέλει να χρησιμοποιήσει σε βραχυπρόθεσμα προγράμματα παρακολούθησης της δημόσιας υγείας και σε μακροπρόθεσμα επιδημιολογικά ερευνητικά προγράμματα.
- **Κυβέρνηση:** Πολλές κυβερνητικές υπηρεσίες ψηφιοποιούν τα δημόσια έγγραφα, όπως πληροφορίες απογραφής, χρήσης ενέργειας, προϋπολογισμοί, εκλογικά δεδομένα και αναφορές επιβολής νόμου.
- **Μέσα ψυχαγωγίας:** Η βιομηχανία ψυχαγωγίας έχει οδηγηθεί στην ψηφιακή εγγραφή, παραγωγή και παράδοση τα τελευταία πέντε χρόνια και τώρα συλλέγει μεγάλες ποσότητες πλούσιου περιεχομένου και συμπεριφορές προβολής χρήστη.
- **Βιοεπιστήμες:** Η αλληλουχία γονιδίων χαμηλού κόστους μπορεί να παράγει δεκάδες terabyte πληροφοριών που πρέπει να αναλυθούν για την αναζήτηση γενετικών παραλλαγών και για αποτελεσματικότερες πιθανές θεραπείες.

Για πολλές επιχειρήσεις, τα πρόσθετα δεδομένα μπορούν να προέρχονται από αγορές self-service, οι οποίες καταγράφουν τη χρήση καρτών συνάφειας και παρακολουθούν τους ιστότοπους επισκεψιμότητας και μπορεί να συνδυαστεί με κοινωνικά δίκτυα και μεταδεδομένα βάσει τοποθεσίας. Αυτό δημιουργεί ένα “χρυσωρυχείο” από ενεργά δεδομένα καταναλωτών για λιανοπωλητές, διανομείς και κατασκευαστές συσκευασμένων προϊόντων.

Πρόσθετες πληροφορίες και μεγάλα σύνολα δεδομένων μπορούν, επίσης, να βρεθούν στους κοινωνικού ιστότοπους όπως το Facebook, το Foursquare και το Twitter. Ένας αριθμός από νέες επιχειρήσεις δημιουργούν, πλέον, περιβάλλοντα Big Data, βασισμένα σε συστάδες κλιμάκωσης χρησιμοποιώντας υψηλής απόδοσης επεξεργαστές πολλαπλών πυρήνων που επεξεργάζονται δεδομένα σύμφωνα με την συμπεριφορά και την καθημερινότητα των χρηστών(π.χ. likes, τοποθεσίες και απόψεις).

Χάρη στην αποτελεσματικότητα του δικτύου των επιτυχημένων ιστότοπων, τα συνολικά δεδομένα που δημιουργούνται μπορούν να επεκταθούν με εκθετικό ρυθμό. Ορισμένες εταιρείες έχουν συλλέξει και αναλύσει πάνω από 4 δισεκατομμύρια σημεία δεδομένων (π.χ. λειτουργίες αποκοπής και επικόλλησης ιστοτόπων) από τότε που ξεκίνησε η συλλογή πληροφοριών και εντός ενός χρόνου η διαδικασία αυτή έχει επεκταθεί σε 20 δισεκατομμύρια σημεία δεδομένων.

Επιπρόσθετα παραδείγματα πηγών άντλησης δεδομένων αποτελούν τα ακόλουθα:

- **Οικονομικές συναλλαγές:** Χάρη στην ενοποίηση του παγκόσμιου εμπορικά περιβάλλοντα και την αυξημένη χρήση προγραμματισμένων διαπραγμάτευση, ο όγκος των συναλλαγών που συλλέγονται και αναλύονται διπλασιάζεται ή τριπλασιάζεται. Οι όγκοι συναλλαγών κυμαίνονται επίσης πολύ πιο γρήγορα, πολύ ευρύτερα και πολύ πιο απρόβλεπτα. Ο ανταγωνισμός μεταξύ των επιχειρήσεων δημιουργεί περισσότερα δεδομένα, απλώς και μόνο επειδή η δειγματοληψία για αποφάσεις διαπραγμάτευσης πραγματοποιείται συχνότερα και σε γρηγορότερα διαστήματα
- **Κινητή τηλεφωνία:** Με τις εξελίξεις στα smartphones και στα συνδεδεμένα PDAs (Προσωπικός ψηφιακός οδηγός), τα κύρια δεδομένα που δημιουργούνται από αυτές τις συσκευές έχουν αναπτυχθεί πέρα τις βασικές λειτουργίες του καλούνται, του δέκτη και της διάρκειας της κλήσης. Επιπλέον δεδομένα συλλέγονται τώρα με εκθετικούς ρυθμούς, συμπεριλαμβανομένων στοιχείων όπως της γεωγραφικής θέσης, των μηνυμάτων κειμένου, της περιήγησης ιστορικού, αλλά και (χάρη στην προσθήκη επιταχυνσιόμετρων) κινήσεων, καθώς και δημοσιεύσεων κοινωνικών δικτύων και χρήσης εφαρμογών.

2.8 Εφαρμογές των Big Data

Healthcare: Τα big data είναι χρήσιμα στον τομέα της υγειονομικής περίθαλψης καθώς βοηθά στην ανάπτυξη ενός ισχυρού συστήματος υγείας, το οποίο μειώνει το κόστος αλλά και βελτιώνει την αποτελεσματικότητα της θεραπείας. Αναλυτές δεδομένων συλλέγουν πληροφορίες από πολλές πηγές, μεταξύ άλλων, γραπτές σημειώσεις και συνταγές γιατρού, ηλεκτρονικά αρχεία ασθενών, κλινικά αρχεία, δεδομένα αισθητήρων ιατρικών μηχανημάτων καθώς και φαρμακεία και εργαστήρια.

Παράλληλα, εκτός από τις εσωτερικές πηγές στον τομέα της υγείας, υπάρχουν και οι εξωτερικές πηγές όπως τα social media από τα οποία η απόκτηση πληροφοριών συμβάλει στην έγκαιρη πρόβλεψη επιδημιών και αντιμετώπιση τους. Για παράδειγμα, αφότου έγινε ο σεισμός στην Αϊτή τον Ιανουάριο του 2010, οι αναλύσεις των tweets βοήθησαν στην παρακολούθηση της εξάπλωσης της χολέρας στην χώρα.

Τέλος, τα περισσότερα ιατροτεχνολογικά προϊόντα είναι προσανατολισμένα στα big data. Η χρήση των μεγάλων δεδομένων έχει αναπτυχθεί τόσο που οι γιατροί σήμερα μπορούν να παρακολουθήσουν την καρδιά και την θερμοκρασία του ασθενούς μέσω

ειδικών ρολογιών τύπου smartwatch και να του συνταγογραφήσουν τα ανάλογα φάρμακα.

Education: Η συμβολή των big data στον τομέα της εκπαίδευσης είναι τεράστια. Μέσω του διαδικτύου η εκπαίδευση ποια, έχει γίνει πιο εύκολη και αποτελεσματική, με τη δημιουργία πολλών εφαρμογών, να βοηθούν δασκάλους και μαθητές να κάνουν κάθε πιθανό μάθημα οπουδήποτε και οποιαδήποτε χρονική στιγμή τηρώντας πάντα την ιδιωτικότητα των προσωπικών δεδομένων τους. Για αυτό, τα εκπαιδευτικά ιδρύματα πρέπει να προσαρμοστούν στα νέα εργαλεία διαχείρισης και ανάλυσης δεδομένων. Επιπλέον, τα big data βοηθούν στην αντιμετώπιση πολλών προβλημάτων που μπορεί να προκύψουν στην ψηφιακή μάθηση.

Telecommunication: Τα τελευταία χρόνια οι εταιρίες τηλεπικοινωνιών προσπαθούν ολοένα και περισσότερο να καταπολεμήσουν οποιαδήποτε μορφή απάτης στις τηλεπικοινωνίες. Ωστόσο τα συστήματα διαχείρισης και εντοπισμού αυτών των κρουσμάτων είναι ανεπαρκή για την καταπολέμηση νέων τύπων απάτης. Για αυτό το λόγο, οι πάροχοι υπηρεσιών κινητής τηλεφωνίας άρχισαν να χρησιμοποιούν τα Big Data και συγκεκριμένα να αναλύουν δεδομένα για τον καλύτερο εντοπισμό οποιαδήποτε απάτης. Επίσης, τα Network Analytics είναι μία άλλη σημαντική τεχνολογία στις τηλεπικοινωνίες, όπου οι πάροχοι υπηρεσιών κινητής τηλεφωνίας έχουν την δυνατότητα να παρακολουθούν και να διαχειρίζονται ολόκληρο το δίκτυο και να επιλύουν κάθε πρόβλημα που μπορεί να παρουσιαστεί σε μόλις λίγα λεπτά, αλλά και να βελτιώσουν την ποιότητα των υπηρεσιών και της εμπειρίας χρήσης των πελατών τους.

Retail: Η εξέλιξη του ηλεκτρονικού εμπορίου, οι διαδικτυακές αγορές, οι συνομιλίες μέσω κοινωνικού δικτύου έχουν συμβάλει στην τεχνολογική εξέλιξη του λιανικού εμπορίου με κατεύθυνση τα Μεγάλα δεδομένα. Τα περισσότερα καταστήματα λιανικής τοποθετούν σύστημα CCTV για την παρακολούθηση όχι μόνο κλοπών αλλά και των αγοραστικών συνηθειών των πελατών τους με βάση την ηλικία και το φύλλο τους. Με βάση αυτές τις αγοραστικές συνήθειες οι λιανοπωλητές χρησιμοποιώντας το data mining και συγκεκριμένα την μέθοδο Market Basket Analysis δημιουργούν ομάδες προϊόντων με σκοπό να οδηγήσουν τους καταναλωτές να αγοράσουν παραπάνω προϊόντα απ' ότι χρειάζονται. Σήμερα, οι εταιρίες ηλεκτρονικού εμπορίου χρησιμοποιούν το Market Basket Analysis και άλλα συστήματα για να στοχεύσουν τους κατάλληλους πελάτες, με σκοπό να παρακολουθήσουν τα “click” τους, την συμπεριφορά τους και να τους προτείνουν τα κατάλληλα προϊόντα. Επίσης σημαντικό ρόλο έχουν και τα analytics στην διαχείριση του αποθέματος των εταιριών λιανικής. Για παράδειγμα η εταιρία Stage Stores Inc., χρησιμοποιεί τα analytics για την πρόβλεψη παραγγελιών για διαφορετικές περιοχές.

Banking: Τα Big Data είναι ωφέλιμα στον τραπεζικό τομέα γιατί βοηθούν στην ανίχνευση απάτης και οποιαδήποτε παράνομης δραστηριότητας που μπορεί να συμβεί, καθώς και για τη διαχείριση και συντήρηση κάθε τραπεζικού συστήματος.

Με λίγα λόγια, τα Big Data μπορούν να προσδιορίσουν την κατάχρηση πιστωτικών και χρεωστικών καρτών, την τροποποίηση δημόσιων αναλυτικών στοιχείων για επιχειρήσεις αλλά και την ανάλυση της επενδυτικής αξίας πελατών χρησιμοποιώντας δημογραφικά στοιχεία, την συμπεριφορά τους και την οικονομική τους κατάσταση.

Government: Σήμερα πολλές κυβερνήσεις όπως η Ηνωμένες Πολιτείες και η Ινδία, χρησιμοποιούν την εξόρυξη δεδομένων για να παρακολουθήσουν τις πολιτικές τάσεις της χώρας, την ανάλυση των συναισθημάτων του πληθυσμού, καθώς και τον εντοπισμό εθνικών αλλά και τοπικών ζητημάτων. Επίσης, κάνουν χρήση συστημάτων Big Data για την καλύτερη διαχείριση και χρησιμοποίηση πολύτιμων πόρων και βοηθητικών προγραμμάτων της χώρας. Για παράδειγμα, μπορούν να τοποθετήσουν ειδικούς αισθητήρες σε αγωγούς των αλυσίδων ύδρευσης για να παρακολουθήσουν κατά πόσο σωστή είναι η ροή του νερού σε μεγάλα δίκτυα, με σκοπό να προβλέψουν και να εντοπίσουν διαρροές αλλά και παράνομες συνδέσεις στο σύστημα ύδρευσης πόλεων.

Marketing: Τα Marketing Analytics σήμερα, είναι το κλειδί της επιτυχίας στον συγκεκριμένο τομέα, βοηθώντας τις επιχειρήσεις, να αξιολογήσουν την απόδοση του μάρκετινγκ, να αναλύσουν τις αγοραστικές συνήθειες και την συμπεριφορά των καταναλωτών, με σκοπό να βελτιώσουν τις στρατηγικές μάρκετινγκ που μπορούν να εφαρμόσουν, όπως η τοποθέτηση διαφημίσεων σε μια ιστοσελίδα, η εφαρμογή δυνητικής τιμολόγησης και η προσφορά προϊόντων ανάλογα με τις προτιμήσεις των πελάτων.

2.9 Νομικά ζητήματα των Big Data

Ένα θέμα συζήτησης γύρω από τα “Big Data” είναι η ιδιωτικότητα των δεδομένων που αντλούνται. Η κινητήρια δύναμη των “μεγάλων δεδομένων” είναι αυτή η συλλογή ολοένα και περισσότερων προσωπικών και κρυφών πληροφοριών του καθενός με σκοπό την έρευνα ή την ανάλυση δεδομένων.

Το μεγάλο πρόβλημα για την ιδιωτικότητα σήμερα δεν είναι τόσο η αύξηση του κινδύνου της ιδιωτικότητας που προκαλείται από την χρήση των “Big Data”, αλλά το πόσο αλλάζει αυτός ο κίνδυνος. Δηλαδή, το πρόβλημα είναι ότι ο κίνδυνος μεταλλάσσεται διότι δεν βρίσκεται στον σκοπό για τον οποίο συλλέχθηκαν όλες αυτές οι πληροφορίες, ούτε στο ζήτημα της παράνομης επεξεργασίας από αυτόν που συλλέχθηκαν, αλλά στην προσβολή του δικαιώματος της προσωπικής ζωής του καθενός.

Οι υποστηρικτές των “Big Data” πιστεύουν ότι ο θόρυβος γύρω από την ιδιωτικότητα είναι υπερβολικός αφού τα δεδομένα ανωνυμοποιούνται πριν χρησιμοποιηθούν. Όμως οι ειδικοί αναλυτές υποστηρίζουν ότι η ανωνυμοποίηση είτε δεν γίνεται όπως πρέπει ή σε κάθε περίπτωση είναι εύκολος τελικά ο επαναπροσδιορισμός της ταυτότητας των υποκειμένων της επεξεργασίας.

Έτσι, π.χ. οι πληροφορίες που σχετίζονται με την υγεία είναι εύκολο να χρησιμοποιηθούν για τον εντοπισμό και τη δημιουργία προφίλ του ασθενούς . Αυτό θολώνει τη λεπτή γραμμή ανάμεσα στα προσωπικά και μη προσωπικά δεδομένα, καθώς επιτρέπει τη σύνδεση τμημάτων δεδομένων στην πραγματική ταυτότητα ενός προσώπου. Υποστηρίζεται ότι ερευνώντας μια βάση δεδομένων συχνά και θέτοντας διαφορετικές ερωτήσεις είναι δυνατόν να ανακαλύψεις που κατοικεί κάποιος, ή να βρεις τον ιατρικό φάκελο ενός συναδέλφου ή φίλου.

Γι' αυτό έπρεπε να δημιουργηθεί ένα νομικό πλαίσιο το οποίο θα προστατεύει όλες αυτές τις χρήσεις δεδομένων και θα δημιουργεί υποχρεώσεις σε όσους και ότι εμπλέκεται σε αυτή την διαδικασία. Τον Απρίλιο του 2016, η Ευρωπαϊκή Ένωση (ΕΕ) ενέκρινε τον Γενικό Κανονισμό για την Προστασία των Προσωπικών Δεδομένων γνωστό ως General Data Protection Regulation (GDPR), με την ουσιαστική εφαρμογή του να ξεκινά το Μάιο του 2018.

Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) είναι ο ισχυρότερος νόμος περί προστασίας της ιδιωτικής ζωής και της ασφάλειας στον κόσμο. Αν και ψηφίστηκε από την Ευρωπαϊκή Ένωση (ΕΕ), επιβάλλει υποχρεώσεις σε οργανισμούς οπουδήποτε, αρκεί να στοχεύουν ή να συλλέγουν δεδομένα που σχετίζονται με άτομα στην ΕΕ. Ο GDPR επιβάλει μεγάλα χρηματικά πρόστιμα, ύψους δεκάδων χιλιάδων ευρώ, σε όσους παραβιάζουν τα πρότυπα προστασίας της ιδιωτικής ζωής και της ασφάλειας.

Η μεγαλύτερη πρόκληση που καλείται να αντιμετωπίσει ο γενικός κανονισμός προστασίας δεδομένων, είναι η παρουσία των μεγάλων δεδομένων τα οποία ασχολούνται με τεχνικές δημιουργίας και ανάλυσης μεγάλων συνόλων δεδομένων, οι οποίες κάποιες φορές περιλαμβάνουν προσωπικές πληροφορίες. Μια διαδικασία ανάλυσης των “μεγάλων δεδομένων” μπορεί να καταλήξει επιβλαβής για τα άτομα των οποίων τα δεδομένα αναλύονται ή και να επηρεάσει έμμεσα κάποια άλλα μέσα από τις αναλύσεις αυτές, γεγονός που συνθέτει ένα νέο πλαίσιο εφαρμογής εννοιολογικών και πρακτικών αλλαγών, στο ισχύον νομικό καθεστώς.

3. Τεχνολογίες των Big Data

3.1 Data Analytics

Τα Big Data Analytics είναι η επιστήμη της εξέτασης ή της ανάλυσης μεγάλων συνόλων δεδομένων με μια ποικιλία τύπων δεδομένων, δηλαδή, δομημένα, ημι-δομημένα ή μη δομημένα δεδομένα, τα οποία μπορεί να είναι δεδομένα ροής ή δέσμης δεδομένων. Τα Data analytics επιτρέπουν τη λήψη καλύτερων αποφάσεων, την εύρεση νέων επιχειρηματικών ευκαιριών, τον ανταγωνισμό ενάντι επιχειρηματικών αντιπάλων, τη βελτίωση της απόδοσης και της αποτελεσματικότητας και τη μείωση του κόστους χρησιμοποιώντας προηγμένες τεχνικές ανάλυσης δεδομένων.

Τα Big Data analytics επικεντρώνονται στην εξαγωγή σημαντικών πληροφοριών χρησιμοποιώντας αποτελεσματικούς αλγόριθμους στα δεδομένα που συλλαμβάνονται για την επεξεργασία, την ανάλυση και την οπτικοποίηση τους. Αυτό περιλαμβάνει τη διαμόρφωση του αποτελεσματικού αλγορίθμου και συστήματος για την ενσωμάτωση δεδομένων, αναλύοντας τις πληροφορίες που παράγονται έτσι ώστε να δημιουργηθούν άμεσες επιχειρηματικές λύσεις. Για παράδειγμα, στο ηλεκτρονικό εμπόριο, η ανάλυση του τεράστιου όγκου δεδομένων που παράγεται από διαδικτυακές συναλλαγές είναι το κλειδί για την ενίσχυση της αντίληψης των εμπόρων για τη συμπεριφορά των πελατών και των προτύπων αγορών για τη λήψη κατάλληλων επιχειρηματικών αποφάσεων. Ομοίως, σε πολλές σελίδες του Facebook εμφανίζονται κατάλληλες διαφημίσεις, μέσω της ανάλυσης αναρτήσεων στο Facebook, φωτογραφιών και ούτω καθεξής. Όταν χρησιμοποιείτε πιστωτικές κάρτες, οι πάροχοι πιστωτικών καρτών χρησιμοποιούν έναν έλεγχο ανίχνευσης απάτης για να επιβεβαιώσουν ότι η συναλλαγή είναι νόμιμη. Η βαθμολογία πίστωσης των πελατών αναλύεται από χρηματοπιστωτικά ιδρύματα για να προβλέψει εάν ο αιτών θα χρεοκοπήσει. Συνοψίζοντας, το αντίκτυπο και η σημασία των analytics έχει φτάσει σε μεγάλο βαθμό με περισσότερα δεδομένα να συλλέγονται συνεχώς. Τα Analytics θα συνεχίσουν να αυξάνονται έως ότου υπάρξει κάποιο στρατηγικό αντίκτυπο στην αντίληψη των κρυφών πληροφοριών από τα δεδομένα.

Τα analytics χωρίζονται σε 4 κατηγορίες. Αυτές οι κατηγορίες είναι οι εξής:

- Descriptive Analytics: Πληροφορίες για το παρελθόν.
- Diagnostic Analytics: Κατανόηση του τι συμβαίνει και γιατί συμβαίνει.
- Predictive Analytics: Κατανόηση του μέλλοντος.
- Prescriptive Analytics: Συμβουλές για πιθανά αποτελέσματα.

3.1.1 Apache Spark

Το Apache Spark αναπτύχθηκε αρχικά για το Hadoop, αλλά η χρήση του από εταιρείες όλων των μεγεθών, από τους μικρότερους έως τους γίγαντες της τεχνολογίας, όπως: Apple, Facebook, IBM κ.λπ. έγινε τόσο διαδεδομένη που ανήκει πλέον σε μια ξεχωριστή κατηγορία.

Το Apache Spark είναι ένα ευέλικτο εργαλείο ανάλυσης για μεγάλες ποσότητες δεδομένων που λειτουργεί με όλες τις μεγάλες γλώσσες προγραμματισμού, συμπεριλαμβανομένων των Java, Python, R και SQL.

Αυτό το εργαλείο παρέχει ολοκληρωμένη λειτουργικότητα για ανάλυση δεδομένων σε πραγματικό χρόνο, SQL, Machine Learning, επεξεργασία γραφημάτων και πολλά άλλα. Έχει βελτιστοποιηθεί για λειτουργία στη μνήμη και μπορεί επίσης να επιτρέψει τη συνεχή διαδραστική ανάλυση. Σε αντίθεση με τη μαζική επεξεργασία, μεγάλες ποσότητες ιστορικών δεδομένων μπορούν να αναλυθούν παράλληλα με πραγματικά δεδομένα για να ληφθούν αποφάσεις σε πραγματικό χρόνο. Επίσης είναι πολύ χρήσιμο για προγνωστικά αναλυτικά στοιχεία, ανίχνευση απάτης, ανάλυση συναισθημάτων κ.λπ.

Αυτή η γνωστή πλατφόρμα ανάλυσης εγγυάται γρήγορο, εύκολο στη χρήση και ευέλικτο υπολογισμό με αποτέλεσμα να εκτελεί προγράμματα γρηγορότερα από τα πρότυπα Hive και Apache Hadoop / (MapReduce).

Το εύρος εφαρμογών που επιτρέπει το Spark είναι τεράστιο και από τότε που έγινε το έργο Apache πλέον είναι ατελείωτο. Η ενσωμάτωση του spark με το apache είχε ως αποτέλεσμα ο πηγαίος κώδικας του να αυξηθεί σημαντικά και πολλές βιβλιοθήκες και πακέτα τρίτων να συμπεριληφθούν για την εκτέλεση εργασιών σε συγκεκριμένους τομείς ή για την απλοποίηση της χρήσης των υπαρχόντων API.

Συγκεκριμένα, κάποιες από τις λειτουργίες του Spark, όπως API και βιβλιοθήκες δίνουν την δυνατότητα στους επιστήμονες δεδομένων να πραγματοποιήσουν μετασχηματισμό, επεξεργασία, ανάλυση δεδομένων και να πραγματοποιήσουν εξερεύνηση ad-hoc κατά τη διάρκεια της οποίας πρέπει να δοκιμάσουν νέους αλγόριθμους ή να επαληθεύσουν τα αποτελέσματα στο λιγότερο χρονικό διάστημα. Μόλις πραγματοποιηθεί η φάση εξερεύνησης, η λύση παράγεται από μηχανικούς που ενσωματώνουν τις εργασίες ανάλυσης δεδομένων σε μια πιο πολύπλοκη επιχειρηματική εφαρμογή.

Παραδείγματα εφαρμογών που κατασκευάστηκαν χρησιμοποιώντας το Apache Spark περιλαμβάνουν ανάλυση δεδομένων από κινητές συσκευές και Internet of Things (IoT), αναλυτικά στοιχεία γραφημάτων κλίμακας ιστού, ανωμαλία ανίχνευσης συμπεριφοράς χρηστών και επισκεψιμότητας δικτύου για ασφάλεια πληροφοριών, μηχανική μάθηση σε πραγματικό χρόνο, αγωγούς επεξεργασίας ροής δεδομένων, φόρτος εργασίας μηχανικής, επεξεργασία γεωχωρικών δεδομένων και πολλά άλλα.

3.1.2 R

Η R είναι ένα δωρεάν πρόγραμμα υπολογιστή, ανοιχτού κώδικα που μας επιτρέπει να αναλύσουμε δεδομένα σε οποιοδήποτε λειτουργικό σύστημα Windows, Apple και Linux. Με τον όρο «ανάλυση» εννοούμε, πρώτα, διαβάστε τα δεδομένα στο πρόγραμμα και στη συνέχεια λειτουργήστε σε αυτό, σχεδιάζοντας γραφήματα και διαγράμματα, χειρίζοντας τιμές, εφαρμόζοντας στατιστικά μοντέλα και ούτω καθεξής. Επίσης η R είναι τόσο στατιστικό «περιβάλλον» όσο και γλώσσα προγραμματισμού και χρησιμοποιείται ευρέως τόσο σε εμπορικά όσο και σε ακαδημαϊκά περιβάλλοντα. Συντηρείται από μια ομάδα εθελοντών που απελευθερώνουν τακτικά διορθώσεις σφαλμάτων και νέες δυνατότητες.

Η R ξεκίνησε ως εργαλείο για στατιστικούς, εξελίχθηκε από μια γλώσσα που ονομάζεται S που δημιουργήθηκε τη δεκαετία του 1970. Σήμερα, η R παραμένει η πρωταρχική γλώσσα των ακαδημαϊκών στατιστικών, και έχει επίσης εξέχουσα θέση μεταξύ των αναλυτών στις επιχειρήσεις και στην κυβέρνηση. Χρησιμοποιείται όχι μόνο για την κατασκευή στατιστικών μοντέλων, αλλά και για το χειρισμό και τον καθαρισμό δεδομένων και για την ανάπτυξη νέων στατιστικών μεθόδων, προσομοίωσης κτιρίων, για οπτικοποίηση και γενικά για όλα τα εργαλεία χειρισμού δεδομένων που απαιτούν οι στατιστικοί και οι επιστήμονες δεδομένων. Λόγω της ευκολίας με την οποία οι χρήστες μπορούν να αναπτύξουν και να διανείμουν νέες μεθόδους, η R έχει γίνει επίσης το εργαλείο επιλογής σε ορισμένα γρήγορα αναπτυσσόμενα πεδία όπως η βιοστατιστική και η γενετική. Τα άρθρα σχετικά με τις «έρευνες των κορυφαίων εργαλείων που χρησιμοποιούν οι επιστήμονες δεδομένων» αναφέρουν αναπόφευκτα την R ως ένα από τα σημαντικά εργαλεία με τα οποία οι επιστήμονες δεδομένων, καθώς και οι στατιστικοί, πρέπει να είναι εξοικειωμένοι. Επιπλέον, η δημοτικότητα της R είναι τέτοια που υπάρχουν επεκτάσεις στην R που σας επιτρέπουν να συνδεθείτε σε άλλα προγράμματα όπως οι γλώσσες Python και Java, το σύστημα μηχανικής μάθησης H2O, το σύστημα γεωγραφικών πληροφοριών ArcGIS και πολλά άλλα.

3.2 Data Storage

Το Data Storage αποτελείται από μεθόδους και τεχνολογίες που λαμβάνουν και διατηρούν ψηφιακές πληροφορίες σε ηλεκτρομαγνητικά, οπτικά ή πυριτικά μέσα αποθήκευσης. Η αποθήκευση είναι ένα βασικό κομμάτι των ψηφιακών συσκευών, εξαιτίας της μεγάλης ανάγκης που υπάρχει από καταναλωτές και επιχειρήσεις για την διατήρηση προσωπικών τους αρχείων, ή και σημαντικών πληροφοριών για τις επιχειρήσεις.

Η αποθήκευση χρησιμοποιείται συχνά για την περιγραφή των συσκευών και των δεδομένων που είναι συνδεδεμένα στον υπολογιστή μέσω λειτουργιών εισόδου/εξόδου (I / O), συμπεριλαμβανομένων σκληρών δίσκων, συσκευών flash και άλλων τύπων μέσων.

Η σημαντικότητα του χώρου αποθήκευσης οδηγεί στη δημιουργία όλο και περισσότερων νέων δεδομένων, η οποία οφείλεται στα Big Data και στην αφθονία των συσκευών Internet of Things (IoT). Παράλληλα, με την εξέλιξη των συστημάτων αποθήκευσης, η λήψη και η ανάλυση αυτών των δεδομένων έχει την ανάγκη εξελιγμένων τεχνολογιών όπως την τεχνητή νοημοσύνη (AI) με δυνατότητα μηχανικής μάθησης.

Μεγαλύτερα σενάρια εφαρμογών και αναλυτικά στοιχεία βάσεων δεδομένων σε πραγματικό χρόνο έχουν συμβάλει στην εμφάνιση πολύ πυκνών και επεκτάσιμων συστημάτων αποθήκευσης, όπως αποθήκευση υπολογιστών υψηλής απόδοσης, συγκλίνουσα υποδομή, συστήματα αποθήκευσης με δυνατότητα σύνθεσης, αποθήκευση συνδεδεμένη στο δίκτυο (NAS) και πλατφόρμες αποθήκευσης αντικειμένων.

Μέχρι το 2025, αναμένεται ότι θα δημιουργηθούν 163 zettabytes (ZB) νέων δεδομένων, σύμφωνα με έκθεση της εταιρείας αναλυτών ITC IDC. Αυτή η εκτίμηση αντιπροσωπεύει μια πιθανή δεκαπλάσια αύξηση από τα 16 ZB που παράγονται έως το 2016.

