

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

## Big data and Predictive Analytics – GDPR and new legislations



Διπλωματική Εργασία  
ΓΡΟΥΜΠΑΣ ΝΙΚΟΛΑΟΣ

ΕΠΙΒΛΕΠΟΥΣΑ: ΜΗΤΡΟΥ ΕΥΑΓΓΕΛΙΑ

ΣΑΜΟΣ, ΙΑΝΟΥΑΡΙΟΣ 2020

<b>Εισαγωγή</b>	<b>4</b>
<b>Κεφάλαιο 1<sup>ο</sup>: Μεγάλα δεδομένα (Big Data)</b>	<b>5</b>
1.1 Ορισμός	5
1.2 Η ιστορία των μεγάλων δεδομένων	6
1.3 Τι είναι τα μεγάλα δεδομένα	11
1.4 Χαρακτηριστικά των Big data - αναλύοντας τα V's	13
1.5 Πρακτικές συλλογής Μεγάλων δεδομένων – έπαινοι και κριτική	15
1.5.1 Χρήσεις των Μεγάλων Δεδομένων	16
1.6 Πως τα μεγάλα δεδομένα αποθηκεύονται και επεξεργάζονται	20
1.7 Μεγάλα Δεδομένα και νέες θέσεις εργασίας.	23
1.8 Ανοιχτά Δεδομένα (open data) και Μεγάλα Δεδομένα.	24
1.8.1 Τι είναι τα Ανοιχτά Δεδομένα.	24
1.8.2 Η σχέση μεταξύ των Big Data και των Ανοιχτών Δεδομένων.	25
<b>Κεφάλαιο 2<sup>ο</sup>: Προγνωστική Ανάλυση (Predictive analytics) και Μεγάλα Δεδομένα</b>	<b>27</b>
2.1 Εισαγωγή	27
2.2 Τι είναι Predictive Analytics	28
2.3 Γιατί BIG DATA και analytics μαζί	32
2.4 Social Media Analytics	34
2.4.1 Social Media Analytics και Intelligence Research	36
2.4.2 Ανταγωνιστική ευφυΐα	38
2.5 Ανάλυση συναισθημάτων ( Sentiment Analysis )	39
2.5.1 Τύποι ανάλυσης συναισθήματος	40
2.5.2 Αυτόματες Προσεγγίσεις Αλγορίθμων	41
2.5.3 Ανάλυση συναισθήματος σε αγορές	44
2.6 Εφαρμογές ανάλυσης μεγάλων δεδομένων	44
2.6.1 Ευκαιρίες και προκλήσεις ανάλυσης μεγάλων δεδομένων	45
2.6.2 Ευκαιρίες	45
2.6.3 Μεγάλο μέγεθος δείγματος για μεγάλα δεδομένα	46
2.6.4 Πρόοδος στην τεχνολογία	46
2.6.5 Σε πραγματικό χρόνο	47
2.6.6 Προκλήσεις με τα Big Data	47
2.6.7 Μέγεθος και ποιότητα δεδομένων	48
2.7 Μελέτη περίπτωσης	48
2.7.1 Περιγραφή εταιρείας	48
2.7.2 Σκοπός των Big Data στην περίπτωση της Lesara	49
2.7.3 Οφέλη της Lesara από την χρήση των Big Data	49
2.7.4 Τρία προβλήματα που λύνει η Lesara με την χρήση των Big Data	50
2.7.5 Πηγές της Lesara	50
2.7.6 Σύνοψη	51
<b>ΚΕΦΑΛΑΙΟ 3ο : Εξόρυξη γνώσης (data mining) και μεγάλων δεδομένων</b>	<b>52</b>
3.1 Ορισμός Εξόρυξης γνώσης και Δεδομένων.	52

<b>3.2 Εξόρυξη δεδομένων και ανεύρεση γνώσης.</b>	<b>52</b>
3.2.1 Εξόρυξη δεδομένων σε Supervised Data	54
3.2.2 Εξόρυξη δεδομένων σε Supervised Data	55
<b>3.3 Θεωρητικό υπόβαθρο</b>	<b>57</b>
<b>3.4 Στόχοι της εξόρυξης δεδομένων</b>	<b>58</b>
<b>3.5 Που εφαρμόζεται η εξόρυξη γνώσης.</b>	<b>60</b>
3.5.1 Εξόρυξη δεδομένων και οικονομία.	61
3.5.2 Εφαρμογές εξόρυξης γνώσης σε επιστημονικά πεδία.	62
<b>3.6 Αποτυγχάνουμε να αντλήσουμε γνώση από τα big data, συνεχίζουμε την εξόρυξη</b>	<b>63</b>
<b>ΚΕΦΑΛΑΙΟ 4ο : Κανονισμοί - Νομοθεσία GDPR</b>	<b>66</b>
<b>4.1 Κανονισμοί</b>	<b>66</b>
<b>4.1 ΑΡΘΡΑ</b>	<b>75</b>
<b>Βιβλιογραφικές Αναφορές</b>	<b>85</b>
Κεφάλαιο 1:	85
Κεφάλαιο 2:	86
Κεφάλαιο 3:	87
Κεφάλαιο 4:	88

# Εισαγωγή

Τα δεδομένα αναδεικνύονται ως η νέα πηγή στον κόσμο για ανταγωνιστικό πλεονέκτημα μεταξύ των εθνών, των οργανισμών και των επιχειρήσεων. Εκτιμάται ότι κάθε μέρα δημιουργούμε 2,5 πεντάκις εκατομμύρια bytes δεδομένων από διάφορες πηγές. Αυτά τα δεδομένα κυμαίνονται από σημειώσεις υπολογιστών ως αναρτήσεις σε ιστότοπους κοινωνικών μέσων, αρχεία συναλλαγών αγοράς και αγοράς εικόνων. Αυτή η συλλογή συνόλων δεδομένων, που είναι τόσο μεγάλα και σύνθετα και είναι δύσκολο να επεξεργασθούν χρησιμοποιώντας συμβατικά εργαλεία διαχείρισης των βάσεων δεδομένων, είναι γνωστά ως Big Data. Οι προκλήσεις στα Big Data περιλαμβάνουν την καταγραφή, επεξεργασία, αποθήκευση, αναζήτηση, κοινή χρήση, μεταφορά, ανάλυση και οπτικοποίηση των δεδομένων.

Παρ' όλα αυτά η συλλογή τους μόνο δεν προσφέρει κάποιο ιδιαίτερο όφελος στις επιχειρήσεις και τους οργανισμούς και για αυτό τεχνολογίες όπως Predictive Analytics, Machine Learning, Intelligent Research και Data Mining αναπτύχθηκαν ώστε να γίνεται η συλλογή των δεδομένων και με τη βοήθεια εξελιγμένων εργαλείων και μοντέλων πρόβλεψης, κάθε οργανισμός μπορεί τώρα να χρησιμοποιήσει προηγούμενα και τρέχοντα δεδομένα για να προβλέψει αξιόπιστα τάσεις και συμπεριφορές σε χιλιοστά του δευτερολέπτου, ημέρες ή χρόνια στο μέλλον δίνοντας του ένα μεγάλο ανταγωνιστικό πλεονέκτημα.

Παρακάτω σε αυτήν την εργασία θα επικεντρωθούμε και θα αναλύσουμε αυτές τις νέες τεχνολογίες και τις τεχνικές καθώς θα τις μελετήσουμε και από την σκοπιά του νομοθετικού πλαισίου του οποίου υπόκεινται.

# Κεφάλαιο 1<sup>ο</sup>: Μεγάλα δεδομένα (Big Data)

## 1.1 Ορισμός

Τα μεγάλα δεδομένα είναι ένας όρος που περιγράφει τον μεγάλο όγκο δεδομένων - τόσο δομημένων όσο και μη δομημένων - που κατακλύζουν τις επιχειρήσεις, τους χρήστες και γενικότερα όλο το ίντερνετ καθημερινά. Δεν είναι το ποσό των δεδομένων που είναι σημαντικό παρ' όλα αυτά αλλά η ανάλυση και η αξιοποίηση τους από τους οργανισμούς ώστε να εξαχθούν κάποια χρήσιμα συμπεράσματα και ιδέες οι οποίες θα οδηγήσουν σε καλύτερες αποφάσεις και στρατηγικές επιχειρηματικές κινήσεις.

Τα μεγάλα δεδομένα έχουν κάποια βασικά χαρακτηριστικά τα οποία διακρίνονται στις εξής κατηγορίες: Volume (όγκος), Velocity (Ταχύτητα), Variety (Ποικιλομορφία) ή αλλιώς γνωστά ως τα 3 V's.

**Volume.** Ο όρος Volume αναφέρεται στην ποσότητα των δεδομένων που παράγονται μέσω ιστοτόπων, πυλών και ηλεκτρονικών εφαρμογών. Οι οργανισμοί συλλέγουν δεδομένα από διάφορες πηγές, συμπεριλαμβανομένων επιχειρηματικών συναλλαγών, κοινωνικών μέσων και πληροφοριών από δεδομένα αισθητήρων ή μηχανημάτων, δηλαδή IoT (Internet of Things) Devices. Ειδικά για τις εταιρείες B2C (Business to Customer), ο Όγκος περιλαμβάνει τα διαθέσιμα δεδομένα που βρίσκονται εκεί έξω και πρέπει να αξιολογηθούν ως προς τη συνάφεια. Για την μοντελοποίηση του όγκου των δεδομένων που επεξεργάζονται ως σκεφτούμε τον όγκο από την άποψη των κοινωνικών μέσων ενημέρωσης, αφού τα κοινωνικά μέσα έχουν τεράστιο αντίκτυπο στα δεδομένα. Από το 2016, υπάρχουν πάνω από 2 τρισεκατομμύρια θέσεις και 250 δισεκατομμύρια φωτογραφίες που φορτώθηκαν. Το Facebook έχει πληθώρα προσωπικών δεδομένων και τα 2,2 δισεκατομμύρια χρήστες μοιράζονται δεδομένα ανά δευτερόλεπτο. Αυτό θα ήταν αδύνατο αν δεν ήταν για την ανάπτυξη μεγάλων δεδομένων.

**Velocity.** Με τον όρο Velocity αναφερόμαστε στην ταχύτητα με την οποία συλλέγουμε και μας έρχονται τα δεδομένα. Τα δεδομένα ρέουν με πρωτοφανή ταχύτητα και πρέπει να αντιμετωπίζονται έγκαιρα και σε πραγματικό χρόνο. Τεχνολογικές λύσεις για αυτό το πρόβλημα έχουν βρεθεί και έχουν πάρει την μορφή έξυπνων τεχνολογιών όπως Artificial Intelligence (Τεχνητή Νοημοσύνη) Deep learning και Machine learning (εκμάθηση μηχανής ή μηχανική εκμάθηση). Παραμένοντας στο παράδειγμα των κοινωνικών μέσων ενημέρωσης, 900 εκατομμύρια φωτογραφίες φορτώνονται καθημερινά στο Facebook, 500 εκατομμύρια tweets δημοσιεύονται στο Twitter, 0,4 εκατομμύρια ώρες βίντεο μεταφορτώνονται στο Youtube και 3,5 δισεκατομμύρια αναζητήσεις πραγματοποιούνται στο Google. Αυτό είναι σαν μια έκρηξη πυρηνικών δεδομένων. Τα Big Data βοηθούν μια εταιρεία ή οργανισμό να κρατήσει αυτή την έκρηξη, να δεχτεί την εισερχόμενη ροή δεδομένων και ταυτόχρονα να την επεξεργαστεί γρήγορα, ώστε να μην δημιουργεί σημεία συμφόρησης.

**Variety.** Τα μεγάλα δεδομένα είναι γρήγορα, είναι μεγάλα, αλλά είναι εξαιρετικά διαφορετικά. Η ποικιλία στα μεγάλα δεδομένα αναφέρεται σε όλα τα δομημένα και αδόμητα δεδομένα που έχουν τη δυνατότητα να παραχθούν είτε από ανθρώπους είτε από μηχανές. Τα πιο συχνά προστιθέμενα δεδομένα είναι δομημένα -απλά κείμενα, tweets, εικόνες & βίντεο. Εντούτοις, τα στοιχεία που δεν έχουν δομηθεί όπως τα μηνύματα ηλεκτρονικού ταχυδρομείου, τα φωνητικά μηνύματα, το χειρόγραφο κείμενο, η ανάγνωση φωνής, η ηχογράφηση κ.λπ., είναι επίσης σημαντικά στοιχεία στην ποικιλομορφία. Η ποικιλία αφορά στη δυνατότητα ταξινόμησης των εισερχόμενων δεδομένων σε διάφορες κατηγορίες.

Τα τελευταία χρόνια έχουν προστεθεί δύο έξτρα υποκατηγορίες, πιο εξειδικευμένες, για την καλύτερη περιγραφή των Big Data.

**Variability.** Εκτός από τις αυξανόμενες ταχύτητες και ποικιλίες δεδομένων, οι ροές δεδομένων μπορεί να είναι σε μεγάλο βαθμό ασυμβίβαστες με τις περιοδικές κορυφές. Υπάρχει κάτι που τείνει στα κοινωνικά μέσα; Τα ημερήσια, εποχιακά και τα γεγονότα που πυροδότησαν τα κορυφαία φορτία δεδομένων μπορεί να είναι δύσκολο να διαχειριστούν. Ακόμη περισσότερο με μη δομημένα δεδομένα.

**Complexity.** Τα σημερινά δεδομένα προέρχονται από πολλές πηγές, πράγμα που καθιστά δύσκολη τη σύνδεση, αντιστοίχιση, καθαρισμό και μετατροπή δεδομένων μεταξύ των συστημάτων. Ωστόσο, είναι απαραίτητο να συνδυαστούν και να συσχετιστούν οι σχέσεις, οι ιεραρχίες και οι πολλαπλές συνδέσεις δεδομένων ειδικά τα δεδομένα μπορούν γρήγορα να ξεφύγουν από τον έλεγχο.

## 1.2 Η ιστορία των μεγάλων δεδομένων

Παρά το γεγονός πως η έννοια των Μεγάλων Δεδομένων εμφανίστηκε πρόσφατα, τα θεμέλια πάνω στα οποία στηρίχθηκε η εμφάνιση και εξέλιξή τους, τέθηκαν πολλά χρόνια πριν. Πολύ πριν την εμφάνιση των ηλεκτρονικών υπολογιστών, υπήρξε η ανάγκη να αποθηκεύουμε, να αναπαράγουμε και αναλύουμε τις πληροφορίες που κατείχαμε. Τον τελευταίο αιώνα ο όγκος των πληροφοριών που διαχειριζόμαστε αυξήθηκε ταχύτατα και αναπτύχθηκαν τεχνικές και μέσα αποθήκευσης και ανάλυσης που διευκόλυναν τη διαχείρισή τους, όπως το ίντερνετ και η ψηφιακή αποθήκευση.

Τα πρώτα αρχεία της χρήσης δεδομένων για την παρακολούθηση και τον έλεγχο των επιχειρήσεων χρονολογούνται από 7.000 χρόνια πριν, όταν η λογιστική εισήχθη στη Μεσοποταμία για να καταγράψει την ανάπτυξη των καλλιεργειών και των αγελάων.

Οι λογιστικές αρχές συνέχισαν να βελτιώνονται και το 1663 ο John Graunt κατέγραψε και εξέτασε όλες τις πληροφορίες σχετικά με τους ρόλους θνησιμότητας στο Λονδίνο. Ήθελε να αποκτήσει μια κατανόηση και να χτίσει ένα προειδοποιητικό σύστημα για τη συνεχιζόμενη παρωδία. Στην πρώτη καταγραφή της στατιστικής ανάλυσης των δεδομένων, συγκέντρωσε τα ευρήματά του στο βιβλίο Φυσικές και πολιτικές παρατηρήσεις που έγιναν από τα νομοσχέδια θνησιμότητας, το οποίο παρέχει μεγάλη εικόνα για τις αιτίες θανάτου τον

δέκατο έβδομο αιώνα. Λόγω του έργου του, ο Graunt μπορεί να θεωρηθεί ο πατέρας των στατιστικών.

Από εκεί και πέρα, οι λογιστικές αρχές βελτιώθηκαν, αλλά δεν συνέβη κάτι θεαματικό. Η παλαιότερη μνήμη των σύγχρονων δεδομένων είναι από το 1887 όταν ο Herman Hollerith εφηύρε μια υπολογιστική μηχανή που μπορούσε να διαβάσει τρύπες διάτρητες σε χαρτί για να οργανώσει δεδομένα απογραφής.

## **Ο 20ός αιώνας**

Το πρώτο σημαντικό έργο δεδομένων δημιουργήθηκε το 1937 και διατάχθηκε από τη διοίκηση του Franklin D. Roosevelt στις ΗΠΑ. Μετά την ψήφιση του νόμου περί κοινωνικής ασφάλισης το 1937, η κυβέρνηση έπρεπε να παρακολουθήσει τη συμβολή 26 εκατομμυρίων Αμερικανών και περισσότερων από 3 εκατομμυρίων εργοδοτών. Η IBM πήρε το συμβόλαιο για την ανάπτυξη της μηχανής ανάγνωσης καρτών διάτρησης για αυτό το τεράστιο έργο λογιστικής.