3.2.1 Apache Hadoop

Μία γνωστή τεχνολογία, η οποία έχει δημιουργηθεί για να αποφευχθεί η χαμηλή απόδοση και η πολυπλοκότητα που υπάρχει κατά την επεξεργασία και την ανάλυση Μεγάλων Δεδομένων είναι το Apache Hadoop. Ένα θετικό του Hadoop είναι η ικανότητα του να επεξεργάζεται, γρήγορα, μεγάλα σύνολα δεδομένων, χάρη στις παράλληλες συστάδες και το καταναμημένο σύστημα αρχείων. Πιο συγκεκριμένα, σε αντίθεση με άλλες τεχνολογίες, το Hadoop δεν αντιγράφει στη μνήμη ολόκληρα τα απομακρυσμένα δεδομένα για να εκτελέσει υπολογισμούς, αλλά εκτελεί τις εργασίες εκεί όπου αποθηκεύονται τα δεδομένα. Κατά αυτό τον τρόπο, το Hadoop απαλλάσσει το δίκτυο και τους διακομιστές από ένα σημαντικό φορτίο επικοινωνίας, κάνοντας το ικανό να αναζητήσει terabyte δεδομένων σε μόνο λίγα δευτερόλεπτα. Ένα άλλο θετικό του Hadoop είναι η δυνατότητα να εκτελεί προγράμματα και να διασφαλίζει παράλληλα την ανοχή σφαλμάτων, που συναντάται συνήθως σε καταναμημένο περιβάλλον. Για να επιτευχθεί αυτό, αποτρέπει την απώλεια δεδομένων με την αναπαραγωγή δεδομένων σε διακομιστές.

Η ισχύς της πλατφόρμας Hadoop βασίζεται σε δύο κύρια υποσυστήματα: το Hadoop Distributed File System (HDFS) και το πλαίσιο MapReduce. Επιπλέον, οι χρήστες μπορούν να προσθέσουν λειτουργικές μονάδες πάνω στο Hadoop, όπως απαιτείται, σύμφωνα με τους στόχους τους καθώς και τις απαιτήσεις εφαρμογής τους (π.χ. χωρητικότητα, επιδόσεις, αξιοπιστία, επεκτασιμότητα, ασφάλεια). Στην πραγματικότητα, το Hadoop διαθέτει μια σημαντική κοινότητα υποστήριξης, η οποία συνέβαλε στον εμπλουτισμό του οικοσυστήματος με διάφορες ενότητες ανοιχτού κώδικα.

Για την αποθήκευση δεδομένων, το Hadoop βασίζεται στο σύστημα αρχείων του HDFS, αλλά και στην μη σχεσιακή βάση δεδομένων που ονομάζεται Apache HBase.

Hadoop Distributed File System (HDFS)

Το HDFS είναι ένα σύστημα αποθήκευσης δομημένων και μη δομημένων δεδομένων, με δυνατότητα να διατηρεί αρχεία τεράστιου μεγέθους (μεγαλύτερα από ένα terabyte). Υποστηρίζει έως και εκατοντάδες κόμβους σε ένα σύμπλεγμα και παρέχει οικονομική και αξιόπιστη δυνατότητα αποθήκευσης. Ωστόσο, πρέπει να είναι γνωστό ότι το HDFS δεν αποτελεί σύστημα αρχείων γενικής χρήσης. Αυτό συμβαίνει επειδή το HDFS έχει σχεδιαστεί για επεξεργασία παρτίδων υψηλής καθυστέρησης.. Το κύριο πλεονέκτημα του HDFS είναι η φορητότητα του σε ετερογενείς πλατφόρμες υλικού και λογισμικού. Επιπλέον, το HDFS συμβάλλει στη μείωση της συμφόρησης δικτύου και στην αύξηση της απόδοσης του συστήματος μετακινώντας υπολογισμούς κοντά στην αποθήκευση δεδομένων. Τέλος, εξασφαλίζει την αναπαραγωγή δεδομένων για ανεκτικότητα σε σφάλματα. Αυτά τα χαρακτηριστικά εξηγούν την ευρεία υιοθέτησή του.

Apache HBase

Το HBase είναι μια κατανεμημένη μη σχεσιακή βάση δεδομένων. Πρόκειται για ένα έργο ανοιχτού κώδικα που βασίζεται σε HDFS. Το HBase βασίζεται σε μοντέλο δεδομένων κλειδιού / τιμής προσανατολισμένο στη στήλη. Έχει σχεδιαστεί για λειτουργίες χαμηλού λανθάνοντος χρόνου και παρέχει μια ευέλικτη δομημένη φιλοξενία για πολύ μεγάλους πίνακες σε μορφή τύπου BigTable.

Το κύριο πλεονέκτημα του HBase είναι η μεγάλη ευελιξία που παρέχει έναντι των σχεσιακών βάσεων δεδομένων. Πιο συγκεκριμένα, επιτρέπει σε πολλά χαρακτηριστικά να ομαδοποιούνται σε οικογένειες στηλών έτσι ώστε τα στοιχεία μιας οικογένειας στηλών να αποθηκεύονται όλα μαζί. Αυτή η προσέγγιση είναι διαφορετική σε μια σχεσιακή βάση δεδομένων προσανατολισμένη στη σειρά, όπου όλες οι στήλες μιας σειράς αποθηκεύονται μαζί. Επιπλέον, το HBase έχει το πλεονέκτημα να επιτρέπει στους χρήστες να εισάγουν ενημερώσεις για καλύτερη διαχείριση των αλλαγών των απαιτήσεων των εφαρμογών. Ωστόσο, ένα αρνητικό του HBase είναι ο περιορισμός να μην μπορεί να υποστηρίξει μια δομημένη γλώσσα ερωτημάτων όπως το SQL.

Το HBase παρέχει πολλές δυνατότητες, όπως ερωτήματα σε πραγματικό χρόνο, αναζήτηση φυσικής γλώσσας, συνεπή πρόσβαση σε πηγές Big Data, γραμμική και αρθρωτή επεκτασιμότητα, αυτόματη και διαμορφώσιμη θραύση πινάκων (Dimiduk et al., 2013). Το HBase περιλαμβάνεται σε πολλές λύσεις Big Data και ιστότοπους βάσει δεδομένων, όπως για παράδειγμα η πλατφόρμα ανταλλαγής μηνυμάτων του Facebook.

3.2.2 MongoDB

Το MongoDB είναι μια βάση δεδομένων που βασίζεται σε έγγραφα που αναπτύχθηκε από την 10gen και διαχειρίζεται τη συλλογή JSON όπως μορφή εγγράφων που ονομάζεται BSON ή απλά «Binary JSON». Τα έγγραφα στο MongoDB αποθηκεύονται σε μια συλλογή που αποθηκεύεται σε μια βάση δεδομένων.

Κάποια από τα χαρακτηριστικά του MongoDB:

- **Map reduce based Aggregation Framework:** Αυτή η δυνατότητα του MongoDB είναι παρόμοια με το "Group By" που προσφέρεται στη MySQL. Το MongoDB χρησιμοποιεί το πρότυπο MapReduce για την εκτέλεση της συγκέντρωσης. Το MapReduce χρησιμοποιείται γενικά για την παράλληλη επεξεργασία μεγάλου όγκου δεδομένων διανέμοντάς τα σε ομάδες.
- **Schema-Less Database:** Το schema αναφέρεται στη δομή στην οποία πρέπει να αποθηκεύονται τα δεδομένα. Με το schema-less αναφερόμαστε σε δυναμικά schemas σε αντίθεση με στατικά schemas σε σχεσιακές βάσεις δεδομένων. Δεδομένου ότι δεν υπάρχει περιορισμός στα δεδομένα και κάθε έγγραφο στη συλλογή μπορεί να έχει διαφορετικά χαρακτηριστικά το ένα από το άλλο, το ονομάζουμε schema-less.
- **Ad-hoc Querying:** Το MongoDB υποστηρίζει σύνθετα ερωτήματα όπως το regex. Παρομοίως, μπορούμε επίσης να γράψουμε ερωτήματα για τη λήψη δεδομένων μικρότερων ή μεγαλύτερων από μια τιμή ή να χρησιμοποιήσουμε κανονικές εκφράσεις για αντιστοίχιση μοτίβων.
- **Replication and fail-over support:** Το MongoDB υποστηρίζει την αναπαραγωγή διανέμοντας δεδομένα σε διάφορες ομάδες. Αυτό επιτυγχάνεται χρησιμοποιώντας το σύνολο αντιγράφων που ουσιαστικά είναι μια ομάδα περιπτώσεων που φιλοξενούν τα ίδια δεδομένα. Σε ένα σύνολο αντιγράφων, ένας κόμβος ορίζεται ως ο κύριος κόμβος ενώ όλοι οι άλλοι κόμβοι ταξινομούνται ως δευτερεύοντες. Όλες οι λειτουργίες εγγραφής αντιστοιχίζονται στον πρωτεύοντα κόμβο ενώ οι δευτερεύοντες κόμβοι μπορούν να εκτελούν λειτουργίες ανάγνωσης.

3.3 Data Mining

Για μια επιχείρηση τα δεδομένα της ίδιας της εταιρίας και των πελατών της, είναι ένα στρατηγικό πλεονέκτημα. Η δυνατότητα εξαγωγής σημαντικών πληροφοριών οι οποίες βρίσκονται σε αυτά τα δεδομένα, με εξειδικευμένες τεχνικές και μεθόδους και η δραστηριοποίηση σε αυτές τις πληροφορίες, ονομάζεται Εξόρυξη Δεδομένων και θεωρείται όλο και πιο αναγκαία στον κόσμο των επιχειρήσεων.

Η δουλειά ενός επιστήμονα είναι να κατανοήσει τα δεδομένα, να ανακαλύψει τα πρότυπα που καθορίζουν τον τρόπο με τον οποίο λειτουργεί ο φυσικός κόσμος και να

τα ενσωματώσει σε θεωρίες οι οποίες οδηγούν στην πρόβλεψη μελλοντικών νέων καταστάσεων. Η δουλειά του επιχειρηματία είναι να εντοπίσει ευκαιρίες, δηλαδή μοτίβα στην συμπεριφορά και εκμεταλεύοντας τα να οδηγήσουν σε ανταγωνιστικό πλεονέκτημα και κερδοφορία της επιχείρησής του.

Στην εξόρυξη δεδομένων, τα δεδομένα αποθηκεύονται ηλεκτρονικά και η αναζήτηση είναι αυτοματοποιημένη από έναν υπολογιστή. Οι οικονομολόγοι, οι στατιστικοί και οι μηχανικοί επικοινωνίας έχουν από καιρό εργαστεί με κριτήριο ότι τα πρότυπα στα δεδομένα μπορούν να αναζητηθούν αυτόματα, να αναγνωριστούν, να επικυρωθούν και να χρησιμοποιηθούν για πρόβλεψη. Όμως αυτό που πρέπει να αναφερθεί είναι η εντυπωσιακή αύξηση των ευκαιριών για την εύρεση προτύπων στα δεδομένα. Η ασταμάτητη ανάπτυξη των βάσεων δεδομένων τα τελευταία χρόνια, για καθημερινές δραστηριότητες, φέρνουν την εξόρυξη δεδομένων στην πρώτη γραμμή των νέων επιχειρησιακών τεχνολογιών. Καθώς η δημιουργία δεδομένων διπλασιάζεται και οι μηχανές που μπορούν να αναλάβουν την αναζήτηση γίνονται συνηθισμένες, οι δυνατότητες της εξόρυξης δεδομένων αυξάνονται. Υπολογίζεται ότι ο όγκος των δεδομένων που αποθηκεύονται στις παγκόσμιες βάσεις δεδομένων διπλασιάζεται κάθε 20 μήνες, καθιστώντας απαραίτητη την έξυπνη ανάλυση τους, με αποτέλεσμα να δημιουργούνται επιχειρηματικές ευκαιρίες από την διευκρίνιση των μοτίβων που στηρίζουν την εξόρυξη δεδομένων.

3.3.1 RapidMiner

Το Rapid Analytics είναι ένας διακομιστής ανοικτού κώδικα για την εξόρυξη δεδομένων και επιχειρησιακών αναλύσεων το οποίο αποθηκεύει τα δεδομένα, τις διαδικασίες και τα μεταδεδομένα. Συγκεκριμένα το Rapid Miner περιλαμβάνει σε έναν διακομιστή, ETL (Extract – Transform – Load) διαδικασίες, υποβολή αναφορών, γραφήματα, καθώς και το βασικό, το οποίο είναι η εξόρυξη δεδομένων.

Το Rapid Miner είναι από τις πιο διαδεδομένες ανοιχτού κώδικα, λύσεις εξόρυξης δεδομένων ενώ και είναι η πιο δυνατή γραφική διεπαφή χρήστη για το σχεδιασμό διαδικασιών ανάλυσης. Το έργο αυτό δημιουργήθηκε στο Πανεπιστήμιο του Dortmund το 2001 και ύστερα αναπτύχθηκε από την Rapid-I GmbH το 2007 και από τότε χρησιμοποιείται τόσο σε επιχειρήσεις όσο και σε πανεπιστήμια και ερευνητές από πολλές ειδικότητες και κλάδους.

Πιο συγκεκριμένα, το Rapid Miner χρησιμοποιείται από πληροφορικούς, στατιστικούς κ.τ.λ. οι οποίοι ενδιαφέρονται για τις τεχνικές εξόρυξης δεδομένων, μηχανικής μάθησης και στατιστικές μεθόδους διότι αποτελεί ένα ικανό εργαλείο για την ανάπτυξη νέων μεθόδων και τρόπων αναλύσεων. Από την άλλη, το Rapid Miner χρησιμοποιείται και σε άλλους κλάδους της επιστήμης όπως η ιατρική, η φυσική κ.τ.λ, λόγω της ανάγκης για ευέλικτα εργαλεία ανάλυσης. Το συγκεκριμένο εργαλείο προσφέρει ένα ευρύ φάσμα μεθόδων από απλές στατιστικές αξιολογήσεις όπως ανάλυση συσχέτισης με τις διαδικασίες παλινδρόμησης, κατηγοριοποίησης και συσταδοποίησης, καθώς και μείωση διάστασης και βελτιστοποίηση παραμέτρου. Όλα

τα παραπάνω έχουν ως πεδίο εφαρμογής: εικόνες, κείμενα, ήχους και αναλύσεις χρονοσειρών τα οποία αυτοματοποιούνται πλήρως και τα αποτελέσματα τους απεικονίζονται με διάφορους τρόπους.

Τέλος, το Rapid Miner μπορεί να διαβάσει όλα τα είδη αρχείων που παράγει (αρχεία εισόδου, μοντέλα, σύνολα παραμέτρων και χαρακτηριστικών). Επίσης, αναγνωρίζει πολλούς τύπους τιμών όπως: κατηγορικός, αριθμητικός, ακέραιος, πραγματικός, κείμενο, διωνυμικός, πολυωνυμικός, ημερομηνία και χρόνος.

3.3.2 Orange

Η Orange είναι μία βιβλιοθήκη που περιλαμβάνει αντικείμενα της C++ και περιέχει μία μεγάλη ποικιλία αλγορίθμων Machine Learning & Data Mining όπου μπορούν και διαχειρίζονται μεγάλο πλήθος δεδομένων. Παράλληλα, με τη βοήθεια μιας συλλογής από modules της Python μπορεί να προβεί στην δημιουργία κώδικα για γρήγορη προτυποποίηση νέων αλγορίθμων και έλεγχο συστημάτων.

Επιπροσθέτως, η Orange αποτελείται από ένα μεγάλο αριθμό widget (συστατικά) γραφικών που χρησιμοποιούν μια συλλογή από modules της Orange και μπορούν να συγκεντρωθούν με μια εφαρμογή μέσω ενός οπτικού προγραμματισμού εργαλείου που ονομάζεται Orange Canvas.

Έτσι η Orange, αποτελεί ένα χρήσιμο εργαλείο για μηχανική εκμάθηση και εξόρυξη δεδομένων το οποίο μπορούν να χρησιμοποιήσουν έμπειροι χρήστες και ερευνητές στην μηχανική εκμάθηση ώστε να αναπτύξουν τους δικούς τους αλγόριθμους, αλλά και αρχάριοι χρήστες που μπορούν να χρησιμοποιήσουν ένα ισχυρό και εύκολο εργαλείο οπτικού προγραμματισμού.

Η Orange περιέχει ένα ποικιλόμορφο περιβάλλον για προγραμματιστές, ερευνητές αλλά και για όσους θέλουν να εμβαθύνουν στην εξόρυξη δεδομένων. Με την Python η οποία είναι μια γλώσσα συγγραφής σεναρίων και προγραμματισμού, η εξόρυξη δεδομένων γίνεται, πια, με πιο αξιοποιήσιμο και απλούστερο τρόπο.

Στην Orange, τα widgets χρησιμοποιούν ένα ειδικά σχεδιασμένο μηχανισμό επικοινωνίας για αντικείμενα όπως σύνολα δεδομένων, λίστες χαρακτηριστικών, τεχνικές εκμάθησης, ταξινομητές, κ.ο.κ, προσφέροντας μια αρκετά εύκολη κατασκευή πολύπλοκων σχημάτων εξόρυξης δεδομένων, που χρησιμοποιεί προσεγγίσεις και τεχνικές τελευταίας τεχνολογίας. Η βασική αρχή της Orange δεν είναι να καλύψει όλες τις μεθόδους της μηχανικής εκμάθησης και εξόρυξης δεδομένων αλλά να καλύψει μόνο αυτές που χρησιμοποιούνται μέσω συστατικών όπου έμπειροί χρήστες μπορούν να αλλάξουν ή και να αντικαταστήσουν με καινούργια

3.4 Data Visualization

Το Data Visualization είναι η οπτική αναπαράσταση δεδομένων, που έχει ως στόχο την παρουσίαση ενός συνόλου πληροφοριών με επεξηγηματικό τρόπο, μέσω της βοήθειας γραφικών μέσων. Επίσης, επιδιώκει να μεταδώσει αποτελεσματικά την πληροφορία με όμορφο τρόπο δίνοντας μία ωραία αίσθηση στο “μάτι”, δημιουργώντας μια ισορροπία μεταξύ της λειτουργίας και της αισθητικής.

Επιπλέον οι οπτικοποιήσεις ανάλογα με τον σκοπό για τον οποίο δημιουργήθηκαν, μπορούν να βοηθήσουν τα άτομα που τις χρησιμοποιούν να συγκρίνουν τιμές και αξίες, να βρίσκουν την άνοδο ή την πτώση αυτών μέσα σε χρονικές περιόδους καθώς και να αναδεικνύουν τις σχέσεις μεταξύ των δεδομένων που αναλύουν.

Εδώ και αρκετά χρόνια, πολλές εφαρμογές οπτικοποίησης έχουν χρησιμοποιηθεί για την παρουσίαση των δεδομένων με στόχο την καλύτερη εξυπηρέτηση του ανθρώπου (διαγράμματα, ιστογράμματα, πίτες κλπ.). Με την άνοδο των νέων τεχνολογιών, η οπτικοποίηση εξελίχθηκε και χάρις την βοήθεια των δυναμικών εφαρμογών και των μεγάλων συνόλων δεδομένων, άρχισαν να δημιουργούνται υπερσύγχρονες εικόνες οι οποίες μπορεί να είναι δισδιάστατες, τρισδιάστατες, να είναι στατικές ή δυναμικές ή και να επιτρέπουν αλληλεπίδραση με το χρήστη, με σκοπό την ανάδειξη νέων εννοιών, που παλαιότερα δεν θα γινόντουσαν ποτέ αντιληπτές.

Το Data Visualization εφαρμόζεται σε πολλά πεδία όπως στο χώρο της τέχνης, της εκπαίδευσης, της επιστήμης ή και των επιχειρήσεων. Στο επιχειρηματικό χώρο συγκεκριμένα, υπάρχουν πολλές δυνατότητες εφαρμογής που συμβάλουν στην καλύτερη αποδοτικότητα της εταιρίας καθώς και στην κατανόηση της συμπεριφοράς των καταναλωτών, προσαρμόζοντας έτσι τα προϊόντα και τις υπηρεσίες τους.

3.4.1 Tableau

Το Tableau είναι ένα πρόγραμμα επιχειρηματικής ευφυΐας, που χρησιμοποιείται ιδιαίτερα ως ένα εργαλείο οπτικοποίησης ικανό να παρέχει διαδραστικές απεικονίσεις άμεσα με τη χρήση του drag and drop. Προσφέρει μια μεγάλη ποικιλία επιλογών, όπως πίτα, γραφήματα, χάρτες, ιστογράμματα κτλ. που χρησιμοποιούν ενημερωτικούς πίνακες ελέγχου, οι οποίοι μπορούν να δημιουργηθούν άμεσα από διαφορετικά σύνολα δεδομένων. Εκτελεί συγκεντρώσεις, επισημαίνοντας ή αναλύοντας τα διαγράμματα με μεγάλη ευκολία, όπου ακόμη και αρχάριοι χρήστες μπορούν να δημιουργήσουν οπτικοποιήσεις για να επισημάνουν γεγονότα σε ένα τεράστιο σύνολο δεδομένων. Το Tableau μπορεί εύκολα να συνδεθεί με τα δεδομένα που είναι αποθηκευμένα σε excel, csv και αρχεία κειμένου και να αναγνωρίσει πεδία και μορφές. Ωστόσο, για την πρόσβαση σε δεδομένα που είναι αποθηκευμένα σε πολλές βάσεις δεδομένων, οι χρήστες του Tableau χρειάζονται κατάλληλους συνδέσμους βάσεων δεδομένων και εξειδίκευσης.

Το Tableau μπορεί να χρησιμοποιηθεί για τον καθορισμό και τον υπολογισμό νέων μεταβλητών και την εκτέλεση απλών χειρισμών δεδομένων με χρήση μαθηματικών τύπων όπως το Excel, αλλά είναι πολύ πιο ισχυρό και μπορεί να διαχειριστεί εκατομμύρια δίσκους όπου το Excel αποτυγχάνει να διαχειριστεί. Η περιορισμένη αναλυτική ικανότητα του Tableau μπορεί να αντικατασταθεί από την ικανότητά του να ενσωματώνεται σε διάφορες πλατφόρμες, συμπεριλαμβανομένης της πλατφόρμας Big Data Hadoop, του στατιστικού εργαλείου ανάλυσης R και της υποστήριξης του για το Google Big Query API. Παρ' όλα αυτά, το Tableau παρέχει εγγενή υποστήριξη για ιδιόκτητους παίκτες Big Data όπως το Amazon Red Shift και το TeraData, αλλά δεν έχει ακόμη την άμεση υποστήριξη πολλών διαδεδομένων βάσεων δεδομένων ανοιχτού κώδικα NoSQL όπως το MongoDB.

Για την κάλυψη ποικίλων αναγκών διαφορετικών χρηστών, το Tableau Software προσφέρει διάφορες επιλογές που περιλαμβάνουν τα Tableau Desktop, Tableau Server, Tableau Mobile ή Tableau Public για να διαλέξετε. Το Tableau Desktop είναι ελκυστικό για άτομα και μικρά οργανισμούς, ενώ το Tableau Server είναι κατάλληλο για οργανισμούς με πολλούς χρήστες. Οι καταναλωτές tablet και κινητών τηλεφώνων υποστηρίζονται από το Tableau Mobile, το οποίο πρόσθεσε μεγάλη αξία στο λογισμικό λόγω του πολλαπλασιασμού της χρήσης smartphone μεταξύ ατόμων. Επίσης, αν και το Tableau Public μπορεί να χρησιμοποιηθεί για απεικονίσεις μικρής κλίμακας, έχει περιορισμένες δυνατότητες σε σύγκριση με το Tableau Desktop, καθιστώντας τον ακατάλληλο για απεικονίσεις μεγάλων δεδομένων.

Με τον Tableau Server, μπορεί κανείς να μοιραστεί οπτικοποιήσεις μεταξύ διαφορετικών χρηστών, επιτρέποντας ταυτόχρονα περιορισμένη πρόσβαση σε διαφορετικές προβολές και υποκείμενα δεδομένα εφαρμόζοντας το φίλτρο χρήστη ή τη λειτουργία ασφαλείας επιπέδου γραμμής. Το Tableau παρέχει σημαντική αλληλεπίδραση με αλλαγές στην εμφάνιση στα κλικ των χρηστών και στο ποντίκι πάνω από ορισμένα μέρη της οπτικοποίησης.

Πιο συγκεκριμένα για τις επιλογές του Tableau Software που αναφερθήκαμε παραπάνω, περιλαμβάνονται:

Tableau Desktop

Το Tableau Desktop είναι ένα εργαλείο γεμάτο συγκεκριμένες λειτουργίες. Είναι ο τύπος όπου όλοι οι τύποι διαγραμμάτων, γραφημάτων, αναφορών κ.λπ. δημιουργούνται για να σχηματίσουν ένα πίνακα εργαλείων. Ο πίνακας εργαλείων μπορεί να κοινοποιηθεί σε οποιονδήποτε διακομιστή και μπορεί να διαβαστεί στο Tableau Server.

Το Tableau Desktop χρησιμοποιείται για διαφορετικούς σκοπούς ως συνδεσιμότητα και πηγή δεδομένων με τον χρήστη. Το Tableau Desktop χρησιμοποιείται είτε προσωπικά είτε επαγγελματικά.

- Το Προσωπικό Tableau Desktop χρησιμοποιείται για ιδιωτική λειτουργία, το βιβλίο εργασίας του δεν μπορεί να κοινοποιηθεί δημόσια και έχει περιορισμένη πρόσβαση.
- Το Επαγγελματικό Tableau Desktop χρησιμοποιείται μόνο για επαγγελματική εργασία, μπορεί να κοινοποιηθεί ή να δημιουργηθεί από οποιονδήποτε και δημοσιεύεται στον Tableau Server με πλήρη πρόσβαση σε όλα τα είδη των δεδομένων

Tableau Public

Αυτή είναι η έκδοση Tableau ειδικά κατασκευασμένη για οικονομικά αποδοτικούς χρήστες, που χρησιμοποιείται δημόσια, το φύλλο εργασίας που δημιουργήθηκε σε αυτό αποθηκεύεται στο cloud του Tableau Server, για χρήση ή κοινή χρήση από όλους. Έχει εύκολη και απεριόριστη πρόσβαση για τους χρήστες. Αν κάποιος θέλει να μάθει κάτι για το Tableau ή θέλει να μοιραστεί το έργο του δημοσίως, τότε, το Tableau Public είναι ευνοϊκό.

Tableau Server

Ο Tableau Server χρησιμοποιείται ιδιαίτερα για κοινή χρήση πινάκων ελέγχου, οπτικοποίησης και φύλλων εργασίας που δημιουργούνται στο Tableau Desktop μέσω οποιασδήποτε εταιρείας. Τα φύλλα εργασίας πρέπει να δημοσιεύονται στο Tableau Desktop για να μοιράζονται με το Tableau Server, όπου είναι μόνο τότε προσβάσιμα από τον χρήστη. Ο Tableau Server είναι ένα πολύ ασφαλές λογισμικό για γρήγορη και πραγματική κοινή χρήση δεδομένων εντός ενός οργανισμού.

Tableau Online

Το Tableau Online είναι ένα διαδικτυακό εργαλείο κοινής χρήσης στο οποίο τα δεδομένα αποθηκεύονται στους διακομιστές, τακτοποιούνται πλήρως στο cloud και συντηρούνται από την ομάδα Tableau. Δεν υπάρχει όριο αποθήκευσης των δεδομένων που δημοσιεύονται στο Tableau Online.

Παρόμοια με το Tableau Server, το Tableau Online χρειάζεται επίσης τα φύλλα εργασίας που δημιουργήθηκαν από το Tableau Desktop για τη μετάδοση των δεδομένων.

Όλα τα είδη δεδομένων μπορούν να υποστηριχθούν από το Tableau Online, όπως δεδομένα από διαφορετικές εφαρμογές ιστού, Google Analytics, καταστήματα ηλεκτρονικού εμπορίου όπως Bigbasket, JioMart, Tesco κ.λπ..

3.4.2 D3.js

Το D3.js πρόκειται για μια βιβλιοθήκη ανοιχτού κώδικα JavaScript που αναπτύχθηκε από τον Mike Bostock για τη δημιουργία προσαρμοσμένων διαδραστικών οπτικοποιήσεων δεδομένων στο φυλλομετρητή χρησιμοποιώντας SVG, HTML και CSS.

Με το D3, τα αυθαίρετα δεδομένα μπορούν να συνδεθούν με ένα Document Object Model (DOM) με αποτέλεσμα ο χρήστης να μπορεί να χειριστεί τα δεδομένα καθώς και να μετασχηματίσει βάσεις δεδομένων. Για παράδειγμα, το D3 μπορεί να δημιουργήσει έναν HTML πίνακα με μία σειρά από αριθμούς, ή και να δημιουργήσει ένα γράφημα ράβδων σε SVG με ομαλές μεταβάσεις και αλληλεπιδράσεις.

Το D3 δεν προσπαθεί να παράσχει κάθε πιθανή λειτουργία οπτικοποίησης και να αντιμετωπίσει κάθε πιθανό πρόβλημα. Αντίθετα, το D3 προσπαθεί να επεξεργαστεί, αποδοτικά, κείμενα με βάση τα δεδομένα. Αυτό συνεπάγεται ελαστικότητα, καθώς μπορεί να παρουσιάσει πολλές από τις δυνατότητες των προτύπων του Διαδικτύου, όπως η HTML5 και η SVG.

Με μικρό κόστος το D3 μπορεί να υποστηρίξει αρκετά γρήγορα μεγάλα σύνολα δεδομένων καθώς και δυναμικές τακτικές για αλληλεπίδραση και animation. Επίσης, το D3 υποστηρίζει όλους τους σύγχρονους φυλλομετρητές, καθώς και τα λειτουργικά συστήματα Android και IOS.

Το D3 έχει πολλά πλεονεκτήματα σε σύγκριση με άλλα frameworks. Το μεγαλύτερο του πλεονέκτημα είναι ότι μπορεί να εντοπίσει σφάλματα κάνοντας χρήση του ενσωματωμένου ελεγκτή στοιχείων. Αν και το D3 είναι σχετικά καινούριο framework και δεν υπάρχουν, ακόμα, πολλά βοηθήματα, υπάρχει ενεργή κοινότητα στην Google για βοήθεια με το D3. Επίσης, οι περισσότεροι χρήστες τέτοιων τεχνολογιών γνωρίζουν την JavaScript και την jQuery και η βιβλιοθήκη αυτή χρησιμοποιεί πολλά στοιχεία από τον τρόπο προγραμματισμού σε jQuery. Ένα άλλο πλεονέκτημα της D3 είναι η απλή επεξεργασία των δεδομένων. Συγκεκριμένα, δίνεται στους προγραμματιστές η δυνατότητα να καταλάβουν τι πρέπει να προσθέσουν ή να αφαιρέσουν κάθε φορά που το σύνολο δεδομένων μεταβάλλεται. Τέλος, το D3 μπορεί να προβεί στην δημιουργία στατικών αλλά και αλληλεπιδραστικών γραφικών.