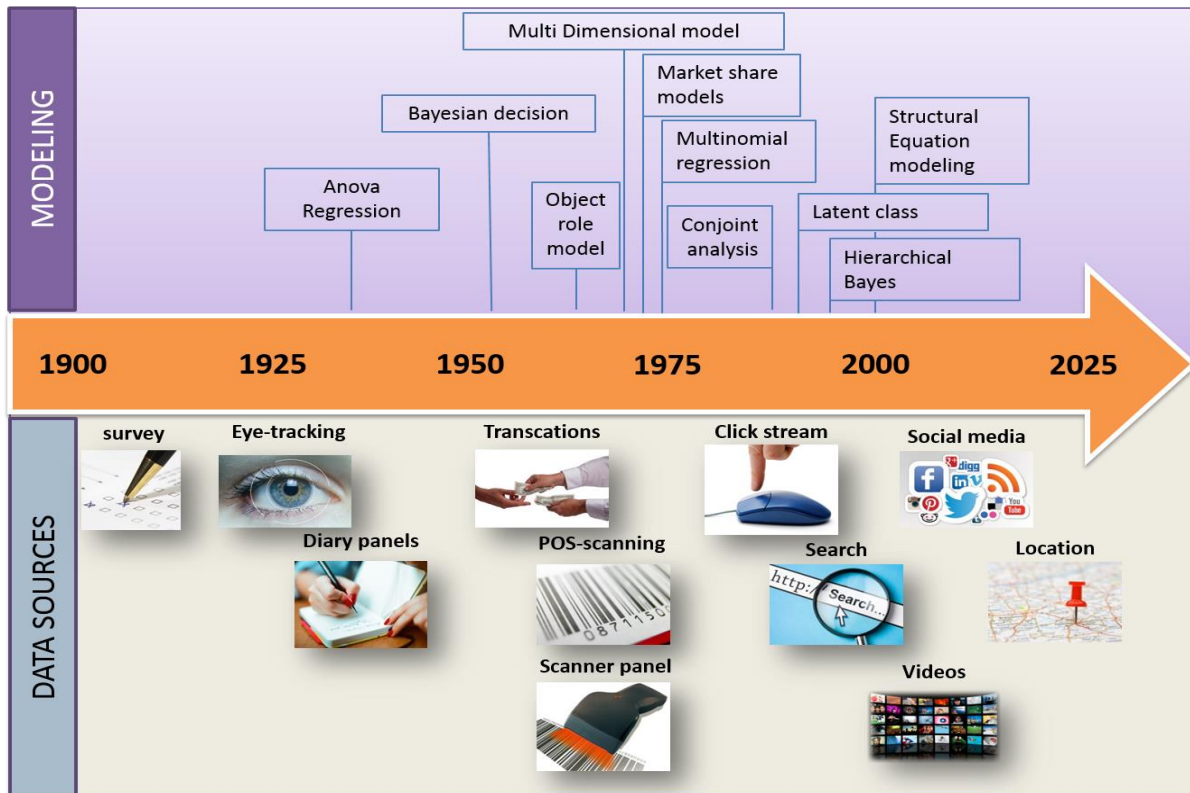
Η πρώτη μηχανή επεξεργασίας δεδομένων εμφανίστηκε το 1943 και αναπτύχθηκε από τους Βρετανούς για να αποκρυπτογραφήσει τους ναζιστικούς κώδικες κατά τη διάρκεια του Β' Παγκοσμίου Πολέμου. Αυτή η συσκευή, που ονομάζεται Colossus, έψαξε για πρότυπα στα ληφθέντα μηνύματα με ρυθμό 5.000 χαρακτήρων ανά δευτερόλεπτο. Με τον τρόπο αυτό μειώνοντας την εργασία από εβδομάδες σε ώρες.

Το 1952 δημιουργήθηκε ο Οργανισμός Εθνικής Ασφάλειας (NSA) και μέσα σε 10 χρόνια σύναψε περισσότερους από 12.000 κρυπτολόγους. Αντιμετωπίζουν την υπερφόρτωση πληροφοριών κατά τη διάρκεια του Ψυχρού Πολέμου καθώς ξεκινούν τη συλλογή και επεξεργασία σημάτων πληροφοριών αυτόματα.

Το 1965 η Ηνωμένη Κυβέρνηση αποφάσισε να κατασκευάσει το πρώτο κέντρο δεδομένων για την αποθήκευση πάνω από 742 εκατομμυρίων φορολογικών δηλώσεων και 175 εκατομμυρίων συνόλων δακτυλικών αποτυπωμάτων μεταφέροντας όλα αυτά τα αρχεία σε μαγνητική μαγνητοταινία που έπρεπε να αποθηκευτεί σε μια ενιαία τοποθεσία. Το έργο εγκαταλείφθηκε αργότερα από φόβο για τον «Big Brother», αλλά είναι γενικά αποδεκτό ότι ήταν η αρχή της εποχής της ηλεκτρονικής αποθήκευσης δεδομένων.

Το 1989 ο Βρετανός επιστήμονας ηλεκτρονικών υπολογιστών Tim Berners-Lee εφηύρε τελικά το World Wide Web. Ήθελε να διευκολύνει την ανταλλαγή πληροφοριών μέσω ενός συστήματος «υπερκειμένου». Λίγο θα μπορούσε να ξέρει τη στιγμή της επίδρασης της εφεύρεσης του.

Από τη δεκαετία του '90, η δημιουργία δεδομένων προωθείται καθώς όλο και περισσότερες συσκευές συνδέονται με το διαδίκτυο. Το 1995 κατασκευάστηκε ο πρώτος υπολογιστής, ο οποίος ήταν σε θέση να κάνει όσο το δυνατόν περισσότερο τη δουλειά σε ένα δευτερόλεπτο από ότι μια αριθμομηχανή που λειτουργεί από ένα μόνο άτομο μπορεί να το κάνει σε 30.000 χρόνια.



Εικόνα 1.1: Ιστορία των Big Data.

## Ο 21ος αιώνας

Το 2005, ο Roger Mougallas από την O'Reilly Media εφάρμοσε για πρώτη φορά τον όρο Big Data, μόνο ένα χρόνο μετά τη δημιουργία του όρου Web 2.0. Αναφέρεται σε ένα μεγάλο σύνολο δεδομένων που είναι σχεδόν αδύνατο να διαχειριστεί και να επεξεργαστεί χρησιμοποιώντας παραδοσιακά εργαλεία επιχειρηματικής ευφυΐας.

Το 2005 είναι επίσης η χρονιά που ο Hadoop δημιουργήθηκε από το Yahoo! που είναι χτισμένο πάνω από το MapReduce της Google. Σκοπός του ήταν η ευρετηρίαση ολόκληρου του Παγκόσμιου Ιστού και σήμερα το Hadoop ανοικτού κώδικα χρησιμοποιείται από πολλούς οργανισμούς για να τρυπήσει μέσω τεράστιων ποσοτήτων δεδομένων

Καθώς αρχίζουν να εμφανίζονται όλο και περισσότερα κοινωνικά δίκτυα και το Web 2.0 παίρνει πτήση, όλο και περισσότερα δεδομένα δημιουργούνται σε καθημερινή βάση. Οι καινοτόμες νεοσύστατες επιχειρήσεις αρχίζουν σιγά σιγά να σκάβουν σε αυτό το τεράστιο όγκο δεδομένων και επίσης οι κυβερνήσεις αρχίζουν να δουλεύουν σε έργα Big Data. Το 2009, η ινδική κυβέρνηση αποφασίζει να προβεί σε σάρωση ίριδας, δακτυλικά αποτυπώματα και φωτογραφία όλων των 1,2 δισεκατομμυρίων κατοίκων. Όλα αυτά τα δεδομένα αποθηκεύονται στη μεγαλύτερη βιομετρική βάση δεδομένων στον κόσμο.

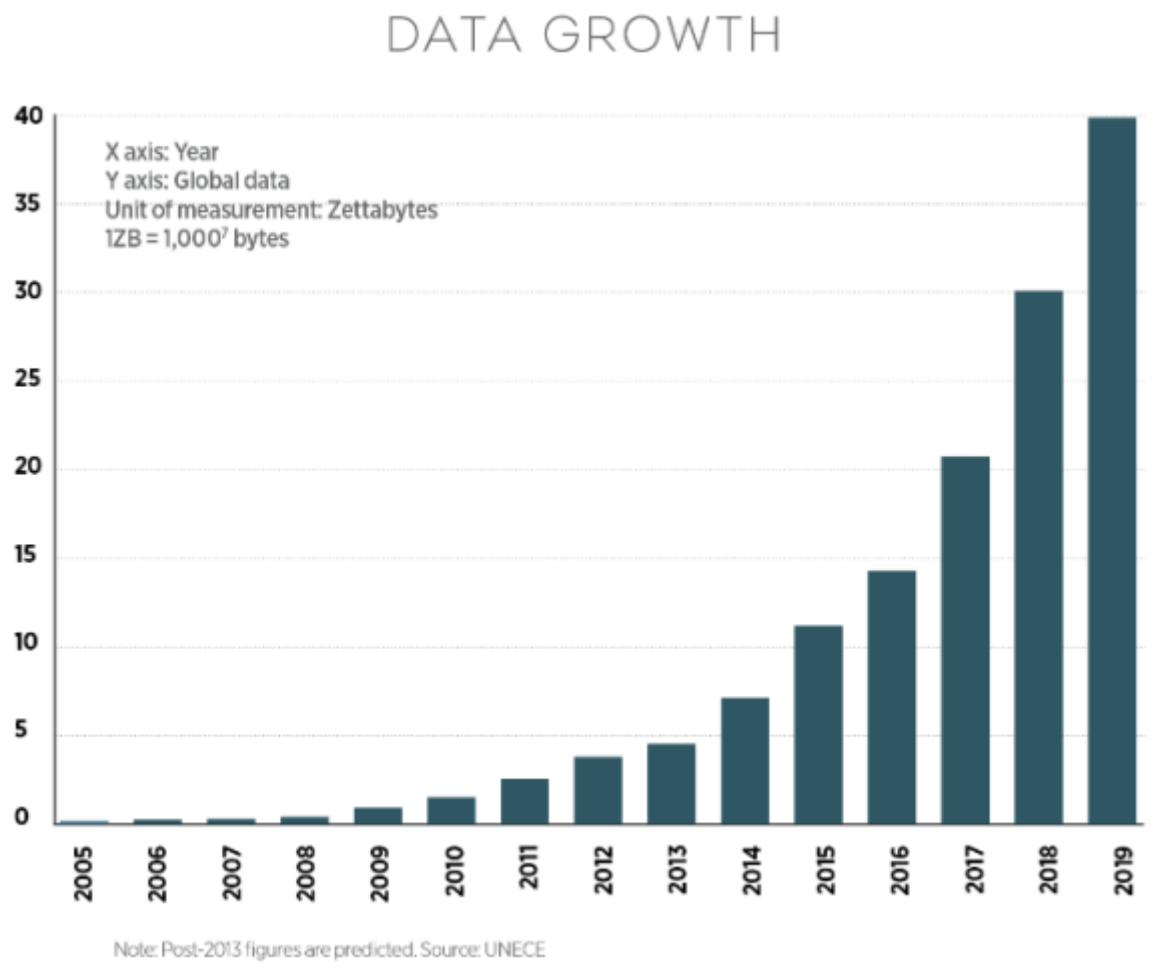
Το 2010 ο Eric Schmidt μιλάει στη διάσκεψη Techonomy στη λίμνη Tahoe στην Καλιφόρνια και δηλώνει ότι «υπήρχαν 5 exabytes πληροφοριών που δημιουργήθηκαν από



όλο τον κόσμο μεταξύ της αυγής του πολιτισμού και του 2003. Τώρα το ίδιο ποσό δημιουργείται κάθε δύο μέρες».

Το 2011 η έκθεση McKinsey για τα μεγάλα δεδομένα: Τα επόμενα σύνορα για την καινοτομία, τον ανταγωνισμό και την παραγωγικότητα, δηλώνει ότι το 2018 μόνο οι ΗΠΑ θα αντιμετωπίσουν έλλειψη 140.000 - 190.000 επιστημόνων δεδομένων καθώς και 1.5 εκατομμύρια διαχειριστές δεδομένων.

Τα τελευταία χρόνια, σημειώθηκαν τεράστιες αυξήσεις στις διοργανώσεις Big Data, όλοι προσπαθώντας να αντιμετωπίσουν το Big Data και βοηθώντας τους οργανισμούς να κατανοήσουν τα Big Data και όλο και περισσότερες εταιρείες υιοθετούν αργά και κινούνται προς τα Μεγάλα Δεδομένα. Ωστόσο, παρόλο που φαίνεται ότι το Big Data είναι εδώ και πολύ καιρό, στην πραγματικότητα το Big Data είναι όσο το Internet ήταν το 1993. Η μεγάλη επανάσταση Big Data είναι ακόμα μπροστά μας, έτσι πολλά θα αλλάξουν τα επόμενα χρόνια. Και έτσι ξεκίνησε η εποχή των μεγάλων δεδομένων.



Εικόνα 1.2: Ποσοστιαία αύξηση δεδομένων – Στοιχεία έως το 2019.

Τα Big Data δημιουργούνται ουσιαστικά από τα πάντα γύρω μας ανά πάσα στιγμή. Κάθε ψηφιακή αλληλεπίδραση με μέσα μαζικής δικτύωσης παράγει δεδομένα, από το browsing στον υπολογιστή και το online λιανεμπόριο μέχρι τις αγορές στο iTunes και τα likes στο Facebook. Τα δεδομένα αυτά αλιεύονται από πολλαπλές πηγές, με τρομακτική ταχύτητα, όγκο και ποικιλία. Για να εξαγάγει όμως κάποιος ουσιαστική αξία από αυτά, θα πρέπει να έχει στην κατοχή του τη βέλτιστη επεξεργαστική ισχύ, τα κατάλληλα εργαλεία ανάλυσης και φυσικά, τις ανάλογες δεξιότητες.

Το εύρος των δεδομένων που συλλέγονται από επιχειρήσεις σήμερα είναι σχεδόν εξωπραγματικό. Σύμφωνα με την IBM, δημιουργούνται πάνω από 2,5 τετράκις εκατομμύρια bytes δεδομένων ανά έτος, ενώ παράλληλα η ποσότητα των δεδομένων που παράγεται αυξάνεται σε ένα τέτοιο εκπληκτικό ποσοστό ώστε το 90% εξ αυτών να έχει δημιουργηθεί μόλις μέσα στα τελευταία δύο χρόνια!



Εικόνα 1.3: Πηγές από τις οποίες πηγάζουν τα Big Data

## 1.3 Τι είναι τα μεγάλα δεδομένα

Με απλά λόγια, 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. Ωστόσο, αυτοί οι τεράστιοι όγκοι δεδομένων μπορούν να χρησιμοποιηθούν για την αντιμετώπιση επιχειρηματικών προβλημάτων που δεν θα μπορούσαν να αντιμετωπιστούν πριν.

Ισχύει ευρέως πως τα τελευταία χρόνια υπάρχει ένα αυξανόμενο ενδιαφέρον για την Διαχείριση Μεγάλων Δεδομένων (ή αλλιώς Big Data) κινούμενο από πραγματικές ανάγκες της παραγωγής. Η πρώτη πραγματική ανάγκη παρουσιάστηκε το 1997 από τους επιστήμονες της NASA. Ανέφεραν ότι αδυνατούσαν να αναπαραστήσουν γραφικά (visualization) τα σύνολα δεδομένων που κατείχαν (data sets), καθώς ήταν τόσο μεγάλα που ήταν ακατόρθωτο να τα αποθηκεύσουν στη κύρια μνήμη, στον τοπικό δίσκο και σε εξωτερικό σκληρό δίσκο. Έτσι δήλωσαν ότι αντιμετώπιζον πρόβλημα Μεγάλων Δεδομένων. Οι τελευταίες τεχνολογικές εξελίξεις κυρίως στον τομέα των επικοινωνιών και των ολοκληρωμένων κυκλωμάτων έχουν δώσει την δυνατότητα να δημιουργηθούν μηχανισμοί παρακολούθησης των λειτουργιών ενός οργανισμού σε πολύ λεπτομερές επίπεδο. Η λεπτομερής αυτή ψηφιοποίηση των διαδικασιών παραγωγής έχουν καταστήσει μεγάλους οργανισμούς αλλά και εταιρείες μικρού μεγέθους ικανούς να παράγουν τεράστιους όγκους δεδομένων με πολύ ταχείς ρυθμούς. Τα δεδομένα αυτά κρύβουν πολύτιμη γνώση καθώς η ανάλυση τους μπορεί να οδηγήσει σε σημαντικές βελτιστοποιήσεις της παραγωγής αλλά και προβλήματα, αφού οι υπάρχουσες τεχνολογικές λύσεις για την διαχείριση δεδομένων δεν ανταποκρίνονται πλήρως στον όγκο αλλά και στην φύση τους.

Ψηφιακά δεδομένα πλέον συναντιούνται παντού. Έχουν ισχύ και "λέγουν" σε πολλούς τομείς όπως, σε κάθε οργανισμό, στην οικονομία στο ίντερνετ και σε κάθε χρήστη της ψηφιακής τεχνολογίας όπως τονίζει ο Philip Russom, ( 2011 ). Τα μεγάλα δεδομένα ελκύουν όλο και περισσότερο το ενδιαφέρον των ηγετών από όλους τους τομείς, ενώ οι καταναλωτές προϊόντων και υπηρεσιών αναμένεται να ωφεληθούν από την αξιοποίησή τους. Η ικανότητα αποθήκευσης, συγκέντρωσης, συνδυασμού δεδομένων και η χρήση των αποτελεσμάτων για την εκπόνηση λεπτομερών αναλύσεων έχει γίνει πολύ πιο προσιτή και εφικτή.

Τα Μεγάλα δεδομένα μπορούν να βοηθήσουν να αντιμετωπιστούν μια σειρά από επιχειρηματικές δραστηριότητες, από την εμπειρία των πελατών έως την ανάλυση. Κάποια από αυτά είναι,

### **Ανάπτυξη προϊόντων**

Εταιρείες όπως η Netflix και η Procter & Gamble χρησιμοποιούν μεγάλα δεδομένα για να προβλέψουν τη ζήτηση των πελατών. Κατασκευάζουν μοντέλα πρόβλεψης για νέα προϊόντα και υπηρεσίες, ταξινομώντας τα βασικά χαρακτηριστικά των παρελθόντων και των σημερινών προϊόντων ή υπηρεσιών και διαμορφώνοντας τη σχέση μεταξύ αυτών των χαρακτηριστικών και της εμπορικής επιτυχίας των προσφορών. Επιπλέον, η P & G χρησιμοποιεί δεδομένα και αναλύσεις από ομάδες εστίασης, κοινωνικά μέσα ενημέρωσης,

δοκιμαστικές αγορές και πρόωμη ανάπτυξη των καταστημάτων για το σχεδιασμό, την παραγωγή και την κυκλοφορία νέων προϊόντων.

### **Προγνωστική Συντήρηση**

Παράγοντες που μπορούν να προβλέψουν μηχανικές βλάβες μπορεί να είναι βαθιά θαμμένοι σε δομημένα δεδομένα, όπως το έτος κατασκευής, το μοντέλο μιας μηχανής, καθώς και σε μη δομημένα δεδομένα που καλύπτουν εκατομμύρια καταχωρήσεις καταγραφής, δεδομένα αισθητήρων, μηνύματα σφάλματος και θερμοκρασία κινητήρα. Αναλύοντας αυτές τις ενδείξεις πιθανών ζητημάτων πριν συμβούν τα προβλήματα, οι οργανισμοί μπορούν να αναπτύξουν τη συντήρηση πιο αποδοτικά και να μεγιστοποιήσουν το χρόνο λειτουργίας των εξαρτημάτων και εξοπλισμού.

### **Εμπειρία πελατών**

Ο αγώνας για τους πελάτες είναι ενεργοποιημένος. Μια σαφέστερη εικόνα της εμπειρίας των πελατών είναι πιο δυνατή τώρα από ποτέ. Μεγάλα δεδομένα επιτρέπουν την συλλογή δεδομένων από κοινωνικά μέσα, επισκέψεις ιστού, αρχεία καταγραφής κλήσεων και άλλες πηγές δεδομένων για την βελτίωση της εμπειρίας, της αλληλεπίδρασης και την μεγιστοποίηση της παραδοθείσας αξίας. Μπορούν να προσφερθούν εξατομικευμένες προσφορές, να μειωθεί το χτύπημα των πελατών και να γίνεται χειρισμός προληπτικά οποιονδήποτε ζητημάτων μπορεί να προκύψουν.