3.5 Αναδυόμενες τεχνολογίες των Big Data

Τα Big Data μπορούν να χρησιμοποιηθούν σε διάφορες εφαρμογές και τομείς, για παράδειγμα, υγειονομική περίθαλψη, μάρκετινγκ, τηλεπικοινωνίες, εκπαίδευση. Αυτές είναι οι περιοχές με τις περισσότερες έρευνες. Η χρήση εργαλείων για την αξιολόγηση και κατανόηση των Big Data θα βοηθήσει οργανισμούς σε διαφορετικούς τομείς να κατανοήσουν τα πλεονεκτήματα της ενσωμάτωσης των Big Data. Αυτό βοηθά τις βιομηχανίες να κινηθούν προς μια πιο εκτεταμένη τεχνολογική κίνηση

καθώς η ενσωμάτωση των Big Data μετατρέπεται σε μέρος των τυπικών επιχειρηματικών πρακτικών. Όπως υποδεικνύεται από τη Susan Hauser, σύμβουλος στρατηγικού σχεδιασμού και προηγούμενη αντιπρόεδρος της Microsoft: "Τα Big Data μπορούν δυνητικά να αλλάξουν τον τρόπο με τον οποίο οι κυβερνήσεις, οι οργανισμοί και τα ιδρύματα υλοποιούν τις δουλειές τους και κάνουν δημοσιοποιήσεις και είναι πιθανό να αλλάξει τον τρόπο με τον οποίο βιώνουν όλοι την καθημερινή τους ζωή".

Η ταχεία επέκταση των δεδομένων έχει επιφέρει μια γρήγορη ανάπτυξη ψηφιοποιημένων πληροφοριών και έχει επίσης αποκτήσει πολλές ερευνητικές ευκαιρίες όσον αφορά τα αναλυτικά στοιχεία (analytics) των Big Data. Τα Big Data δεν έχουν καμία αξία στο κενό. Η δυνητική τους αξία αντιλαμβάνεται μόνο όταν χρησιμοποιούνται για τη λήψη τεκμηριωμένων αποφάσεων.

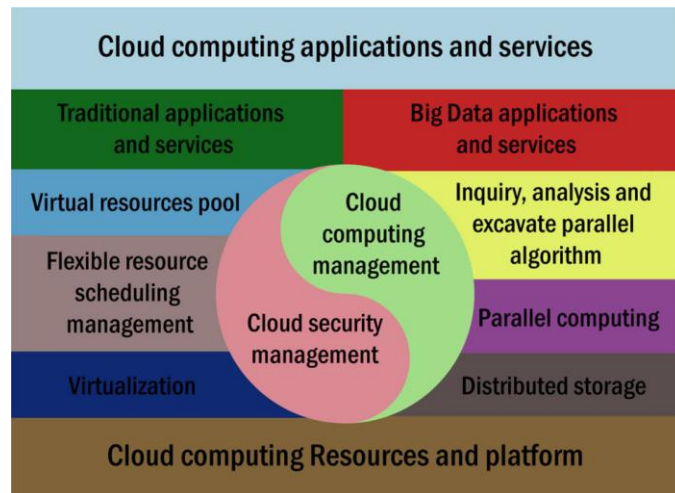
Προς το παρόν, υπάρχουν πολλές τεχνολογίες που χρησιμοποιούν τα Big Data για τον έλεγχο, τη συγκέντρωση, την οπτικοποίηση και την ανάλυση δεδομένων. Αυτά τα εργαλεία και οι τεχνολογίες προέρχονται από πεδία όπως η στατιστική, η μηχανική λογισμικού, η τεχνητή νοημοσύνη, τα εφαρμοσμένα μαθηματικά και ούτω καθεξής. Τα Big Data χρειάζονται εξαιρετικές τεχνολογίες για να επεξεργαστούν αποτελεσματικά εκτεταμένες ποσότητες δεδομένων εντός μιας χρονικής περιόδου που είναι ανεκτή. Η χρήση υπαρχουσών συσκευών και μεθόδων για τη διαχείριση των Big Data οδηγεί σε απώλεια παραγωγικότητας και δημιουργεί πολλές πολυπλοκότητες. Έτσι, οι σημερινές τεχνολογίες δεν είναι σε θέση να επιλύσουν πλήρως τα προβλήματα των Big Data.

Νέες τεχνολογίες αποθήκευσης, επεξεργασίας και ανάλυσης απαιτούνται ολοένα και περισσότερο. Τα Big Data Analytics μετατρέπονται προοδευτικά σε μια τάση που πολλές εταιρείες αγκαλιάζουν με στόχο την ανάπτυξη σημαντικών πληροφοριών χρησιμοποιώντας Big Data.

Παρακάτω βλέπουμε μερικές από τις αναδυόμενες τεχνολογίες που βοηθούν και θα βοηθήσουν μελλοντικά τα Big Data να επιλύσουν πλήρως σημαντικά τους προβλήματα.

1. Cloud Computing

Το Cloud Computing σχετίζεται στενά με τα Big Data. Τα βασικά συστατικά του cloud computing παρουσιάζονται στο Σχήμα 4. Ο κύριος στόχος του cloud computing είναι η χρήση τεράστιων υπολογιστικών και αποθηκευτικών πόρων υπό συγκεντρωμένη διαχείριση, έτσι ώστε να παρέχονται εφαρμογές Big Data με πολύ καλή ικανότητα υπολογισμού. Η ανάπτυξη του cloud computing παρέχει λύσεις για την αποθήκευση και την επεξεργασία των Big Data. Από την άλλη πλευρά, η εμφάνιση των Big Data επιταχύνει επίσης την ανάπτυξη του cloud computing. Η κατανεμημένη τεχνολογία αποθήκευσης που βασίζεται στο cloud computing μπορεί να διαχειριστεί αποτελεσματικά τα Big Data, όπως και η παράλληλη υπολογιστική ικανότητα χάρη στο cloud computing μπορεί να βελτιώσει την αποτελεσματικότητα της απόκτησης και ανάλυσης των Big Data.



Σχήμα 4: Key components of cloud computing
 Πηγή: Min Chen, Shiwen Mao, Yunhao Liu (2014)

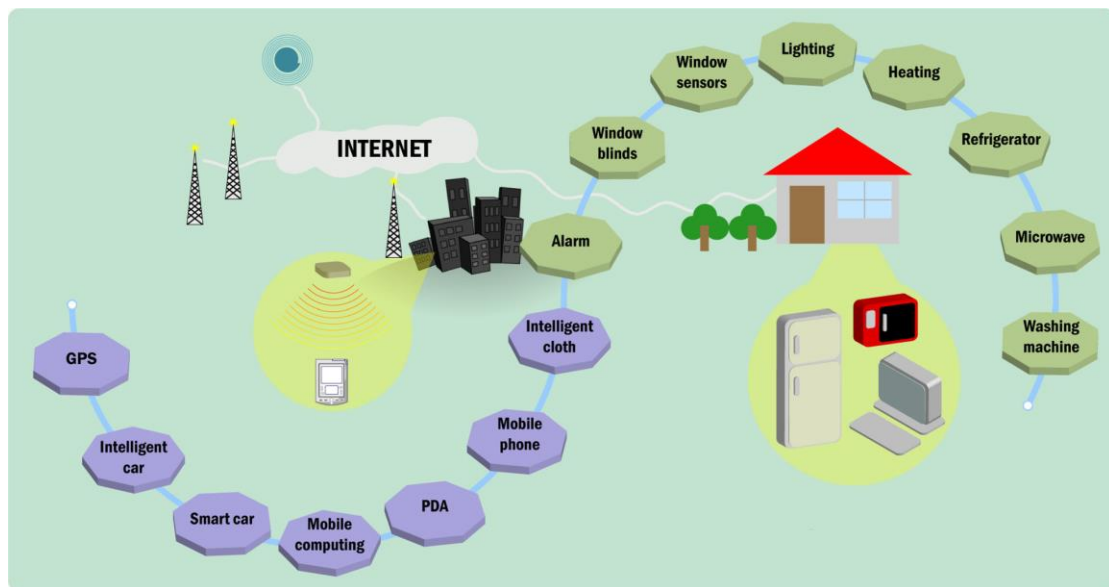
Παρόλο που υπάρχουν πολλές αλληλεπικαλυπτόμενες τεχνολογίες στο cloud computing και τα Big Data, διαφέρουν στις ακόλουθες δύο πτυχές. Πρώτον, οι έννοιες διαφέρουν σε κάποιο βαθμό. Το cloud computing αλλάζει την αρχιτεκτονική υπολογιστών, ενώ τα Big Data επηρεάζουν τη λήψη επιχειρηματικών αποφάσεων. Ωστόσο, τα Big Data εξαρτώνται από το cloud computing ως την βασική υποδομή για ομαλή λειτουργία.

Δεύτερον, τα Big Data και το cloud computing έχουν διαφορετικούς πελάτες-στόχους. Το cloud computing είναι μια τεχνολογία και προϊόν που στοχεύει τους Chief Information Officers (CIO) ως μια προηγμένη λύση πληροφορικής. Τα Big Data είναι ένα προϊόν που στοχεύει τους Διευθύνοντες Συμβούλους (CEO) που εστιάζουν στις επιχειρηματικές δραστηριότητες. Δεδομένου ότι οι υπεύθυνοι λήψης αποφάσεων μπορεί να αισθάνονται άμεσα την πίεση από τον ανταγωνισμό της αγοράς, πρέπει να νικήσουν τους επιχειρηματικούς αντιπάλους με πιο ανταγωνιστικούς τρόπους. Με την πρόοδο των Big Data και του cloud computing, αυτές οι δύο τεχνολογίες είναι σίγουρα και ολοένα και περισσότερο αλληλένδετες. Το cloud computing, με λειτουργίες παρόμοιες με εκείνες των υπολογιστών και των λειτουργικών συστημάτων, παρέχει πόρους σε επίπεδο συστήματος. Από την άλλη, τα Big Data λειτουργούν στο ανώτερο επίπεδο που υποστηρίζεται από το cloud computing και παρέχουν λειτουργίες παρόμοιες με αυτές των βάσεων δεδομένων, καθώς και κατέχουν αποτελεσματική ικανότητα για επεξεργασία δεδομένων. Ο Kissinger, Πρόεδρος της EMC, ανέφερε ότι η εφαρμογή των Big Data πρέπει να βασίζεται, οπωσδήποτε, στο cloud computing.

Η εξέλιξη των Big Data καθοδηγείται από την ταχεία εξέλιξη των απαιτήσεων των εφαρμογών και του cloud computing που αναπτύχθηκε από εικονικοποιημένες τεχνολογίες. Επομένως, το cloud computing όχι μόνο παρέχει καλύτερη υπολογιστική ικανότητα και επεξεργασία για Big Data, αλλά και από μόνο του είναι μια λειτουργία υπηρεσίας. Σε κάποιο βαθμό, οι εξελίξεις του cloud computing προωθούν επίσης την ανάπτυξη Big Data, τα οποία αλληλοσυμπληρώνονται.

2. Internet of Things

Σχετικά με το IoT, ένας τεράστιος αριθμός αισθητήρων δικτύωσης ενσωματώνεται σε διάφορες συσκευές και μηχανές στον πραγματικό κόσμο. Τέτοιοι αισθητήρες που αναπτύσσονται σε διαφορετικά πεδία μπορεί να συλλέγουν διάφορα είδη δεδομένων, όπως περιβαλλοντικά δεδομένα, γεωγραφικά δεδομένα, αστρονομικά δεδομένα καθώς και λογιστικά δεδομένα. Οι φορητοί εξοπλισμοί, οι εγκαταστάσεις μεταφορών, οι δημόσιες εγκαταστάσεις και οι οικιακές συσκευές θα μπορούσαν όλοι να είναι εξοπλισμοί απόκτησης δεδομένων στο IoT, όπως φαίνεται στο Σχήμα 5.



Σχήμα 5: Illustration of data acquisition equipment in IoT

Πηγή: Min Chen, Shiwen Mao, Yunhao Liu (2014)

Τα Big Data που παράγονται από το IoT έχουν διαφορετικά χαρακτηριστικά σε σύγκριση με τα γενικά Big Data λόγω των διαφορετικών τύπων δεδομένων που συλλέγονται. Αν και τα τρέχοντα δεδομένα IoT δεν είναι το κυρίαρχο μέρος των Big Data, μέχρι το 2030, η ποσότητα των αισθητήρων θα φτάσει ένα τρισεκατομμύριο και στη συνέχεια τα δεδομένα IoT θα είναι το πιο σημαντικό μέρος των Big Data, σύμφωνα με τις προβλέψεις της HP. Μια αναφορά από την Intel επεσήμανε ότι τα Big Data στο IoT έχουν τρία χαρακτηριστικά που ανταποκρίνονται στο παράδειγμα Big Data: (i) άφθονα τερματικά που παράγουν μάζες δεδομένων, (ii) τα δεδομένα που δημιουργούνται από το IoT είναι συνήθως ημιδομημένα ή μη δομημένα, (iii) τα δεδομένα του IoT είναι χρήσιμα μόνο όταν αναλύονται.

Προς το παρόν, η ικανότητα επεξεργασίας δεδομένων του IoT έχει μείνει πίσω από τα συλλεγόμενα δεδομένα και είναι εξαιρετικά επείγον να επιταχυνθεί η εισαγωγή τεχνολογιών Big Data για την προώθηση της ανάπτυξης του IoT. Πολλοί χειριστές του IoT αρχίζουν να συνειδητοποιούν τη σημαντικότητα των Big Data, καθώς η

επιτυχία του IoT εξαρτάται από την αποτελεσματική ενσωμάτωση της τεχνολογίας των Big Data και του cloud computing. Η εκτεταμένη ανάπτυξη του IoT θα φέρει επίσης πολλές πόλεις στην εποχή των Big Data.

Υπάρχει επιτακτική ανάγκη υιοθέτησης των Big Data για εφαρμογές IoT, ενώ η ανάπτυξη τους έχει ήδη καθυστερήσει. Έχει αναγνωριστεί ευρέως ότι αυτές οι δύο τεχνολογίες είναι αλληλεξαρτώμενες και πρέπει να αναπτυχθούν από κοινού: από τη μία πλευρά, η εκτεταμένη ανάπτυξη του IoT οδηγεί στην υψηλή ανάπτυξη δεδομένων τόσο σε ποσότητα όσο και σε ποιότητα, παρέχοντας έτσι την ευκαιρία για την εφαρμογή και την ανάπτυξη των Big Data. Από την άλλη πλευρά, η εφαρμογή της τεχνολογίας Big Data στο IoT επιταχύνει επίσης την πρόοδο της έρευνας και τα επιχειρηματικά μοντέλα του IoT.

3. Artificial Intelligence

Τα τελευταία χρόνια, οι επιχειρηματικές τάσεις έχουν μετατοπιστεί προς τη χρήση της τεχνολογικής ανάπτυξης και των νέων πληροφοριών στους τομείς τεράστιων βάσεων δεδομένων πληροφοριών και των συσχετισμών τους με πιθανές επιχειρηματικές ευκαιρίες, όπως τα Big Data με την Τεχνητή Νοημοσύνη.

Τα Big Data με την Τεχνητή Νοημοσύνη αναπτύσσονται εκτενώς για την απόκτηση πληροφοριών και την εξερεύνηση πιθανών ευκαιριών, ώστε εταιρίες να αποκτήσουν πλεονέκτημα έναντι άλλων ανταγωνιστικών επιχειρήσεων. Τα Big Data μαζί με την Τεχνητή Νοημοσύνη γίνονται η νέα “μητέρα της καινοτομίας”, επιτρέποντας στις επιχειρήσεις να λαμβάνουν στρατηγικά αποφάσεις λαμβάνοντας υπόψη τα επιχειρηματικά analytics για να ευθυγραμμιστούν με τους επιχειρηματικούς τους στόχους, να αποκτήσουν ανταγωνιστικό πλεονέκτημα και να αξιολογούν συνεχώς ποιες βελτιώσεις πρέπει να ενσωματωθούν για να διασφαλιστεί η επιτυχία τους.

Τα Big Data με την Τεχνητή Νοημοσύνη έχουν την ικανότητα να παράγουν και να δημιουργούν χρήσιμες πληροφορίες. Αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν για την απόκτηση κερδών, των οποίων η έντονη ανάπτυξη ονομάζεται γενικά ως οικονομία Big Data.

Τα Big Data μαζί με την Τεχνητή Νοημοσύνη έχουν αξιοσημείωτη επίδραση στις επιχειρήσεις επηρεάζοντας σε μεγάλο βαθμό τον τρόπο με τον οποίο δημιουργούνται. Την τελευταία δεκαετία σημειώθηκε η φαινομενική στροφή προς τις διαδικτυακές αγορές για προϊόντα και υπηρεσίες, συμπεριλαμβανομένων όλων των προϊόντων κοινής ωφελείας, αναλώσιμων ειδών, ενδυμάτων και αξεσουάρ, ακόμη και υπηρεσιών, όπως μεταφορές, τραπεζικές, ασφαλιστικές, κρατικές διεπαφές, εξειδικευμένες συμβουλές κ.λπ.

Οι εφαρμογές των Big Data και της Τεχνητής Νοημοσύνης μπορούν, επίσης, να είναι εξαιρετικά χρήσιμες για τις επιχειρήσεις για να τις βοηθήσουν να βελτιώσουν τις δυνατότητες λήψης αποφάσεων και απόδοσης τους. Τα Big Data και η Τεχνητή Νοημοσύνη σχετίζονται με την πραγματική συμπεριφορά ενός ατόμου, η οποία ποτέ δεν προοριζόταν να αποκαλυφθεί ή να εκφραστεί σκόπιμα από αυτό το άτομο. Αυτό είναι διαφορετικό από το άτομο που αναζητά αποτελέσματα αναζήτησης μέσω

Google ή μέσω δημοσίευσης στο Facebook, η οποία μπορεί να περιγράψει ή όχι τα ακριβή χαρακτηριστικά της προσωπικότητας του.

Αυτό το τεχνολογικό εργαλείο έχει καταστεί στρατηγικά σημαντικό, καθώς επιβεβαιώνει τις πληροφορίες που σχετίζονται με τις αγοραστικές συμπεριφορές των ατόμων, απεικονίζει τις συνήθειες αναζήτησης ατόμων όταν, πού, τι και ούτω καθεξής, τα οποία μπορούν να αποκαλυφθούν από τον επιστήμονα δεδομένων με την χρήση της Τεχνητής Νοημοσύνης. Οι επιστήμονες δεδομένων διαθέτουν επίσης τις δυνατότητες πρόβλεψης στο βαθμό που τα άτομα π.χ. είτε θα αποπληρώσουν το δάνειο, θα αναπτύξουν οποιαδήποτε ασθένεια ή θα αγοράσουν εισιτήρια. Ως εκ τούτου, η σημασία του είναι εξαιρετικά σημαντική καθώς αντικατοπτρίζει την αντίληψη, τις επιθυμίες και τις προθέσεις των ατόμων που συνήθως αλλάζουν γρήγορα και απαιτεί από έναν αναλυτή δεδομένων να εξαγάγει συμπεράσματα με ουσιαστικό τρόπο για να αποφευχθούν οι ατέλειες δεδομένων κατά τη λήψη αποφάσεων.

4. Blockchain

Ως μία από τις πρωτοπόρες βιομηχανίες που αγκάλιασαν την εποχή των Big Data από το αρχικό της στάδιο, η χρηματοοικονομική βιομηχανία έχει τη φήμη ότι είναι πλούσια από χρήσιμα δεδομένα και κάνει καινοτόμες αλλαγές για να συμβαδίζει με τις τεχνολογικές εξελίξεις. Η σύγχρονη ψηφιακή οικονομία καθώς και τεχνολογίες όπως το Blockchain και το Cryptocurrency ώθησαν τη νέα γενιά χρηματοοικονομικών υπηρεσιών σε έναν ευρύτερο ορίζοντα, ενσωματώνοντάς την σε σχεδόν όλες τις πτυχές της ανθρώπινης ζωής. Είναι ενδιαφέρον ότι και οι αναπτυσσόμενες χώρες βελτιώνονται ραγδαία όσον αφορά τη δημιουργία της τεχνολογικής και χρηματοοικονομικής υποδομής για να αγκαλιάσουν και να επωφεληθούν από τη νέα γενιά ψηφιακών χρηματοοικονομικών υπηρεσιών (Demirguc Kunt et al. 2018).

Το FinTech αναφέρεται στην Χρηματοοικονομική Τεχνολογία, η οποία είναι ένας όρος ομπρέλας για τεχνολογικές λύσεις, υπηρεσίες ή καινοτομίες επιχειρηματικών μοντέλων στον χρηματοοικονομικό τομέα (Arner et al. 2015; Philippon 2016). Η FinTech δραστηριοποιείται σήμερα σε πέντε τομείς: χρηματοδότηση και επενδύσεις, διεθνής λειτουργία και διαχείριση κινδύνων, πληρωμές και υποδομές, ασφάλεια δεδομένων και δημιουργία εσόδων και διεπαφή πελατών (Arner et al. 2015). Για όλους αυτούς τους τομείς, το Big Data και τα αναλυτικά του στοιχεία, καθώς και η τεχνολογία Blockchain, έπαιξαν σημαντικούς ρόλους (Gai et al. 2018).

Αν και η νέα γενιά της εξέλιξης της FinTech με τα Big Data έφερε αμέτρητες ευκαιρίες και προόδους χρηματοοικονομικών υπηρεσιών και προϊόντων, υπάρχουν μεγάλες προκλήσεις που συνοδεύουν τα μεγάλα οφέλη. Για παράδειγμα, οι γενικές ανησυχίες για την ασφάλεια, το απόρρητο, τον περιορισμό του υλικού (hardware) και

η βιωσιμότητα είναι μερικά παραδείγματα από τις προκλήσεις που αντιμετωπίζουν τα Big Data.

Για το σκοπό αυτό, η τεχνολογία Blockchain αναμένεται να είναι μια βασική λύση για την αντιμετώπιση αυτών των προκλήσεων και χρησιμεύει ως συμπληρωματική τεχνολογία για τα Big Data. Όπως ισχυρίζεται η Fedak (2018), το Blockchain όχι μόνο κάνει τα Big Data ακόμη μεγαλύτερα, αλλά συμβάλλει κάνοντας τα Big Data πιο ασφαλή και πολύτιμα.

4. Τα Big Data στην επιχειρηματικότητα

4.1 Τι είναι η επιχειρηματικότητα;

Η έννοια της επιχειρηματικότητας έχει στόχο την επίτευξη οικονομικού κέρδους ως αποτέλεσμα εκτέλεσης συγκεκριμένων ενεργειών. Παρόλα αυτά η έννοια της επιχειρηματικότητας εξαιτίας των μεταβολών των αναγκών της κάθε εποχής, συνεχίζει να παραμένει ασαφής. Για αυτό τον λόγο, πρώτος ο ακαδημαϊκός Alfred Marshall, το 1890 αναγνώρισε την σημασία της επιχειρηματικότητας, αναφέροντας μέσα στο έργο του, «Principles of Economics», ότι η επιχειρηματικότητα αποτελεί την κινητήρια δύναμη της οργάνωσης που συνδυάζει τους τρεις συντελεστές της παραγωγικής διαδικασίας δηλαδή γη, εργασία και κεφάλαιο. Επιπλέον, ο πατέρας και θεμελιωτής του Management P Drucker, συνείσφερε σημαντικά στην έννοια της επιχειρηματικότητας, δηλώνοντας ότι η επιχειρηματικότητα είναι η πρακτική στην οποία, η γνώση, αποτελεί το μέσο για την επίτευξη του σκοπού της.

Επίσης μια σχετική δημοσίευση από το Harvard Business School, παρουσιάζει την επιχειρηματικότητα, ως μια διαδικασία δημιουργίας μιας ευκαιρίας και την εκμετάλλευση της.

Βάσει οικονομικών και κοινωνιολογικών μελετών, η επιχειρηματικότητα, χαρακτηρίζεται σαν μια πλήρη διαδικασία και όχι σαν ένα μετρήσιμο οικονομικό μέγεθος. Οπότε, όταν γίνεται αναφορά στην επιχειρηματικότητα, εξυπακούεται ως «μια δυναμική διαδικασία, όπου τα άτομα προσδιορίζουν συνεχώς τις οικονομικές ευκαιρίες που εμφανίζονται και τις εκμεταλλεύονται πουλώντας αγαθά και υπηρεσίες.»

Κάτω από αυτές τις συνθήκες, έχουμε την μετατροπή όλων των καινοτόμων ιδεών σε οικονομικές ευκαιρίες, γεγονός που αποτελεί τον ακρογωνιαίο λίθο της επιχειρηματικότητας.

Επιπλέον ένας άλλος ορισμός που χρησιμοποιεί ο Οργανισμός Οικονομικής Συνεργασίας & Ανάπτυξης (Ο.Ο.Σ.Α) είναι ότι: *«οι επιχειρηματίες είναι οι φορείς της αλλαγής και της ανάπτυξης της οικονομίας της αγοράς. Είναι εκείνοι που με την δράση τους, δημιουργούν, διαχέουν και εφαρμόζουν καινοτόμες ιδέες, δεν επιδιώκουν να αξιοποιήσουν μόνο τις οικονομικές ευκαιρίες, αλλά έλκονται από την ανάληψη κινδύνου και αρέσκονται να βλέπουν την επιβεβαίωση του προαισθήματός τους».*

Τέλος, μπορούμε να δώσουμε στην επιχειρηματικότητα, έναν ορισμό όσον αφορά την οπτική, οικονομική ή διοικητική πλευρά. Έτσι, από την οικονομική πλευρά, ο επιχειρηματίας, είναι το άτομο που παίρνει την ευθύνη του κινδύνου, λαμβάνοντας αποφάσεις, που είναι καθοριστικές για την κατανομή, τον τύπο, την χρήση των αγαθών, των πόρων, ή των επιχειρήσεων.

Από την πλευρά της διοίκησης, οι επιχειρηματίες χωρίζονται σε δύο κατηγορίες. Σε εκείνους που κυνηγάνε τις ευκαιρίες λαμβάνοντας το ρίσκο και σε εκείνους που διοικούν αποφεύγοντας το ρίσκο.

Επιπλέον, σύμφωνα με τον Joseph Alois Schumpeter η έννοια της καινοτομίας αποτελεί μέρος του ορισμού της επιχειρηματικότητας, υποθέτοντας ότι ο σημαντικότερος σκοπός του επιχειρηματία είναι η δημιουργία, αλλά και η απόκριση σε οικονομικά αδιέξοδα.

Επομένως η αξιοποίηση των ευκαιριών και η μετατροπή των ιδεών σε καινοτομία, είναι η νέα οικονομική δραστηριότητα που προκύπτει. Άρα, η επιχειρηματικότητα θα μπορούσε να οριστεί ως η διαδικασία μέσω της οποίας δημιουργούνται νέες ευκαιρίες για την ανακάλυψη, και αξιοποίηση νέων υπηρεσιών και προϊόντων.

Εξαρτάται βέβαια από τις ικανότητες του επιχειρηματία πως μπορεί να εκμεταλλευτεί τις ευκαιρίες που παρουσιάζονται, καθώς και πως θα οργανώσει την οικονομική τους εκμετάλλευση.

Καταλήγοντας συμπεραίνουμε ότι με όλα τα παραπάνω, πως η δημιουργικότητα και η καινοτομία, απαρτίζουν βασικές έννοιες για την κατανόηση της επιχειρηματικότητας. Έτσι: *«Η επιχειρηματικότητα είναι η ικανότητα και προθυμία των ατόμων, από μόνοι τους, ή σε ομάδες, εντός και εκτός υφιστάμενων επιχειρήσεων, να αντιληφθούν και να δημιουργήσουν νέες οικονομικές ευκαιρίες (νέα προϊόντα, νέες μεθόδους παραγωγής, νέα οργανωτικά συστήματα και νέους συνδυασμούς προϊόντος-αγοράς), και να εισαγάγουν τις ιδέες τους στην αγορά, εν όψει της αβεβαιότητας και άλλων εμποδίων, με τη λήψη αποφάσεων σχετικά με την τοποθεσία, τη μορφή και τη χρήση των πόρων και των θεσμικών οργάνων».* (Papulovaa & Papulab 2015)

4.2 Πως συμβάλλουν τα Big Data στην επιχειρηματικότητα;

Τα προηγούμενα χρόνια, οι επιχειρήσεις, για την διαχείριση και επεξεργασία των πληροφοριών χρησιμοποιούσαν σχεσιακές βάσεις δεδομένων. Οι τεχνικές και μέθοδοι που χρησιμοποιούνταν στα συγκεκριμένα συστήματα στηρίζονταν σε απάντηση απλών ερωτημάτων, αναζητήσεων, διαδικτυακή επεξεργασία συναλλαγών και διαδραστικών οπτικοποιήσεων. Ωστόσο, η μοναδική ευκαιρία που παρείχαν τα παγκόσμια δίκτυα και το World Wide Web (WWW) στις αρχές του 21^{ου} αιώνα, με τις ηλεκτρονικές εφαρμογές, δίνοντας την δυνατότητα στις επιχειρήσεις να αλληλοεπιδρούν άμεσα με τους πελάτες και σε συνδυασμό με την εξέλιξη των Big Data προέτρεψε τις επιχειρήσεις στην εφαρμογή τους για την βέλτιστη αξιοποίηση των συνόλων δεδομένων. Γενικά, η εφαρμογή των Big Data στις επιχειρήσεις μπορεί να ενισχύσει την αποτελεσματικότητα της παραγωγής τους και την ανταγωνιστικότητα σε πολλές πτυχές. Συγκεκριμένα, για τον εμπορικό κλάδο, με τη ανάλυση συσχέτισης των Big Data, οι επιχειρήσεις μπορούν προβλέψουν με μεγαλύτερη ακρίβεια τη συμπεριφορά των καταναλωτών και να δημιουργήσουν νέες λειτουργίες και προϊόντα. Σχετικά με τον προγραμματισμό των πωλήσεων, μετά από σύγκριση των μαζικών δεδομένων, οι επιχειρήσεις μπορούν να βελτιστοποιήσουν βασικά προϊόντα τους σε επίπεδο τιμών, κόστους και ποιότητας. Αναφορικά με τη λειτουργία τους βελτιώνεται η αποτελεσματικότητα, η ικανοποίηση, η βελτιστοποίηση του εργατικού δυναμικού, η ακρίβεια πρόβλεψης απαιτήσεων στην Διπλωματική Εργασία

κατανομή του προσωπικού, αποφεύγεται η πλεονάζουσα παραγωγική ικανότητα και μειώνεται το κόστος εργασίας. Επιπλέον η εφαρμογή των Big Data στις επιχειρήσεις επιτρέπει την ευκολότερη απογραφή, την υλικοτεχνική βελτιστοποίηση και τη συνεργασία με τους προμηθευτές για τον περιορισμό του χάσματος μεταξύ προσφοράς και ζήτησης ενώ και η χρηματοδότηση επιχειρήσεων που στρέφονται στα Big Data έχει αποκτήσει άλλη υπόσταση.

4.3 Ο ρόλος των Big Data στην καινοτομία

Πρόσφατες ερευνητικές μελέτες έδειξαν πώς τα Big Data θα μπορούσαν να φέρουν διαφορετικά είδη καινοτομιών, από τη διαδικασία στο προϊόν, από μια αρχιτεκτονική έως μια αρθρωτή βάση (e.g. Caputo et al., 2016; George and Lin, 2017). Από τη μία πλευρά, τα Big Data μπορούν να βοηθήσουν τις εταιρείες να βελτιώσουν τη χρήση των υπάρχοντων περιουσιακών στοιχείων. Επομένως, τα δεδομένα μπορούν να χρησιμοποιηθούν και να επαναχρησιμοποιηθούν εντός του χώρου όπου έχουν αναπτυχθεί προωθώντας μια λεγόμενη καινοτομία υποδομής βάσει δεδομένων (Tempini, 2017) ή τα Big Data μπορούν να χρησιμοποιηθούν για να αυξήσουν το επίπεδο εξατομίκευσης στις υπηρεσίες, αξιοποιώντας δεδομένα ως παράδειγμα (e.g. Ng and Wakenshaw, 2017).

Από την άλλη, τα Big Data μπορούν να βοηθήσουν στην επέκταση των δυνατοτήτων της εταιρείας σε ένα νέο πλαίσιο, ξεπερνώντας έτσι τα όρια της εταιρείας (Pellegrini, 2017). Τα Big Data μπορούν να οδηγήσουν σε μια ανοιχτή στρατηγική καινοτομίας για την απόκτηση νέων επιχειρηματικών ευκαιριών (Del Vecchio et al., 2018) ή ακόμη και την επιδίωξη ενός επιχειρηματικού μοντέλου που αφορά την καινοτομία για τη διεύρυνση του λειτουργικού τομέα της εταιρείας (Trabucchi et al., 2018). Οι ερευνητές αρχίζουν επίσης να εμβαθύνουν στους οργανωτικούς παράγοντες που μπορούν να διευκολύνουν αυτό το είδος καινοτομίας, όπως η κεντρικότητα των πελατών, η διαχείριση προσανατολισμού δεδομένων και η εφαρμογή ευέλικτων πρακτικών (Troilo et al., 2017).

Ακόμα κι αν η βιβλιογραφία για το θέμα αναπτύσσεται γρήγορα, εξακολουθεί να υπάρχει έλλειψη έρευνας σχετικά με το πώς τα Big Data μπορούν να προκαλέσουν τους υποκείμενους μηχανισμούς δημιουργίας αξίας (Huberty, 2015).

Πράγματι, όλες οι προηγούμενες ερευνητικές μελέτες σχετικά με το θέμα θεώρησαν τα Big Data ως το προϊόν κάποιας διαδικασίας. Η εστίαση μέχρι στιγμής ήταν στις στρατηγικές που μπορούν να επιτρέψουν στις εταιρείες να αποκαλύψουν και να αποτυπώσουν αυτήν την ενσωματωμένη αξία.

Συγκεκριμένα, πρόσφατη έρευνα έδειξε πώς αυτό το υποπροϊόν θα μπορούσε να αποτελέσει την πηγή ενός νέου μηχανισμού πίστωσης αξίας που αξιοποιεί ένα διαφορετικό είδος στρατηγικών. Έχουν περιγραφεί τρεις στρατηγικές για τη λήψη αξίας μέσω Big Data: βελτίωση υπάρχοντων μηχανισμών όπως η διαφήμιση, η κατανόηση με πιο βαθύ τρόπο των τρεχόντων πελατών ή ακόμη και η εύρεση ενός νέου είδους πελάτη που μπορεί να δει μια ευρύτερη αξία σε αυτά τα δεδομένα για να

την αναζητήσει (Trabucchi et al., 2017). Στη βάση αυτού του μοντέλου, υπάρχει η αμφίδρομη θεωρία της αγοράς. Αυτό το είδος αγορών χαρακτηρίζεται από την παρουσία δύο (ή περισσότερων) διαφορετικών ομάδων πελατών που συνδέονται μεταξύ τους μέσω μιας πλατφόρμας που εσωτερικεύει τις εξωτερικές πλευρές του δικτύου που υπάρχουν μεταξύ των διαφόρων ομάδων (π.χ. Rochet και Tirole, 2003).

Αυτές οι αγορές έχουν αναγνωριστεί σε διαφορετικούς κλάδους, όπως πιστωτικές κάρτες ή κινούμενα βίντεο, και έχουν ταξινομηθεί σύμφωνα με τον τύπο συναλλαγής που ενεργοποιείται από την πλατφόρμα (π.χ. Rysman, 2009; Filistrucchi et al., 2014).

Η πρόσφατη ανάπτυξη του θέματος έδειξε πώς τα Big Data θα μπορούσαν να προκαλέσουν τις παραδοσιακές αμφίδρομες αγορές βάσει των διαφημιστικών μηχανισμών. Οι επιχειρήσεις που βασίζονται σε διαφημίσεις αξιοποιούν μια δομή δύο όψεων, καθώς μια πλευρά (οι διαφημιστές) ανυπομονεί να φτάσει στην άλλη πλευρά (οι αναγνώστες ή οι θεατές) μέσω μιας προοπτικής Client-As-a-Target για να τραβήξει την προσοχή τους (Filistrucchi et al., 2014; Trabucchi et al., 2017).

Τα δεδομένα που συλλέγονται μέσω μιας ψηφιακής πλατφόρμας (π.χ. μια εφαρμογή για κινητά) μπορούν να αντιστρέψουν τη θεμελιώδη σχέση μεταξύ των δύο πλευρών, καθώς η πλατφόρμα μπορεί να αξιοποιήσει μία από τις παραπάνω στρατηγικές για να προσελκύσει ένα συγκεκριμένο είδος πελατών στη δεύτερη πλευρά, οι οποίες δείχνουν ενδιαφέρον στα δεδομένα που συγκεντρώθηκαν, αξιοποιώντας μια προοπτική Client-As-a-Source (Trabucchi et al., 2017). Αυτό μπορεί να μεταφραστεί σε διαφορετικές στρατηγικές: ενίσχυση της διαφήμισης (αντιστοίχιση στρατηγικής Client-As-a-Target και Client-As-a-Source για την παροχή διαφημιστικών μηνυμάτων με καλύτερη στόχευση), e-Ethnography (η οποία επιτρέπει στον πάροχο της εφαρμογής να συλλέγει δεδομένα σχετικά με συγκεκριμένες συμπεριφορές των τελικών χρηστών) και εμπορία δεδομένων (η οποία βασίζεται στην ανταλλαγή δεδομένων με τρίτους).

Οι ερευνητές άρχισαν πρόσφατα να επισημαίνουν τα διαφορετικά είδη καινοτομίας που μπορούν να επιδιωχθούν μέσω των Big Data. Ο Chang (2018) επεσήμανε την ανάγκη για περαιτέρω έρευνα σχετικά με το πώς τα Big Data μπορούν να παρέχουν κίνητρα για την προώθηση της καινοτομίας, ειδικά ενώ η τεχνολογική εξέλιξη δίνει τη δυνατότητα συλλογής συγκεκριμένων δεδομένων με έγκαιρο και οικονομικά αποδοτικό τρόπο (Gandomi and Haider, 2015). Στην ίδια προοπτική, ο Rindfleisch et al. (2017) έδειξε πώς τα Big Data μπορούν να επιτρέψουν στις εταιρείες να εκμεταλλευτούν την αξία τους (καινοτομία από δεδομένα), καθώς και να επιτρέψουν ένα συγκεκριμένο είδος καινοτομίας προϊόντων που θα έχει τη δυνατότητα να συλλέξει και να αναλύσει δεδομένα (καινοτομία ως δεδομένα).

4.4 Τα Big Data σε κλάδους των επιχειρήσεων/ Τα Big Data στην λειτουργικότητα των επιχειρήσεων:

4.4.1 Big Data στα Logistics

Οι μεγάλες εταιρείες στον κλάδο των logistics έχοντας, σε καθημερινή βάση μεγάλο όγκο δεδομένων, είναι ιδανικές για να αξιοποιήσουν τις τεχνολογικές και μεθοδολογικές εξελίξεις των Big Data. Με την χρήση προγραμμάτων συλλογής δεδομένων (software) αλλά και εισαγωγής αυτών (data entry applications), τους δίνεται η δυνατότητα να λάβουν γνώση για όλες τις δραστηριότητες – διαδικασίες τους αλλά και πώς να τις κάνουν ποιο αποτελεσματικές.

Στις μέρες μας οι πάροχοι υπηρεσιών logistics (logistics providers) διαχειρίζονται ταυτόχρονα μαζικές ροές προϊόντων οι οποίες με την σειρά τους παράγουν ένα τεράστιο σύνολο από δεδομένα (data sets). Πιο συγκεκριμένα, μερικά από τα δεδομένα που συλλέγονται καθημερινά είναι από εκατομμύρια αποστολές, σημεία προέλευσης και προορισμού, ώρες παράδοσης, διαστάσεις, βάρος, περιεχόμενα, στοιχεία οδηγών, οχημάτων καθώς και tracking συστήματα τα οποία ανανεώνουν συνεχώς την ακριβή τοποθεσία των αποστολών κ.τ.λ.

Παρ' όλα αυτά, οι εταιρείες δεν αξιοποιούν κατάλληλα αυτές τις πληροφορίες. Σύμφωνα με μια έρευνα που διεξήχθη το 2013 (Robert, et al., 2013) σχετικά με τις τάσεις στην εφοδιαστική αλυσίδα, το εξήντα τις εκατό (60%) των ερωτηθέντων σχεδιάζουν να επενδύσουν σε τεχνολογίες των Big Data και Business Analytics μέσα στα επόμενα πέντε χρόνια, βελτιώνοντας την επιχειρησιακή αποδοτικότητα καθώς και την εμπειρία του καταναλωτή (customer experience) (Martin, et al., 2013).

Φανταστείτε πόσα πολλά θα ήταν τα πλεονεκτήματα που θα αποκόμιζε μια εφοδιαστική αλυσίδα με την προσθήκη πολλών logistics providers η οποία θα πρόσφερε μια συνεχόμενη ροή πληροφοριών σε όλα τα μέρη της. Αυτό θα είχε ως αποτέλεσμα να δημιουργηθούν νέες ευνοϊκές συνεργασίες μεταξύ επιχειρήσεων αλλά και νέες καινοτόμες υπηρεσίες. Έτσι οι εταιρείες μπορούν να επεξεργαστούν τα μεγάλα δεδομένα σε ακριβείς προβλέψεις στην ζήτηση της αγοράς, με σκοπό την απόκτηση ενός ανταγωνιστικού πλεονεκτήματος.

4.4.2 Big Data στο Λιανικό Εμπόριο

Το λιανικό εμπόριο είναι ένας κλάδος της αγοράς ο οποίος με την σωστή αξιοποίηση των Big Data και των Business Analytics, μπορεί να προσκομίσει αρκετά οφέλη. Πιο συγκεκριμένα, οι επιχειρήσεις του λιανικού εμπορίου, με την κατάλληλη αξιοποίηση των Big Data και ευκαιριών όπως υποστήριξη πελατών, στοχευμένη διαφήμιση, παρακολούθηση παραγγελιών κ.ο.κ μπορούν να αποκτήσουν σημαντική αύξηση στα έσοδα τους. Ωστόσο, για να επιτευχθεί αυτή η κερδοφορία θα πρέπει

κάθε επιχείρηση να προβεί σε συγχώνευση διοίκησης της εφοδιαστικής αλυσίδας, των καταναλωτών, της εξυπηρέτησης, της διαφήμισης κ.τ.λ.

Το μέγεθος των δεδομένων των μεγάλων επιχειρήσεων του λιανικού εμπορίου υπολογίζεται στα Exabyte (10006 bytes). Αυτό οφείλεται στο ότι οι μεγάλες εταιρείες αντλούν ποικίλα (variety) και μεγάλου όγκου δεδομένα (volume) από τις συναλλαγές των καταναλωτών, την διαχείριση των αποθεμάτων, τις κινήσεις των προϊόντων, τις κάμερες ασφαλείας, τα στοιχεία των διαφημίσεων, αλλά και από τα στατιστικά στοιχεία, τα ιστορικά δεδομένα και με βασικότερο, τα οικονομικά δεδομένα της επιχείρησης. Παράλληλα, η συλλογή των δεδομένων πραγματοποιείται από τα συστήματα Radio Frequency Identification (RFID) τα οποία παρακολουθούν το απόθεμα και την αυτόματη ενημέρωση των προμηθευτών καθώς και καταγράφουν τις συνήθειες των πελατών μέσω προγραμμάτων επιβράβευσης loyalty. Εφόσον οι επιχειρήσεις καταφέρουν να τα εφαρμόσουν, θα μπορέσουν να μεγιστοποιήσουν την αποτελεσματικότητα και την αποδοτικότητα τους, μέσω ενός ενοποιημένου πληροφοριακού συστήματος ERP που θα περιέχει όλα τα παραπάνω. Σύμφωνα με έρευνες, ο όγκος των δεδομένων που επεξεργάζονται οι επιχειρήσεις παγκοσμίως σχεδόν διπλασιάζεται κάθε 1 με 2 χρόνια. Συγκεκριμένα, ως παράδειγμα η Wal-Mart τα τελευταία χρόνια έχει αντλήσει πάνω από 4 Petabyte δεδομένων μέσω της βοήθειας Machine Learning (Chen & Zhang, 2014) πετυχαίνοντας έτσι βελτιωμένη διαχείριση των αποθεμάτων και της αλυσίδας εφοδιασμού.

4.4.3 Big Data στο Ηλεκτρονικό Εμπόριο

Ένας κλάδος ο οποίος συνδέεται με την βέλτιστη αξιοποίηση των Big Data Business Analytics είναι το ηλεκτρονικό εμπόριο (eCommerce). Στο eCommerce μια εταιρία για να πετύχει τους στόχους της, θα πρέπει να προσελκύσει τους πελάτες της μέσω στοχευμένου μάρκετινγκ στα social media και email, εκπτώσεων, διαφημιστικές καμπάνιες και πολλά άλλα. Όλα αυτά, οδηγούν καθημερινά σε μια πληθώρα δεδομένων τα οποία συνήθως χωρίζονται σε δύο κατηγορίες. Τα Δομημένα (structured) και τα Αδόμητα (unstructured) δεδομένα. Πιο συγκεκριμένα, τα δομημένα δεδομένα είναι το όνομα, η διεύθυνση, το φύλο, η ηλικία, οι προτιμήσεις του πελάτη κ.τ.λ, τα οποία συλλέγονται από διάφορες φόρμες συμπλήρωσης. Από την άλλη, τα αδόμητα δεδομένα λόγω της σημαντικότητας τους είναι πιο δύσκολο να συλλεχθούν και να επεξεργαστούν, και συνήθως βρίσκονται στα "κλικ" στις ιστοσελίδες, στο engagement στα social media (like, κοινοποιήσεις) κ.ο.κ. Επιπλέον, η χρησιμοποίηση των αδόμητων δεδομένων, οδηγεί τις eCommerce επιχειρήσεις σε χρήσιμες πληροφορίες που τις βοηθούν σε αύξηση των κερδών τους. Τέλος, οι επιχειρήσεις που ασχολούνται με το eCommerce, λόγω της συνεχής αύξησης του ανταγωνισμού έχουν κατανοήσει την σημαντικότητα των Big Data με αποτέλεσμα όλο και περισσότερες να στρέφονται στην χρήση των Big Data Business Analytics.

4.4.4 Big Data στον Τουρισμό

Η χρήση των Big Data στον τομέα του τουρισμού μπορεί να ενισχύσει τις τουριστικές υπηρεσίες, προσφέροντας την καλύτερη δυνατή εξυπηρέτηση στους πελάτες-τουρίστες μέσω της εξαγωγής συμπερασμάτων σχετικά με την ανθρώπινη δραστηριότητα και συμπεριφορά. Αυτό επιτυγχάνεται, μέσω της χρήσης των μέσων κοινωνικής δικτύωσης (Facebook, Instagram και Twitter) και ιστοσελίδων (booking, trivago κ.τ.λ) από τους ταξιδιώτες, καθ' όλη την διάρκεια του ταξιδιού τους, τα οποία μας παρέχουν πλήθος δεδομένων αντλώντας χρήσιμες πληροφορίες από την ανάλυση τους. Για παράδειγμα, ένας ταξιδιώτης για την διοργάνωση ενός ταξιδιού του θα χρησιμοποιήσει το διαδίκτυο για να αγοράσει εισιτήρια, να βρει τόπο διαμονής, και χρήσιμες πληροφορίες για το μέρος που πρόκειται να επισκεφθεί. Έτσι μέσω των Big Data δίνεται η ευκαιρία στις εταιρείες να αναλύσουν τα μοτίβα αγορών των καταναλωτών σχετικά με ένα ταξιδιωτικό προορισμό, με σκοπό την προσέλκυση τους από την εφαρμογή στοχευμένου μάρκετινγκ/στοχευμένης διαφήμισης. Τέλος, η αποθήκευση και η αξιολόγηση του τεράστιου όγκου δεδομένων που αφορά τους πελάτες/ταξιδιώτες αποτελεί πλέον αντικείμενο εξειδίκευσης για πολλές επιχειρήσεις του τουριστικού κλάδου που σκοπό έχουν την παροχή πιο αποτελεσματικής και υψηλής ποιότητας υπηρεσιών.

4.4.5 Big Data στην Φαρμακοβιομηχανία

Η φαρμακευτική βιομηχανία πρέπει να είναι σε θέση να ελαχιστοποιήσει το κόστος επεξεργασίας και διαχείρισης δεδομένων καθώς και να μεγιστοποιήσει την αξία τους ώστε να αναπτυχθούν καινοτόμες θεραπείες μέσα στα πλαίσια των προκλήσεων της παγκοσμιοποίησης. Αυτό οδηγεί σε έναν αγώνα για την ενσωμάτωση νέων τύπων και πηγών δεδομένων δίχως την ανάγκη για συλλογή από μη δομημένα δεδομένα. Σημαντικό είναι οι φαρμακευτικές εταιρίες να είναι υπάκουες στους κανονισμούς που διέπουν την ασφάλεια των ναρκωτικών, την ασφάλεια της εφοδιαστικής αλυσίδας, το απόρρητο των ασθενών και άλλες ευαίσθητες πληροφορίες.

Τα μεγάλα δεδομένα ενισχύουν την φαρμακευτική βιομηχανία μέσω της βελτιστοποίησης της έρευνας και των κλινικών δοκιμών καθώς και της αποτελεσματικότητας των θεραπειών που παρέχουν. Έτσι, η επένδυση στον εμπλουτισμό και στην διαχείριση των δεδομένων θα προσφέρει στον κλάδο πλήθος πληροφοριών σε πραγματικό χρόνο για βαθύτερη γνώση αλλά και ένα ισχυρό ανταγωνιστικό πλεονέκτημα.

Η φαρμακευτική βιομηχανία είναι μια πηγή με πλήθος δεδομένων, με τους κλάδους της να δημιουργούν και να επεξεργάζονται διάφορους τύπους δεδομένων οποίος ο όγκος τους αυξάνεται εκθετικά κάθε μέρα. Στον κλάδο της υγειονομικής περίθαλψης και των φαρμακευτικών προϊόντων, τα ανεπεξέργαστα δεδομένα παράγονται από ποικίλες εσωτερικές και εξωτερικές πηγές, όπως για παράδειγμα ερευνητικά εργαστήρια, κλινικές δοκιμές και από γιατρούς, ακαδημαϊκούς και ασθενείς. Η ανάλυση και η επεξεργασία αυτών των δεδομένων θα προσφέρει στις φαρμακευτικές

εταιρείες χρήσιμες πληροφορίες για τον εντοπισμό συστατικών που οδηγούν σε θεραπευτικά αποτελέσματα άρα και την ανάπτυξη αποτελεσματικών φαρμάκων. Η ανάπτυξη φαρμάκων είναι μια διαδικασία που απαιτεί πολύ χρόνο και έξοδα. Ωστόσο με την χρήση των Big Data παρέχεται βαθύτερη γνώση σε νέες τεχνικές με αποτέλεσμα την μείωση του κόστους και του χρόνου. Επίσης, με την πάροδο του χρόνου το μοντέλο τιμολόγησης Cost Vs Benefit γίνεται όλο και πιο σημαντικό για τις φαρμακευτικές εταιρείες, καθώς τα νοσοκομεία και οι ασθενείς αρχίζουν να απαιτούν την ανάγκη πραγματικών αποδεικτικών στοιχείων που μπορεί να καταστούν αναγκαία στο μέλλον. Οι φαρμακευτικές εταιρείες μπορούν να ανταποκριθούν σε αυτήν την πίεση κόστους-οφέλους χρησιμοποιώντας τα Big Data, μέσω των οποίων μπορούν να εξορύξουν τα δεδομένα από τα κοινωνικά μέσα και μπορούν να παρουσιάσουν τέτοια στοιχεία στον πραγματικό κόσμο διαφοροποιημένα από τον ανταγωνισμό. Επιπλέον οι ρυθμιστικές αρχές μπορούν να αποφύγουν οποιαδήποτε ανεπιθύμητη ενέργεια ή ιατρικό λάθος μέσω αυτής της εξόρυξης δεδομένων παρέχοντας έτσι μεγαλύτερη ασφάλεια των δεδομένων των ασθενών. Συγχρόνως τα Big Data χρησιμοποιούνται στη βελτίωση της παρακολούθησης βοηθώντας τις εταιρίες να αποθηκεύουν ή να μετακινούν φάρμακα σε ατυχή γεγονότα όπως φυσική καταστροφή ή να προβλέπουν την εμφάνιση πανδημίας που θα σώσει πολλές ζωές σε σύντομο χρονικό διάστημα. Η επικύρωση της αποτελεσματικότητας, της ασφάλειας των ναρκωτικών, η σύγκριση των διαφοροποιημένων προϊόντων, ο σχεδιασμός στρατηγικών μάρκετινγκ κ.λπ. θα γίνει ακόμη πιο σημαντική με τη χρήση των Big Data. Επιπροσθέτως, θα βοηθήσει τις εταιρείες στη δημιουργία νέων ιδεών και καινοτόμων προϊόντων. Αν και οι βιοεπιστήμες υστερούν στη χρήση των Big Data αναδύονται συστήματα αξίας τα οποία οι εταιρείες δεν μπορούν να αγνοούν. Παρόλο, που τα Big Data στον φαρμακευτικό κλάδο αποτελούν ένα απαιτητικό μέλλον, τα οφέλη στις εταιρίες που θα επιτύχουν, θα είναι πολλά και σημαντικά για το μέλλον.

4.4.6 Big Data στον Τραπεζικό Κλάδο

Τα Big Data και η ανάλυση τους στον τραπεζικό κλάδο προσφέρουν την ευκαιρία για βελτιστοποίηση των συναλλαγών, την εκτέλεση στοχευμένου μάρκετινγκ και την παροχή συμβουλευτικής διαχείρισης πλούτου και όλα αυτά με μοναδικό στόχο την διαφοροποίηση από τον ανταγωνισμό.

Επιπλέον οι τραπεζικές εταιρείες με τη χρήση των δεδομένων, έχουν τη δυνατότητα να μοντελοποιούν τη χρηματοοικονομική απόδοση των πελατών τους, να συμβάλουν στη διαχείριση κινδύνων, όπως απάτες σχετικά με την χρήση χρεωστικών και πιστωτικών καρτών, την συμμόρφωση στα χρηματοοικονομικά εγκλήματα και την πιστοληπτική ικανότητα του πελάτη (Siddiqui and Qureshi, 2017). Επιπροσθέτως επωφελούνται στον εντοπισμό αιτιών & αποτυχιών, στον εντοπισμό σημαντικών πελατών και τη διατήρησή τους, στην ανάλυση κινδύνου και στην ορθή χρήση των ανθρωπίνων πόρων.

Τέλος, με τη χρήση των Big Data, οι τραπεζικές επιχειρήσεις μπορούν να καταφέρουν τους στόχους τους, οι οποίοι μπορεί να είναι η μείωση κόστους, η

ελαχιστοποίηση του χρόνου ανάπτυξη ενός νέου προϊόντος (Siddiqui and Qureshi, 2017) κ.τ.λ τα οποία όλα αυτά οδηγούν στην βέλτιστη λήψη αποφάσεων.

5. Μελέτες Περίπτωσης των Big Data στις επιχειρήσεις

5.1 Walmart

Η Walmart είναι ο μεγαλύτερος λιανοπωλητής στον κόσμο και η μεγαλύτερη εταιρεία στον κόσμο με έσοδα, με περισσότερους από δύο εκατομμύρια υπαλλήλους και 20.000 καταστήματα σε 28 χώρες.

Με λειτουργίες σε αυτήν την κλίμακα, δεν αποτελεί έκπληξη το γεγονός ότι έχουν ανακαλύψει από καιρό την αξία των δεδομένων ανάλυσης. Το 2004, όταν ο τυφώνας Sandy έπληξε τις ΗΠΑ, διαπίστωσαν ότι απρόσμενες πληροφορίες θα μπορούσαν να έρθουν στο φως, όταν τα δεδομένα μελετήθηκαν στο σύνολό τους, και όχι ως μεμονωμένα σύνολα. Προσπαθώντας να προβλέψει τη ζήτηση για προμήθειες έκτακτης ανάγκης ενόψει του πλησιέστερου τυφώνα Sandy, η CIO Linda Dillman ανακάλυψε μερικά εκπληκτικά στατιστικά στοιχεία. Εκτός από τους φακούς και τον εξοπλισμό έκτακτης ανάγκης, ο αναμενόμενος άσχημος καιρός οδήγησε σε αύξηση των πωλήσεων φράουλας Pop Tarts σε πολλές άλλες τοποθεσίες.

Η Walmart έχει αναπτύξει το τμήμα Big Data και Analytics από τότε, παραμένοντας συνεχώς στην αιχμή. Το 2015, η εταιρεία ανακοίνωσε ότι βρισκόταν στη διαδικασία δημιουργίας του μεγαλύτερου ιδιωτικού cloud δεδομένων στον κόσμο, για να επιτρέψει την επεξεργασία 2,5 petabytes πληροφοριών κάθε ώρα.

Ποιο πρόβλημα τα Big Data συμβάλουν να επιλύσουν;

Τα supermarket πωλούν εκατομμύρια προϊόντα σε εκατομμύρια ανθρώπους κάθε μέρα. Είναι μια εξαιρετικά ανταγωνιστική βιομηχανία στην οποία βασίζεται μεγάλο μέρος των ανθρώπων που ζουν στον ανεπτυγμένο κόσμο για να τους παρέχει καθημερινά προϊόντα πρώτης ανάγκης. Τα supermarket ανταγωνίζονται όχι μόνο στην τιμή αλλά και στην εξυπηρέτηση πελατών και, ουσιαστικά, στην ευκολία. Έχοντας τα σωστά προϊόντα στο σωστό μέρος τη σωστή στιγμή, έτσι ώστε οι σωστοί άνθρωποι να μπορούν να τα αγοράσουν, παρουσιάζει τεράστια υλικοτεχνικά προβλήματα. Τα προϊόντα πρέπει να τιμολογούνται αποτελεσματικά στο λεπτό, για να παραμένουν ανταγωνιστικά. Και αν οι πελάτες βρουν ότι δεν μπορούν να πάρουν ό, τι χρειάζονται κάτω από μια στέγη, θα ψάξουν κάπου αλλού για να αγοράσουν κάτι που ταιριάζει καλύτερα στο πολυάσχολο πρόγραμμά τους.

Πώς τα Big Data επιλύουν το πρόβλημα;

Το 2011, με την αυξανόμενη επίγνωση για το πώς τα δεδομένα θα μπορούσαν να χρησιμοποιηθούν για να κατανοήσουν τις ανάγκες των πελατών τους και να τους παράσχουν τα προϊόντα που ήθελαν να αγοράσουν, η Walmart δημιούργησε το @WalmartLabs και την ομάδα Fast Big Data για να ερευνήσει και να αναπτύξει νέες πρωτοβουλίες που βασίζονται σε δεδομένα όλης της επιχείρησης.

Το αποκορύφωμα αυτής της στρατηγικής αναφέρθηκε ως Data Cafe - ένας υπερσύγχρονος κόμβος αναλυτικών στοιχείων στην έδρα τους στο Bentonville του Arkansas. Στο Cafe, η ομάδα αναλύσεων μπορεί να παρακολουθεί 200 ροές εσωτερικών και εξωτερικών δεδομένων σε πραγματικό χρόνο, συμπεριλαμβανομένης μιας βάσης δεδομένων 40 petabyte για όλες τις συναλλαγές πωλήσεων των προηγούμενων εβδομάδων. Η έγκαιρη ανάλυση δεδομένων σε πραγματικό χρόνο θεωρείται κλειδί για την επίδοση της επιχειρηματικής απόδοσης.

Ομάδες από οποιοδήποτε τμήμα της επιχείρησης καλούνται να επισκεφθούν το Cafe με τα προβλήματα δεδομένων τους και να συνεργαστούν με τους αναλυτές για να βρουν μια λύση. Υπάρχει επίσης ένα σύστημα που παρακολουθεί τους δείκτες απόδοσης σε ολόκληρη την εταιρεία και ενεργοποιεί αυτοματοποιημένες ειδοποιήσεις όταν φτάσουν σε ένα συγκεκριμένο επίπεδο, καλώντας τις ομάδες που είναι υπεύθυνες για αυτές να μιλήσουν στην ομάδα δεδομένων για πιθανές λύσεις.

Έτσι, μπορούμε να πούμε πως το Data Cafe αποτελεί μια πολύ σημαντική στρατηγική Big Data που επιλύει ένα μεγάλο πρόβλημα που έχουν πολλές αλυσίδες supermarket σήμερα, όπως η Walmart, καθώς έχει οδηγήσει σε μείωση του χρόνου που απαιτείται για την επίλυση κάποιου προβλήματος που εντοπίζεται στην επιχείρηση από, κατά μέσο όρο, 2-3 εβδομάδες σε περίπου 20 λεπτά.

5.2 Google

Περισσότερο από οποιαδήποτε άλλη εταιρεία, η Google είναι πιθανότατα η πιο κατάλληλη για να μας παρουσιάσει τα οφέλη της ανάλυσης και της ερμηνείας των Big Data στην καθημερινή μας ζωή.