### **Απάτη και συμμόρφωση**

Όταν πρόκειται για την ασφάλεια, δεν είναι μόνο μερικοί αδίστακτοι χάκερ. Τα τοπία ασφαλείας και οι απαιτήσεις συμμόρφωσης εξελίσσονται συνεχώς. Τα μεγάλα δεδομένα μπορούν να βοηθήσουν να προσδιοριστούν τα πρότυπα στα δεδομένα που υποδεικνύουν απάτη και συγκεντρώνουν μεγάλους όγκους πληροφοριών για να γίνουν πολύ πιο γρήγορα οι ρυθμιστικές αναφορές.

### **Μηχανική μάθηση**

Η μηχανική μάθηση είναι ένα καυτό θέμα τον τελευταίο καιρό και τα δεδομένα - ειδικά μεγάλα δεδομένα - είναι ένας από τους λόγους. Τώρα μπορεί πλέον να διδαχτούν μηχανές αντί να προγραμματιστούν. Η διαθεσιμότητα μεγάλων δεδομένων για την εκμάθηση μοντέλων μηχανικής μάθησης κάνει αυτό πραγματικότητα.

### **Επιχειρησιακή απόδοση**

Η λειτουργική αποτελεσματικότητα μπορεί να μην είναι το πιο εμφανές πλεονέκτημα των Big Data, αλλά είναι ένας τομέας στον οποίο τα μεγάλα δεδομένα έχουν τον μεγαλύτερο αντίκτυπο. Με τα μεγάλα δεδομένα, μπορεί να αναλυθεί και να αξιολογηθεί η παραγωγή, οι ανατροφοδοτήσεις και οι αποδόσεις των πελατών και άλλοι παράγοντες για τη μείωση των διακοπών και την πρόβλεψη των μελλοντικών απαιτήσεων. Μεγάλα δεδομένα μπορούν

επίσης να χρησιμοποιηθούν για τη βελτίωση της διαδικασίας λήψης αποφάσεων σύμφωνα με την τρέχουσα ζήτηση της αγοράς.

### **Κίνηση της καινοτομίας**

Τα μεγάλα δεδομένα μπορούν να βοηθήσουν μια επιχείρηση να καινοτομήσει μελετώντας τις αλληλεξαρτήσεις μεταξύ ανθρώπων, θεσμών, οντοτήτων και διαδικασίας και στη συνέχεια να καθορίσει νέους τρόπους χρήσης αυτών των στοιχείων. Χρησιμοποιεί στοιχεία για να βελτιώσει τις αποφάσεις σχετικά με οικονομικούς και προγραμματιστικούς λόγους. Εξετάζει τις τάσεις και τι θέλουν οι πελάτες να παραδώσουν σε νέα προϊόντα, υπηρεσίες έως και εφαρμογή δυναμικής τιμολόγησης. Υπάρχουν ατελείωτες δυνατότητες.

## **1.4 Χαρακτηριστικά των Big data - αναλύοντας τα V's**

Αυτά τα ογκώδη στοιχεία που μπορεί να προέρχονται από μυριάδες διαφορετικές πηγές, όπως τα συστήματα εμπορικής συναλλαγής, βάσεις δεδομένων πελατών, ιατρικά αρχεία, αρχεία καταγραφής διαδικτύου, κινητές εφαρμογές, κοινωνικά δίκτυα, συλλεχθέντα αποτελέσματα επιστημονικών πειραμάτων, τα στοιχεία των 'μηχανών' που δημιουργούνται από αισθητήρες δεδομένων σε πραγματικό χρόνο, χρησιμοποιήθηκαν στο διαδίκτυο των πραγμάτων, (IoT) περιβάλλον. Τα δεδομένα ενδέχεται να παραμείνουν στην ακατέργαστη μορφή τους ή να προεπεξεργαστούν χρησιμοποιώντας εργαλεία εξόρυξης δεδομένων ή λογισμικό προετοιμασίας δεδομένων πριν αναλυθούν ( By Philip Russom, 2011 )

Τα μεγάλα δεδομένα περιλαμβάνουν επίσης μια μεγάλη ποικιλία τύπων δεδομένων, συμπεριλαμβανομένων δομημένων δεδομένων σε βάσεις δεδομένων SQL και αποθήκες δεδομένων, μη δομημένα δεδομένα, όπως αρχεία κειμένου και εγγράφων που διατηρούνται σε ομάδες Hadoop ή συστήματα NoSQL και ημιδομημένα δεδομένα, όπως αρχεία καταγραφής διακομιστή ιστού ή τη ροή δεδομένων από αισθητήρες. Περαιτέρω, τα μεγάλα δεδομένα περιλαμβάνουν πολλαπλές, ταυτόχρονες πηγές δεδομένων, οι οποίες δεν μπορούν να ενσωματωθούν διαφορετικά. Για παράδειγμα, ένα μεγάλο έργο ανάλυσης δεδομένων μπορεί να επιχειρήσει να μετρήσει την επιτυχία ενός προϊόντος και τις μελλοντικές πωλήσεις του, συσχετίζοντας προηγούμενα δεδομένα πωλήσεων, δεδομένα επιστροφής και ηλεκτρονικά δεδομένα αγοραστών για το συγκεκριμένο προϊόν.

Η ταχύτητα (Velocity) αναφέρεται στην ταχύτητα με την οποία παράγονται μεγάλα δεδομένα και πρέπει να υποβληθεί σε επεξεργασία και ανάλυση. Σε πολλές περιπτώσεις, τα σύνολα μεγάλων δεδομένων ενημερώνονται σε πραγματική ή σχεδόν πραγματική βάση, σε σύγκριση με καθημερινές, εβδομαδιαίες ή μηνιαίες ενημερώσεις σε πολλές παραδοσιακές αποθήκες δεδομένων. Τα μεγάλα προγράμματα ανάλυσης δεδομένων καταναλώνουν, συσχετίζουν και αναλύουν τα εισερχόμενα δεδομένα και μετά δίνουν μια απάντηση ή αποτέλεσμα βασισμένα σε ένα γενικό ερώτημα. Αυτό σημαίνει ότι οι επιστήμονες δεδομένων και άλλοι αναλυτές δεδομένων πρέπει να έχουν πλήρη κατανόηση των διαθέσιμων δεδομένων και να έχουν κάποια αίσθηση των απαντήσεων που αναζητούν για να βεβαιωθούν ότι οι πληροφορίες που λαμβάνουν είναι έγκυρες και ενημερωμένες. Η ταχύτητα είναι επίσης σημαντική καθώς η μεγάλη ανάλυση δεδομένων επεκτείνεται σε τομείς όπως η μάθηση

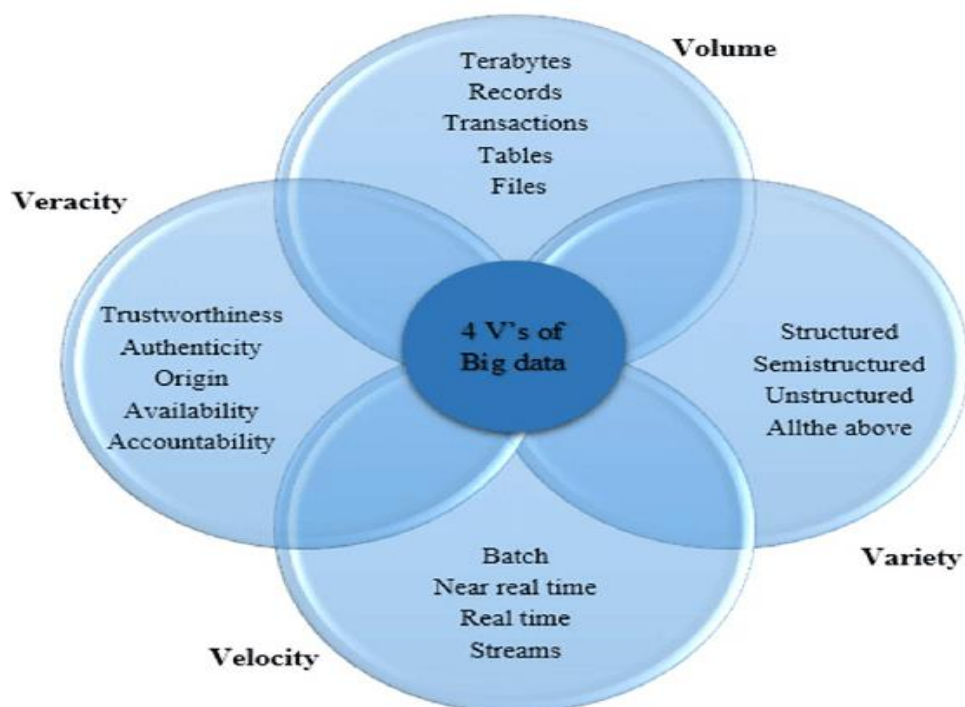
μηχανών και η τεχνητή νοημοσύνη (AI), όπου οι αναλυτικές διεργασίες εντοπίζουν αυτόματα τα πρότυπα στα δεδομένα που συλλέγονται και τα χρησιμοποιούν για να δημιουργήσουν ιδέες.

Η ακρίβεια (Veracity) των δεδομένων αναφέρεται στον βαθμό βεβαιότητας στα σύνολα δεδομένων. Τα αβέβαια πρωτογενή δεδομένα που συλλέγονται από πολλαπλές πηγές, όπως οι πλατφόρμες κοινωνικών μέσων και οι ιστοσελίδες, μπορούν να προκαλέσουν σοβαρά προβλήματα ποιότητας δεδομένων που μπορεί να είναι δύσκολο να εντοπιστούν. Για παράδειγμα, μια εταιρεία που συλλέγει δεδομένα από εκατοντάδες πηγές ενδέχεται να είναι σε θέση να εντοπίσει ανακριβή δεδομένα, αλλά οι αναλυτές της χρειάζονται πληροφορίες σχετικά με τις γενεαλογικές παραμέτρους για τον εντοπισμό των αποθηκευμένων δεδομένων ώστε να μπορέσουν να διορθώσουν τα προβλήματα.

Τα κακά δεδομένα οδηγούν σε ανακριβή ανάλυση και ενδέχεται να υπονομεύσουν την αξία των επιχειρηματικών αναλυτικών στοιχείων, διότι μπορεί να αναγκάσουν τα στελέχη να μην εμπιστεύονται τα δεδομένα στο σύνολό τους. Ο αριθμός των αβέβαιων δεδομένων σε έναν οργανισμό πρέπει να λογίζεται πριν χρησιμοποιηθεί σε μεγάλες εφαρμογές ανάλυσης δεδομένων. Οι ομάδες τεχνολογιών πληροφορικής και αναλύσεων πρέπει επίσης να διασφαλίσουν ότι διαθέτουν αρκετά ακριβή δεδομένα για να παράγουν έγκυρα αποτελέσματα.

Μερικοί επιστήμονες δεδομένων προσθέτουν επίσης μια πέμπτη τιμή V - στη λίστα των χαρακτηριστικών των μεγάλων δεδομένων. Όπως εξηγείται παραπάνω, δεν συλλέγονται όλα τα δεδομένα με πραγματική επιχειρησιακή αξία και η χρήση ανακριβών δεδομένων μπορεί να αποδυναμώσει τις πληροφορίες που παρέχουν οι εφαρμογές ανάλυσης. Είναι κρίσιμο οι οργανισμοί να χρησιμοποιούν πρακτικές όπως η εκκαθάριση δεδομένων και να επιβεβαιώνουν ότι τα δεδομένα σχετίζονται με συναφή επιχειρηματικά ζητήματα προτού τα χρησιμοποιήσουν σε ένα μεγάλο έργο ανάλυσης δεδομένων.

Η μεταβλητότητα (Variability) ισχύει επίσης συχνά για σύνολα μεγάλων δεδομένων, τα οποία είναι λιγότερο συνεπή από τα συμβατικά δεδομένα συναλλαγών και μπορεί να έχουν πολλαπλές έννοιες ή να μορφοποιηθούν με διαφορετικούς τρόπους από μία πηγή δεδομένων σε άλλη - πράγματα που περιπλέκουν ακόμη περισσότερο τις προσπάθειες επεξεργασίας και ανάλυσης των δεδομένων. Μερικοί άνθρωποι αποδίδουν ακόμη περισσότερα σε μεγάλα δεδομένα - οι επιστήμονες δεδομένων και οι σύμβουλοι έχουν δημιουργήσει διάφορους καταλόγους με μεταξύ επτά και 10 V's.



Εικόνα 1.4: Τα κύρια V's των Big Data

## 1.5 Πρακτικές συλλογής Μεγάλων δεδομένων – έπαινοι και κριτική

Για πολλά χρόνια, οι εταιρείες είχαν λίγους περιορισμούς όσον αφορά το είδος των δεδομένων που συλλέγουν από τους πελάτες τους. Οι εταιρείες χρησιμοποιούν τα μεγάλα δεδομένα που συσσωρεύονται στα συστήματά τους για να βελτιώσουν τις λειτουργίες τους, να προσφέρουν καλύτερη εξυπηρέτηση πελατών, να δημιουργήσουν εξατομικευμένες εκστρατείες μάρκετινγκ με βάση συγκεκριμένες προτιμήσεις των πελατών και, τελικά, να αυξήσουν την κερδοφορία. Τα μεγάλα δεδομένα χρησιμοποιούνται επίσης από ιατρικούς ερευνητές για τον εντοπισμό των παραγόντων κινδύνου ασθενειών. Τα δεδομένα που προέρχονται από τα ηλεκτρονικά αρχεία υγείας, τα κοινωνικά μέσα, τον ιστό και άλλες πηγές παρέχουν ενημερωμένες πληροφορίες σχετικά με απειλές ή εκδηλώσεις μολυσματικών ασθενειών.

Όμως, καθώς η συλλογή και χρήση δεδομένων έχει αυξηθεί, υπάρχει επίσης κακή χρήση δεδομένων. Οι πολίτες που εκφράζουν ανησυχία και έχουν βιώσει την κακή μεταχείριση των δεδομένων τους ή έχουν πέσει θύματα παραβίασης δεδομένων ζητούν τη θέσπιση νόμων σχετικά με τη διαφάνεια της συλλογής δεδομένων και την ιδιωτικότητα των δεδομένων των καταναλωτών.

Η κατακραυγή για τις παραβιάσεις προσωπικών δεδομένων οδήγησε την Ευρωπαϊκή Ένωση στην έγκριση του Γενικού Κανονισμού Προστασίας Δεδομένων (GDPR), ο οποίος τέθηκε σε ισχύ τον Μάιο του 2018. περιορίζει τους τύπους δεδομένων που μπορούν να

συλλέξουν οι οργανισμοί και απαιτεί συγκατάθεση από ιδιώτες. Παρόλο που δεν υπάρχουν παρόμοιοι νόμοι στις Η.Π.Α., κυβερνητικοί αξιωματούχοι διερευνούν τις πρακτικές επεξεργασίας δεδομένων, συγκεκριμένα μεταξύ εταιρειών που συλλέγουν δεδομένα καταναλωτών και τις πουλάνε σε άλλες εταιρείες για άγνωστη χρήση.

### **1.5.1 Χρήσεις των Μεγάλων Δεδομένων**

Ο πρωταρχικός στόχος των μεγάλων αναλυτικών στοιχείων είναι να βοηθήσουν τις επιχειρήσεις να λαμβάνουν επιχειρηματικές αποφάσεις με γνώμονα την πληροφόρηση, επιτρέποντας σε επιστήμονες δεδομένων, προγνωστικούς διαμορφωτές και άλλους επαγγελματίες αναλυτών να αναλύουν μεγάλους όγκους δεδομένων συναλλαγών καθώς και άλλες μορφές δεδομένων που δεν μπορούν να αξιοποιηθούν από συμβατικό Business Intelligence (BI). Αυτό θα μπορούσε να περιλαμβάνει τα αρχεία καταγραφής διακομιστή και τα δεδομένα ροής δεδομένων στο Internet, περιεχόμενο κοινωνικών μέσων και αναφορές δραστηριοτήτων κοινωνικού δικτύου, κείμενα από ηλεκτρονικά μηνύματα πελάτη και απαντήσεις έρευνας, αρχεία κινητών τηλεφωνικών κλήσεων και δεδομένα μηχανών που συλλαμβάνονται από αισθητήρες και συνδέονται με το Internet of Things.

Τα μεγάλα δεδομένα έχουν βρει πολλές εφαρμογές σε διάφορους τομείς σήμερα. Οι κύριοι τομείς στους οποίους χρησιμοποιούνται μεγάλα δεδομένα έχουν ως εξής.

#### **Κυβέρνηση**

Οι μεγάλες αναλύσεις δεδομένων έχουν αποδειχθεί πολύ χρήσιμες στον κυβερνητικό τομέα. Η μεγάλη ανάλυση δεδομένων διαδραμάτισε σημαντικό ρόλο στην επιτυχημένη εκστρατεία επανεκλογής του 2012 του Μπαράκ Ομπάμα. Επίσης, πρόσφατα, η μεγάλη ανάλυση δεδομένων ήταν υπεύθυνη για το BJP και τους συμμάχους του να κερδίσουν μια εξαιρετικά επιτυχημένη ινδική γενική εκλογή 2014. Η ινδική κυβέρνηση χρησιμοποιεί πολλές τεχνικές για να διαπιστώσει πώς το ινδικό εκλογικό σώμα ανταποκρίνεται στις κυβερνητικές ενέργειες, καθώς και ιδέες για πολιτική αύξηση.

#### **Social Media Analytics**

Η έλευση των κοινωνικών μέσων ενημέρωσης οδήγησε σε έκρηξη μεγάλων δεδομένων. Διάφορες λύσεις έχουν σχεδιαστεί για την ανάλυση της δραστηριότητας των κοινωνικών μέσων όπως η Cognos Consumer Insights της IBM, μια λύση σημείο που τρέχει στην πλατφόρμα BigInsights της Big Data της IBM, μπορεί να έχει νόημα για την φλυαρία. Τα κοινωνικά μέσα ενημέρωσης μπορούν να παράσχουν πολύτιμες πληροφορίες σε πραγματικό χρόνο σχετικά με το πώς η αγορά ανταποκρίνεται σε προϊόντα και εκστρατείες. Με τη βοήθεια αυτών των στοιχείων, οι εταιρείες μπορούν να προσαρμόσουν ανάλογα τις τοποθετήσεις τιμολόγησης, προώθησης και καμπάνιας τους. Πριν από τη χρήση των



μεγάλων δεδομένων χρειάζεται να γίνει κάποια επεξεργασία των μεγάλων δεδομένων, προκειμένου να αντληθούν κάποια έξυπνα και πολύτιμα αποτελέσματα. Έτσι, για να γνωρίζουμε τη νοοτροπία των καταναλωτών, είναι απαραίτητη η εφαρμογή έξυπνων αποφάσεων που προέρχονται από μεγάλα δεδομένα.