Όταν πραγματοποιούμε μια αναζήτηση στο Google, χειριζόμαστε τα Big Data. Το μέγεθος του ευρετηρίου της Google – δηλαδή το αρχείο κάθε ιστοσελίδας που μπορεί να βρει, το οποίο χρησιμοποιείται για την επιστροφή των αποτελεσμάτων αναζήτησης - εκτιμάται ότι ανέρχεται σε περίπου 100 petabytes (ή 100 εκατομμύρια gigabytes) - σίγουρα Μεγάλα Δεδομένα, σύμφωνα με τα πρότυπα όλων.

Ωστόσο την τελευταία δεκαετία, η συγκέντρωση όλων των πληροφοριών στο Διαδίκτυο για να διευκολύνεται η εύρεση πραγμάτων, ήταν μόνο η αρχή του σχεδίου τους. Η Google προχώρησε στην κυκλοφορία προγραμμάτων περιήγησης στο Web, Email, λειτουργικά συστήματα κινητών τηλεφώνων και του μεγαλύτερου διαδικτυακού διαφημιστικού δικτύου στον κόσμο - όλα στηριζόμενα στην τεχνολογία Big Data με την οποία δημιούργησαν ένα τόσο δημοφιλές όνομα στην αγορά.

Ποιο πρόβλημα τα Big Data συμβάλουν να επιλύσουν;

Το Διαδίκτυο είναι ένα μεγάλο μέρος, το οποίο αναπτύσσεται με εκπληκτικό ρυθμό καθημερινά, χωρίς να δείχνει σημάδια επιβράδυνσης. Ωστόσο αυτό το μέγεθος είναι ένα πρόβλημα από μόνο του. Όταν έχουμε πρόσβαση σε σχεδόν όλα όσα έχει γνωρίσει ποτέ κάποιος, πώς βρίσκουμε αυτό που χρειαζόμαστε για να μας βοηθήσει να λύσουμε τα προβλήματά μας;

Όχι μόνο είναι μεγάλο, το Διαδίκτυο είναι πολύ διαδεδομένο. Οι πληροφορίες μεταφορτώνονται σε διακομιστές που ενδέχεται να βρίσκονται οπουδήποτε στον κόσμο, πράγμα που σημαίνει ότι όποιος θέλει να περιηγηθεί στα δεδομένα που είναι διαθέσιμα, συνδέεται σε υπολογιστές που μερικές φορές βρίσκονται χιλιάδες μίλια μακριά. Η λήψη μεμονωμένων ψηφίων συγκεκριμένων δεδομένων στον χρήστη δεν απαιτεί πολύ χρόνο, με την ταχύτητα με την οποία οι πληροφορίες μπορούν να ταξιδεύουν κατά μήκος καλωδίων χαλκού ή οπτικών ινών σε λίγα δευτερόλεπτα. Ωστόσο, υποθέτουμε ότι ο χρήστης γνωρίζει πού βρίσκονται τα δεδομένα στην πρώτη θέση. Η αναζήτηση σε ολόκληρο το Διαδίκτυο, ακόμη και για μια πολύ απλή πληροφορία, αν δεν γνώριζε την ακριβή διεύθυνση IP του υπολογιστή στον οποίο είναι αποθηκευμένη η συγκεκριμένη πληροφορία, θα χρειαζόταν πολύ, πολύς καιρός, εάν δεν είχε ευρητήριο.

Ωστόσο, με δισεκατομμύρια σελίδες πληροφοριών διαθέσιμες στο διαδίκτυο, η δημιουργία ευρητηρίου δεν είναι ασήμαντη. Βέβαια θα χρειαζόταν έναν στρατό ανθρώπων για μια αιωνιότητα για να βρει οτιδήποτε πλησίαζε μια ολοκληρωμένη βάση δεδομένων με τα περιεχόμενα του Διαδικτύου. Έτσι έπρεπε να γίνει αυτόματα, από υπολογιστές. Αυτό έθεσε ένα άλλο πρόβλημα: πώς θα γνωρίζουν οι υπολογιστές ποιες ήταν καλές πληροφορίες και ποιες όχι; Από προεπιλογή, οι υπολογιστές δεν μπορούν να το προσδιορίσουν μόνοι τους. Δεν έχουν καμία έννοια για τη διαφορά μεταξύ χρήσιμων και άχρηστων, εκτός εάν τους διδάξουμε και, ούτως ή άλλως, αυτό που είναι άχρηστο για ένα άτομο μπορεί να είναι κρίσιμο για ένα άλλο άτομο για την επίλυση των προβλημάτων του.

Πώς τα Big Data επιλύουν το πρόβλημα;

Η Google δεν εφεύρε την έννοια της μηχανής αναζήτησης ή ενός ευρητηρίου Ιστού. Αλλά πολύ γρήγορα αφότου ξεκίνησε το 1997, το καθιέρωσαν ως το κορυφαίο τους εργαλείο - έναν τίτλο που συνεχίζει να κατέχει για σχεδόν 20 χρόνια.

Η ιδέα που την καθιέρωσε τόσο γνωστή σε κάθε γωνιά του κόσμου, ενώ οι πρώτοι ανταγωνιστές της, όπως η Alta Vista είναι γνωστή ως Google PageRank. Το PageRank αναπτύχθηκε από τους ιδρυτές της Google, Larry Page και Sergey Brin, προτού ιδρύσουν την εταιρεία, κατά τη διάρκεια έρευνας τους στο Πανεπιστήμιο του Στάνφορντ. Η αρχή του PageRank είναι ότι όσο περισσότερες σελίδες συνδέονται με μια συγκεκριμένη σελίδα, τόσο υψηλότερο είναι το «κύρος» της συγκεκριμένης σελίδας. Η Google δημιούργησε τους πρώτους αλγορίθμους αναζήτησης για να

εκχωρήσει σε κάθε σελίδα στο ευρετήριο της μια κατάταξη με βάση τον αριθμό άλλων ιστότοπων που χρησιμοποιούν παρόμοιες λέξεις-κλειδιά που συνδέονται με αυτήν, καθώς και κατά πόσο έγκυρες είναι αυτές οι σελίδες. Με άλλα λόγια, αυτή είναι μια διαδικασία που περιλαμβάνει τη μετατροπή των μη δομημένων δεδομένων (τα περιεχόμενα των ιστοσελίδων) σε δομημένα δεδομένα που απαιτούνται για την ποσοτικοποίηση αυτών των πληροφοριών και την κατάταξή τους, για χρησιμότητα.

Η Google δημιουργεί το ευρετήριο του Ιστού της στέλνοντας ρομπότ λογισμικού που συλλέγουν όλο το κείμενο και άλλες πληροφορίες, όπως εικόνες ή ήχους, που περιέχονται σε έναν ιστότοπο και τα αντιγράφουν στα τεράστια αρχεία της Google. Με τις πληροφορίες να βρίσκονται πλέον αποθηκευμένες σε ένα μέρος, μπορούν να αναζητηθούν πολύ πιο γρήγορα από τους χρήστες. Σε συνδυασμό με το PageRank και τις μεταγενέστερες εξελίξεις όπως το Γράφημα Γνώσεων, κάνει ό, τι καλύτερο μπορεί για να συνδυάσει το ερώτημά μας με πληροφορίες που θα βρούμε χρήσιμες.

Αποτέλεσμα όλης αυτής της ιδέας είναι πως, σήμερα, η Google αντιπροσωπεύει το 89% της χρήσης αναζήτησης στο Διαδίκτυο. Μεταξύ αυτών, οι πλησιέστεροι ανταγωνιστές Yahoo, Bing και Baidu αντιπροσώπευαν σχεδόν το υπόλοιπο 11%, μόνο.

5.3 Sprint

Η Sprint είναι ένας από τους τέσσερις μεγάλους παρόχους υπηρεσιών κινητής τηλεφωνίας στις ΗΠΑ με περισσότερους από 57 εκατομμύρια συνδρομητές. Αυτό τους δίνει πρόσβαση σε ένα τεράστιο ποσό δεδομένων για τους πελάτες τους, οι οποίοι βασίζονται όλο και περισσότερο στις κινητές τους συσκευές στην καθημερινή τους ζωή.

Το 2012, ίδρυσαν τη θυγατρική Pinsight Media, με σκοπό τη λήψη αυτών των πηγών δεδομένων και τη χρήση τους για την τμηματοποίηση του κοινού για στοχευμένες, πλατφόρμες διαφημίσεων για κινητά.

Ποιο πρόβλημα τα Big Data συμβάλουν να επιλύσουν;

Πολλοί από εμάς πιστεύουμε ότι η διαφήμιση είναι μία ενόχληση ή εισβολή. Γενικά, οι διαφημιστές έχουν πολύ λίγη ιδέα σχετικά με το ποιος διαβιβάζει το μήνυμά τους και, ως αποτέλεσμα, ξοδεύουν πολλά χρήματα μεταφέροντας ένα μήνυμα σε άτομα που απλά δεν ενδιαφέρονται ή δεν μπορούν να αντέξουν οικονομικά, οτιδήποτε είναι αυτό που πωλείται. Όταν συμβεί αυτό, η διαφήμιση καθίσταται άσχετη και η προσπάθεια και τα έξοδα που έχει κάνει ο διαφημιστής για να στείλει το μήνυμά προς αυτό το άτομο έχει χαθεί εντελώς.

Η στοχευμένη διαφήμιση είναι η λύση, καθώς έχει εμφανιστεί στον κλάδο του άμεσου μάρκετινγκ και έχει εξελιχθεί σε όλη την ψηφιακή εποχή. Προσπαθεί να τμηματοποιήσει με όσο το δυνατόν πιο λεπτομερή τρόπο, λαμβάνοντας υπόψη

δημογραφικά, συμπεριφορικά και τοπικά δεδομένα. Υπάρχει όμως ένα μεγάλο πρόβλημα, καθώς πολλές μέθοδοι τμηματοποίησης κοινού βασίζονται σε μεγάλο βαθμό σε αυτοαναφερόμενα δεδομένα. Οι χρήστες μπορούν εύκολα να δημιουργήσουν προφίλ κοινωνικών μέσων με ψευδείς πληροφορίες, για λόγους ανωνυμίας, και πολλά από τα δεδομένα που δημιουργούνται στο διαδίκτυο αποκόπονται από οτιδήποτε μπορεί να τον συνδέσει με έναν πραγματικό δυνητικό πελάτη.

Πώς τα Big Data επιλύουν το πρόβλημα;

Η Pinsight Media χρησιμοποίησε δεδομένα ελέγχου ταυτότητας δικτύου για τη δημιουργία πιο ακριβή και αξιόπιστων (και επομένως πιο πολύτιμων) προφίλ συμπεριφοράς των καταναλωτών, τα οποία τους επιτρέπουν να προσφέρουν πιο ακριβή στοχευμένο κοινό στους διαφημιστές. Αυτό σημαίνει ότι υπάρχουν λιγότερες πιθανότητες να τοποθετήσουν μια διαφήμιση την οποία το κοινό θα θεωρήσει βαρετή ή άσχετη και μια μεγαλύτερη πιθανότητα, η διαφήμιση να βρει το κατάλληλο κοινό το οποίο θα εξετάσει το ενδεχόμενο να ξοδέψει χρήματα.

Αυτό είναι παρόμοιο με τις στοχευμένες διαφημιστικές υπηρεσίες, που είναι κοινές σήμερα, χάρη στο Facebook και το Google, αλλά με τη μεγάλη διαφορά ότι βασίζονται κυρίως σε δεδομένα φορέα δικτύου.

Η Pinsight Media έχει αναπτύξει το δικό της εργαλείο, γνωστό ως πλατφόρμα διαχείρισης δεδομένων (DMP), το οποίο χρησιμοποιείται για τη δημιουργία στοχευμένων διαφημιστικών προφίλ χρησιμοποιώντας αυτά τα μοναδικά δεδομένα, στα οποία έχει πρόσβαση μόνο η Sprint. Αυτό το συνδυάζει με αγορασμένα και ελεύθερα διαθέσιμα εξωτερικά σύνολα δεδομένων για να βελτιώσει περαιτέρω την ακρίβεια με την οποία οι διαφημιστές μπορούν να στοχεύουν τις καμπάνιες τους.

Επιπλέον, αναπτύσσουν τις δικές τους εφαρμογές, όπως εφαρμογές καιρού, αθλητικές εφαρμογές και ένα πρόγραμμα περιήγησης για το κοινωνικό μέσο και υπηρεσία συζήτησης Reddit. Αυτό τους επιτρέπει να συλλέγουν περισσότερες πληροφορίες, οι οποίες μπορούν να συνδεθούν με ένα αναγνωριστικό διαφήμισης χρήστη βασιζόμενο σε "πραγματικό" άτομο, όπως πιστοποιείται μέσω των δεδομένων χρήστη της Sprint.

Έτσι, σε μόλις τρία χρόνια από τότε που κυκλοφόρησε το Pinsight Media, η Sprint σταμάτησε να μην έχει παρουσία στην αγορά διαφημίσεων για κινητές συσκευές και εξυπηρετεί περισσότερες από έξι δεσεκατομμύρια διαφημιστικές εμφανίσεις κάθε μήνα, καθιστώντας την έναν σημαντικό παίκτη στο online διαφημιστικό παιχνίδι για κινητά.

5.4 LinkedIn

Το LinkedIn είναι το μεγαλύτερο διαδικτυακό επαγγελματικό δίκτυο στον κόσμο, με περισσότερα από 410 εκατομμύρια μέλη σε περισσότερες από 200 χώρες. Το LinkedIn συνδέει επαγγελματίες επιτρέποντάς τους να δημιουργήσουν ένα δίκτυο διασυνδέσεων, που αλληλοεπιδρούν μεταξύ τους. Ο ιστότοπος ξεκίνησε από τον Reid Hoffman το 2003, καθιστώντας τον ένα από τα παλαιότερα δίκτυα κοινωνικών μέσων στον κόσμο.

Ποιο πρόβλημα τα Big Data συμβάλουν να επιλύσουν;

Ο ανταγωνισμός μεταξύ των κοινωνικών δικτύων είναι πιο έντονος από ποτέ και εκείνο που είναι στα trends τον έναν χρόνο μπορεί να μην είναι τον επόμενο. Το LinkedIn πρέπει να διασφαλίσει ότι ο ιστότοπος του παραμένει ουσιαστικό εργαλείο για πολυάσχολους επαγγελματίες, βοηθώντας τους να γίνουν πιο παραγωγικοί και επιτυχημένοι, είτε χρησιμοποιούν την premium υπηρεσία (επί πληρωμή) είτε την δωρεάν υπηρεσία. Ως εκ τούτου, τα Big Data βρίσκονται στον πυρήνα των λειτουργιών και της λήψης αποφάσεων του LinkedIn, βοηθώντας τους να παρέχουν την καλύτερη δυνατή υπηρεσία για τα εκατομμύρια μέλη του ιστότοπου.

Πώς τα Big Data επιλύουν το πρόβλημα;

Το LinkedIn παρακολουθεί κάθε κίνηση που κάνουν οι χρήστες στον ιστότοπο: κάθε κλικ, κάθε προβολή σελίδας, κάθε αλληλεπίδραση. Με 410 εκατομμύρια μέλη, αυτός είναι ένας τεράστιος όγκος για επεξεργασία κάθε μέρα. Οι επιστήμονες δεδομένων και οι ερευνητές στο LinkedIn αναλύουν αυτό το μεγάλο όγκο δεδομένων προκειμένου να βοηθήσουν στη λήψη αποφάσεων και να σχεδιάσουν προϊόντα και λειτουργίες που βασίζονται σε δεδομένα.

Όπως τα άλλα κοινωνικά δίκτυα, το LinkedIn χρησιμοποιεί δεδομένα για να κάνει προτάσεις για τους χρήστες του, όπως "άτομα που ίσως γνωρίζετε". Αυτές οι προτάσεις βασίζονται σε διάφορους παράγοντες, για παράδειγμα εάν κάνετε κλικ στο προφίλ κάποιου ατόμου, εάν εργαζόσασταν στην ίδια εταιρεία κατά τη διάρκεια της ίδιας περιόδου ή εάν μοιράζεστε ορισμένες διασυνδέσεις. Επίσης, επειδή οι χρήστες μπορούν να ανεβάσουν τις επαφές του ηλεκτρονικού ταχυδρομείου τους, το LinkedIn χρησιμοποιεί αυτές τις πληροφορίες για να κάνει προτάσεις, όχι μόνο για τα άτομα που ίσως γνωρίζετε στον ιστότοπο, αλλά και για άτομα που οι επαφές σας μπορεί να γνωρίζουν όταν συνδέονται στον ιστότοπο. Το LinkedIn μπορεί επίσης να αντλήσει δεδομένα για τους χρήστες από άλλους ιστότοπους, όπως το Twitter, για να κάνει προτάσεις για άτομα που ίσως γνωρίζετε.

Μία από τις δυνατότητες που ξεχωρίζουν το LinkedIn από άλλες πλατφόρμες κοινωνικών μέσων όπως το Facebook, είναι ο τρόπος που σας επιτρέπει να δείτε

ποιος έχει δει το προφίλ σας. Και αυτή η λειτουργία πρόσφατα έγινε πολύ πιο λεπτομερής: ενώ κάποτε μπορούσατε να δείτε πόσους είδαν το προφίλ σας και ποιοι ήταν οι πιο πρόσφατοι θεατές, τώρα μπορείτε επίσης να δείτε από ποιες περιοχές και κλάδους προέρχονται οι θεατές, για ποιες εταιρίες εργάζονται και ποιες λέξεις-κλειδιά τους έφερε στο προφίλ σας. Αυτές οι πληροφορίες, που έγιναν δυνατές από τα Big Data, βοηθούν τους χρήστες να αυξήσουν την αποτελεσματικότητά τους στον ιστότοπο.

Το LinkedIn χρησιμοποιεί τεχνολογία επεξεργασίας ροής για να διασφαλίσει ότι οι πιο ενημερωμένες πληροφορίες εμφανίζονται όταν οι χρήστες βρίσκονται στον ιστότοπο: από πληροφορίες σχετικά με το ποιος έχει εγγραφεί στον ιστότοπο και ποιος πήρε μια νέα δουλειά έως χρήσιμα άρθρα που άρεσαν ή μοιράστηκαν οι επαφές. Με λίγα λόγια, ο ιστότοπος συλλέγει και εμφανίζει συνεχώς νέα δεδομένα για τους χρήστες. Όχι μόνο αυτή η συνεχής ροή δεδομένων καθιστά τον ιστότοπο πιο ενδιαφέρον για τους χρήστες, αλλά επιταχύνει την διαδικασία ανάλυσης. Παραδοσιακά, μια εταιρεία θα συλλάβει δεδομένα και θα τα αποθηκεύει σε μια βάση δεδομένων ή σε μία αποθήκη δεδομένων όπου και θα αναλυθούν αργότερα. Ωστόσο, με την τεχνολογία επεξεργασίας ροής, το LinkedIn έχει τη δυνατότητα να μεταδίδει δεδομένα απευθείας από την πηγή (όπως η δραστηριότητα των χρηστών) και να τα αναλύει κατά την διάρκεια της μετάδοσης.

Επίσης, τα Big Data μπορούν να διαδραματίσουν σημαντικό ρόλο στην αύξηση των εσόδων καθώς και στη βελτίωση της εμπειρίας του χρήστη. Για παράδειγμα, στη διαφήμιση, που αντιπροσωπεύει το 20-25% των ετήσιων εσόδων του LinkedIn, οι αναλυτές συνεργάζονται με το δυναμικό πωλήσεων του LinkedIn για να κατανοήσουν γιατί τα μέλη κάνουν κλικ σε συγκεκριμένες διαφημίσεις και όχι σε άλλες. Αυτές οι πληροφορίες μεταφέρονται στη συνέχεια στους διαφημιστές, προκειμένου να κάνουν τις διαφημίσεις τους πιο αποτελεσματικές.

Αποτέλεσμα της χρήσης των Big Data στις στρατηγικές των λήψεων αποφάσεων του LinkedIn είναι πως τα έσοδα και ο αριθμός των μελών συνεχίζουν να αυξάνονται κάθε χρόνο. Το LinkedIn κέρδισε 40 εκατομμύρια νέα μέλη το πρώτο εξάμηνο του 2015, καθώς και τα έσοδα της εταιρείας ξεπέρασαν τα 700 εκατομμύρια δολάρια. Δεν υπάρχει αμφιβολία ότι τα Big Data παίζουν μεγάλο ρόλο στη συνεχή επιτυχία της εταιρείας.

6. Ερευνητικό Μέρος

6.1 Σκοπός της έρευνας

Σκοπός της παρούσας έρευνας είναι η διερεύνηση και αξιολόγηση της στάσης των ελληνικών επιχειρήσεων απέναντι στην χρήση των Big Data και των τεχνολογιών τους. Η συγκεκριμένη έρευνα αναζητά κατά πόσο οι ελληνικές επιχειρήσεις, σήμερα, χρησιμοποιούν ή όχι την τεχνολογία των Big Data, και τον ρόλο που έχουν στην εταιρία. Επίσης, διερευνά τον τρόπο με τον οποίο συλλέγουν και αναλύουν τα δεδομένα τους, οι εταιρίες που κάνουν χρήση των Big Data, καθώς και το όφελος που προσκομίζουν από την χρήση τους. Η παράθεση των αποτελεσμάτων κατά το τελικό στάδιο διεξαγωγής της έρευνας, θα βοηθήσει σε μια πιο ολοκληρωμένη εικόνα της ελληνικής πραγματικότητας απέναντι στο κομμάτι αυτό.

6.2 Επιλογή δείγματος της έρευνας

Κατά την αναζήτηση και καταγραφή των επιχειρήσεων και φορέων που αποτέλεσαν το δείγμα της έρευνάς μας, λήφθηκαν υπόψιν ορισμένες παράμετροι με βάση τις οποίες συντάξαμε και τη λίστα αυτών. Στόχος μας ήταν, να είναι το δείγμα μας όσο πιο αντιπροσωπευτικό γίνεται, προσεγγίζοντας 65 επιχειρήσεις όλων των κλάδων και οικονομικών μεγεθών. Έτσι, καταλήξαμε πως το δείγμα μας πρέπει να αποτελείται από: α) εταιρίες νεοσύστατες (Startups), διότι η πλειοψηφία τους χρησιμοποιεί το διαδίκτυο αλλά και αξιοποιεί τα δεδομένα που βρίσκονται σε αυτό. β) Μικρομεσαίες επιχειρήσεις, οι οποίες αν και δεν διακρίνονται για τα οικονομικά τους μεγέθη, την πρωτοπορία τους κτλ., διαθέτουν έναν μεγάλο όγκο δεδομένων τον οποίο μερικές από αυτές, ήδη, αξιοποιούν ή και θα αξιοποιήσουν στο μέλλον. γ) Μεγάλες επιχειρήσεις ή πολυεθνικές, από τις οποίες, οι περισσότερες αξιοποιούν στο μέγιστο τα Big Data και αποτελούν πρωτοπόρες στον τομέα αυτό. Για την κατηγορία των πολυεθνικών, ζητήσαμε μόνο για τα δεδομένα που ισχύουν στην Ελλάδα. Παράλληλα, απευθυνθήκαμε σε στελέχη και γνώμονες της συγκεκριμένης τεχνολογίας, κυρίως, από κλάδους της α) Υγείας, β) Τραπεζικής, γ) Τηλεπικοινωνιών και δ) Λιανικού εμπορίου. Ωστόσο δεν παραλείψαμε να απευθυνθούμε και σε άλλους κλάδους που έχουν άμεση σχέση με την τεχνολογία των Big Data. Επιπροσθέτως, λόγω της ελευθερίας που δώσαμε στους ερωτηθέντες να προσθέσουν τις δικιές τους απαντήσεις ανεξαρτήτως των προκαθορισμένων επιλογών, ομαδοποιήσαμε τους κλάδους στις κατηγορίες που αναφέρονται στο κεφάλαιο 6.5, σχετικά με τα αποτελέσματα της έρευνας μας, για διευκόλυνση και απλοποίηση των απαντήσεων του δείγματος μας.

Παράλληλα, την ίδια μέθοδο ακολουθήσαμε και στο ερώτημα 1 σχετικά με την θέση των ερωτηθέντων στην επιχείρηση.

6.3 Επιλογή ερευνητικού εργαλείου

Η συλλογή των δεδομένων της έρευνας έγινε με την μορφή του ερωτηματολογίου, το οποίο δημιουργήθηκε με το Google Drive. Το συγκεκριμένο εργαλείο επιλέχθηκε καθώς είναι εύκολο στο χειρισμό. Είναι ανοιχτό και κυρίως πρόκειται για ένα εύκολο πρόγραμμα τόσο για τον ερευνητή όσο και για αυτόν που απαντάει στο ερωτηματολόγιο.

Το ερωτηματολόγιο περιλαμβάνει 17 ερωτήσεις, από τις οποίες, 9 αναφέρονται σε δημογραφικά στοιχεία και στον βαθμό εξοικείωσης που έχουν οι εταιρίες με τα Big Data, και 8 που αναφέρονται στον τρόπο συλλογής και ανάλυσης των δεδομένων των εταιριών.

Η αποστολή του ερωτηματολογίου και η συλλογή των απαντήσεων διήρκεσε από τις 21/02/2021 μέχρι τις 31/03/2021. Ο διαμοιρασμός των ερωτηματολογίων έγινε με ηλεκτρονικό τρόπο και με αποστολή στις εταιρίες, είτε σε επαγγελματικά ηλεκτρονικά ταχυδρομεία είτε και σε προσωπικούς λογαριασμούς των ανώτατων στελεχών. Πριν τη συμπλήρωση του ερωτηματολογίου, οι συμμετέχοντες είχαν ενημερωθεί για την προστασία των προσωπικών τους δεδομένων και ότι η συμμετοχή τους είναι εθελοντική.

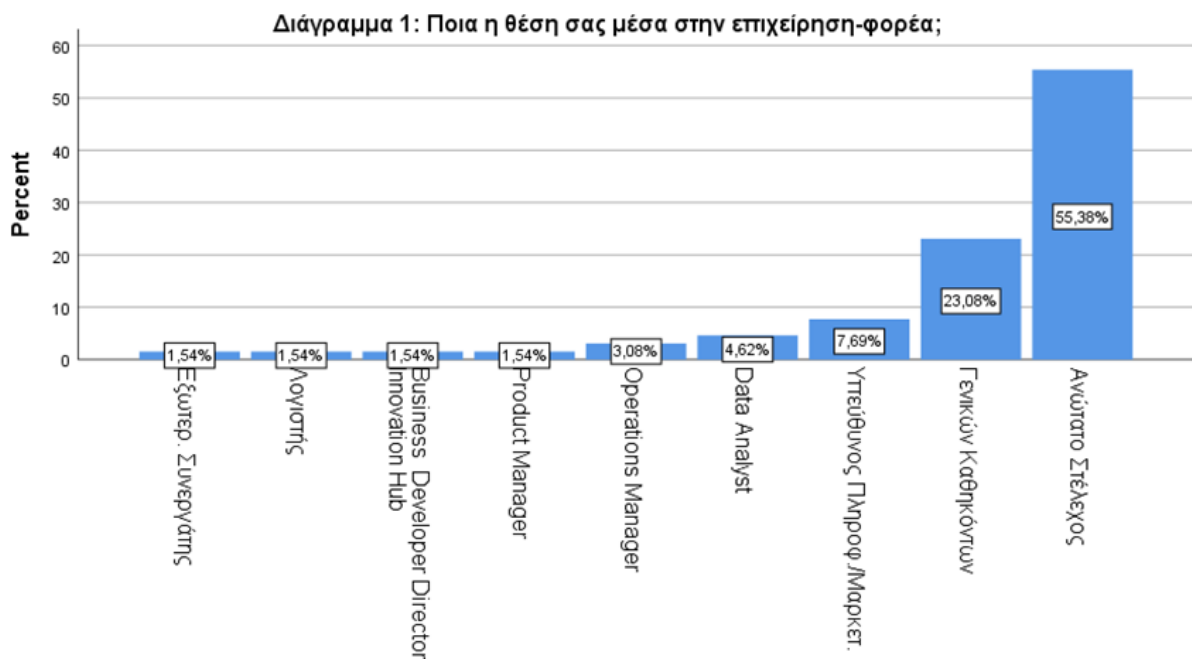
6.4 Επιλογή των ερωτήσεων της έρευνας

Η σύνταξη των 17 ερωτήσεων έγινε μετά από εκτενή βιβλιογραφική ανασκόπηση των τελευταίων 4 ετών και των βιβλιογραφιών που έχουμε χρησιμοποιήσει για την συγγραφή της πτυχιακής μας εργασίας. Οι πρώτες 9 ερωτήσεις ήταν γενικότερης φύσεως και στόχος ήταν η σκιαγράφηση του προφίλ της κάθε ερωτώμενης εταιρίας και τον βαθμό εξοικείωσης που έχει με την τεχνολογία των Big Data. Οι ερωτήσεις αυτές αφορούσαν την θέση που κατείχε το άτομο που απάντησε το ερωτηματολόγιο μέσα στην εταιρία, τον κλάδο στον οποίο δραστηριοποιείται η εταιρία και το φυσικό μέγεθος της εταιρίας (αριθμός προσωπικού). Επίσης, αφορούσαν τις γνώσεις που είχαν πάνω στην τεχνολογία των Big Data και αν κάνουν χρήση αυτής της τεχνολογίας στην εταιρίας τους και για ποιο λόγο. Έπειτα, οι επόμενες 7 ερωτήσεις ήταν πιο ειδικές ερωτήσεις που αφορούσαν τον τρόπο που συλλέγουν και αναλύουν τα δεδομένα τους οι εταιρίες, καθώς και τον βαθμό επίδρασης που έχουν τα Big Data στο ανταγωνιστικό πλεονέκτημα της εταιρίας. Τέλος, έγινε αναφορά και στον ρόλο και την σημαντικότητα που πιστεύουν πως θα έχουν τα Big Data στις εταιρίες των ερωτώμενων.

6.5 Αποτελέσματα της έρευνας

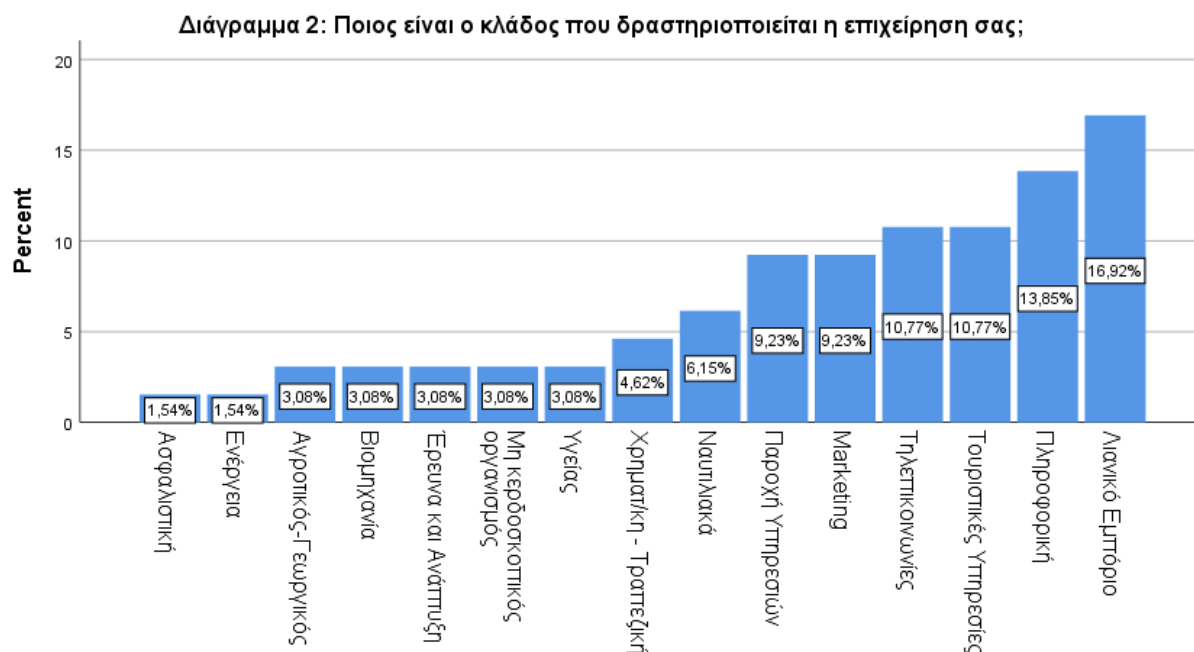
Η έρευνα που διεξήγαμε μετά το πέρας περίπου 40 ημερών που βρισκόταν σε ισχύει, και την ανάκτηση 65 ολοκληρωμένων απαντήσεων, ολοκληρώθηκε και παραθέτουμε τα αποτελέσματά της.

Η πρώτη ερώτηση αφορούσε την θέση του ατόμου που μας απαντά στο ερωτηματολόγιο μέσα στο φορέα ή την επιχείρηση. Από τους συμμετέχοντες της έρευνας, το 55,38%(N=36) είναι ανώτατα στελέχη στην επιχείρηση τους και το 23,08%(N=15) είναι γενικών καθηκόντων. Επιπλέον, το 7,69%(N=5) είναι υπεύθυνοι πληροφορικής και μάρκετινγκ και το 4,62%(N=3) είναι Data Analyst. Τέλος, τις λιγότερες απαντήσεις μας δώσανε οι operation manager με ποσοστό 3,08%(n=2) και ακολουθούν οι Product Manager, Business Developer Director, Λογιστής και Εξωτερικός Συνεργάτης με ποσοστό 1,54%(N=1) αντίστοιχα. (Διάγραμμα 1)



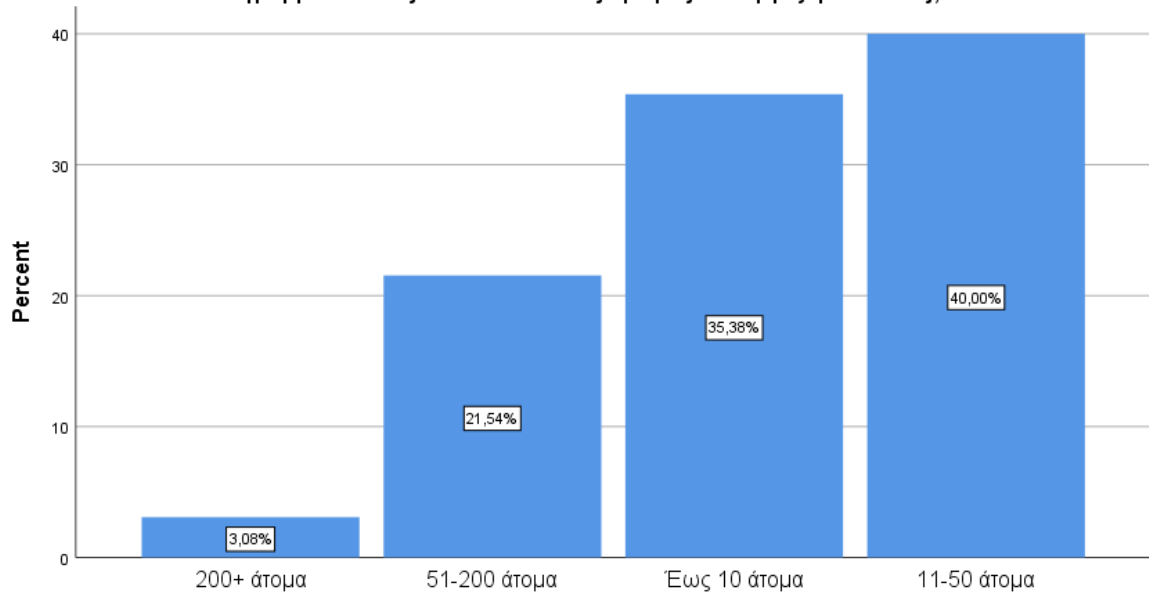
Εν συνέχεια, προσπαθήσαμε να προσδιορίσουμε τον κλάδο στον οποίο εντάσσεται η κάθε επιχείρηση καθώς και το μέγεθος της. Αρχικά, βλέπουμε ότι ο κλάδος του Λιανικού Εμπορίου έχει το μεγαλύτερο ποσοστό στην έρευνα μας με 16,92%(N=11) και ακολουθεί ο κλάδος της πληροφορικής με 13,85%(N=9). Επίσης στο 10,77%(N=7) είναι οι τηλεπικοινωνίες και οι τουριστικές υπηρεσίες αντίστοιχα. Έπειτα ακολουθεί το μάρκετινγκ και η παροχή υπηρεσιών με ποσοστό 9.23%(N=6) . Επιπλέον, έχουμε τα ναυτιλιακά με 6.15%(N=4) και τον κλάδο της χρηματοοικονομικής-τραπεζικής με 4.62%(N=3). Παράλληλα, με ποσοστό Διπλωματική Εργασία

3.08%(N=2) βρίσκονται αντίστοιχα οι κλάδοι υγείας, έρευνας και ανάπτυξης, βιομηχανίας, γεωργίας και μη κερδοσκοπικοί οργανισμοί. Στο τέλος του διαγράμματος βρίσκεται ο κλάδος ενέργειας και ασφαλιστικής με ποσοστό 1.54%(N=1). (Διάγραμμα 2)

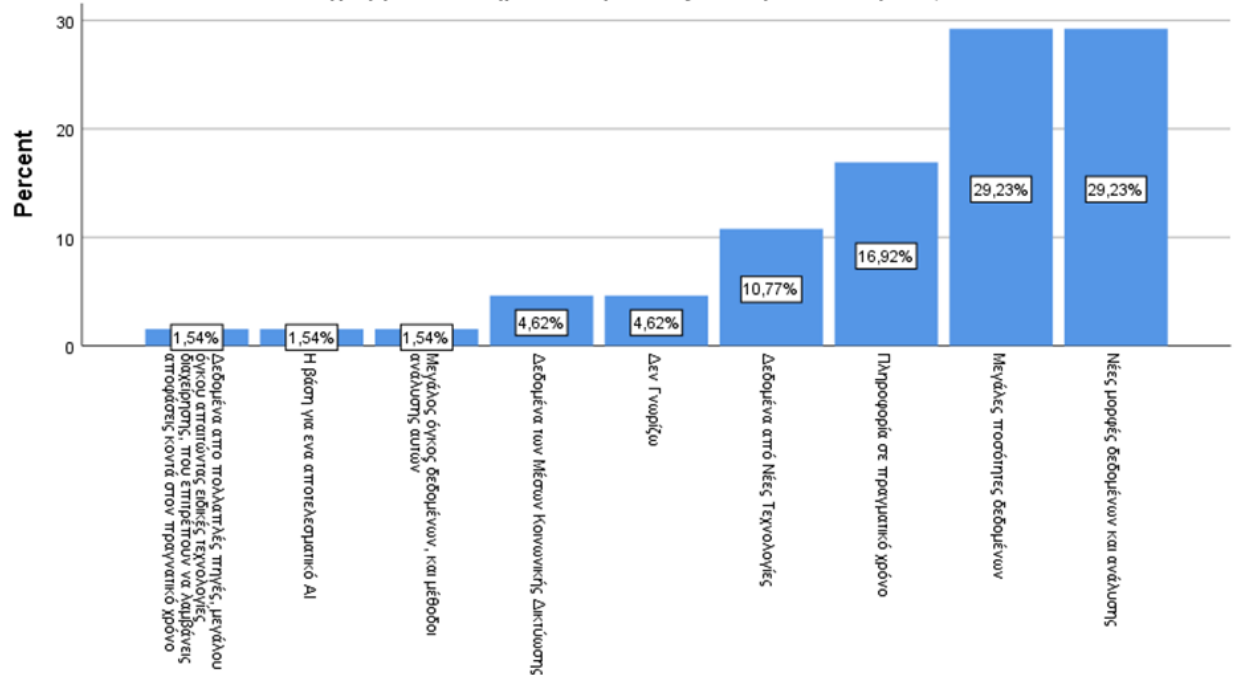


Έπειτα για να προσδιορίσουμε το μέγεθος των επιχειρήσεων, ερωτήσαμε τους συμμετέχοντες σχετικά με το ανθρώπινο δυναμικό της επιχείρησής στην οποία ανήκουν. Αρχικά, παρατηρούμε ότι οι περισσότερες επιχειρήσεις με ποσοστό 40%(n=26) απασχολούν από 11 έως 50 άτομα. Έπειτα, στο 35.8%(N=23) βρίσκονται οι μικρές επιχειρήσεις οι οποίες απασχολούν μέχρι και 10 άτομα. Τέλος, στο 21.54%(N=14) βρίσκονται οι επιχειρήσεις από 51 έως 200 άτομα στο ανθρώπινο δυναμικό τους και μόλις με 3.048%(N=2) βρίσκονται οι οργανισμοί που απασχολούν πάνω από 200 εργαζομένους. (Διάγραμμα 3)

Διάγραμμα 3: Ποιος είναι ο συνολικός αριθμός των εργαζομένων σας;



Διάγραμμα 4: Τί σημαίνουν για εσάς τα Μεγάλα Δεδομένα;

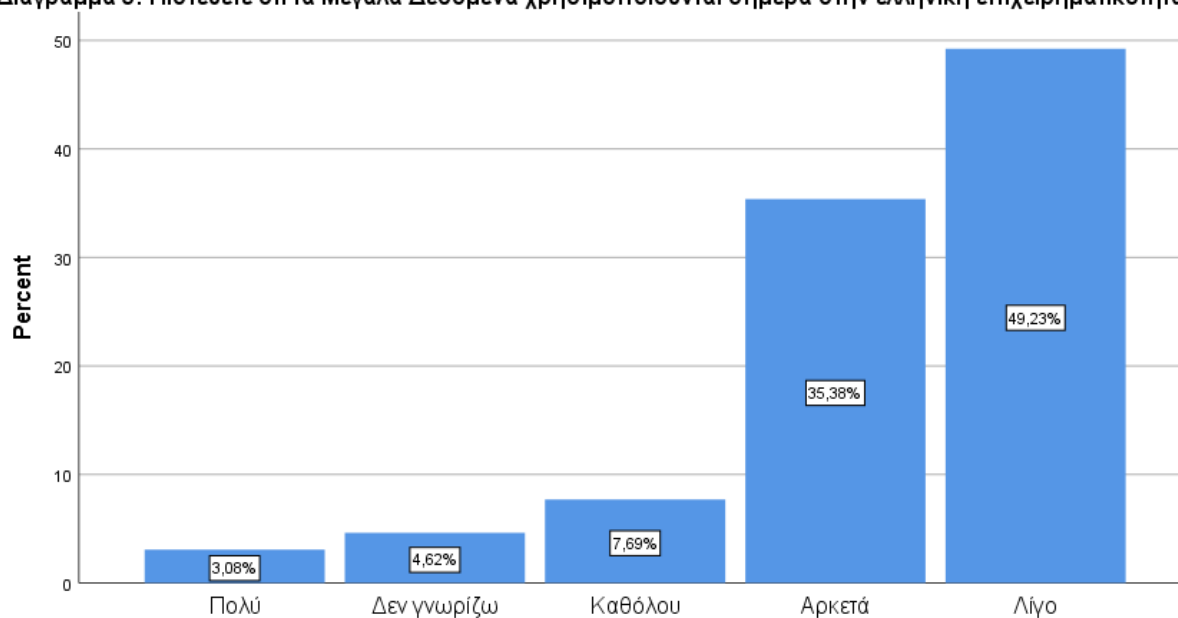


Στην συγκεκριμένη ερώτηση (Διάγραμμα 4), θέλαμε να μάθουμε από τους συμμετέχοντες, τι σημαίνουν για τους ίδιους και τις επιχειρήσεις τους τα Μεγάλα Δεδομένα. Αρχικά, το 29,23%(N=19) του δείγματος θεωρεί πως τα μεγάλα δεδομένα είναι νέες μορφές δεδομένων και ανάλυσης ενώ συγχρόνως το ίδιο ποσοστό πιστεύει πως είναι μεγάλες ποσότητες δεδομένων. Όμοια, ένα 16,92%(N=11) θεωρεί πως αποτελούν πληροφορία σε πραγματικό χρόνο, καθώς και ένα 10,77%(N=7) απάντησε

ότι συμβολίζουν δεδομένα από νέες τεχνολογίες. Επιπλέον, υπάρχουν πολλοί ακόμη ορισμοί που έχουν αποδοθεί από το δείγμα μας στο παραπάνω διάγραμμα(Διάγραμμα 4). Τέλος, πρέπει να σημειωθεί ότι ένα 4,62%(N=3) δεν γνωρίζει τι σημαίνουν τα Big Data.

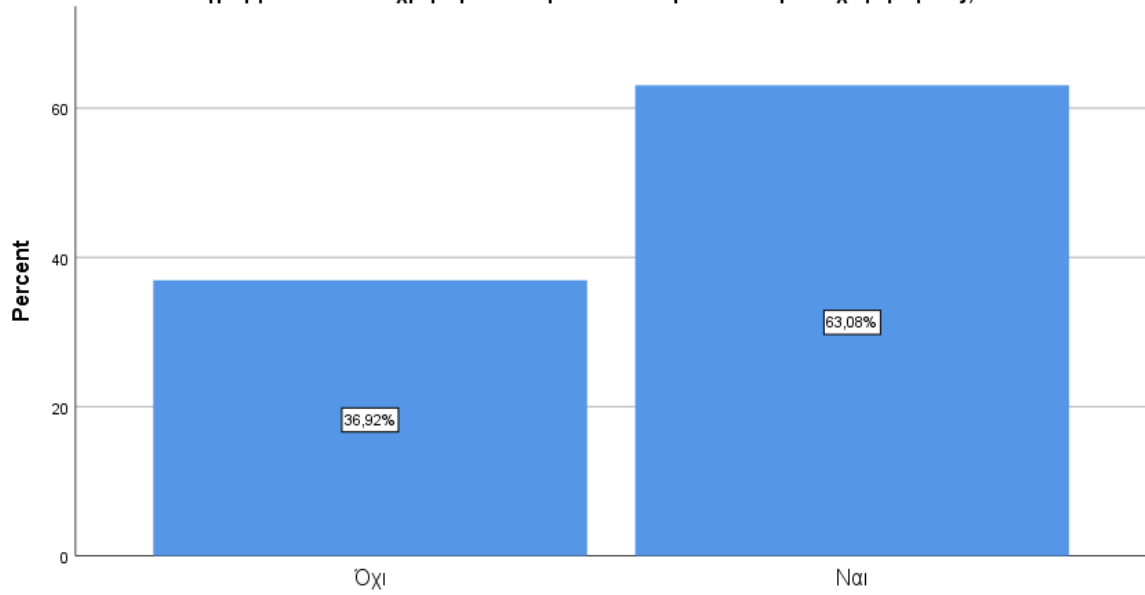
Ύστερα, το δείγμα μας ερωτήθηκε για το πόσο χρησιμοποιούνται στην ελληνική επιχειρηματικότητα τα μεγάλα δεδομένα. Οι απαντήσεις που λάβαμε είναι ότι το 49,23%(N=32) θεωρούν ότι τα Big Data χρησιμοποιούνται λίγο στον ελληνικό επιχειρηματικό κόσμο με το αρκετά να ακολουθεί με 35,38%(N=23). Έπειτα έχουμε το 7,69%(N=5) να πιστεύει ότι δεν χρησιμοποιούνται καθόλου σήμερα και τέλος το 3,06%(N=2) από το δείγμα μας να θεωρεί ότι χρησιμοποιούνται πολύ. Το 4,62%(N=3) μας απάντησε ότι δεν γνωρίζει καθόλου. (Διάγραμμα 5)

Διάγραμμα 5: Πιστεύετε ότι τα Μεγάλα Δεδομένα χρησιμοποιούνται σήμερα στην ελληνική επιχειρηματικότητα;



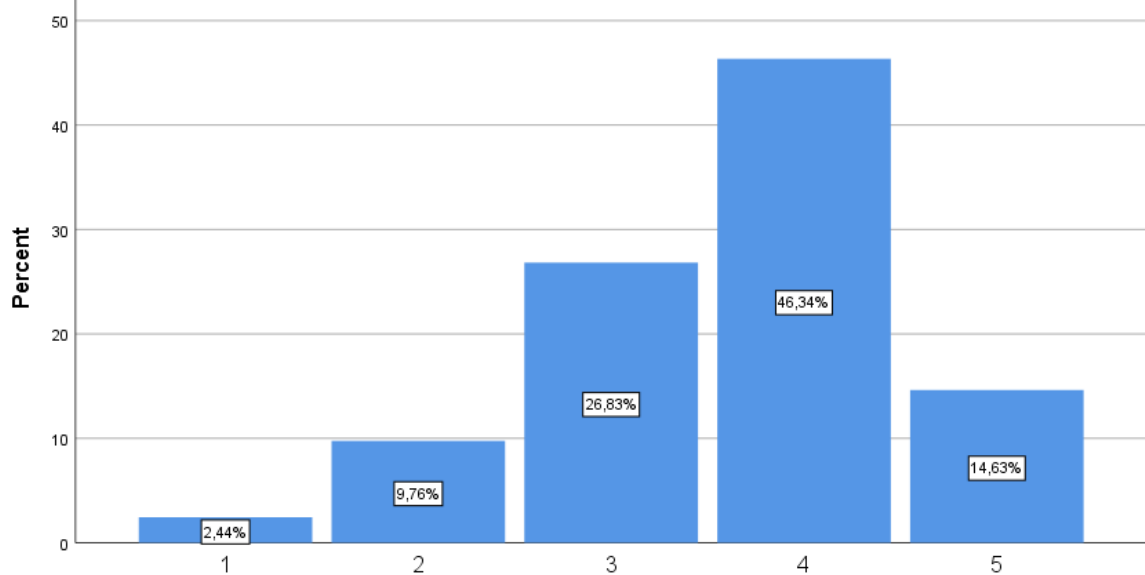
Έπειτα, στόχος μας ήταν να μάθουμε από το δείγμα μας για το αν κάνουν χρήση Μεγάλων Δεδομένων στην επιχείρησή τους. Τα αποτελέσματα ήταν ότι το 63,08%(N=41) χρησιμοποιούν Big Data και το υπόλοιπο 36,92%(N=24) δεν χρησιμοποιεί. (Διάγραμμα 6)

Διάγραμμα 6: Κάνετε χρήση των Μεγάλων Δεδομένων στην επιχείρησή σας;

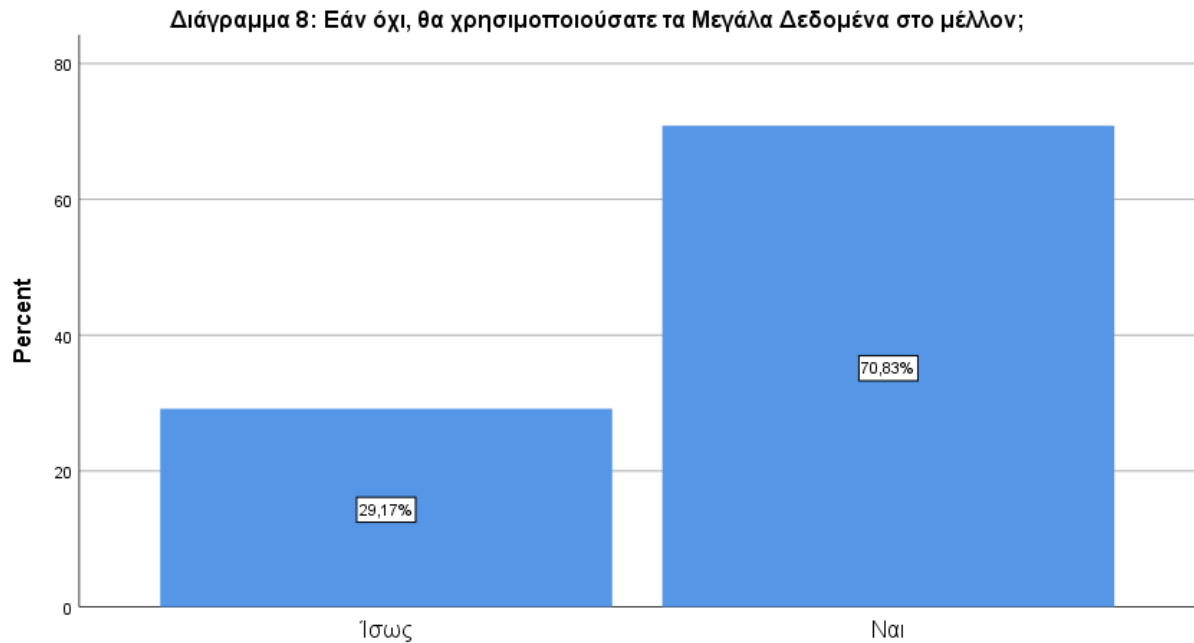


Επιπλέον, για όσους απάντησαν ότι χρησιμοποιούν Big Data (N=41) τους ερωτήσαμε σχετικά με το πόσο έμπειροι είναι με αυτήν την τεχνολογία. Για την καλύτερη απόδοση αποτελεσμάτων χρησιμοποιήσαμε την κλίμακα Likert. Αρχικά, οι περισσότεροι με ποσοστό 46,34%(N=19) απάντησαν ότι έχουν αρκετά καλό βαθμό εμπειρίας με τα Big Data, έπειτα ένα 26,83%(N=11) μας δείχνει ότι η εμπειρία τους είναι μέτρια. Ακολουθούν, ένα 14,63%(N=6) που έχει πολύ καλό βαθμό εμπειρίας και ένα 9,76%(N=4) του οποίου ο βαθμός εμπειρίας είναι ελάχιστος. Σημαντικό είναι ότι μόνο το 2,44%(N=1) δεν έχει καθόλου εμπειρία. (Διάγραμμα 7)

Διάγραμμα 7: Εάν χρησιμοποιείτε, αξιολογίστε τον βαθμό εμπειρίας σας με τα Μεγάλα Δεδομένα.

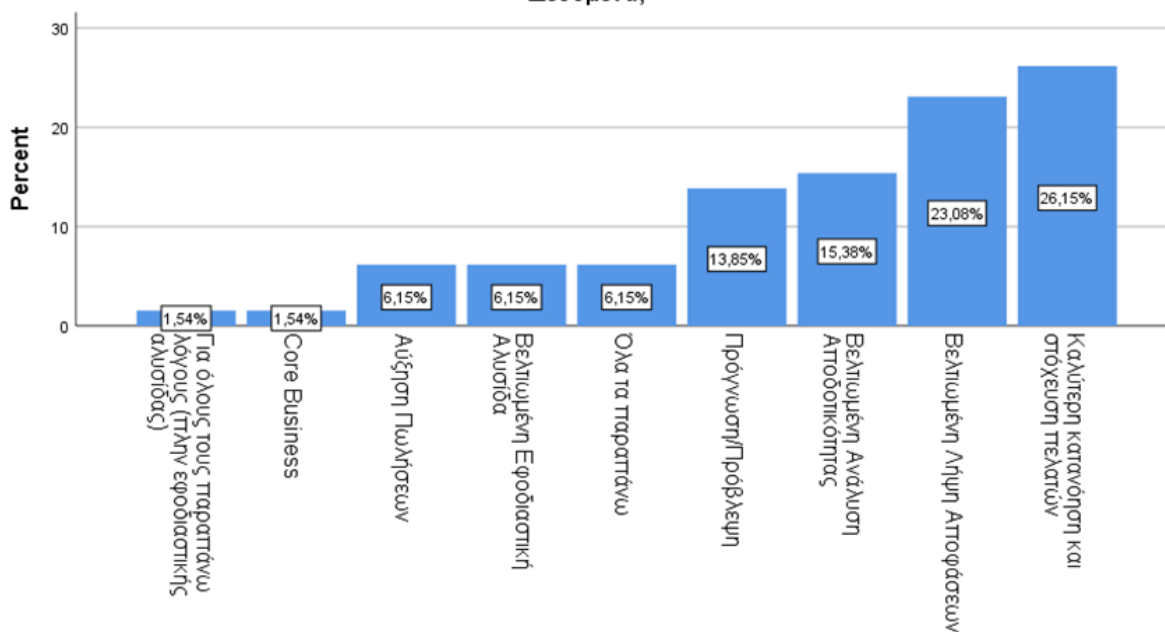


Ωστόσο για εκείνους που δεν χρησιμοποιούν Μεγάλα Δεδομένα (N=24) τέθηκε η επιπλέον ερώτηση για το αν θα χρησιμοποιούσαν στο μέλλον. Οι απαντήσεις που λάβαμε ήταν θετικές στο να χρησιμοποιήσουν με 70.38%(N=17) και με το 29.17%(N=7) από το δείγμα να απαντάνε πως ίσως και να χρησιμοποιούσαν στο μέλλον. Αξιοσημείωτο ήταν ότι στην επιλογή του 'ΟΧΙ' που δώσαμε στην συγκεκριμένη ερώτηση κανένας από το δείγμα μας δεν την επέλεξε. (Διάγραμμα 8)



Έπειτα σκοπός της παραπάνω ερώτησης είναι να μάθουμε τον λόγο για τον οποίο χρησιμοποιούνται τα Big Data. Ο κυριότερος λόγος, που βλέπουμε στο διάγραμμα είναι η καλύτερη κατανόηση και στόχευση πελατών με ποσοστό 26,15%(N=17). Ακολουθεί, η βελτιωμένη λήψη αποφάσεων με 23,08%(N=15) καθώς και η βελτιωμένη ανάλυση αποδοτικότητας με 15,38% (N=10) αντίστοιχα. Επιπλέον ένα 13,85%(N=9) χρησιμοποιεί ή/και θα χρησιμοποιούσε τα Big Data για πρόγνωση/πρόβλεψη. Στο διάγραμμα βλέπουμε και τους υπόλοιπους λόγους του δείγματος μας(Διάγραμμα 9).