## **Τεχνολογία**

Οι τεχνολογικές εφαρμογές μεγάλων δεδομένων αποτελούνται από τις ακόλουθες εταιρείες που ασχολούνται καθημερινά με τεράστια ποσά δεδομένων και τις θέτουν σε χρήση για επιχειρηματικές αποφάσεις. Για παράδειγμα, το eBay.com χρησιμοποιεί δύο αποθήκες δεδομένων στα 7.5 petabytes και 40PB καθώς και ένα cluster Hadoop 40PB για αναζήτηση, συστάσεις καταναλωτών και merchandising. Το Amazon.com χειρίζεται εκατομμύρια λειτουργίες back-end κάθε μέρα, καθώς και ερωτήματα από περισσότερους από μισό εκατομμύριο πωλητές τρίτων. Η βασική τεχνολογία που διατηρεί το Amazon σε λειτουργία είναι το Linux και από το 2005 είχαν τις τρεις μεγαλύτερες βάσεις δεδομένων Linux με χωρητικότητες 7,8 TB, 18,5 TB και 24,7 TB. Το Facebook διαχειρίζεται 50 δισεκατομμύρια φωτογραφίες από τη βάση χρηστών του. Η Windermere Real Estate χρησιμοποιεί ανώνυμα σήματα GPS από περίπου 100 εκατομμύρια οδηγούς για να βοηθήσει τους νέους αγοραστές κατοικιών να καθορίσουν τους τυπικούς χρόνους οδήγησης τους προς και από την εργασία σε διάφορες ώρες της ημέρας.

## **Ανίχνευση απάτης**

Για τις επιχειρήσεις των οποίων οι εργασίες αφορούν τυχόν αξιώσεις ή επεξεργασία συναλλαγών, η ανίχνευση απάτης είναι ένα από τα πιο συναρπαστικά παραδείγματα εφαρμογών μεγάλων δεδομένων. Ιστορικά, η ανίχνευση απάτης σε πραγματικό χρόνο έχει αποδείξει ότι είναι κάτι μη εφικτό. Στις περισσότερες περιπτώσεις, η απάτη ανακαλύπτεται πολύ καιρό μετά το γεγονός, σε ποιο σημείο έχει γίνει η ζημιά και το μόνο που έχει απομείνει είναι να ελαχιστοποιηθεί τη βλάβη και να προσαρμοστούν οι πολιτικές για να αποφευχθεί η επανάληψή της. Οι πλατφόρμες Big Data, οι οποίες μπορούν να αναλύσουν αξιώσεις και συναλλαγές σε πραγματικό χρόνο, να εντοπίσουν μεγάλης κλίμακας σχέδια σε πολλές συναλλαγές ή να ανιχνεύσουν ανώμαλη συμπεριφορά από έναν μεμονωμένο χρήστη, μπορούν να αλλάξουν το παιχνίδι ανίχνευσης απάτης.

## **Ανάλυση τηλεφωνικού κέντρου**

Κάνοντας στροφή προς τα παραδείγματα εφαρμογών μεγάλων δεδομένων που αντιμετωπίζουν οι πελάτες, όπως του τηλεφωνικού κέντρου, είναι ιδιαίτερα ισχυρά. Αυτό που συμβαίνει στο τηλεφωνικό κέντρο ενός πελάτη είναι συχνά ένα μεγάλο βαρόμετρο και επηρεάζει το συναίσθημα της αγοράς, αλλά χωρίς μια λύση Big Data, μεγάλο μέρος της διορατικότητας που μπορεί να προσφέρει ένα τηλεφωνικό κέντρο θα παραληφθεί ή θα ανακαλυφθεί πολύ αργά. Οι λύσεις Big Data μπορούν να βοηθήσουν στον εντοπισμό των επαναλαμβανόμενων προβλημάτων ή των μοντέλων συμπεριφοράς των πελατών και των υπαλλήλων, όχι μόνο με την έννοια της μέτρησης χρόνου / ποιότητας της ανάλυσης, αλλά και με την ίδια την καταγραφή και επεξεργασία περιεχομένου κλήσεων.

## **Τραπεζικές εργασίες**

Η χρήση των δεδομένων των πελατών δημιουργεί πάντα ζητήματα προστασίας της ιδιωτικής ζωής. Με την αποκάλυψη κρυφών συνδέσεων μεταξύ φαινομενικά άσχετων στοιχείων, τα μεγάλα στοιχεία ανάλυσης δεδομένων θα μπορούσαν ενδεχομένως να αποκαλύψουν ευαίσθητα προσωπικά δεδομένα. Οι έρευνες δείχνουν ότι το 62% των τραπεζιτών είναι προσεκτικοί στη χρήση μεγάλων δεδομένων λόγω προβλημάτων ιδιωτικότητας. Επιπλέον, η εξωτερική ανάθεση δραστηριοτήτων ανάλυσης δεδομένων ή η διανομή δεδομένων πελατών μεταξύ των τμημάτων για τη δημιουργία πλουσιότερων στοιχείων ενισχύει επίσης τους κινδύνους ασφάλειας. Όπως τα κέρδη των πελατών, οι αποταμιεύσεις, τα στεγαστικά δάνεια και τα ασφαλιστήρια συμβόλαια κατέληξαν σε λάθος χέρια. Τέτοια περιστατικά ενισχύουν τις ανησυχίες σχετικά με την ιδιωτική ζωή των δεδομένων και αποθαρρύνουν τους πελάτες να μοιράζονται προσωπικές πληροφορίες με αντάλλαγμα προσαρμοσμένες προσφορές.

## **Γεωργία**

Μια εταιρεία βιοτεχνολογίας χρησιμοποιεί στοιχεία αισθητήρων για τη βελτιστοποίηση της αποδοτικότητας των καλλιεργειών. Εγκαθιστά τις δοκιμαστικές καλλιέργειες και εκτελεί προσομοιώσεις για να μετρήσει τον τρόπο με τον οποίο τα φυτά αντιδρούν στις διάφορες αλλαγές της κατάστασης. Το περιβάλλον δεδομένων του προσαρμόζεται συνεχώς στις μεταβολές των χαρακτηριστικών των διαφόρων δεδομένων που συλλέγει, συμπεριλαμβανομένων της θερμοκρασίας, των επιπέδων των υδάτων, της σύνθεσης του εδάφους, της ανάπτυξης, της παραγωγής και της αλληλουχίας των γονιδίων κάθε μονάδας στην κλίνη εξέτασης. Αυτές οι προσομοιώσεις επιτρέπουν την ανακάλυψη των βέλτιστων περιβαλλοντικών συνθηκών για συγκεκριμένους τύπους γονιδίων.

## **Εμπορία**

Οι έμποροι έχουν αρχίσει να χρησιμοποιούν το λογισμικό αναγνώρισης προσώπου για να μάθουν πόσο καλά η διαφήμισή τους επιτυγχάνει ή αποτυγχάνει στην τόνωση του ενδιαφέροντος για τα προϊόντα τους. Μια πρόσφατη μελέτη που δημοσιεύτηκε στο περιοδικό Harvard Business Review εξέτασε τι είδους διαφημίσεις υποχρέωσαν τους θεατές να συνεχίσουν να παρακολουθούν και τι απέτρεψε τους θεατές. Μεταξύ των εργαλείων τους ήταν ένα "σύστημα που αναλύει τις εκφράσεις του προσώπου για να αποκαλύψει τι βλέπουν οι θεατές". Η έρευνα σχεδιάστηκε για να ανακαλύψει τι είδους προωθήσεις προκάλεσε παρατηρητές να μοιραστούν τις διαφημίσεις με το κοινωνικό τους δίκτυο, βοηθώντας τους διαφημιζόμενους να δημιουργήσουν διαφημίσεις που είναι πιθανότερο να "Και τη βελτίωση των πωλήσεων.

## Smart Phones

Ίσως πιο εντυπωσιακή, οι άνθρωποι τώρα φέρουν την τεχνολογία αναγνώρισης προσώπου στις τσέπες τους. Οι χρήστες του smartphone I-Phone και του Android διαθέτουν αιτήματα στα χέρια τους που χρησιμοποιούν τεχνολογία αναγνώρισης προσώπου για διάφορες εργασίες. Για παράδειγμα, οι χρήστες Android με την εφαρμογή που θυμάται μπορούν να τραβήξουν μια φωτογραφία κάποιου χρήστη και στη συνέχεια να εμφανίσουν αποθηκευμένες πληροφορίες σχετικά με το συγκεκριμένο άτομο, με βάση την εικόνα τους, όταν η δική τους μνήμη τους αφήνει ένα πιθανό όφελος για τους πωλητές.

## Τηλεπικοινωνίες

Τώρα τα μεγάλα δεδομένα μιας ημέρας χρησιμοποιούνται σε διαφορετικά πεδία. Στην τηλεπικοινωνία παίζει επίσης πολύ καλό ρόλο. Οι φορείς εκμετάλλευσης αντιμετωπίζουν μια ανησυχητική πρόκληση όταν χρειάζονται νέες, συναρπαστικές υπηρεσίες δημιουργίας εσόδων, χωρίς να επιβαρύνουν τα δίκτυά τους και να διατηρήσουν το κόστος λειτουργίας τους υπό έλεγχο. Η αγορά απαιτεί νέα σύνολα δυνατοτήτων διαχείρισης και ανάλυσης δεδομένων που μπορούν να βοηθήσουν τους παρόχους υπηρεσιών να λαμβάνουν ακριβείς αποφάσεις, λαμβάνοντας υπόψη τον πελάτη, το περιβάλλον δικτύου και άλλες κρίσιμες πτυχές των επιχειρήσεων τους. Οι περισσότερες από αυτές τις αποφάσεις πρέπει να λαμβάνονται σε πραγματικό χρόνο, ασκώντας πρόσθετη πίεση στους φορείς εκμετάλλευσης. Οι αναλυτικές μέθοδοι πρόβλεψης σε πραγματικό χρόνο μπορούν να συμβάλουν στην αξιοποίηση των δεδομένων που βρίσκονται στα πολλαπλά τους συστήματα, να είναι άμεσα προσβάσιμα και να βοηθούν στη συσχέτιση αυτών των δεδομένων ώστε να παράγουν γνώση που μπορεί να τους βοηθήσει να οδηγήσουν τις επιχειρήσεις τους προς τα εμπρός.

## Φροντίδα υγείας

Παραδοσιακά, ο κλάδος της υγειονομικής περίθαλψης έχει καθυστερήσει από άλλες βιομηχανίες στη χρήση μεγάλων δεδομένων, ένα μέρος του προβλήματος προέρχεται από την αντίσταση στους παρόχους αλλαγών που έχουν συνηθίσει να λαμβάνουν αποφάσεις θεραπείας ανεξάρτητα, χρησιμοποιώντας τη δική τους κλινική κρίση, αντί να βασίζονται σε πρωτόκολλα βασισμένα σε μεγάλα δεδομένα. Άλλα εμπόδια έχουν πιο διαρθρωτικό χαρακτήρα. Ακόμη και στο εσωτερικό ενός νοσοκομείου, ενός πληρωτή ή μιας φαρμακευτικής εταιρείας, σημαντικές πληροφορίες παραμένουν συχνά παραμορφωμένες σε μία ομάδα ή τμήμα, επειδή οι οργανώσεις δεν διαθέτουν διαδικασίες για την ενσωμάτωση δεδομένων και την κοινοποίηση των ευρημάτων. Οι ενδιαφερόμενοι για την υγειονομική περίθαλψη έχουν τώρα πρόσβαση σε υποσχόμενες νέες γνώσεις. Οι πληροφορίες αυτές είναι μια μορφή "μεγάλων δεδομένων", που αποκαλούνται όχι μόνο για τον τεράστιο όγκο της, αλλά για την πολυπλοκότητα, την ποικιλομορφία και τα χρονοδιαγράμματα. Οι εξαγωγές φαρμακευτικής βιομηχανίας, οι πληρωτές και οι πάροχοι αρχίζουν τώρα να αναλύουν μεγάλα δεδομένα για να αποκτήσουν πληροφορίες. Οι πρόσφατες τεχνολογικές εξελίξεις στη βιομηχανία έχουν βελτιώσει την ικανότητά τους να συνεργάζονται με τέτοια δεδομένα, παρόλο που τα αρχεία είναι τεράστια και συχνά έχουν διαφορετικές δομές βάσεων δεδομένων και τεχνικά χαρακτηριστικά.

## 1.6 Πως τα μεγάλα δεδομένα αποθηκεύονται και επεξεργάζονται

Η ανάγκη αντιμετώπισης μεγάλης ταχύτητας δεδομένων επιβάλλει μοναδικές απαιτήσεις στη υποκείμενη υπολογιστική υποδομή. Η υπολογιστική ισχύς που απαιτείται για να επεξεργάζεται γρήγορα τεράστιους όγκους και ποικιλίες δεδομένων μπορεί να κατακλύσει ένα διακομιστή ή σύμπλεγμα διακομιστών. Οι οργανισμοί πρέπει να εφαρμόζουν επαρκή ικανότητα επεξεργασίας σε μεγάλα καθήκοντα δεδομένων για την επίτευξη της απαιτούμενης ταχύτητας. Αυτό μπορεί δυνητικά να απαιτήσει εκατοντάδες ή χιλιάδες διακομιστές που μπορούν να διανείμουν τις εργασίες επεξεργασίας και να λειτουργούν σε συνεργασία σε μια συστοιχία αρχιτεκτονικής.

Η επίτευξη τέτοιας ταχύτητας με οικονομικά αποδοτικό τρόπο είναι επίσης μια πρόκληση. Πολλοί ηγέτες των επιχειρήσεων είναι ανυπόμονοι να επενδύσουν σε μια εκτεταμένη υποδομή αποθήκευσης για να υποστηρίξουν μεγάλο φόρτο εργασίας δεδομένων, ιδιαίτερα εκείνων που δεν λειτουργούν 24/7. Ως αποτέλεσμα, το δημόσιο υπολογιστικό νέφος είναι πλέον ένα βασικό μέσο για τη φιλοξενία μεγάλων συστημάτων δεδομένων. Ένας δημόσιος πάροχος σύννεφων μπορεί να αποθηκεύσει petabytes δεδομένων και να αυξήσει τον απαιτούμενο αριθμό διακομιστών για αρκετό καιρό για να ολοκληρώσει ένα μεγάλο έργο ανάλυσης δεδομένων. Η επιχείρηση πληρώνει μόνο για τον χρόνο αποθήκευσης και τον υπολογισμό που πραγματικά χρησιμοποιείται και οι περιπτώσεις cloud μπορούν να απενεργοποιηθούν μέχρι να χρειαστούν ξανά.

Για να βελτιώσουν ακόμη περισσότερο τα επίπεδα υπηρεσιών, οι δημόσιοι προμηθευτές cloud προσφέρουν μεγάλες δυνατότητες δεδομένων μέσω διαχειριζόμενων υπηρεσιών που περιλαμβάνουν εξαιρετικά καταναμημένες υποθέσεις υπολογιστών Apache Hadoop, μηχανή επεξεργασίας Apache Spark και σχετικές μεγάλες τεχνολογίες δεδομένων. Το Amazon Elastic MapReduce (EMR) από την υπηρεσία Amazon Web Services (AWS) είναι ένα παράδειγμα μιας μεγάλης υπηρεσίας δεδομένων που τρέχει σε ένα δημόσιο σύννεφο. Άλλοι περιλαμβάνουν το Azure HDInsight της Microsoft και το Google Cloud Dataproc. Σε περιβάλλοντα Cloud, μεγάλα δεδομένα μπορούν να αποθηκευτούν στο Hadoop Distributed File System (HDFS) ή σε αποθηκευτικό αντικείμενο σύννεφο χαμηλού κόστους, όπως η υπηρεσία απλής αποθήκευσης Amazon (S3). Οι βάσεις δεδομένων NoSQL είναι μια άλλη επιλογή στο σύννεφο για εφαρμογές που είναι κατάλληλες για αυτούς.

Για τους οργανισμούς που επιθυμούν να αναπτύξουν μεγάλα συστήματα δεδομένων στο χώρο, οι κοινές τεχνολογίες ανοιχτού κώδικα της Apache εκτός από τον Hadoop και τον Spark περιλαμβάνουν ακόμη έναν άλλο διαπραγματευτή πόρων (YARN), τον ενσωματωμένο διαχειριστή πόρων και τον προγραμματιστή εργασίας του Hadoop. το πλαίσιο προγραμματισμού MapReduce. Kafka, πλατφόρμα μηνυμάτων εφαρμογής και εφαρμογής και ροή δεδομένων. τη βάση δεδομένων HBase. και μηχανές αναζήτησης SQL-on-Hadoop όπως Drill, Hive, Impala και Presto. Οι χρήστες μπορούν να εγκαταστήσουν τις εκδόσεις ανοιχτού κώδικα των τεχνολογιών τους ή να μεταβούν σε εμπορικές πλατφόρμες μεγάλων δεδομένων που προσφέρονται από τις τεχνολογίες Cloudera, Hortonworks και MapR, οι οποίες υποστηρίζονται επίσης από το cloud. Ωστόσο, οι Cloudera και Hortonworks συμφώνησαν να συγχωνευθούν τον Οκτώβριο του 2018, γεγονός που πιθανόν θα μειώσει τον αριθμό των διαθέσιμων πλατφόρμων εντός των εγκαταστάσεων σε δύο.

Υπάρχει ακόμα όμως η προκατάληψη και ένας σταθερός αλλά κατανοητός φόβος ότι τα μεγάλα δεδομένα μπορούν να διαχειριστούν και να αξιοποιηθούν μόνο από μεγάλες επιχειρήσεις. Επιπλέον, ενώ ο όρος "μεγάλα δεδομένα" μπορεί επίσης να είναι ένα ισχυρό παράδειγμα κενής ρητορικής του μάρκετινγκ, αποτελεί όμως ουσιαστικό μέρος κάθε εταιρικής επιχείρησης. Εν τέλει όμως, είναι η συσσώρευση επιχειρηματικών εγγράφων, σχεδίων, δεδομένων χρηστών και οικονομικών πληροφοριών και βασικών συστατικών κάθε επιχειρηματικής υποδομής.

Φυσικά, οι λεπτομέρειες μιας επιχείρησης εμφανίζονται συχνά ως μια σειρά από συντριπτικές προοπτικές και ευθύνες, με φαινομενικά λίγες μεθόδους στη διαχείριση όλων αποτελεσματικά. Ωστόσο, πρόσφατα, οι μικρές επιχειρήσεις επωφελούνται και από τα μεγάλα δεδομένα, οδηγώντας σημαντικά τις δαπάνες και τις επενδύσεις, επιτρέποντάς τους να ανταγωνιστούν με μεγαλύτερες επιχειρήσεις. Τελικά, η αποθήκευση και διαχείριση όλων των γεγονότων και αριθμών έχει διευρυνθεί σταδιακά. Πιο τεχνικά, οι τρόποι με τους οποίους αποθηκεύονται και επεξεργάζονται τα δεδομένα σε μια επιχείρηση, ακόμα και μικρού βεληνεκούς, εξηγούνται παρακάτω.

## **Συνεχής κρυπτογράφηση**

Δεδομένα οποιουδήποτε είδους είναι κρίσιμα για οποιαδήποτε επιχείρηση και συχνά θεωρούνται ιδιωτικά και ασφαλή με τα δικά τους μέσα. Ωστόσο, η πειρατεία καλύπτεται συχνά από επιχειρηματικές αποτυχίες, με τις τελευταίες προσπάθειες σε επιθέσεις στον κυβερνοχώρο να κυκλοφορούν συνεχώς και παντού. Κατά συνέπεια, πολλές εταιρείες δυσκολεύονται να αισθάνονται ασφαλείς, ειδικά όταν τα μεγάλα ονόματα είναι συχνά στόχοι. Καθώς οι επιχειρήσεις εργάζονται για την προστασία των περιουσιακών στοιχείων σε όλο το εύρος, η κρυπτογράφηση αυξάνεται στην επιφάνεια ως βιώσιμος τρόπος για την καταπολέμηση των απειλών στον κυβερνοχώρο. Μετατρέποντας τα πάντα σε κώδικα, η γίνεται μια κρυπτογράφηση των πληροφοριών και των δεδομένων, έτσι ώστε κανένας εκτός από τον παραλήπτη να μπορεί να την αποκωδικοποιήσει. Πέρα από όλα τα άλλα, η κρυπτογράφηση προστατεύει τη μεταφορά δεδομένων, ενισχύοντας τις πιθανότητες να φτάσει αποτελεσματικά στους σωστούς ανθρώπους ενώ βρίσκεται σε ψηφιακή διαμετακόμιση.

## **Αποθήκευση των δεδομένων σε ασφαλείς τόπους**

Τα μεγάλα δεδομένα μπορεί να φαίνονται ανεξέλεγκτα, μια περιπλεγμένη πολυπλοκότητα στατιστικών που δεν τελειώνουν ποτέ. Συνεπώς, φαίνεται λογικό να εξορθολογιστούν οι πληροφορίες σε μια ξεχωριστή εταιρική τοποθεσία, μια αποθήκη όπου όλα τα δεδομένα και οι εξυπηρετητές μπορούν να χαρακτηριστούν με άφθονο σχεδιασμό. Ωστόσο, ορισμένες αναφορές υποδεικνύουν ένα επιχείρημα κατά αυτής της μεθόδου, αναφέροντας ότι η εκθετική ανάπτυξη μεγάλων δεδομένων δεν μπορεί πλέον να διατηρηθεί ακόμα και από τα μεγαλύτερα κέντρα αποθήκευσης.

Βέβαια, σε ορισμένες περιπτώσεις, μια επιχείρηση μπορεί να νοικιάσει μια αποθήκη για την αποθήκευση μεγάλων δεδομένων, μια προσωρινή επιδιόρθωση πριν τα μεγάλα δεδομένα της επεκταθούν πέρα από αυτά τα μέσα και αρμόδιες εταιρείες, όπως η LCP (London & Cartridge Properties) προσφέρουν κάποιες καλές ευκαιρίες εκεί. Εξάλλου, οι επιχειρήσεις δεν επιβαρύνονται άμεσα με συντριπτικά κύματα πληροφοριών, οπότε η μίσθωση μιας αποθήκης για τα φυσικά μηχανήματα είναι, τουλάχιστον βραχυπρόθεσμα, βιώσιμος τρόπος προς τα εμπρός. Είναι αποτελεσματική ως προς την τιμή για μια σύντομη λύση, χωρίς να υπονομεύεται από μια δέσμευση μόνιμης δαπάνης.

## **Υπηρεσίες Back-Up, το νέφος**

Φυσικά, είναι αναμφισβήτητο ότι η διαχείριση και αποθήκευση μεγάλων δεδομένων κινείται αμέσως μακριά από τα φυσικά μηχανήματα και γρήγορα στην ψηφιακή σκηνή. Παράλληλα με όλες τις εξελίξεις στην τεχνολογία, τα μεγάλα δεδομένα μεγαλώνουν ταχύτερα, με τέτοιο ρυθμό ώστε όλα τα μηχανήματα και οι αποθήκες στον κόσμο δεν μπορούν να τα περιέχουν όλα.

Κατά συνέπεια, το υπολογιστικό νέφος έχει βιώσει μια άνθηση χάρη στις υπηρεσίες αποθήκευσης σύννεφων που προωθούν την ψηφιακή μετάβαση. Τα δεδομένα δεν ελέγχονται πλέον σε ένα σημείο, αλλά είναι προσβάσιμα οπουδήποτε, με μεγάλες εταιρίες σύννεφων, όπως το Google Drive, που οδηγούν στην χρέωση για ταχύτερη πρόσβαση στα βασικά στατιστικά στοιχεία. Μπορούν επίσης να υποστηριχθούν δεδομένα για αυτές τις υπηρεσίες, πράγμα που σημαίνει ότι μια ενιαία επίθεση στον κυβερνοχώρο δεν εξαλείφει τα χρόνια ανάπτυξης και ανάπτυξης των επιχειρήσεων. Τελικά, σε περίπτωση Cyber hack στον κυβερνοχώρο, το Cloud όλες τις λύσεις σε μια μοναδική υπηρεσία.

## 1.7 Μεγάλα Δεδομένα και νέες θέσεις εργασίας.

Τα δεδομένα αποκτούν αληθινή δύναμη, μόνο όταν είναι συγκεντρωμένα, ξεκαθαρισμένα, έχουν υποστεί ανάλυση και χειρισμό και όταν μεταφράζονται στην γλώσσα αυτών που ηγούνται της επιχείρησης και θέτουν τους κανόνες. Και έτσι η έκρηξη των δεδομένων έχει οδηγήσει στην ανάγκη δημιουργίας τομέων που ασχολούνται με την μετατροπή των απλών δεδομένων σε χρήσιμα δεδομένα.

Σήμερα βλέπουμε ολόκληρους τομείς και αρχές επιχειρήσεων που έχουν γεννηθεί από την ανάγκη απόκτησης ιδεών από πληροφορίες που υπό άλλες συνθήκες θα ήταν μη αποκωδικοποιήσιμες. Μέχρι το 2017, η αγορά “μεγάλων δεδομένων” αναμενόταν να φτάσει τα 50 δισεκατομμύρια δολάρια και να δημιουργήσει μέχρι και 3,75 εκατομμύρια καινούριες θέσεις εργασίας. Συγκεκριμένα ένα νέο επάγγελμα, αυτό των επιστημόνων δεδομένων έχει δημιουργηθεί για να ικανοποιήσει την συνεχώς αυξανόμενη αυτή ανάγκη - το να δώσουν μορφή σε μεγάλες ποσότητες ακατέργαστων δεδομένων. Το επάγγελμα αυτό βρίσκεται ακόμη σε νηπιακή ηλικία- ο όρος “Επιστήμονας Δεδομένων” χρησιμοποιήθηκε πρώτη φορά το 2008 από τους επικεφαλής της ανάλυσης δεδομένων του LinkedIn και του Facebook- αλλά ακόμη και έτσι, μέσα σε λίγα χρόνια μόλις, το επάγγελμα έχει ανθίσει.

Το 2012, η κριτική επιχειρήσεων του Harvard, αποκάλυψε αυτή την θέση εργασίας, ως το “πιο σέξι επάγγελμα του 21ου αιώνα.” Το 2015, ήταν το πιο σέξι επάγγελμα σύμφωνα με το Mashable. Με αναγνωρίσεις όπως αυτές, δεν θα έπρεπε να αποτελεί έκπληξη το γεγονός ότι φοιτητές Μάστερ και Διδακτορικών αλλά και χωρίς καθόλου εργασιακή εμπειρία καταλήγουν να λαμβάνουν εξανήφιους μισθούς. Οι επιστήμονες δεδομένων με εμπειρία, πολύ συχνά λαμβάνουν μισθούς παρόμοιους με τους επικεφαλής διευθυντές επιχειρήσεων.

Η επιστήμη των δεδομένων είναι πλέον ένα εργαλείο απαραίτητο για τις επιχειρήσεις. Σύμφωνα με μία πρόσφατη έρευνα από την Accenture, το 87% των εταιριών συμφωνούν. Πιστεύουν ότι η ανάλυση μεγάλων δεδομένων, μέσα σε 3 χρόνια, θα επαναπροσδιορισθεί από τις σημαντικές εταιρίες οι οποίες θα επενδύουν πιο σωστά στο αντικείμενο αυτό. Μάλιστα, το 73% των εταιριών ξοδεύουν πάνω από το 1/5 του τεχνολογικού τους προϋπολογισμού στην ανάλυση δεδομένων. Όχι μόνο αυτό, αλλά δεν υπάρχουν και αρκετοί επιστήμονες για να γεμίσουν οι θέσεις αυτές. Οι εταιρίες ψάχνουν πλέον ταλέντα για τις θέσεις αυτές σε άτομα που έχουν πτυχίο στην Φυσική, στην Μηχανική και στα Εφαρμοσμένα Μαθηματικά, αλλά κάτι τέτοιο απαιτεί αρκετές εξετάσεις για να επιβεβαιωθεί πως μπορούν να προσαρμοστούν στις δύσκολες απαιτήσεις του τομέα της επιστήμης των δεδομένων.

Για ποιόν όμως λόγο ακριβώς χρειάζεται ένας επιστήμονας δεδομένων τις δεξιότητες αυτές; Τι ακριβώς κάνει ένας επιστήμονας δεδομένων; Και τι κάνει το επάγγελμα του ξεχωριστό σε σχέση με τους μαθηματικούς και τους στατιστικούς αναλυτές; Η μεγάλη διαφορά είναι η ικανότητα των επιστημόνων δεδομένων να σκεφτούν ως επιχειρηματίες και όχι απλά να μπορούν να εξετάσουν ευρείες και ποικίλες τράπεζες δεδομένων, αλλά να παραθέσουν τα ευρήματα πιο ξεκάθαρα σε όσους είναι υπεύθυνοι για την λήψη αποφάσεων. Όπως το ορίζει η βιομηχανία, “οι Επιστήμονες δεδομένων έχουν την ικανότητα να μεταφέρουν τα ευρήματά τους στους επικεφαλής επιχειρήσεων και πληροφοριακών

συστημάτων, με τέτοιο τρόπο που να επηρεάζει το πως προσεγγίζουν τις προκλήσεις οι επιχειρήσεις αυτές.”

Ακόμη μία μεγάλη διαφορά είναι ο αριθμός των υψηλά τεχνικών και ποσοτικών ικανοτήτων που είναι απαραίτητα για την επιτυχία. Υπάρχει πολύ μεγάλη ζήτηση για ικανότητες εκμάθησης μηχανών, ικανότητες ανάπτυξης των δεδομένων σε Python και Java, και διαχείρισης ανάλυσης ανοιχτής πηγής και δεδομένων.

## 1.8 Ανοιχτά Δεδομένα (open data) και Μεγάλα Δεδομένα.

### 1.8.1 Τι είναι τα Ανοιχτά Δεδομένα.

Και οι δύο τύποι δεδομένων μπορούν να μεταμορφώσουν τον κόσμο, αλλά όταν οι κυβερνήσεις μετατρέπουν τα Big Data σε ανοικτά δεδομένα, αυτό έχει ιδιαίτερο αντίκτυπο. Τα Big Data και το νέο φαινόμενο των ανοικτών δεδομένων είναι στενά συνδεδεμένα, αλλά δεν είναι το ίδιο. Τα Ανοιχτά Δεδομένα φέρνουν μια προοπτική που μπορεί να κάνουν τα Big Data πιο χρήσιμα, πιο δημοκρατικά και λιγότερο απειλητικά.

Δεν προέρχονται όλα τα μεγάλα, ανοικτά δεδομένα από τις κυβερνήσεις, όπως π.χ η έρευνα των επιστημόνων στην αστρονομία. Ενώ τα Big Data καθορίζονται από το μέγεθος, τα ανοικτά δεδομένα ορίζονται από τη χρήση τους. Τα Big Data είναι ο όρος που χρησιμοποιείται για να περιγράψει τα πολύ μεγάλα, σύνθετα, ραγδαία μεταβαλλόμενα σύνολα δεδομένων. Αλλά αυτές οι κρίσεις είναι υποκειμενικές και εξαρτώνται από την τεχνολογία: τα σημερινά Big Data μπορεί να μην φαίνονται τόσο μεγάλα σε μερικά χρόνια, όταν η ανάλυση των δεδομένων και η τεχνολογία των υπολογιστών βελτιωθεί.

Τα ανοικτά δεδομένα είναι δημόσια προσβάσιμα δεδομένα που οι άνθρωποι, οι επιχειρήσεις και οι οργανισμοί μπορούν να χρησιμοποιούν για την έναρξη νέων επιχειρήσεων, την ανάλυση προτύπων και τάσεων, να λαμβάνουν αποφάσεις που βασίζονται σε δεδομένα, και να επιλύουν σύνθετα προβλήματα.

Όλοι οι ορισμοί των ανοικτών δεδομένων περιλαμβάνουν δύο βασικά χαρακτηριστικά:

- i) τα δεδομένα πρέπει να είναι διαθέσιμα στο κοινό για οποιονδήποτε θελήσει να τα χρησιμοποιήσει
- ii) πρέπει να διαθέτουν άδεια κατά τέτοιο τρόπο που να επιτρέπει την επαναχρησιμοποίησή τους.

Τα Ανοιχτά Δεδομένα θα πρέπει επίσης να είναι σχετικά εύκολα στη χρήση, αν και υπάρχουν διαβαθμίσεις του «ανοίγματος». Και υπάρχει γενική συμφωνία ότι τα ανοικτά δεδομένα θα πρέπει να διατίθενται δωρεάν ή με ελάχιστο κόστος.



## 1.8.2 Η σχέση μεταξύ των Big Data και των Ανοικτών Δεδομένων.

Υπάρχουν μερικά σημαντικά σημεία που πρέπει να προσεχθούν, στη σχέση μεταξύ των Big Data και των Ανοικτών Δεδομένων, και πώς αυτές συνδέονται με την ευρεία έννοια της ανοικτής διακυβέρνησης:

Τα Big data που δεν είναι ανοικτά, δεν είναι δημοκρατικά. Υπάρχουν σύνολα δεδομένων που περιλαμβάνουν όλα τα είδη των μεγάλων δεδομένων που τηρούνται από το κοινό – όπως τα στοιχεία που κρατούν οι μεγάλες εταιρείες λιανικού εμπορίου για τους πελάτες τους, ή τα δεδομένα που συλλέγονται για την εθνική ασφάλεια από την Εθνική Υπηρεσία Πληροφοριών. Αυτό το είδος των μεγάλων δεδομένων δίνει ένα πλεονέκτημα στους ανθρώπους που την ελέγχουν, αλλά μπορεί να αποδυναμώσει τους υπόλοιπους από εμάς. Είναι αυτό το είδος των μεγάλων δεδομένων που έχουν γίνει τα πιο αμφιλεγόμενα.

Τα Ανοικτά Δεδομένα δεν πρέπει να είναι μεγάλα, για να έχουν σημασία τα δεδομένα. Οι μέτριες ποσότητες δεδομένων, μπορεί να έχουν μεγάλο αντίκτυπο όταν γίνονται δημόσια. Τα δεδομένα από την τοπική αυτοδιοίκηση, για παράδειγμα, μπορεί να βοηθήσουν τους πολίτες να συμμετέχουν στην τοπική κατάρτιση του προϋπολογισμού, να επιλέξουν την υγειονομική τους περίθαλψη, να αναλύουν την ποιότητα των τοπικών υπηρεσιών, ή να «φτιάξουν» εφαρμογές που βοηθούν τους ανθρώπους να περιηγηθούν με τα μέσα μαζικής μεταφοράς.

Τα Big open data, δεν πρέπει να προέρχονται μόνο από την κυβέρνηση. Όλο και περισσότεροι επιστήμονες μοιράζονται την έρευνά τους στην αστρονομία, τη γονιδιωματική και άλλους τομείς σε ένα νέο, συνεργατικό μοντέλο της έρευνας. Άλλοι ερευνητές χρησιμοποιούν τα μεγάλα δεδομένα που συλλέγονται από τα social media – τα περισσότερα από τα οποία είναι ανοικτά για το κοινό – για την ανάλυση της κοινής γνώμης και τις τάσεις της αγοράς.

Όμως, όταν η κυβέρνηση μετατρέπει τα Big data σε ανοικτά δεδομένα, αυτό είναι ιδιαίτερα ισχυρό: Οι κρατικές υπηρεσίες που έχουν την ικανότητα και τους πόρους για να συγκεντρώσουν πολύ μεγάλες ποσότητες δεδομένων, και το άνοιγμα των εν λόγω συνόλων δεδομένων μπορούν να έχουν σημαντικά οικονομικά οφέλη. Έχουν βρεθεί 500 παραδείγματα εταιρειών με έδρα τις ΗΠΑ που χτίζουν τις επιχειρήσεις τους σε ανοιχτά κυβερνητικά δεδομένα, και πολλά από αυτά τα δεδομένα είναι μεγάλα.

Εφαρμόζοντας τις αρχές των Ανοικτών Δεδομένων στα Big Data, μπορεί κανείς να βοηθήσει στην επίλυση μερικών από τα δυσκολότερα ζητήματα που τα μαζικά δεδομένα έχουν εγείρει. Η μεγαλύτερη απειλή για τη δημόσια ευημερία είναι ο κίνδυνος ότι τα ιδιωτικά, τα προσωπικά δεδομένα μπορούν να συλλεχθούν και να χρησιμοποιηθούν ως Big data, με τρόπους που τα υποκείμενα των δεδομένων – δηλαδή, όλοι μας – μπορεί να μην θέλουν ή να μην εγκρίνουν. Παραδόξως, το άνοιγμα αυτών των ευαίσθητων δεδομένων, με ένα συγκεκριμένο και ελεγχόμενο τρόπο, μπορεί στην πραγματικότητα να γίνει πιο ασφαλές.

Το πρόβλημα τώρα είναι ότι δεν είναι μόνο οι κυβερνητικές υπηρεσίες και ορισμένες επιχειρήσεις που συλλέγουν προσωπικά δεδομένα για όλους μας. Είναι επίσης, ότι εμείς ως άτομα δεν ξέρουμε τι συλλέγεται και μπορεί να μην έχουμε πρόσβαση στις πληροφορίες για τους εαυτούς μας. Αν ξέραμε περισσότερα, θα μπορούσαμε να ελέγχουμε περισσότερα.

Τόσο τα Big data, όσο και τα Ανοικτά Δεδομένα μπορούν να μετατρέψουν τις επιχειρήσεις, την κυβέρνηση και την κοινωνία – καθώς και ένας συνδυασμός και των δύο δεδομένων είναι ιδιαίτερα ισχυρός. Τα Big Data μας δίνουν πρωτοφανή δύναμη να κατανοούμε, να αναλύουμε και, τελικά, να αλλάζουμε τον κόσμο στον οποίο ζούμε. Τα Ανοικτά Δεδομένα διασφαλίζουν ότι η εξουσία θα πρέπει να μοιράζεται και ότι ο κόσμος που θα αλλάξουμε, με λίγη τύχη, θα γίνει πιο δίκαιος και δημοκρατικός.