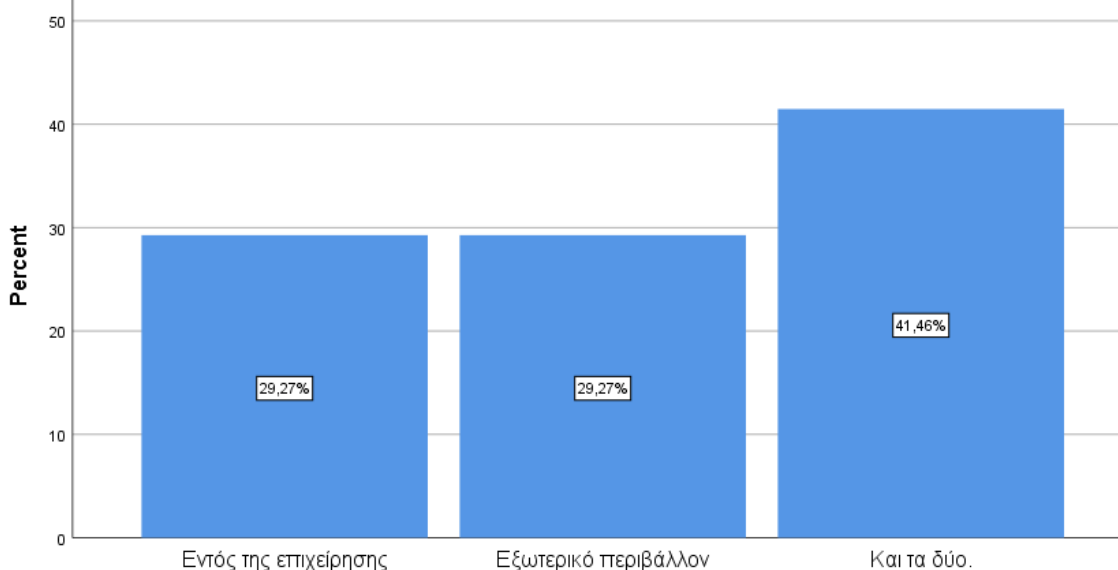
Διάγραμμα 9: Ποιος είναι ο κυριότερος λόγος που χρησιμοποιείτε, ή θα χρησιμοποιούσατε τα Μεγάλα Δεδομένα;



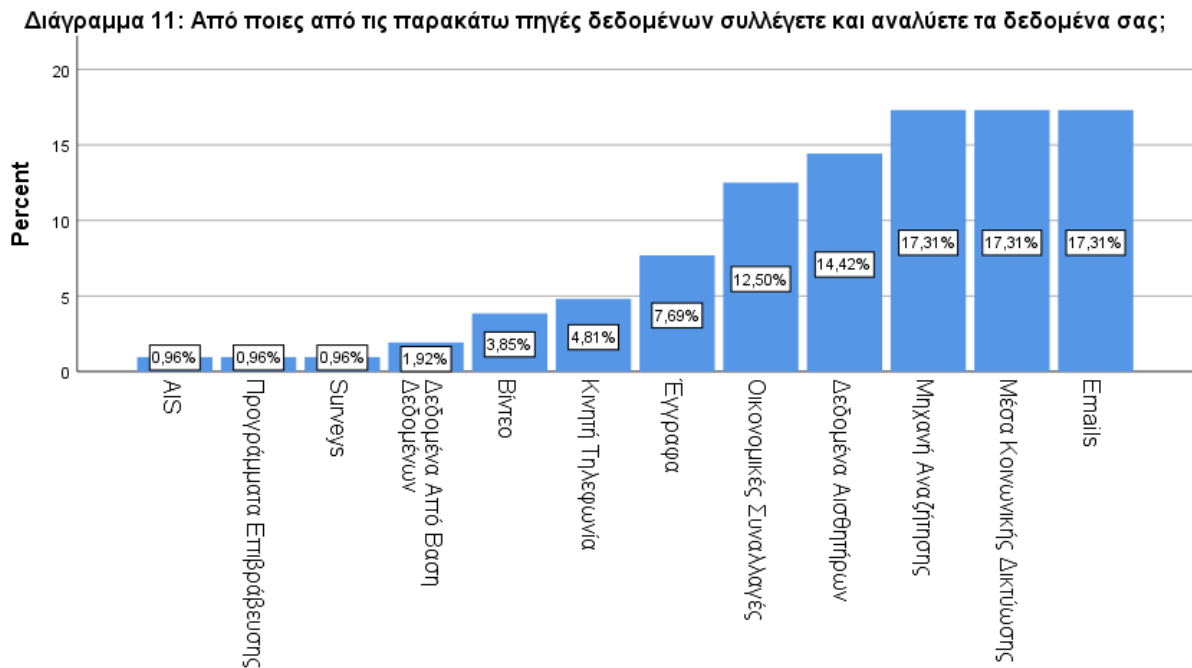
Οι ερωτήσεις που ακολουθούν είναι εξειδικευμένες για αυτούς μόνο που χρησιμοποιούν την τεχνολογία των Μεγάλων Δεδομένων στην επιχείρησή τους. (N=41)

Πιο συγκεκριμένα, μία από τις ερωτήσεις μας αφορούσε το περιβάλλον από το οποίο γίνεται η συλλογή δεδομένων στις επιχειρήσεις. Το μεγαλύτερο ποσοστό με 41,46%(N=17) απάντησε πως τα δεδομένα συλλέγονται και εντός και εκτός των επιχειρήσεων. Έπειτα, το 29,27%(N=12) απάντησε πως η συλλογή των δεδομένων γίνεται εντός της επιχείρησης και ομοίως ένα 29,27%(N=12) συλλέγει τα δεδομένα του από το εξωτερικό περιβάλλον. (Διάγραμμα 10)

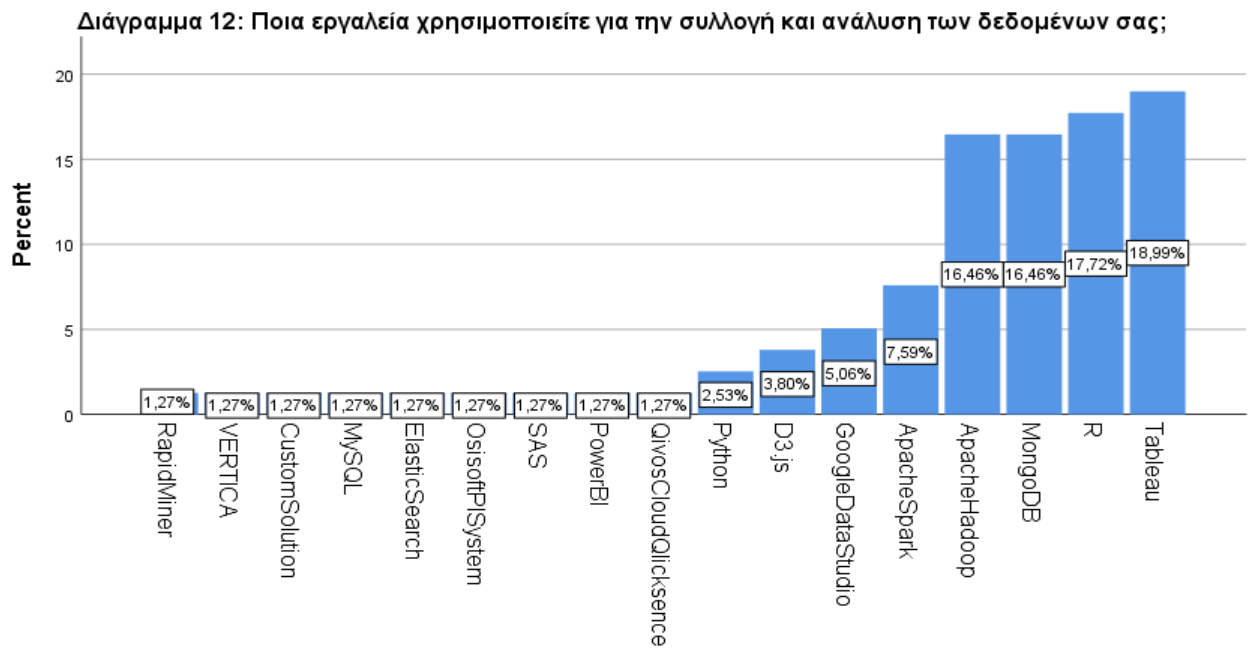
Διάγραμμα 10: Στην επιχείρησή σας, η συλλογή των Μεγάλων Δεδομένων είναι κυρίως από:



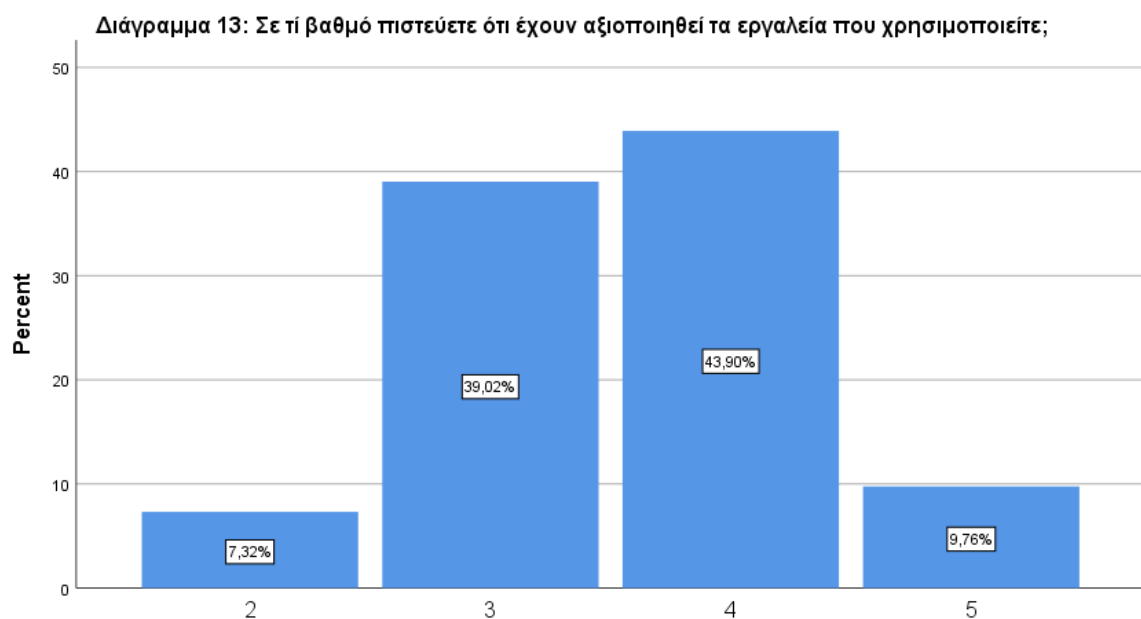
Έπειτα, εμβαθύνουμε από ποιες πηγές γίνεται η συλλογή των δεδομένων. Μια μεγάλη ισοβαθμία με ποσοστό 17,31%(N=18) έχουν ως πηγές, τα emails, τα μέσα κοινωνικής δικτύωσης και την μηχανή αναζήτησης. Κατόπιν ένα 14,42%(N=15) μας ανέφερε ως πηγή δεδομένων τους αισθητήρες και ένα 12,50%(N=13) τις οικονομικές συναλλαγές. Στο διάγραμμα, παρουσιάζονται οι υπόλοιπες πηγές δεδομένων (Διάγραμμα 11).



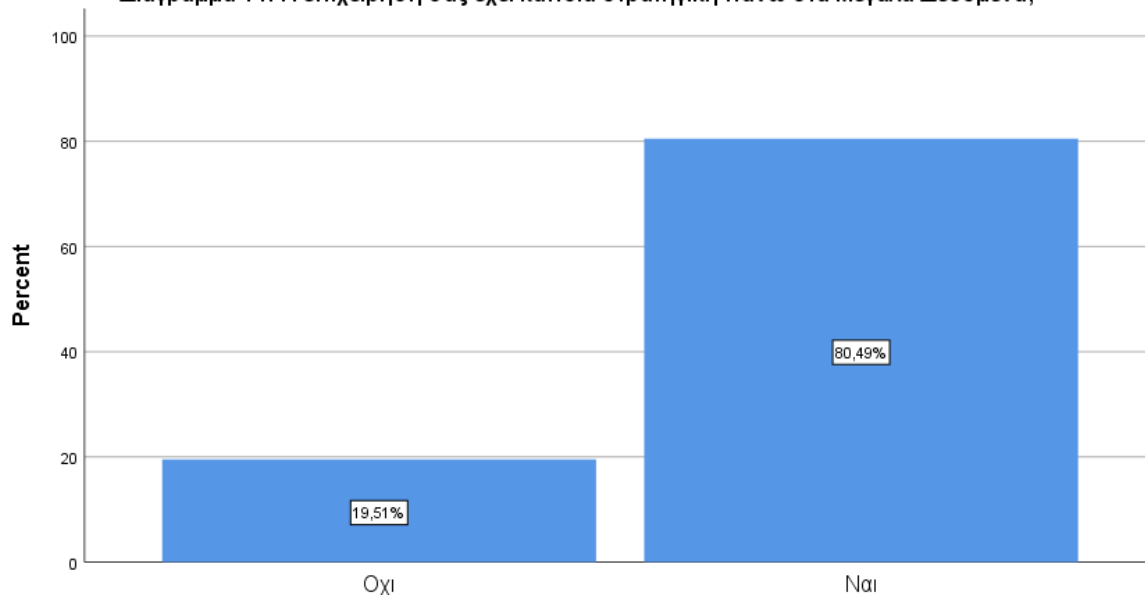
Ύστερα, ζητήσαμε από το δείγμα μας να μας επισημάνει τα εργαλεία με τα οποία γίνεται η συλλογή και η ανάλυση των δεδομένων τους. Το μεγαλύτερο ποσοστό με 18,99%(N=15) μας ανέφερε το Tableau και με την R να ακολουθεί με ποσοστό 17,72%(N=14). Παράλληλα, ένα 16,46%(N=13) χρησιμοποιεί το MongoDB καθώς και το Apache Hadoop αντίστοιχα. Τέλος, υπάρχει μια πληθώρα εργαλείων τα οποία αναφέρονται στο παραπάνω διάγραμμα(Διάγραμμα 12).



Εν συνέχεια της παραπάνω ερώτησης, το δείγμα μας ερωτήθηκε να αξιολογήσουν τον βαθμό αξιοποίησης των εργαλείων που χρησιμοποιούν. Οι περισσότεροι με 43,90%(N=18) θεωρούν ότι αξιοποιούν τα εργαλεία τους σε αρκετά καλό βαθμό. Επιπλέον, το δεύτερο μεγαλύτερο ποσοστό το οποίο είναι 39,02%(N=16) αξιοποιεί τα εργαλεία του σε μέτριο βαθμό. Ακολουθούν με 9,76%(N=4) οι επιχειρήσεις που αξιοποιούν σε πολύ καλό βαθμό τα εργαλεία και με 7,32%(N=3) εκείνες που τα χρησιμοποιούν σε ελάχιστο βαθμό. Αξίζει να αναφερθούμε, πως κανένας από το δείγμα μας δεν πιστεύει ότι δεν έχουν αξιοποιηθεί καθόλου τα εργαλεία τους. (Διάγραμμα 13)



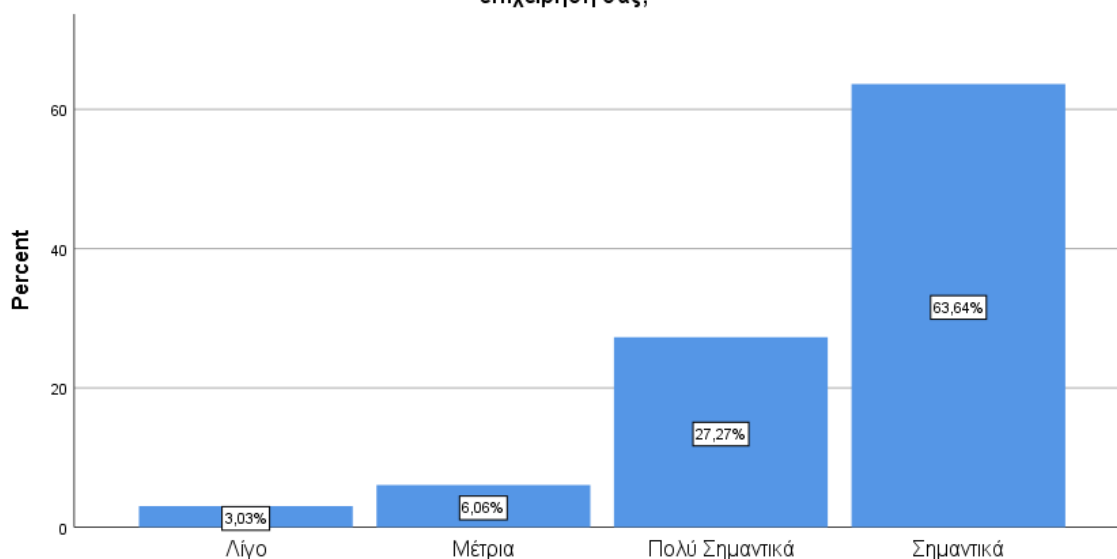
Διάγραμμα 14: Η επιχείρησή σας έχει κάποια στρατηγική πάνω στα Μεγάλα Δεδομένα;



Στην συγκεκριμένη ερώτηση, σκοπός μας ήταν να μάθουμε για το αν έχουν οι επιχειρήσεις του δείγματος μας, κάποια στρατηγική πάνω στα Big Data. Τα αποτελέσματα τα οποία πήραμε είναι ότι για το 80.49%(N=33) υπάρχει κάποια στρατηγική πάνω στα Big Data, αλλά το 19.51%(N=8) δεν χρησιμοποιεί κάποια στρατηγική. (Διάγραμμα 14)

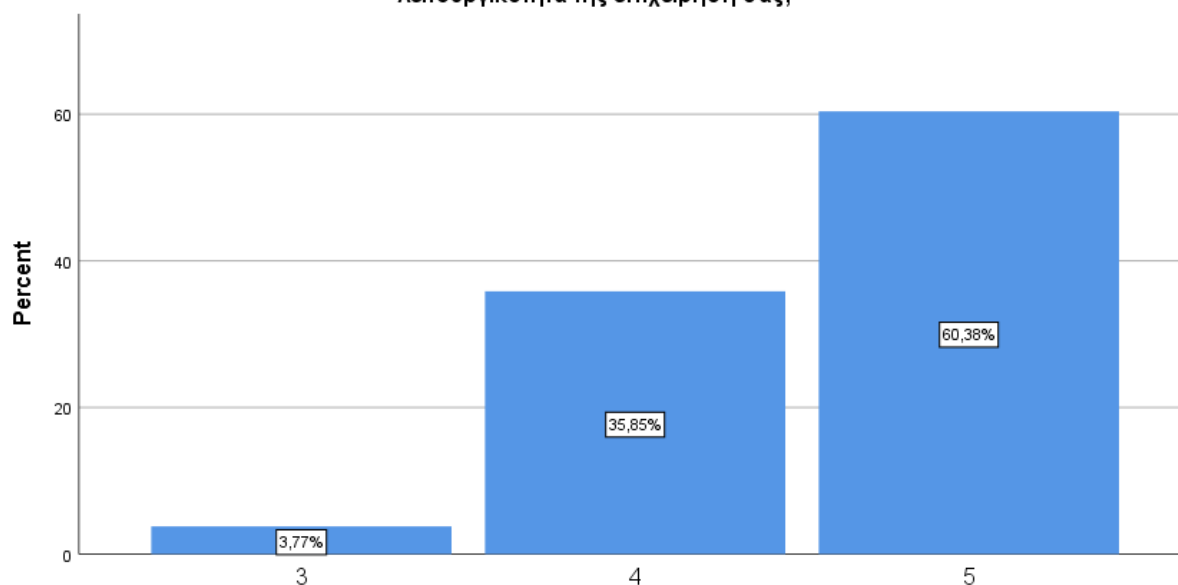
Βάση της παραπάνω ερώτησης, το δείγμα μας ερωτήθηκε για το πόσο σημαντικά έχει αυξηθεί το ανταγωνιστικό πλεονέκτημα της επιχείρησης. Πιο συγκεκριμένα, το 63.64%(N=21) απάντησε, πως το ανταγωνιστικό της πλεονέκτημα έχει αυξηθεί σημαντικά. Ένα 27,27%(N=9) πιστεύει ότι έχει αυξηθεί πολύ σημαντικά. Ακολουθεί το μέτρια και το λίγο με 6,06%(N=2) και 3,03%(N=1) αντίστοιχα. Να σημειώσουμε πως κανένας δεν απάντησε καθόλου. (Διάγραμμα 15)

Διάγραμμα 15: Αν ναι, σε τι βαθμό τα Μεγάλα Δεδομένα έχουν αυξήσει το ανταγωνιστικό πλεονέκτημα της επιχείρησής σας;



Τέλος, η τελευταία ερώτηση είχε μελλοντικό σκοπό, αφορούσε όλο το δείγμα μας και αποσκοπούσε στο πόσο σημαντική θα είναι η χρήση των Big Data στην λειτουργικότητα της επιχείρησής σε 5 χρόνια από τώρα. Αρχικά, το 60,38%(N=32) του δείγματος μας, προβλέπει πως θα είναι πολύ σημαντική η χρήση των Big Data στην λειτουργικότητα της επιχείρησής τους. Ένα 35,85%(N=19) θεωρεί πως να είναι αρκετά σημαντική και μόλις ένα 3,77% (N=2) πιστεύει πως θα είναι μέτρια η χρήση τους στη λειτουργικότητα της επιχείρησής. Σημαντικό είναι ότι κανένας δεν πιστεύει πως δεν θα αξιοποιηθούν τα Μεγάλα Δεδομένα στην επιχείρησή τους σε 5 χρόνια. (Διάγραμμα 16)

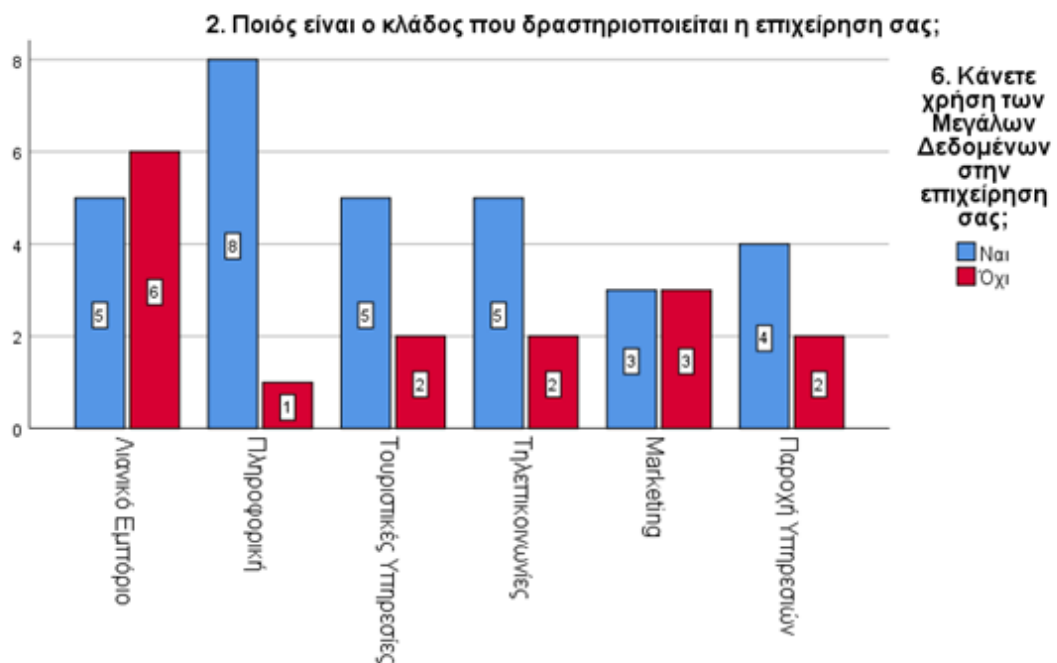
Διάγραμμα 16: Σε 5 χρόνια, πόσο σημαντική πιστεύετε θα είναι η χρήση των Μεγάλων Δεδομένων στην λειτουργικότητα της επιχείρησής σας;



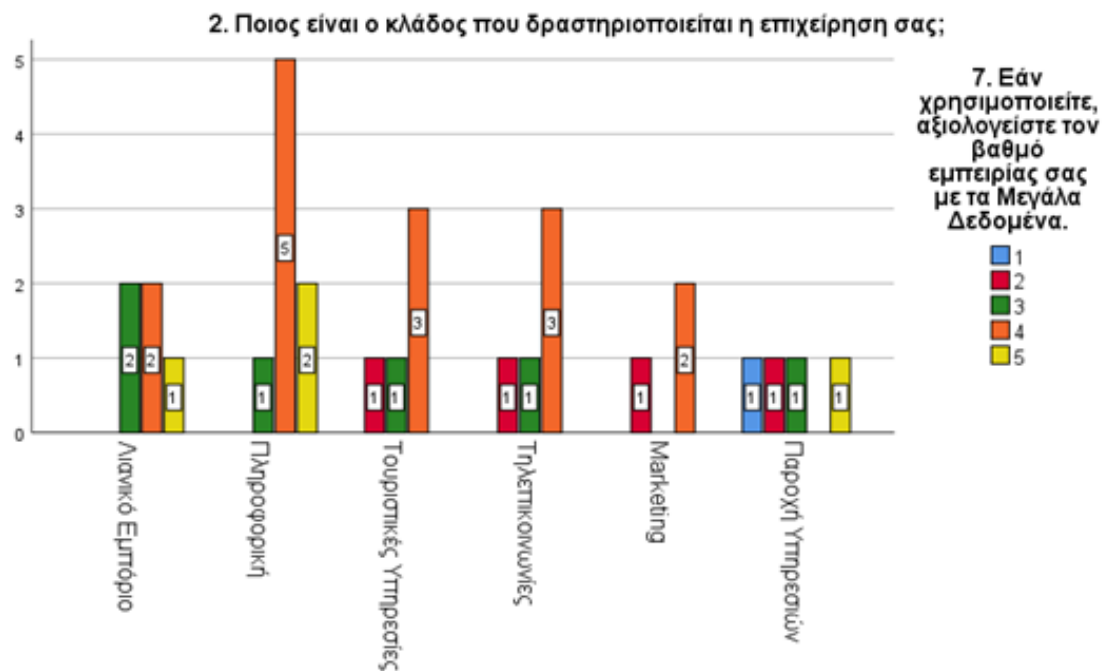
6.6 Συνδυαστικά αποτελέσματα της έρευνας

Αφού αναλύσαμε παραπάνω το γενικό μέρος της έρευνας μας, προσπαθήσαμε να εμβαθύνουμε στις συμπεριφορές του δείγματος μας, σύμφωνα με τους κλάδους στους οποίους ανήκουν και το μέγεθος των εταιρειών. Αυτό το καταφέραμε, απομονώνοντας τις απαντήσεις στις ερωτήσεις 2 & 3 και συνδυάζοντάς τις με τις ερωτήσεις 6, 7 και 14.

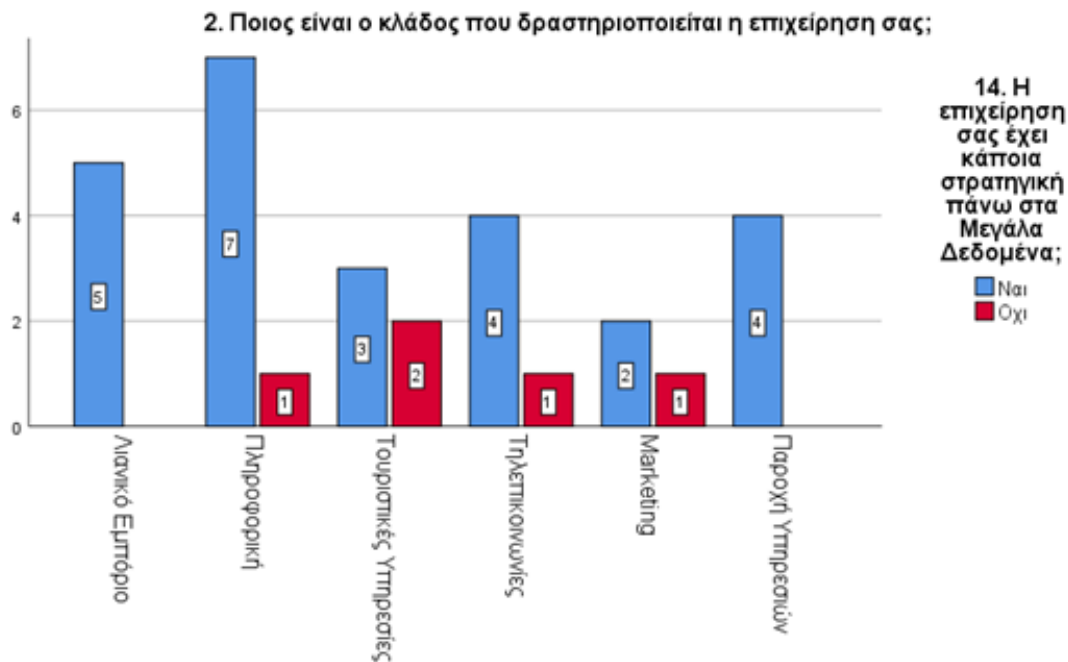
Ως δείγμα μας πήραμε τους πρώτους 6 κλάδους οι οποίοι είχαν τις περισσότερες απαντήσεις στην έρευνα μας. Πιο συγκεκριμένα, μεταξύ αυτών των κλάδων με βάση το αν χρησιμοποιούν Μεγάλα Δεδομένα, παρατηρήσαμε ότι ο κλάδος της πληροφορικής, ποσοστιαία είναι υψηλότερος σε σύγκριση με τους υπόλοιπους εφόσον 8 από τις 9 επιχειρήσεις κάνουν χρήση Big Data(88,9%). Επίσης στους κλάδους των τουριστικών υπηρεσιών και τηλεπικοινωνιών βλέπουμε πως έχουν απαντήσει από τις 7 εταιρείες οι 5 με ποσοστό 71,4% αντίστοιχα. Αξίζει να αναφερθεί ότι ο κλάδος του λιανικού εμπορίου, ο οποίος αποτέλεσε το μεγαλύτερο μέρος του δείγματος μας, μόνο οι 5 από τις 11 εταιρείες κάνουν χρήση Big Data(45,5%). Εδώ συμπεραίνουμε πως το λιανικό εμπόριο σε σύγκριση με τους κλάδους των τουριστικών υπηρεσιών και τηλεπικοινωνιών, παρόλο που έχουν τον ίδιο αριθμό εταιρειών που κάνουν χρήση Μεγάλων Δεδομένων, οι κλάδοι των τουριστικών και των τηλεπικοινωνιών υπερέρχουν ποσοστιαία του λιανικού.



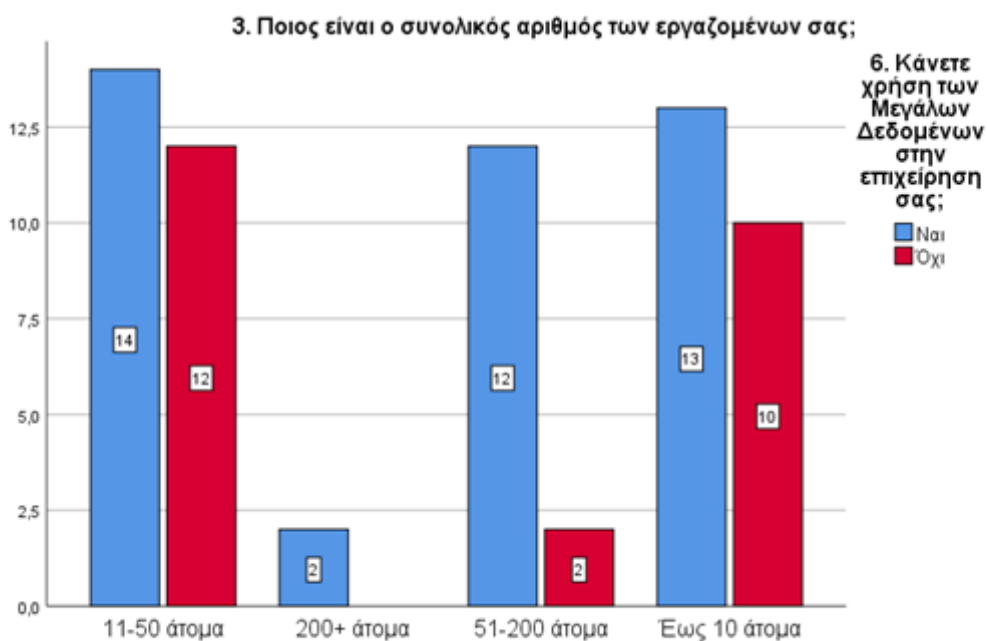
Έπειτα προχωρήσαμε ως προς τον βαθμό εμπειρίας που έχουν οι εταιρείες στην τεχνολογία των Big Data. Αυτό που σημειώσαμε, είναι πως ο κλάδος της πληροφορικής σε σύγκριση με όλους τους υπόλοιπους κλάδους έχει τον μεγαλύτερο βαθμό εμπειρίας σε αυτήν την τεχνολογία με 5 επιχειρήσεις να έχουν απαντήσει πως έχουν αρκετά καλό βαθμός εμπειρίας και μόλις 2 να έχουν απαντήσει ότι έχουν άριστο βαθμό εμπειρίας. Από την άλλη μεριά, ο κλάδος με τον χαμηλότερο βαθμό εμπειρίας είναι η παροχή υπηρεσιών όντας η μοναδική που έχει εταιρεία με καθόλου εμπειρία.