# Κεφάλαιο 2<sup>ο</sup>: Προγνωστική Ανάλυση (Predictive analytics) και Μεγάλα Δεδομένα

## 2.1 Εισαγωγή

Τα Μεγάλα δεδομένα έχουν αναδειχθεί ως ένα σημαντικό πεδίο ενδιαφέροντος για τη μελέτη και την έρευνα μεταξύ επαγγελματιών και επαγγελματιών ακαδημαϊκών. Η εκθετική ανάπτυξη των δεδομένων τροφοδοτείται από την εκθετική ανάπτυξη του διαδικτύου και των ψηφιακών συσκευών. Η πρόοδος στην τεχνολογία καθιστά οικονομικά εφικτή την αποθήκευση και την ανάλυση τεράστιων ποσών δεδομένων.

Τα μεγάλα δεδομένα περιλαμβάνουν ένα μείγμα δομημένων, ημιδομημένων και αδόμητων δεδομένων σε πραγματικό χρόνο που προέρχονται από ποικίλες πηγές. Τα Predictive Analytics παρέχουν μεθοδολογίες στη συλλογή πληροφοριών από μεγάλα σύνολα δεδομένων. Πολλές εταιρίες οραματισμού όπως το Google, η Amazon κλπ. έχουν συνειδητοποιήσει την δυναμική και τις προοπτικές των μεγάλων δεδομένων και των Analytics και τα μελετούν και τα αναπτύσσουν για να αποκτήσουν ανταγωνιστικό πλεονέκτημα. Αυτές οι τεχνικές παρέχουν αρκετές ευκαιρίες όπως η ανακάλυψη μοτίβων ή καλύτερων αλγορίθμων βελτιστοποίησης.

Η διαχείριση και η ανάλυση των δεδομένων προκαλούν επίσης πολλές προκλήσεις - συγκεκριμένα το μέγεθος, την ποιότητα, την αξιοπιστία και την πληρότητα των δεδομένων. Οι οργανισμοί χρησιμοποιούν προγνωστικά αναλυτικά με διάφορους τρόπους, από την πρόβλεψη της αγοράς για αποδοτικότερο μάρκετινγκ και την εξόρυξη δεδομένων έως την εφαρμογή αλγορίθμων μηχανικής μάθησης (Machine Learning) και τεχνητής νοημοσύνης (Artificial Intelligence) για βελτιστοποίηση επιχειρηματικών διαδικασιών και αποκάλυψη νέων στατιστικών προτύπων. Είναι βασικά οι υπολογιστές που μαθαίνουν από προηγούμενη συμπεριφορά για το πώς να κάνουν ορισμένες επιχειρηματικές διαδικασίες καλύτερα και να παρέχουν νέες γνώσεις σχετικά με το πώς λειτουργεί ο οργανισμός σας.

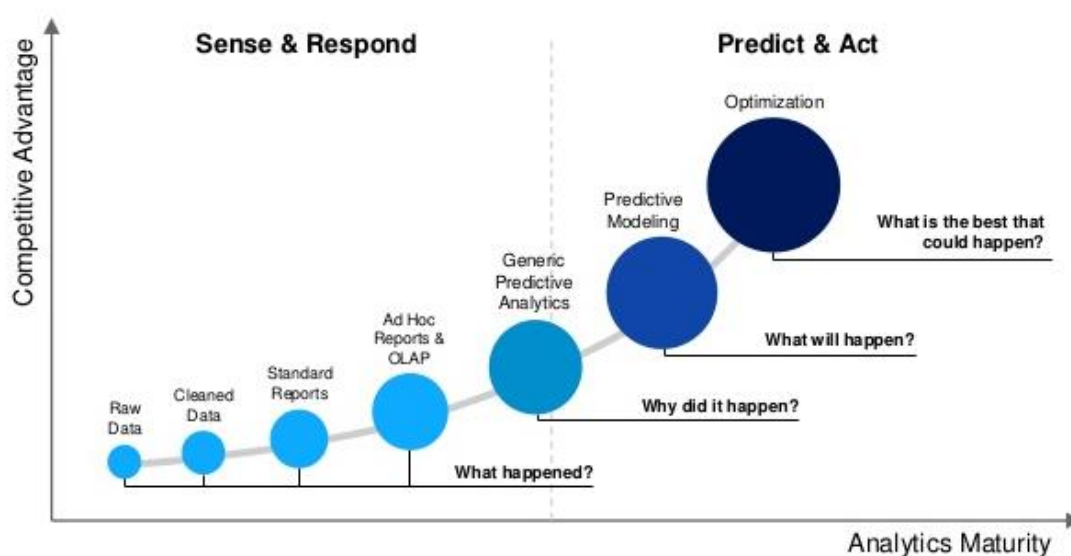
Αλλά πριν φτάσουμε σε όλους τους συναρπαστικούς τρόπους με τους οποίους οι επιχειρήσεις και οι εταιρείες τεχνολογίας απασχολούν προβλέψιμα στοιχεία ανάλυσης για να εξοικονομήσουν χρόνο, να εξοικονομήσουν χρήματα και να κερδίσουν πλεονέκτημα έναντι της υπόλοιπης αγοράς, είναι σημαντικό να μιλήσουμε για το τι ακριβώς είναι τα προγνωστικά και τι δεν είναι.

**Λέξεις κλειδιά :** *Big Data, Big Data and Analytics, Data Science, Predictive Analytics, Advanced Analytics, Social Media Analytics*

## 2.2 Τι είναι Predictive Analytics

Τα Predictive Analytics είναι μια μάρκα προηγμένων αναλυτικών στοιχείων που χρησιμοποιούνται για την πραγματοποίηση προβλέψεων σχετικά με άγνωστα μελλοντικά συμβάντα με βάση τα τρέχοντα σύνολα δεδομένων και τη συμπεριφορά των καταναλωτών. Χρησιμοποιεί διάφορες τεχνικές από την εξόρυξη δεδομένων, τις στατιστικές, την εκμάθηση μηχανών και την τεχνητή νοημοσύνη για την εξάλειψη σημείων δεδομένων που μπορούν να προβλέψουν και να ολοκληρώσουν μελλοντικά γεγονότα.

Όπως πολύ σωστά τονίζει η VANHISHIKHA BHARGAVA (2018), τα καταναλωτικά μοτίβα που βρίσκονται στα ιστορικά δεδομένα και τα δεδομένα συναλλαγών χρησιμοποιούνται για τον εντοπισμό κινδύνων και ευκαιριών για την επιχείρηση στο μέλλον.



Εικόνα 2.1: Στάδια ωρίμανσης των Predictive analytics

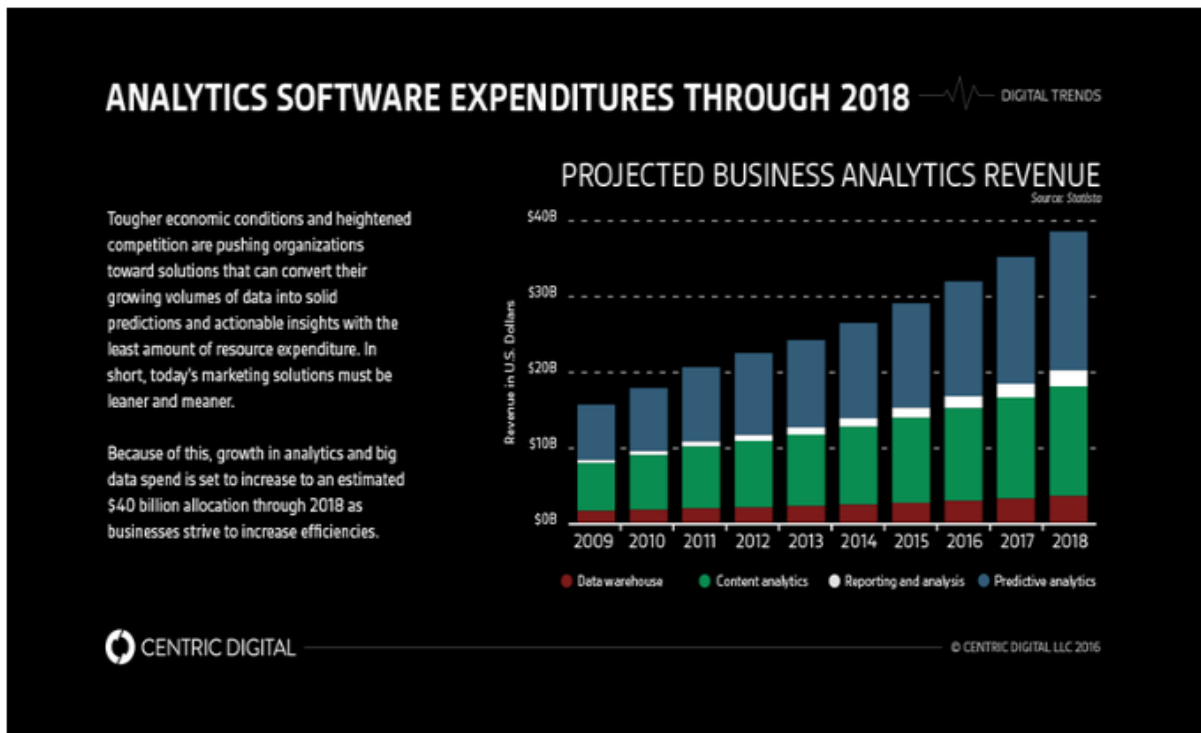
### Για ποιες τεχνολογίες μιλάμε;

Έρευνα της VANHISHIKHA BHARGAVA (2018) δείχνει πως η προγνωστική ανάλυση βασικά χρησιμοποιεί 3 τεχνολογίες:

1. Εξόρυξη δεδομένων: Ένα διεπιστημονικό υποπεδίο της επιστήμης των υπολογιστών, η εξόρυξη δεδομένων αναφέρεται στη διαδικασία ανεύρεσης σχημάτων σε μεγάλα σύνολα δεδομένων.
2. Μηχανική μάθηση: Μια μορφή τεχνητής νοημοσύνης, η μηχανική μάθηση επιτρέπει στους υπολογιστές να μαθαίνουν κάτι μέσω εμπειρίας, δοκιμών και προσαρμογής.
3. Προγνωστική μοντελοποίηση: Ο συνδυασμός της εξόρυξης δεδομένων και η πιθανότητα να γίνουν μελλοντικές προβλέψεις καλείται πρότυπο πρόβλεψης. Αυτά χρησιμοποιούνται για

την πρόβλεψη ενεργειών όπως όταν οι πελάτες είναι πιο πιθανό να αγοράσουν ένα συγκεκριμένο φάσμα προϊόντων, να βγουν περισσότερο, κλπ.

Γιατί θα πρέπει η μια επιχείρηση να χρησιμοποιεί προγνωστικά αναλυτικά στοιχεία;



Εικόνα 2.2: Οφέλη επιχειρήσεων ανά έτος από Predictive Analytics

Οι προγνωστικές αναλύσεις έχουν τεράστιο ρόλο στην ανάπτυξη των επιχειρήσεων. Βοηθάει τους οργανισμούς να συγκεντρώνουν σημεία δεδομένων των πελατών τους σε όλα τα online και offline κανάλια, να εξαγάγουν πληροφορίες από αυτήν και να τους επιτρέπουν να δημιουργήσουν έναν χάρτη πορείας για την ανάπτυξη ανάλογα.

"Μια κοινή σκηνή παίζει μέρα με τη μέρα σε εταιρείες σε ολόκληρο τον κόσμο: Η ομάδα πωλήσεων θρηνεί ότι παρεμποδίζεται από την έλλειψη αποτελεσματικού μάρκετινγκ και το τμήμα μάρκετινγκ υποστηρίζει ότι δεν διαθέτει νέα και πλήρη δεδομένα από τα οποία να δουλέψει".

Τα μεγάλα δεδομένα είναι εξίσου χρήσιμα με τις ευκαιρίες που μπορεί να έχει ένας οργανισμός. Οι σημερινές επιχειρήσεις συλλέγουν ένα πρωτοφανές ποσό δεδομένων σχετικά με τους σημερινούς και δυνητικούς πελάτες τους, αλλά πόσο από αυτά μπορούν να μετατραπούν σε αυξημένα έσοδα; Σύμφωνα με την ερευνητική εταιρεία της International Data Corporation, τα μεγάλα δεδομένα και αναλυτικά στοιχεία αναμένεται να οδηγήσουν τα παγκόσμια έσοδα σε 187 δισεκατομμύρια δολάρια το 2019, αύξηση κατά 50% το 2015.

Οι προγνωστικές αναλύσεις διαδραματίζουν σημαντικό ρόλο σε αυτήν την ανάπτυξη. Λειτουργεί βοηθώντας τις εταιρείες να χρησιμοποιούν τα δεδομένα τους πιο αποτελεσματικά, επιτρέποντάς τους να χρησιμοποιούν πληροφορίες σχετικά με τις προηγούμενες ενέργειες και συμπεριφορές των καταναλωτών για να μάθουν:

- Ποιος θα ενδιαφέρεται περισσότερο για τις προσφορές τους.
- Τι είναι πιθανό να αγοράσουν.
- Πού κάνουν αγορές.
- Όταν είναι πιο πιθανό να κάνουν μια αγορά.
- Γιατί έχουν κίνητρα να αγοράσουν.
- Πόσα θα ξοδέψουν.

Οι εταιρείες οδηγούνται όχι μόνο στην κατανόηση της συμπεριφοράς των πελατών, αλλά και στην αξιοποίηση αυτών των δεδομένων σε ένα στρατηγικό, ολοκληρωμένο σχέδιο. Σε ανταπόκριση, ένας αυξανόμενος αριθμός εμπορικών σημάτων ψάχνει για την ανάλυση πρόβλεψης για να ενισχύσει τις κατώτατες γραμμές τους. Γιατί τώρα; Επειδή τα αναλυτικά στοιχεία έχουν καθιερωθεί ως ο νέος διαφοροποιητής για τις επιχειρήσεις.

## **ΤΙ ΕΙΝΑΙ PREDICTIVE ANALYTICS**

Η προγνωστική ανάλυση είναι η χρήση δεδομένων, στατιστικών αναλύσεων και μηχανικής μάθησης για την εκτίμηση της πιθανότητας μελλοντικών αποτελεσμάτων βάσει δεδομένων από το παρελθόν. Αυτά τα μοντέλα χρησιμοποιούν γνωστά αποτελέσματα για να αναπτύξουν έναν αλγόριθμο που καθιστά προβλέψιμες βαθμολογίες - τα χρυσά αυγά των προγνωστικών αναλύσεων.

Αυτές οι βαθμολογίες είναι αριθμητικές αναπαραστάσεις της πιθανότητας ότι ένας πελάτης θα έχει μια συγκεκριμένη αλληλεπίδραση με μια εταιρεία, όπως κάνοντας κλικ σε μια διαφήμιση ή πραγματοποιώντας μια αγορά, ενημερώνοντας έτσι τον οργανισμό για την πιο αποτελεσματική πορεία δράσης που πρέπει να ακολουθήσει με αυτόν τον πελάτη. Είναι η ανακάλυψη και η επικοινωνία αυτών των σημαντικών προτύπων σε δεδομένα που μπορούν να επηρεάσουν θετικά την επιχείρηση σε επίπεδο εταιρειών, προϊόντων, πελατών και καναλιών.

Για να είναι σαφές, οι αναλυτικές μέθοδοι πρόβλεψης δεν μπορούν να φανταστούν τον τρόπο με τον οποίο ένας καταναλωτής θα ανταποκριθεί σε μια συγκεκριμένη καμπάνια μάρκετινγκ ή μια ασφάλεια πώλησης. Αντ' αυτού, τα διαρκώς εξελισσόμενα στατιστικά μοντέλα εντοπίζουν και παρακολουθούν τα πρότυπα των καταναλωτικών απαιτήσεων και συμπεριφορών και στη συνέχεια υπολογίζουν την πιθανότητα διάφορων αποτελεσμάτων. Η προγνωστική ανάλυση παρέχει στις επιχειρήσεις πολύ ακριβείς προβλέψεις, επιτρέποντας πιο αποτελεσματική λήψη αποφάσεων και επενδύσεις πόρων.

## ΕΦΑΡΜΟΓΕΣ PREDICTIVE ANALYTICS

Οι προγνωστικές αναλύσεις μπορούν να βελτιστοποιήσουν τις κεντρικές επιχειρηματικές λειτουργίες μέσω του μάρκετινγκ, των πωλήσεων και πέραν του, τόσο σε απευθείας σύνδεση όσο και εκτός λειτουργίας. Οι πιο συνηθισμένες χρήσεις για αναλυτικές προβλέψεις περιλαμβάνουν:

**Μάρκετινγκ:** Προσδιορισμός της συμπεριφοράς των πελατών και των προτύπων αγοράς για να προσελκύσουν και να διατηρήσουν τους πιο παραγωγικούς πελάτες και να αξιοποιήσουν στο έπακρο τις δαπάνες μάρκετινγκ.

**Cybersecurity:** Ανίχνευση απάτης πριν εμφανιστεί με τη χρήση λειτουργιών όπως επιχειρηματικοί κανόνες, ανίχνευση ανωμαλιών και αναλύσεις συνδέσεων.

**Ανάλυση κινδύνου:** Αξιολόγηση της πιστοληπτικής ικανότητας και αξιολόγηση της πιθανότητας αθέτησης των αγοραστών.

**Διαχείριση αποθέματος:** Παραδείγματα περιλαμβάνουν αεροπορικές εταιρείες που χρησιμοποιούν μοντέλα πρόβλεψης για να αποφασίσουν για την τιμή και τη διαθεσιμότητα εισιτηρίων και ξενοδοχείων που προβλέπουν τον αριθμό των προσκεκλημένων που αναμένουν για οποιαδήποτε δεδομένη νύχτα και την προσαρμογή των συντελεστών ανάλογα.

Η βασική διαφορά μεταξύ κάθε τρόπου εφαρμογής είναι αυτό που προβλέπεται. Ποια εφαρμογή είναι καλύτερη για την επιχείρησή σας είναι ένα στρατηγικό ερώτημα που εξαρτάται από το ποια βαθμολογία πρόβλεψης θα χρησιμεύσει καλύτερα για να καθοδηγήσει τις αποφάσεις στο συγκεκριμένο οργανισμό σας.

Αλλά πέρα από αυτά τα βασικά, οι αναλυτικές μέθοδοι πρόβλεψης μπορούν επίσης να βοηθήσουν μια επιχείρηση να ξεπεράσει μερικές βασικές προκλήσεις.