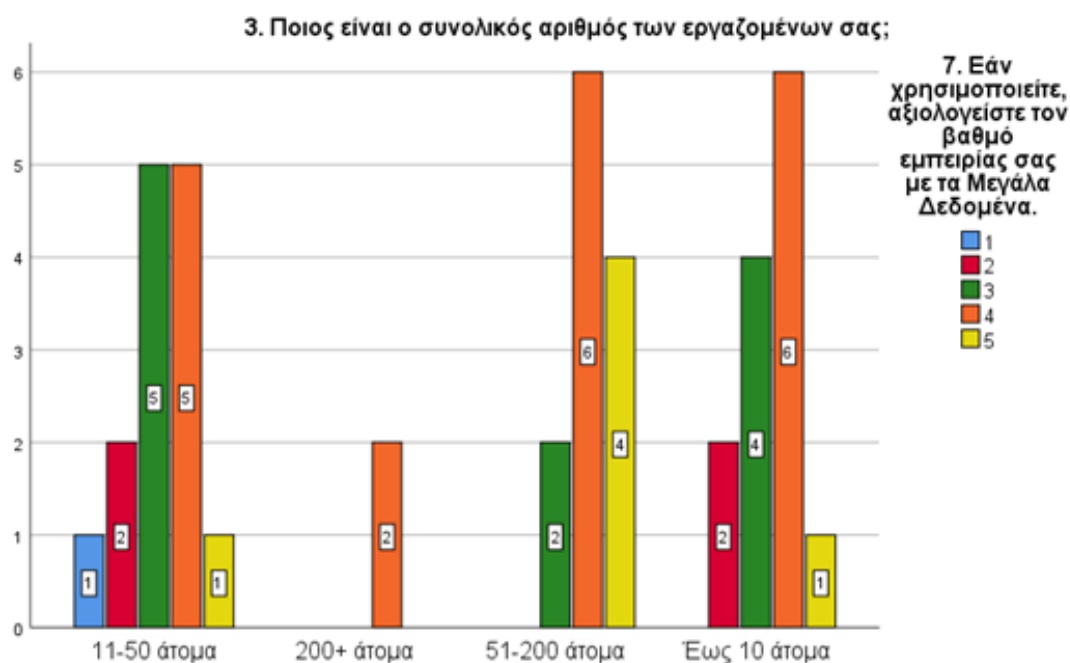
Επίσης διασταυρώσαμε τον κάθε κλάδο με την ερώτηση για το αν έχουν κάποια στρατηγική πάνω στα Big Data. Αρχικά πρέπει να υπενθυμίσουμε πως δείγμα μας σε αυτήν την σύγκριση είναι μόνο οι εταιρείες που χρησιμοποιούν Μεγάλα Δεδομένα. Έτσι, παρατηρήσαμε ότι, στους κλάδους του λιανικού εμπορίου και της παροχής υπηρεσιών όλες οι εταιρείες έχουν απαντήσει ότι χρησιμοποιούν κάποια στρατηγική. Έπειτα, ο κλάδος της πληροφορικής έχει τις περισσότερες εταιρείες οι οποίες έχουν κάποια στρατηγική (N=7) σε σύγκριση με τους υπόλοιπους. Τέλος, ο κλάδος με τις λιγότερες εταιρείες που δεν έχουν κάποια στρατηγική είναι αυτός των τουριστικών υπηρεσιών (N=2).



Επιπλέον, προχωρήσαμε στις συγκρίσεις με βάση το μέγεθος της κάθε εταιρείας. Όσο αφορά το ερώτημα 6, αρχικά παρατηρούμε ότι οι εταιρείες με 51-200 άτομα έχουν το μεγαλύτερο ποσοστό εταιρειών που χρησιμοποιούν Big Data(85,7%) παρόλο που οι εταιρείες με 11-50 και έως 10 άτομα έχουν δώσει περισσότερες απαντήσεις. Οι εταιρείες με 11-50 εργαζομένους, έχουν το μεγαλύτερο ποσοστό εταιρειών που έχουν απαντήσει ότι δεν κάνουν χρήση Big Data(46.2%).

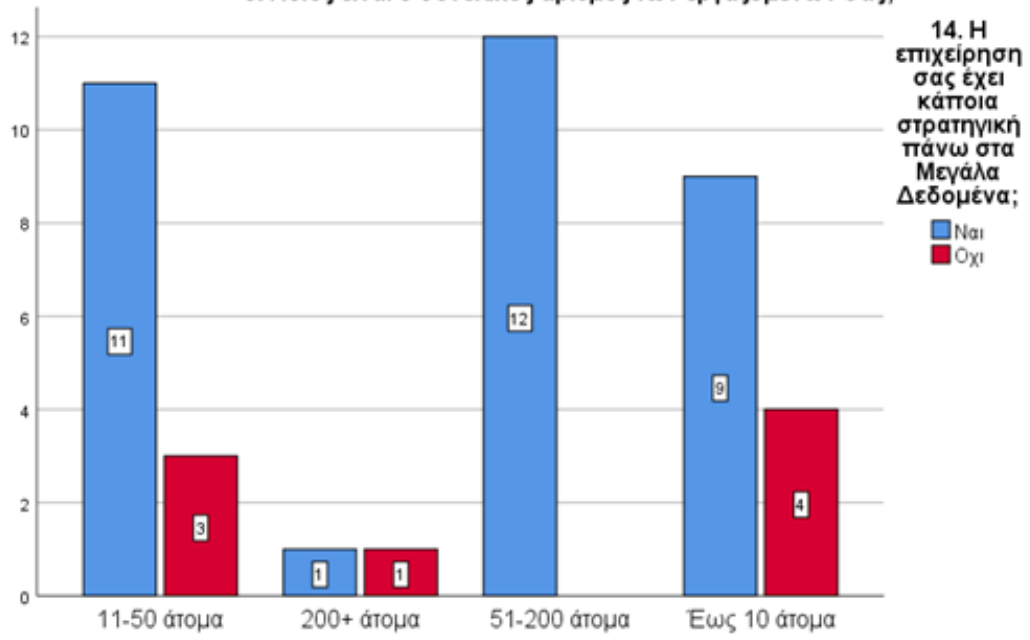


Έπειτα, ως προς τον βαθμό εμπειρίας είδαμε πως αυτή που έχει τον μεγαλύτερο βαθμό εμπειρίας είναι οι μεσαίες επιχειρήσεις που απασχολούν 51 έως 200 άτομα. Επίσης, οι μικρές επιχειρήσεις παρατηρούμε στο σύνολο πως έχουν καλύτερο βαθμό εμπειρίας από τις μικρομεσαίες. Αξίζει να σημειωθεί πως όλοι οι οργανισμοί που απασχολούν πάνω από 200 άτομα κάνουν χρήση των Big Data καθώς και έχουν καλό βαθμό εμπειρίας σε αυτά.



Τέλος, σύμφωνα με το αν έχουν κάποια στρατηγική πάνω στα Big Data βλέπουμε πως όλες οι μεσαίες επιχειρήσεις έχουν κάποια στρατηγική. Ωστόσο ενώ είχαμε δει προηγουμένως ότι οι μικρές επιχειρήσεις έχουν καλύτερο βαθμό εμπειρίας από τις μικρομεσαίες στο σύνολο τους, εδώ παρατηρούμε το αντίστροφο, δηλαδή πως οι περισσότερες μικρομεσαίες επιχειρήσεις έχουν κάποια στρατηγική.

3. Ποιος είναι ο συνολικός αριθμός των εργαζομένων σας;



7. Ανακεφαλαίωση

Η συνεχής εξέλιξη της τεχνολογίας και ιδίως στην διαχείριση μεγάλου όγκου δεδομένων κατέστησαν τα Big Data μία τεχνολογική πρόκληση της εποχής μας για πολλές εταιρείες και οργανισμούς. Σκοπός της πτυχιακής εργασίας ήταν η διερεύνηση της τεχνολογίας των Big Data, οι τεχνολογίες που βοηθούν στην ανάλυση και αποθήκευση μεγάλου όγκου δεδομένων καθώς και οι ρόλοι και οι ευκαιρίες που υπάρχουν για τις ελληνικές εταιρείες μέσω της χρήσης των Big Data στην ελληνική οικονομία.

Η έρευνα που διεξήχθη στην παρούσα εργασία κατέδειξε σημαντικά συμπεράσματα αναφορικά με τα Big Data. Αρχικά πρέπει να αναφερθεί πως η πλειοψηφία των ερωτώμενων εταιρειών κάνει χρήση των Big Data (63,08%) με τις περισσότερες να έχουν αρκετά καλή εμπειρία με την τεχνολογία αυτή. Όσον αφορά τον κλάδο και το μέγεθος της κάθε ερωτώμενης εταιρείας, παρατηρήθηκε πως υπάρχει μια μεγάλη πλειοψηφία από διαφορετικούς κλάδους που κάνουν χρήση των Big Data, γεγονός που δείχνει το πλήθος διαφορετικών εταιρειών που έχουν επηρεαστεί από την τεχνολογία αυτή. Παράλληλα, κατά τον προσδιορισμό του μεγέθους της κάθε επιχείρησης, συναντάμε πως το μεγαλύτερο ποσοστό απαντήσεων έχουν δώσει εταιρείες με 11-50 άτομα (40%) και εταιρείες έως 10 άτομα προσωπικό(35,38%). ποσοστό εταιρειών που κάνει χρήση των Big Data είναι αυτές με προσωπικό 51-200 άτομα. Συμπεραίνουμε πως η τεχνολογία των Big Data υιοθετείται περισσότερο από μεγάλες εταιρείες που έχουν καλύτερη οικονομική δύναμη αλλά και τεχνογνωσία σε σχέση με μικρότερες εταιρείες. Αξιοσημείωτο συμπέρασμα αποτέλεσε το γεγονός πως για την πλειονότητα του δείγματος τα Big Data αντιπροσωπεύουν ως έννοια νέες μορφές δεδομένων και ανάλυσης (29,23%) καθώς και μεγάλες ποσότητες δεδομένων. Παράλληλα, η πλειοψηφία αυτών χρησιμοποιεί ή και θα χρησιμοποιούσε, μελλοντικά, τα Big Data με σκοπό την καλύτερη κατανόηση και στόχευση των πελατών αλλά και ως εργαλείο για μια βελτιωμένη λήψη αποφάσεων.

Το μεγαλύτερο ποσοστό των εταιρειών συλλέγει τα δεδομένα τους από Emails, Μέσα Κοινωνικής Δικτύωσης και Μηχανές Αναζήτησης (π.χ. Google), ενώ τα κυριότερα εργαλεία που χρησιμοποιούν οι εταιρείες αυτές για την συλλογή και ανάλυση των δεδομένων τους αποτελούν το Tableau, η R, το MongoDB και το Apache Hadoop. Η πλειοψηφία των εταιρειών απάντησε πως τα εργαλεία αυτά έχουν αξιοποιηθεί σε αρκετά καλό βαθμό (43,90%), κάτι που δηλώνει πως έχουν βοηθήσει σε ένα σημαντικό βαθμό τις εταιρείες αυτές να επιτύχουν το σκοπό τους, αλλά υπάρχει ένα περιθώριο επιπλέον αξιοποίησης τους. Για αυτό τον λόγο θεωρούν πως δεν έχουν αξιοποιήσει πλήρως αυτά τα εργαλεία. Τέλος, το μεγαλύτερο ποσοστό των ερωτώμενων εταιρειών έχει κάποια στρατηγική πάνω στα Big Data, με τις περισσότερες να έχουν αυξήσει σημαντικά το ανταγωνιστικό τους πλεονέκτημα λόγω αυτής της στρατηγικής. Το γεγονός αυτό δείχνει την αποτελεσματικότητα που έχουν τα Big Data στις ελληνικές επιχειρήσεις παρά τις αντίξοες συνθήκες που υπάρχουν στην σημερινή ελληνική οικονομία.

Τα αποτελέσματα αυτής της έρευνας δείχνουν την σημαντικότητα αλλά και την δύναμη που κατέχουν τα Big Data στην σημερινή ελληνική επιχειρηματικότητα. Παρόλο που τα Big Data δεν είναι, ακόμα, τόσο εδραιωμένα στην χώρα μας, σύμφωνα και με τις απαντήσεις της πλειοψηφίας των ερωτηθέντων που δηλώνουν πως τα Big Data χρησιμοποιούνται ελάχιστα στην ελληνική επιχειρηματικότητα, υπάρχει μία ελπιδοφόρα και ενθαρρυντική στάση των ελληνικών επιχειρήσεων απέναντι σε αυτά. Με το μεγαλύτερο ποσοστό των ερωτώμενων εταιρειών, που δεν κάνουν χρήση των Big Data, να έχουν απαντήσει πως θα επενδύσουν σίγουρα σε αυτή την τεχνολογία στο άμεσο μέλλον και με τις περισσότερες να πιστεύουν πως η χρήση των Big Data θα είναι πολύ σημαντική στην λειτουργικότητα της επιχείρησης στα επόμενα 5 χρόνια, μας δείχνει πως στο μέλλον η τεχνολογία αυτή θα αποτελέσει σημαντικό παράγοντα για τις επιχειρήσεις που επιθυμούν να αναπτυχθούν, να ευδοκιμήσουν αλλά και να καινοτομήσουν.

Τα Big Data αν και υπήρχαν ανέκαθεν στην ζωή μας, ανακαλύφθηκαν και άρχισαν να εξερευνούνται τα τελευταία χρόνια λόγω της εξέλιξης της τεχνολογίας. Η εμφάνιση τους έφερε ριζικές αλλαγές τόσο στον τρόπο λειτουργίας των ανθρώπινων κοινωνιών όσο και στον τρόπο λειτουργίας των επιχειρήσεων. Δεδομένα που θεωρούνταν άχρηστα ή μη διαχειρίσιμα, πλέον θεωρούνται εξαιρετικά χρήσιμα για την εξαγωγή πολύτιμης πληροφορίας. Η συνεχής εξέλιξη των τεχνολογιών σε συνδυασμό με την επιστήμη των Big Data έχει τεράστιο αντίκτυπο στην σημερινή επιχειρηματικότητα, βοηθώντας εταιρείες, μικρές και μεγάλες, να έχουν ένα ισχυρό όπλο για βελτιωμένη λήψη αποφάσεων καθώς και για καλύτερη πληροφόρηση σε πραγματικό χρόνο για όλα τα τμήματα της επιχείρησης. Είναι σίγουρο πως τα Big Data διαθέτουν ένα ευρύ φάσμα δυνατοτήτων με εξαιρετικές προοπτικές για το μέλλον, βοηθώντας τις επιχειρήσεις να εξελίσσονται και αναπτύσσονται συνεχώς, ώστε να πετύχουν το μέγιστο δυνατόν επιχειρηματικό όφελος.

Συνοψίζοντας και με βάση όσα έχουν μελετηθεί τα Big Data είναι και θα παραμείνουν σημαντικά στην ανάπτυξη κάθε επιχείρησης, αρκεί να αξιοποιηθούν με σωστό τρόπο. Επί πρόσθετα, θα υπάρξει στο μέλλον μεγάλη ανάγκη για περαιτέρω έρευνα στον τομέα της Τεχνητής Νοημοσύνης (Artificial Intelligence) αλλά και των τεχνολογιών της, Machine Learning και Deep Learning, που είναι εξαιρετικά σημαντικά στην επιστήμη των Big Data. Η Τεχνητή Νοημοσύνη θα έχει καθοριστικό ρόλο στην εξέλιξη και βελτίωση των τεχνικών αλλά και εφαρμογών των Big Data, καθώς και θα είναι σημαντικός παράγοντας επιτυχίας για πολλές επιχειρήσεις.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Abhay Kumar Bhadani, Dhanya Jothimani (2016). Big data: Challenges, opportunities and realities, Effective Big Data Management and Opportunities for Implementation, pp.1-24.
2. Ardavan Ashabi, Shamsul Bin Sahibuddin, Medhi Salkhordeh Haghighi (2020). Big Data: Current Challenges and Future Scope. Received from <https://ieeexplore.ieee.org/document/9108826>
3. Avita Katal, Mohammad Wazid, R. H. Goudar (2013). Big Data: Issues, Challenges, Tools and Good Practices. Received from <https://ieeexplore.ieee.org/document/6612229/references#references>
4. Almeida Fernando (2018). Big Data: Concept, Potentialities and Vulnerabilities, Vol. 2, No. 1.
5. Ahmed Oussous, Fatima-Zahra Benjelloun, Ayoub Ait Lahcen, Samir Belfkih (2017). Big Data technologies: A survey, Journal of King Saud University – Computer and Information Sciences, 30(2018), pp.431-448.
6. Anna Brzozowska, Leszek Ziora, Robert Sałek, Anna Wiśniewska-Sałek (2016). The Possibilities of Big Data Solutions Application in Logistics. International Multidisciplinary Scientific Conference University of Miskolc.
7. Alexandre da Silva Veith, Marcos Dias de Assunção (2018). Apache Spark. Inria Avalon, LIP Laboratory, ENS Lyon, University of Lyon. Received from https://www.researchgate.net/publication/323447097_Apache_Spark
8. Bernard Marr (2016). BIG DATA IN PRACTICE. United Kingdom: John Wiley and Sons Ltd.
9. Balamurugan Balusamy, R Nandhini Abirami, Seifedine Kadry, Amir H. Gandomi (2021). Big Data: Concepts, Technology, and Architecture. John Wiley and Sons Inc.
10. Big Data Applications in Various Fields. Retrieved from <https://www.insightssuccess.com/big-data-applications-various-fields/>
11. Big Data: What is it and why it matters. Retrieved from https://www.sas.com/el_gr/insights/big-data/what-is-big-data.html

12. ΓΙΑΝΕΣΤΡΑΣ ΓΕΩΡΓΙΟΣ (2019). Οι Επιπτώσεις του ΓΚΠΔ στον Τραπεζικό Τομέα – Με Εξειδίκευση στην Επίδραση της Εταιρικής Κουλτούρας. Χαροκόπειο Πανεπιστήμιο, Αθήνα.
13. Δημήτριος – Γεώργιος Μπακλαγής (2016). Χρήση Αξιοποίηση και Εφαρμογές των Big Data στα Logistics και το Ηλεκτρονικό Εμπόριο. Πανεπιστήμιο Πειραιώς, Πειραιάς.
14. ΔΕΣΠΟΙΝΑΣ Θ. ΜΕΡΙΔΟΥ (2012). Εφαρμογές και Τεχνικές Εξόρυξης και Οπτικοποίησης Γνώσης σε Βιοϊατρικά Δεδομένα. Πανεπιστήμιο Πατρών, Πάτρα.
15. Dr Mark van Rijmenam (2013, August 7). Why The 3V's Are Not Sufficient To Describe Big Data. Retrieved from <https://datafloq.com/read/3vs-sufficient-describe-big-data/166>
16. Daniel Trabucchi, Tommaso Buganza (2018). Data-driven innovation: switching the perspective of Big Data. European Journal of Innovation Management, Vol. 22 (No.1), pp.23-40.
17. Frank Ohlhorst (2012). Big Data Analytics: Turning Big Data into Big Money. Canada: John Wiley & Sons, Inc.
18. Garry Kranz, Dave Raffo (2018, May). Storage (computer storage). Retrieved from <https://searchstorage.techtarget.com/definition/storage>
19. Gema Bello-Orgaz, Jason J. Jung, David Camacho (2015). Social big data: Recent achievements and new challenges, Information Fusion, 28(2016), pp.45-59.
20. Hossein Hassani, Xu Huang, Emmanuel Sirimal Silva (2019). Fusing Big Data, Blockchain and Cryptocurrency. Springer Nature Switzerland AG.
21. Θωμά Λεγκάκου (2014). ΕΞΑΓΩΓΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΪΑΣ ΑΠΟ ΔΕΔΟΜΕΝΑ ΜΕΤΑΝΑΣΤΕΥΣΗΣ. Χαροκόπειο Πανεπιστήμιο, Αθήνα.
22. ΘΕΟΧΑΡΗ ΑΙΚΑΤΕΡΙΝΗ (2019). DATA MINING – ΑΠΟ ΤΗΝ ΑΝΑΚΑΛΥΨΗ ΣΤΗΝ ΕΦΑΡΜΟΓΗ. Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Ελλάδας, Αντίρριο.
23. Θάνος Ευάγγελος (2017). Big Data: Νέοι ρόλοι και ευκαιρίες για τους επαγγελματίες πληροφόρησης. Τεχνολογικό Εκπαιδευτικό Ίδρυμα Αθήνας, Αθήνα.

24. Lekha R. Nair, Sujala D. Shetty, Siddhanth D. Shetty (2016). INTERACTIVE VISUAL ANALYTICS ON BIG DATA: TABLEAU VS D3.JS. Journal of e-Learning and Knowledge Society, Vol. 12, No. 4, pp.139-150.
25. Marr Bernard (2015, February 24). A Brief History of Big Data Everyone Should Read. Retrieved from <https://www.linkedin.com/pulse/brief-history-big-data-everyone-should-read-bernard-marr>
26. Min Chen, Shiwen Mao, Yunhao Liu (2014). Big Data: A Survey. Mobile Networks and Applications, Vol. 19 (No.2), pp.171-209
27. Mousouleas Ioannis (2019). Applying big data in pharmaceutical industry: development, strategy and administration. University of Patras, Patras.
28. Neelam Tyagi (2019, August 21). What is Tableau? Working and Key Features. Retrieved from <https://www.analyticssteps.com/blogs/why-is-tableau-used-for-read-the-working-and-key-features-of-tableau>
29. Press Gil (2013, May 9). A Very Short History Of Big Data. Retrieved from <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data>
30. Philip Russom (2011). Big Data Analytics. TWDI Research
31. Παναγόπουλος Ευστάθιος (2018). Η ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ ΚΑΙ Η ΚΑΙΝΟΤΟΜΙΑ ΣΤΟΥΣ ΦΟΙΤΗΤΕΣ. Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Ελλάδας, Μεσολόγγι.
32. Pedro César Tebaldi Gomes (2014, July 3). Big Data Analytics. Retrieved from <https://www.opservices.com/big-data-analytics/>
33. Προδρομίτη Γερασιμούλα (2017). Μεγάλα Δεδομένα. Η εξόρυξη τους και η συμβολή τους στην Επιχειρηματική Ευφυΐα. Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Αθήνα.
34. Παναγιώτης Κίτσος, Παρασκευή Παππά (2013). Η ευρωπαϊκή νομοθεσία αντιμετώπιη με τα Big Data: Οι επιπτώσεις από την «έκρηξη» πληροφοριών στην προστασία της ιδιωτικής ζωής.
35. Στέφανος Χρ. Μπαλαμώτης (2020). Η επιρροή της τεχνολογίας μεγάλων δεδομένων (Big Data) στη λογιστική. Πανεπιστήμιο Θεσσαλίας, Βόλος.

36. Soraya Sedkaoui, Mounia Khelfaoui (2020). Big Data Analytics: Applied to the Sharing Economy (pp.119-140). Sharing Economy and Big Data Analytics, Great Britain, United States: ISTE Ltd and John Wiley & Sons, Inc.
37. Shingha Chaveesuk, Bilal Khalid, Wornchanok Chaiyasoonthorn (2019). Emergence of New Business Environment with Big Data and Artificial Intelligence (pp.181-185). Prague, Czech Republic: Association for Computing Machinery.
38. Samuel E. Buttrey, Lyn R. Whitaker (2018). A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R (pp.1-19). California, US: John Wiley & Sons Ltd.
39. Sahib Singh (2019). Security Analysis of MongoDB. International Journal of Digital Society (IJDS), Volume 10, Issue 4, pp.1556-1561.
40. Τσιάρρα Ελισσάβητ (2016). Τα Μεγάλα Δεδομένα και η Επιχειρηματική Ευφυΐα. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Θεσσαλονίκη.
41. Tal Z. Zarsky (2017). Incompatible: The GDPR in the Age of Big Data, Seton Hall Law Review, Vol. 47, No. 4(2), pp.995-1020.
42. ΤΖΕΚΑΣ ΑΛΕΞΑΝΔΡΟΣ (2018). Το RapidMiner ως Εργαλείο εφαρμογών Big Data Analytics. Πανεπιστήμιο Πειραιώς, Αθήνα.
43. What is Big Data? Retrieved from <https://www.oracle.com/big-data/guide/what-is-big-data.html>
44. What is GDPR, the EU's new data protection law? Retrieved from <https://gdpr.eu/what-is-gdpr/>
45. Χανδακάρης Ελευθέριος. ΕΦΑΡΜΟΓΗ ΚΑΙ ΑΝΑΛΥΣΗ BIG DATA ΣΤΗΝ ΒΕΛΤΙΩΣΗ ΤΟΥ ΚΟΣΤΟΥΣ ΔΙΑΧΕΙΡΙΣΗΣ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ. Πανεπιστήμιο Πατρών, Πάτρα.

Παράρτημα

Υπόδειγμα Ερωτηματολόγιο

Στο πλαίσιο διεξαγωγής της πτυχιακής μας εργασίας με θέμα "Big Data και Ανάπτυξη Επιχειρηματικότητας", η οποία εκπονείται στο τμήμα Διοικητικής Επιστήμης και Τεχνολογίας του Πανεπιστημίου Πατρών, συντάχθηκε το παρακάτω ερωτηματολόγιο. Η εργασία εκπονείται από τους φοιτητές: Μαυρούλια Παναγιώτη και Λαδά Μάριο.

Το ερωτηματολόγιο έχει ως σκοπό την διεξαγωγή πληροφοριών στην έρευνα σχετικά με την εξοικείωση, αλλά και την χρήση των Big Data στην ελληνική επιχειρηματικότητα. Περιλαμβάνει 17 ερωτήσεις (9 γενικές ερωτήσεις & 8 ειδικά για αυτούς που χρησιμοποιούν Big Data), οι οποίες συγκεντρώνουν πλήθος προσωπικών δεδομένων των χρηστών. Το ερωτηματολόγιο τηρεί με μέριμνα όλα όσα προβλέπονται από την κείμενη νομοθεσία περί προστασίας προσωπικών δεδομένων (GDPR).

Η χρήση των Big Data δεν είναι απαραίτητη για την απάντηση του ερωτηματολογίου. Η συμμετοχή σας στην παρούσα έρευνα θα αποτελούσε ένα ιδιαίτερα χρήσιμο εργαλείο εξαγωγής συμπερασμάτων.

Ευχαριστούμε εκ των προτέρων για τον χρόνο σας!

1. Ποια η θέση σας μέσα στην επιχείρηση-φορέα;
 Γενικών Καθηκόντων Data Analyst Υπεύθυνος Πληροφορ./Μαρκετ.
 Ανώτατο Στέλεχος Εξωτερ. Συνεργάτης Άλλο...
2. Ποιος είναι ο κλάδος που δραστηριοποιείτε η επιχείρησή σας;
 Υγείας Βιομηχανία Χρηματ/κη - Τραπεζική
 Τηλεπικοινωνίες Λιανικό Εμπόριο Άλλο...
3. Ποιος είναι ο συνολικός αριθμός των εργαζομένων σας;
 Έως 10 άτομα 11-50 άτομα 51-200 άτομα 200+ άτομα
4. Τί σημαίνουν για εσάς τα Μεγάλα Δεδομένα;
 Νέες μορφές δεδομένων και ανάλυσης Πληροφορία σε πραγματικό χρόνο
Μεγάλες ποσότητες δεδομένων Δεδομένα των Μέσων Κοινωνικής Δικτύωσης
 Δεδομένα από Νέες Τεχνολογίες Δεν Γνωρίζω Άλλο...
5. Πιστεύετε ότι τα Μεγάλα Δεδομένα χρησιμοποιούνται σήμερα στην ελληνική επιχειρηματικότητα;
 Πολύ Αρκετά Λίγο Καθόλου Δεν γνωρίζω

6. Κάνετε χρήση των Μεγάλων Δεδομένων στην επιχείρησή σας;
 Ναι Όχι
7. Εάν χρησιμοποιείτε, αξιολογείτε τον βαθμό εμπειρίας σας με τα Μεγάλα Δεδομένα.
Ελάχιστη 1 2 3 4 5 Προχωρημένη
8. Εάν όχι, θα χρησιμοποιούσατε τα Μεγάλα Δεδομένα στο μέλλον;
 Ναι Όχι Ίσως
9. Ποιος είναι ο κυριότερος λόγος που χρησιμοποιείτε, ή θα χρησιμοποιούσατε τα Μεγάλα Δεδομένα;
 Αύξηση Πωλήσεων Βελτιωμένη Εφοδιαστική Αλυσίδα
 Καλύτερη κατανόηση και στόχευση πελατών Βελτιωμένη Ανάλυση Αποδοτικότητας Πρόγνωση/Πρόβλεψη Βελτιωμένη Λήψη Αποφάσεων
 Άλλο...
10. Στην επιχείρησή σας, η συλλογή των Μεγάλων Δεδομένων είναι κυρίως από:
 Εντός της επιχείρησης Εξωτερικό περιβάλλον Και τα δύο.
11. Από ποιες από τις παρακάτω πηγές δεδομένων συλλέγετε και αναλύετε τα δεδομένα σας;
 Μέσα Κοινωνικής Δικτύωσης Emails Βίντεο Οικονομικές Συναλλαγές
 Κινητή τηλεφωνία Δεδομένα Αισθητήρων(π.χ. GPS) Έγγραφα
 Μηχανή Αναζήτησης Άλλο...
12. Ποια εργαλεία χρησιμοποιείτε για την συλλογή και ανάλυση των δεδομένων σας;
 Apache Spark R Apache Hadoop MongoDB RapidMiner Orange
 Tableau D3.js Άλλο...
13. Σε τί βαθμό πιστεύετε ότι έχουν αξιοποιηθεί τα εργαλεία που χρησιμοποιείτε;
Καθόλου 1 2 3 4 5 Πολύ
14. Η επιχείρησή σας έχει κάποια στρατηγική πάνω στα Μεγάλα Δεδομένα;
 Ναι Όχι
15. Αν ναι, σε τι βαθμό τα Μεγάλα Δεδομένα έχουν αυξήσει το ανταγωνιστικό πλεονέκτημα της επιχείρησής σας;
 Πολύ Σημαντικά Σημαντικά Μέτρια Λίγο Καθόλου

16. Σε 5 χρόνια, πόσο σημαντική πιστεύετε θα είναι η χρήση των Μεγάλων Δεδομένων στην λειτουργικότητα της επιχείρησή σας;

Ελάχιστα 1 2 3 4 5 Πάρα πολύ

17. Σε 5 χρόνια, ποιος πιστεύετε θα είναι ο ρόλος της χρήσης των Μεγάλων Δεδομένων στην επιχείρησή σας;

Κείμενο μακροσκελούς απάντησης...