## ΑΠΕΛΕΥΘΕΡΩΣΗ ΔΙΑΧΩΡΙΣΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Μια κοινή σκηνή διαδραματίζεται καθημερινά σε εταιρείες σε όλο τον κόσμο: Η ομάδα πωλήσεων βλαστάνει ότι εμποδίζεται από την έλλειψη αποτελεσματικού μάρκετινγκ και το τμήμα μάρκετινγκ υποστηρίζει ότι δεν διαθέτει νέα, ολοκληρωμένα δεδομένα από τα οποία να δουλεύει. Αυτό αφήνει το τμήμα IT να κρατάει την τσάντα, υπεύθυνο για ελλείψεις που συχνά βρίσκονται εκτός του πεδίου του ρόλου.

Ωστόσο, ο πραγματικός ένοχος συνήθως δεν είναι η έλλειψη δεδομένων. Είναι στην πραγματικότητα ο τρόπος με τον οποίο οι πληροφορίες μοιράζονται μέσα στον οργανισμό. Σε πολλές περιπτώσεις, τα διάφορα στοιχεία των πελατών είναι "κύριοι" από οποιονδήποτε αριθμό τμημάτων σε ξεχωριστά σιλό δεδομένων σε όλη την εταιρεία. Αυτή η ρύθμιση καθιστά σχεδόν αδύνατο για οποιοδήποτε τμήμα, όπως το μάρκετινγκ ή τις πωλήσεις, να αποκτήσει μια ολιστική εικόνα των προηγούμενων και μελλοντικών δραστηριοτήτων των πελατών. Τα προγνωστικά συστήματα ανάλυσης έχουν σχεδιαστεί για να συνδέουν αυτές τις

διαφορετικές συλλογές πληροφοριών, καθιστώντας μια βασική λύση σε αυτή την παλιά πτυχή των οργανισμών που βασίζονται σε δεδομένα.

## **ΑΥΞΗΜΕΝΟ ΚΟΣΤΟΣ-ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ**

Οι αυστηρότερες οικονομικές συνθήκες και ο αυξημένος ανταγωνισμός ωθούν επίσης τους οργανισμούς σε λύσεις που μπορούν να μετατρέψουν τον αυξανόμενο όγκο των δεδομένων τους σε στερεές προβλέψεις και αντιλήψεις με την ελάχιστη δυνατή δαπάνη πόρων. Εν ολίγοις, οι σημερινές λύσεις μάρκετινγκ πρέπει να είναι πιο λιτές και πιο έντονες.

Οι προβλέψιμοι αλγόριθμοι μάρκετινγκ επιτρέπουν στις εταιρείες να στοχεύουν τους καταναλωτές με ακρίβεια λείζερ, μετατρέποντας τους αγοραστές σε αγοραστές και αγοραστές σε πιστούς πελάτες της ζωής. Η προληπτική ανάλυση μειώνει και οξύνει τον κύκλο των πωλήσεων και αυξάνει τις ευκαιρίες cross-up και up-selling της εταιρείας σας. Με την πλήρη ενσωμάτωση των προβλέψεων αναλύσεων, ένας οργανισμός μπορεί να προβλέψει τις τάσεις και να αποκαλύψει πληροφορίες που βοηθούν να γίνει καλύτερη κατανόηση από τους πελάτες, να βελτιωθούν οι επιχειρηματικές επιδόσεις και να οδηγήσει τη στρατηγική λήψη αποφάσεων - όλες χωρίς αύξηση του προϋπολογισμού. Το μειωμένο κόστος της αναλυτικής τεχνολογίας πρόβλεψης και των σχετικών εργαλείων, όπως η αποθήκευση σύννεφων, προσθέτουν αυτές τις οικονομίες. Το 68% των στελεχών αναγνωρίζει μια θετική απόδοση επένδυσης (ROI) μετά την εφαρμογή της προβλέψιμης αναλυτικής τεχνολογίας και σχεδόν το ήμισυ αναφέρει αύξηση τουλάχιστον κατά 25% της απόδοσης επένδυσης (ROI), γεγονότα που μιλούν τόμους σχετικά με την ισχυρή επίδραση αυτών των εργαλείων στην κατώτατη γραμμή.

Η αναφορά στο τι συμβαίνει σε μια επιχείρηση αυτή τη στιγμή, σε πραγματικό χρόνο, είναι μόνο το πρώτο βήμα για την λήψη καλύτερων επιχειρηματικών αποφάσεων. Οι οργανισμοί πρέπει να είναι διατεθειμένοι να ξοδέψουν χρόνο για να αλλάξουν τις επιχειρηματικές τους διαδικασίες και τις τεχνολογίες που τους υποστηρίζουν, ώστε να επωφεληθούν πλήρως από τις νέες ιδέες που παρέχει η προγνωστική ανάλυση.

## **2.3 Γιατί BIG DATA και analytics μαζί**

**Τα μεγάλα δεδομένα παρέχουν γιγαντιαία στατιστικά δείγματα, τα οποία ενισχύουν τα αποτελέσματα των εργαλείων ανάλυσης.**

Τα περισσότερα εργαλεία που έχουν σχεδιαστεί για την εξόρυξη δεδομένων ή τη στατιστική ανάλυση τείνουν να βελτιστοποιούνται για μεγάλα σύνολα δεδομένων. Στην πραγματικότητα, ο γενικός κανόνας είναι ότι όσο μεγαλύτερο είναι το δείγμα δεδομένων,



τόσο ακριβέστερα είναι τα στατιστικά στοιχεία και άλλα προϊόντα της ανάλυσης. Αντί να χρησιμοποιούνται data mining και στατιστικά εργαλεία, πολλοί χρήστες δημιουργούν ή φτιάχνουν μόνοι τους περίπλοκη SQL, η οποία αναλύει τα μεγάλα δεδομένα σε αναζήτηση μόνο του σωστού τμήματος πελατών. Τελευταία γενιά εργαλείων οπτικοποίησης δεδομένων και αναλυτικών λειτουργιών βάσεων δεδομένων λειτουργούν επίσης σε μεγάλα δεδομένα.

### **Τα αναλυτικά εργαλεία και οι βάσεις δεδομένων μπορούν τώρα να χειριστούν μεγάλα δεδομένα.**

Μπορούν επίσης να εκτελέσουν μεγάλα ερωτήματα και να αναλύσουν πίνακες σε χρόνο ρεκόρ. Πρόσφατες γενιές εργαλείων και πλατφορμών μας έχουν φέρει σε νέα επίπεδο επιδόσεων που είναι πολύ επιτακτικό για εφαρμογές που αφορούν μεγάλα δεδομένα.

### **Τα οικονομικά των analytics είναι πλέον πιο ανοιχτά από ποτέ.**

Αυτό οφείλεται σε μια απότομη πτώση στο κόστος της αποθήκευσης δεδομένων και του εύρους ζώνης επεξεργασίας. Το γεγονός ότι τα εργαλεία και οι πλατφόρμες για big data analytics είναι σχετικά προσιτές είναι σημαντική, διότι τα μεγάλα δεδομένα δεν είναι μόνο για μεγάλες επιχειρήσεις. Πολλές μικρές έως μεσαίες επιχειρήσεις (ειδικά εκείνες που βρίσκονται βαθιά σε ψηφιακές διαδικασίες για πωλήσεις, αλληλεπιδράσεις πελατών ή βρίσκονται στην αλυσίδα εφοδιασμού) πρέπει επίσης να διαχειρίζονται και να αξιοποιούν τα μεγάλα δεδομένα.

### **Υπάρχουν πολλά που πρέπει να μαθευτούν από ακατάστατα δεδομένα, εφόσον είναι μεγάλα.**

Τα περισσότερα σύγχρονα εργαλεία και τεχνικές για τα προηγμένα στοιχεία ανάλυσης και τα μεγάλα δεδομένα είναι πολύ ανεκτικά για τα δεδομένα πρώτης πηγής (raw data). Αυτό είναι καλό, διότι η ανακάλυψη πληροφοριών και τα predictive analytics εξαρτώνται από πολλές λεπτομέρειες - ακόμη και αμφισβητήσιμα δεδομένα.

### **Τα μεγάλα δεδομένα είναι ένα ειδικό πλεονέκτημα να έχει κάποιος.**

Αυτό είναι το πραγματικό σημείο των Big Data Analytics. Οι νέες τεχνολογίες και οι νέες βέλτιστες πρακτικές είναι συναρπαστικές, ακόμη και γοητευτικές, όταν έχεις να κάνεις και εργάζεσαι με δεκάδες terabytes. Αλλά πολλές φορές δεν γίνεται για την τεχνολογία. Αν συνδυαστούν μεγάλα δεδομένα και Discovery analytics μαζί τα αποτελέσματα θα είναι νέες γνώσεις για την επιχείρηση.

### **Το Analytics που βασίζονται σε μεγάλα δείγματα δεδομένων αποκαλύπτουν και αξιοποιούν την επιχειρηματική αλλαγή.**

Η ύφεση έχει επιταχύνει τον ήδη γρήγορο ρυθμό των επιχειρήσεων. Η ανάκαμψη, αν και ευπρόσδεκτη, φέρνει ακόμη περισσότερες αλλαγές. Στην πραγματικότητα, ο μέσος όρος των επιχειρήσεων έχει αλλάξει πέρα από κάθε αναγνώριση και λόγω της πρόσφατης οικονομικής ύφεσης και ανάκαμψης. Η αλλαγή δεν έχει περάσει απαρατήρητη. Οι επιχειρηματίες πρέπει να διερευνήσουν την αλλαγή ούτως ώστε να κατανοήσουν τη νέα κατάσταση των επιχειρήσεων. Ακόμη πιο επιτακτική όμως είναι η προοπτική να

ανακαλυφθούν προβλήματα που χρειάζονται διόρθωση (όπως νέες μορφές καταγισμού των πελατών και ανταγωνιστική πίεση) και ευκαιρίες που αξίζουν μόχλευσης (όπως νέα πελατειακά τμήματα και προοπτικές πωλήσεων).

## 2.4 Social Media Analytics

Τα Social Media Analytics είναι μια αναδυόμενη έννοια που γίνεται μέρος της mainstream marketing στρατηγικής.

Βασίζονται σε δεδομένα κοινωνικών μέσων που δημιουργήθηκαν σε ιστότοπους όπως το Twitter, το Facebook ή το WhatsApp. Απασχολείται με την ανάπτυξη και αξιολόγηση εργαλείων και πλαισίων πληροφορικής για τη συλλογή, παρακολούθηση, ανάλυση, συνοπτική παρουσίαση και την απεικόνιση των δεδομένων των κοινωνικών μέσων. Τα μεγάλα δεδομένα που δημιουργούνται στα κοινωνικά μέσα συνιστώνται σε μηνύματα κειμένου, τραγούδια, εικόνες, βίντεο κλπ. Οι άνθρωποι μοιράζονται πληροφορίες μέσω μηνυμάτων κειμένου, βίντεο, εικόνων, τραγουδιών κλπ. Συχνά εκφράζουν την πρόθεσή τους να αγοράσουν ένα προϊόν, να υποβάλουν αίτημα για feedback, να μοιραστούν την εμπειρία χρήσης μιας υπηρεσίας ή να υποβάλουν αξιολογήσεις προϊόντων στα κοινωνικά μέσα. Αυτά τα δεδομένα από τα κοινωνικά μέσα ενημέρωσης έχουν πολύτιμες πληροφορίες για τις επιχειρήσεις που μπορούν να αναλύσουν και να εξορύξουν δεδομένα. Το παρακάτω διάγραμμα συνοψίζει μια γενική διαδικασία που ακολουθούν οι καταναλωτές για την αγορά προϊόντων. Αυτή η διαδικασία παρουσιάζει τον τρόπο με τον οποίο οι καταναλωτές ερευνούν προϊόντα χρησιμοποιώντας μηχανές αναζήτησης και κοινωνικά μέσα. Αυτοί επίσης συμβάλλουν τις απόψεις τους, τις απόψεις τους σχετικά με τα κοινωνικά μέσα πριν και μετά την αγορά του προϊόντος. Η τεχνολογία επιτρέπει στις επιχειρήσεις να χρησιμοποιούν αυτό το μέσο για να υποστηρίξουν επιχειρηματικές στρατηγικές και να κάνουν λήψη τακτικών αποφάσεων.



Εικόνα 2.3: Βήματα των Social media analytics

## Εφαρμογή και αντίκτυπος

Οι τεχνολογίες Διαδικτύου και κινητής τηλεφωνίας αποτελούν την κύρια δύναμη πίσω από την άνοδο των κοινωνικών μέσων, παρέχοντας τεχνολογικές πλατφόρμες για τη διάδοση πληροφοριών, την παραγωγή περιεχομένου και τις διαδραστικές επικοινωνίες. Στην πραγματικότητα, τα κύρια συστατικά των κοινωνικών μέσων όπως το περιεχόμενο που δημιουργείται από τους χρήστες ή τα μέσα που παράγονται από τους καταναλωτές έχουν θεωρηθεί ως τα καθοριστικά χαρακτηριστικά του Web 2.0. Από τη σκοπιά ενός εργαλείου, μια σειρά εφαρμογών που βασίζονται στο Web ορίζουν τον τρόπο λειτουργίας των κοινωνικών μέσων. Παραδείγματα περιλαμβάνουν ιστολόγια, microblogs, ηλεκτρονικά φόρουμ, wikis, podcasts, live streams, κοινωνικούς σελιδοδείκτες, κοινότητες του ιστού, κοινωνική δικτύωση και εικονική πραγματικότητα που βασίζεται σε avatar.

Από τη σκοπιά της εφαρμογής, πολλές από τις ιστοσελίδες που απευθύνονται σε κοινωνικά μέσα είναι από τις πιο δημοφιλείς - Wikipedia (συλλογική συλλογή γνώσεων), MySpace και Facebook (κοινωνική δικτύωση), YouTube (κοινωνική δικτύωση και ανταλλαγή πολυμέσων), Digg and Delicious (ταξινόμηση και bookmarking), Second Life (εικονική πραγματικότητα) και το Twitter (κοινωνική δικτύωση και microblogging), για να αναφερθούν μερικά μόνο.

Επειδή τα κοινωνικά μέσα είναι ήδη ένα κρίσιμο κομμάτι του οικοσυστήματος πληροφόρησης και καθώς οι πλατφόρμες και οι εφαρμογές των κοινωνικών μέσων κερδίζουν ευρεία υιοθεσία με πρωτοφανή πρόσβαση στους χρήστες, τους καταναλωτές, τους ψηφοφόρους, τις επιχειρήσεις, τις κυβερνήσεις και τις μη κερδοσκοπικές οργανώσεις, το ενδιαφέρον για τα κοινωνικά μέσα από όλα τα κοινωνικά στρώματα έχει εκτοξευθεί από τις προοπτικές εφαρμογής και έρευνας.

Οι κερδοσκοπικές επιχειρήσεις αξιοποιούν τα κοινωνικά μέσα ενημέρωσης ως πλούσια πηγή πληροφοριών και πλατφόρμα εκτέλεσης επιχειρηματικών δραστηριοτήτων για το σχεδιασμό και την καινοτομία των προϊόντων, τη διαχείριση των σχέσεων των καταναλωτών και των ενδιαφερομένων μερών και το μάρκετινγκ. Για αυτούς, τα κοινωνικά μέσα ενημέρωσης αποτελούν ουσιαστικό στοιχείο της πλατφόρμας επιχειρηματικής ευφυΐας επόμενης γενιάς.

Για τους πολιτικούς, τα πολιτικά κόμματα και τις κυβερνήσεις, τα κοινωνικά μέσα αποτελούν την ιδανική βάση πληροφοριών και πληροφοριών για την εκτίμηση της κοινής γνώμης σχετικά με τις πολιτικές και τις πολιτικές θέσεις, καθώς και για την οικοδόμηση της κοινοτικής υποστήριξης για τους υποψηφίους που ασκούν δημόσια γραφεία. Οι υπάλληλοι της δημόσιας υγείας θα μπορούσαν ενδεχομένως να χρησιμοποιήσουν τα κοινωνικά μέσα ως πολύτιμες, πρώιμες ενδείξεις για την εμφάνιση ασθενειών και να παράσχουν ανατροφοδότηση σχετικά με τις πολιτικές δημόσιας υγείας και τα μέτρα αντίδρασης.

Για τις κοινότητες της εθνικής ασφάλειας και των υπηρεσιών ανάλυσης πληροφοριών, τα κοινωνικά μέσα παρουσιάζουν τεράστιες ευκαιρίες για να μελετήσουν τη συμπεριφορά των τρομοκρατικών ομάδων, συμπεριλαμβανομένων των συστημάτων προσλήψεων και δημόσιων σχέσεων τους, καθώς και των κοινωνικών και πολιτιστικών

πλαισίων. Ακόμη και οι ομάδες προβληματισμού και οι κοινωνικοί επιστήμονες και οι ερευνητές των επιχειρήσεων χρησιμοποιούν τα κοινωνικά μέσα ως ένα αμερόληπτο δίκτυο αισθητήρων και ένα εργαστήριο για φυσικούς πειραματισμούς, παρέχοντας πολύτιμους δείκτες και βοηθώντας να δοκιμάσουν υποθέσεις σχετικά με την κοινωνική παραγωγή και τις αλληλεπιδράσεις, καθώς και τις οικονομικές, πολιτικές και κοινωνικές επιπτώσεις τους.

Για πολλά άτομα, τα κοινωνικά μέσα έχουν γίνει μια μοναδική πηγή πληροφόρησης για την αντιμετώπιση προβλημάτων πληροφόρησης και γνωσιακής υπερφόρτωσης, να βρουν απαντήσεις σε συγκεκριμένες ερωτήσεις και να ανακαλύψουν πιο πολύτιμες ευκαιρίες για κοινωνικές και οικονομικές ανταλλαγές. Επιπλέον, έχει γίνει μια πλατφόρμα για να συνδέσουν και να συνεισφέρουν σε όλους τους δυναμικούς διαλόγους, μοιράζοντας την εμπειρογνωμοσύνη και τις απόψεις τους. Είναι ασφαλές να ισχυριστεί κανείς ότι τα κοινωνικά μέσα έχουν ήδη εισχωρήσει σε ένα φάσμα εφαρμογών με αξιοσημείωτο αντίκτυπο. Δεδομένου του συνεχούς ενδιαφέροντος και των συνεχώς αυξανόμενων πληροφοριών και μετα-πληροφοριών που παράγονται μέσω των κοινωνικών μέσων, αναμένεται να συνεχίσει να δίνει τη δυνατότητα νέων συναρπαστικών εφαρμογών και να φέρει επανάσταση σε πολλά υπάρχοντα.

#### **2.4.1 Social Media Analytics και Intelligence Research**

Η έρευνα για τα κοινωνικά μέσα έχει ενταθεί σημαντικά τα τελευταία χρόνια, δεδομένου του σημαντικού ενδιαφέροντος από την προοπτική της εφαρμογής και των σχετικών μοναδικών τεχνικών και κοινωνικών επιστημών προκλήσεων και ευκαιριών. Αυτή η ερευνητική ατζέντα είναι διεπιστημονικής φύσης και έχει προσελκύσει την προσοχή από τις ερευνητικές κοινότητες σε όλους τους μεγάλους κλάδους. Από την άποψη της τεχνολογίας της πληροφορίας, η έρευνα των κοινωνικών μέσων έχει επικεντρωθεί κυρίως στην ανάλυση των κοινωνικών μέσων και, πιο πρόσφατα, στην κοινωνική μέσων μαζικής ενημέρωσης.

Η ανάλυση κοινωνικών μέσων ασχολείται με την ανάπτυξη και την αξιολόγηση εργαλείων και πλαισίων πληροφορικής για τη συλλογή, παρακολούθηση, ανάλυση, συνοπτική παρουσίαση και οπτικοποίηση των δεδομένων των κοινωνικών μέσων, τα οποία συνήθως οδηγούνται από συγκεκριμένες απαιτήσεις από μια εφαρμογή-στόχο. Η έρευνα για την ανάλυση κοινωνικών μέσων εξυπηρετεί διάφορους σκοπούς:

- διευκολύνει τις συνομιλίες και την αλληλεπίδραση μεταξύ διαδικτυακών κοινοτήτων και
- την εξαγωγή χρήσιμων μοτίβων και πληροφοριών για την εξυπηρέτηση οντοτήτων που περιλαμβάνουν, αλλά δεν περιορίζονται σε, ενεργούς συνεισφέροντες σε διαλόγους που βρίσκονται σε εξέλιξη.

Από τεχνικής απόψεως, η έρευνα για τα κοινωνικά μέσα ενημέρωσης αντιμετωπίζει διάφορες μοναδικές προκλήσεις. Πρώτον, τα κοινωνικά μέσα περιέχουν ένα εμπλουτισμένο σύνολο δεδομένων ή μεταδεδωμένων, τα οποία δεν έχουν αντιμετωπιστεί συστηματικά στη λογοτεχνία δεδομένων και εξόρυξης κειμένων. Παραδείγματα περιλαμβάνουν ετικέτες (σχολιασμούς ή ετικέτες που χρησιμοποιούν λέξεις-κλειδιά ελεύθερης μορφής) εκφρασμένες από το χρήστη, υποκειμενικές απόψεις, ιδέες, αξιολόγηση και προοπτικές, ακροαματικότητα προφίλ χρηστών. Δεύτερον, οι εφαρμογές κοινωνικών μέσων είναι ένα σημαντικό παράδειγμα ανθρώπινων υπολογιστών με τη δική τους μοναδική έμφαση στις κοινωνικές αλληλεπιδράσεις μεταξύ των χρηστών. Ως εκ τούτου, πρέπει να επανεξεταστούν ζητήματα όπως το προφίλ χρήστη που εξαρτάται από το περιβάλλον και η ανάληψη αναγκών, καθώς και διάφορα είδη αλληλεπιδράσεων ανθρώπου-υπολογιστή. Τρίτον, αν και τα κοινωνικά μέσα υποσχέθηκαν μια νέα προσέγγιση για την αντιμετώπιση του προβλήματος του θορύβου και της υπερφόρτωσης της πληροφορίας με την επεξεργασία πληροφοριών μέσω του διαδικτύου, πολλά θέματα όπως σημασιολογική ασυνέπεια, συγκρουόμενα στοιχεία, έλλειψη δομής, ανακρίβειες και δυσκολίες ενσωμάτωσης διαφόρων ειδών σημάτων βρίθουν μέσα στα μέσα κοινωνικής δικτύωσης. Τέταρτον, τα δεδομένα των κοινωνικών μέσων είναι δυναμικές ροές, με τον όγκο τους να αυξάνεται ραγδαία. Η δυναμική φύση αυτών των δεδομένων και το τεράστιο μέγεθος τους δημιουργούν σημαντικές προκλήσεις όσον αφορά τον υπολογισμό γενικότερα και τον σημασιολογικό υπολογισμό ειδικότερα.

Η χρήση της τεχνολογίας, των πλαισίων λύσεων και των εργαλείων από την ανάλυση κοινωνικών μέσων ενημέρωσης, η κοινωνική μέσων μαζικής ενημέρωσης στοχεύει να αντλεί πληροφορίες από τα κοινωνικά μέσα ενημέρωσης πλούσια σε context απο application settings, να αναπτύσσει αντίστοιχα πλαίσια λήψης αποφάσεων ή υποβοήθησης λήψης αποφάσεων και να παρέχει αρχιτεκτονικά σχέδια και πλαίσια λύσεων για υπάρχουσες και νέες εφαρμογές που μπορούν να επωφεληθούν από τη "σοφία των πλήθους" μέσω του διαδικτύου. Σε σύγκριση με την ανάλυση των κοινωνικών μέσων ενημέρωσης, η έρευνα στον τομέα των κοινωνικών μέσων μαζικής ενημέρωσης βρίσκεται ακόμα σε πρώιμο στάδιο ανάπτυξης παρά την αυξημένη προσοχή από τις επιχειρήσεις και άλλες κοινότητες που θα μπορούσαν να επωφεληθούν από τέτοιες μελέτες. Από την άποψη της έρευνας, έχουν γίνει συζητήσεις για διάφορες εννοιολογικές διαστάσεις των τεχνικών προκλήσεων που σχετίζονται με τις πληροφορίες των κοινωνικών μέσων μαζικής ενημέρωσης, καθώς και για επιστημονικά θέματα που θα μπορούσαν ενδεχομένως να δώσουν χρήσιμα εργαλεία για να βοηθήσουν στην αντιμετώπιση αυτών των προκλήσεων. Ωστόσο, εξακολουθεί να λείπει η συστηματική έρευνα και συγκεκριμένα, καλά αξιολογημένα αποτελέσματα.

Παρ' όλα αυτά, η νοημοσύνη των κοινωνικών μέσων παρουσιάζει μεγάλο δυναμικό με σημαντική πρακτική σημασία ως πλούσιο και νέο πεδίο έρευνας, το οποίο μπορεί να βασίζεται σε επιστημονικούς κλάδους τόσο από την τεχνητή νοημοσύνη όσο και από άλλους τομείς. Ορισμένες βασικές προκλήσεις αντιμετωπίζουν επί του παρόντος την έρευνα των κοινωνικών μέσων μαζικής ενημέρωσης. Πρώτον, η έρευνα των κοινωνικών μέσων μαζικής ενημέρωσης απαιτεί πολύ ολοκληρωμένη πολυεπιστημονική έρευνα. Αν και η ανάγκη αυτή έχει επαναληφθεί συχνά σε αυτόν τον αναπτυσσόμενο τομέα, το επίπεδο ολοκλήρωσης στην υπάρχουσα έρευνα τείνει να είναι χαμηλό. Σε πολλές περιπτώσεις, η μεθοδολογία της έρευνας στον τομέα της πληροφορικής και τα ερευνητικά ερωτήματα παίρνουν τον κυρίαρχο ρόλο, ενώ η έρευνα που επικεντρώνεται σε μεθόδους και ζητήματα από άλλους εξίσου

σημαντικούς κλάδους, όπως η κοινωνική ψυχολογία, η θεωρία των μέσων ενημέρωσης, οι πολιτικές επιστήμες και οι κοινωνικές επιστήμες, μεταξύ άλλων έχουν διασπαρθεί. Υπήρξαν ελάχιστα ολοκληρωμένα ερευνητικά προγράμματα που εξετάζουν τόσο την πληροφορική όσο και τις επιστήμες του τομέα.

Δεύτερον, η έρευνα στον τομέα των κοινωνικών μέσων μαζικής ενημέρωσης απαιτεί καλά αρθρωτά και σαφώς καθορισμένα μέτρα απόδοσης, επειδή ένα μεγάλο μέρος της πρέπει να διεξάγεται σε περιβάλλοντα εφαρμογής με στόχο την υποστήριξη αποφάσεων. Ωστόσο, σε ένα ευρύ φάσμα εφαρμογών στις οποίες η κοινωνική ευφυΐα μπορεί να είναι σημαντική, είναι δύσκολο να ποσοτικοποιηθούν αυτά τα μέτρα. Αυτό το πρόβλημα μέτρησης καθιστά ιδιαίτερα δύσκολο να κρίνουμε την απόδοση επένδυσης των κοινωνικών μέσων μαζικής ενημέρωσης (ROI) και οδηγεί σε δυσκολίες μοντελοποίησης.

Τρίτον, από μια προοπτική καθαρά μοντελοποίησης και λήψης αποφάσεων, η νοημοσύνη των κοινωνικών μέσων αποτελεί μια μοναδική κατηγορία προβλημάτων με την ανάγκη για αποτελεσματική και δυναμική λήψη αποφάσεων βάσει δεδομένων, αβεβαιότητα και υποκειμενική ανάλυση κινδύνου, μοντελοποίηση και βελτιστοποίηση σε μεγάλα δυναμικά δίκτυα. Καθώς η έρευνα στον τομέα των κοινωνικών μέσων μαζικής ενημέρωσης ωριμάζει και βρίσκει εφαρμογές σε πραγματικό χρόνο, οι ερευνητές πιθανόν θα πρέπει να αναπτύξουν νέα αναλυτικά και υπολογιστικά πλαίσια και μεθόδους.

#### **2.4.2 Ανταγωνιστική ευφυΐα**

Στις σημερινές άκρως ανταγωνιστικές αγορές, οι υπεύθυνοι λήψης επιχειρηματικών αποφάσεων αναζητούν feedback από τους καταναλωτές για τα δικά τους προϊόντα καθώς και των ανταγωνιστών τους.

Για να γίνει καλύτερα κατανοητή η ανταγωνιστική ευφυΐα θα παρουσιαστεί με την βοήθεια ενός παραδείγματος. Οι Kim et al, (2016) προτείνουν τη χρήση αναλυτικών στοιχείων κοινωνικών μέσων για την συγκέντρωση ανταγωνιστικής νοημοσύνης για το προϊόν της εταιρείας και τα προϊόντα που προσφέρονται από τους ανταγωνιστές στο ίδιο τμήμα της αγοράς. Η συγκριτική ανάλυση των δεδομένων Twitter για 2 ανταγωνιστικά έξυπνα τηλέφωνα (Apple iPhone 6 και Samsung Galaxy S5) για μια χρονική περίοδο αποκάλυψε τη συσχέτιση μεταξύ των συναισθημάτων που εκφράστηκαν στα κοινωνικά μέσα και τη διαφορά στη ζήτηση της αγοράς για τα 2 προϊόντα. Διερεύνησαν 3 μετρήσεις, α) τον όγκο των κοινωνικών μέσων ενημέρωσης, β) την πρόθεση αγοράς και γ) το καταναλωτικό κλίμα για περαιτέρω ανάλυση.

Αυτή η προσέγγιση μπορεί να βοηθήσει τις επιχειρήσεις να προβλέψουν τις επιδόσεις των πωλήσεων στην αγορά και να εκτιμήσουν το χάσμα μεταξύ ανταγωνιστικών προϊόντων. Σαν αποτέλεσμα είχε οι υπεύθυνοι λήψης αποφάσεων να μπορούν να προσαρμόσουν γρήγορα την στρατηγική της αγοράς και να αντισταθμίσουν τις αδυναμίες σε αντίθεση με την στρατηγική των ανταγωνιστών τους.

Υπάρχουν διάφοροι τρόποι με τους οποίους μπορεί να αξιοποιηθεί η νοημοσύνη από τα κοινωνικά μέσα, από επιχειρήσεις να αναπτύξουν ανταγωνιστική νοημοσύνη βοηθώντας

τον οργανισμό να κατανοήσει τους προμηθευτές του, τους ανταγωνιστές, το περιβάλλον και τις γενικές επιχειρηματικές τάσεις.

Η επιχειρηματική ευφυΐα που αποκτάται από τα κοινωνικά μέσα μπορεί να επιτρέψει στους επιχειρηματικούς αναλυτές και τους υπεύθυνους για τη λήψη αποφάσεων να αναπτύξουν πληροφορίες σχετικά με τη συμπεριφορά των καταναλωτών, να ανακαλύψουν νέες ιδέες μάρκετινγκ, να βελτιώσουν την ικανοποίηση του πελάτη και τελικά να βελτιώσουν την απόδοση επένδυσης.

## 2.5 Ανάλυση συναισθημάτων ( Sentiment Analysis )

Ορισμένες εταιρείες συλλέγουν και αναλύουν τα αισθήματα των καταναλωτών που εκφράζονται στα κοινωνικά μέσα ενημέρωσης σχετικά με τα προϊόντα τους ή τις υπηρεσίες τους. Υπάρχουν διάφοροι τρόποι ανάλυσης με τη χρήση της επεξεργασίας φυσικής γλώσσας, λεξιλόγιο συναισθήματος (Sentiment lexicons) ή αλγορίθμων μηχανικής μάθησης σε μεγάλη ποικιλία βιομηχανιών. Υπάρχουν ιστοσελίδες όπου παρέχουν γνώσεις σχετικά με το κλίμα των πελατών σε προϊόντα ή υπηρεσίες. Το Google Analytics παρέχει επίσης διάφορες υπηρεσίες - πιο έξυπνη διαφήμιση (σχετική και αποτελεσματική διαφήμιση με γνώμονα τα δεδομένα), βαθιά γνώση ( Deep learning) πελατών (συλλαμβάνει συμπεριφορά πελατών σε CRM, σημεία πώλησης, τηλεφωνικό κέντρο), κλπ.



Εικόνα 2.4: Τομείς που αποτελούν το Sentiment analysis

Το sentiment analysis είναι ένας τομέας της Επεξεργασίας Φυσικής Γλώσσας (NLP) που αναπτύσσει συστήματα που προσπαθούν να εντοπίσουν και να εξάγουν απόψεις μέσα στο κείμενο. Συνήθως, εκτός από την αναγνώριση της γνώμης, αυτά τα συστήματα εξάγουν χαρακτηριστικά της έκφρασης.

Επί του παρόντος, η ανάλυση συναισθημάτων είναι ένα θέμα ιδιαίτερου ενδιαφέροντος και ανάπτυξης, καθώς έχει πολλές πρακτικές εφαρμογές. Δεδομένου ότι οι πληροφορίες που παρέχονται δημόσια και ιδιωτικά μέσω του διαδικτύου αυξάνονται διαρκώς, μεγάλος αριθμός κειμένων που εκφράζουν απόψεις σε review sites, forum, blogs και social media.

Με τη βοήθεια συστημάτων ανάλυσης συναισθήματος, αυτή η αδόμητη πληροφορία θα μπορούσε να μετατραπεί αυτόματα σε δομημένα δεδομένα των απόψεων για προϊόντα, υπηρεσίες, brands, πολιτική ή οποιοδήποτε θέμα που οι άνθρωποι μπορούν να εκφράσουν τις απόψεις τους. Αυτά τα δεδομένα μπορούν να είναι πολύ χρήσιμα για εμπορικές εφαρμογές, όπως η ανάλυση μάρκετινγκ, οι δημόσιες σχέσεις, οι αναθεωρήσεις προϊόντων, η βαθμολόγηση του καθαρού υποκινητή, η ανατροφοδότηση προϊόντων και η εξυπηρέτηση των πελατών.

## 2.5.1 Τύποι ανάλυσης συναισθήματος

### Ακριβής Ανάλυση Συναισθήματος

Πολλές φορές ενδιαφερόμαστε για μία πιο συγκεκριμένη ανάλυση, οπότε αντί για θετικό, αρνητικό και ουδέτερο μπορούμε να δείξουμε περισσότερες κατηγορίες

1. Πολύ θετικό
2. Θετικό
3. Ουδέτερο
4. Αρνητικό
5. Πολύ αρνητικό

Μερικά συστήματα αναγνωρίζουν και άλλα συναισθήματα όπως θυμός, λύπη, χαρά, ενθουσιασμό και άλλα.

### Aspect-Based Sentiment Analysis

Όταν αναλύεις το συναίσθημα για κάποια προϊόντα, μπορεί να ενδιαφερθείς για συγκεκριμένα χαρακτηριστικά του καθενός και όχι για την γενική θετική ή αρνητική εικόνα του. Για παράδειγμα, σε ένα κινητό μπορεί κάποιος να ενδιαφέρεται για την ποιότητα και διάρκεια της 19



μπαταρίας του και όχι για το σύνολο του σαν συσκευή. Με βάση αυτό το ενδιαφέρον ορίζεται η Aspect-Based Sentiment Analysis.

### Multilingual Sentiment Analysis

Απαιτεί πολλή προεπεξεργασία και καταναλώνει αρκετούς πόρους. Οι πόροι είτε είναι διαθέσιμοι Online (συναισθηματικά λεξικά) είτε χρειάζεται να τους δημιουργήσουμε εμείς (πχ αλγόριθμοι εντοπισμού θορύβου).

### **2.5.2 Αυτόματες Προσεγγίσεις Αλγορίθμων**

Οι αυτόματες προσεγγίσεις βασίζονται σε αλγορίθμους μηχανικής μάθησης. Το κάθε task μοντελοποιείται σε ένα πρόβλημα κατηγοριοποίησης όπου ο classifier λαμβάνει ένα text και ως output εξάγει συναίσθημα.

### Εξαγωγή Χαρακτηριστικών από κείμενο

Αναφέρεται στην μετατροπή του κειμένου σε αριθμητική αναπαράσταση, συνήθως διανύσματα. Κάθε διάνυσμα αναπαριστά την συχνότητα μίας λέξης. Αυτή η διαδικασία καθιστά δυνατό για τις λέξεις με ίδια αναπαράσταση να έχουν και ίδιο συναίσθημα πράγμα το οποίο μπορεί να αυξήσει την απόδοση των classifiers.

### Training-Prediction Processes

Στη διαδικασία εκπαίδευσης, το μοντέλο μας μαθαίνει να συσχετίζει μια συγκεκριμένη είσοδο (δηλαδή ένα κείμενο) με την αντίστοιχη έξοδο με βάση τα test samples που χρησιμοποιούνται για την εκπαίδευση. Ο extractor χαρακτηριστικών μεταφέρει την εισαγωγή κειμένου σε ένα διάνυσμα χαρακτηριστικών. Ζεύγη φορέων χαρακτηριστικών και ετικετών (π.χ. θετικών, αρνητικών ή ουδέτερων) τροφοδοτούνται στον αλγόριθμο μηχανικής μάθησης για τη δημιουργία ενός μοντέλου.

Στη διαδικασία πρόβλεψης, ο extractor χρησιμοποιείται για να μετασχηματίσει τα text inputs σε διανύσματα χαρακτηριστικών. Αυτά τα διανύσματα χαρακτηριστικών τροφοδοτούνται στη συνέχεια στο μοντέλο, το οποίο δημιουργεί προβλεπόμενες ετικέτες (πάλι, θετικές, αρνητικές ή ουδέτερες).

### Sentiment Analysis Metrics and Evaluation

Υπάρχουν πολλοί τρόποι για την λήψη απόδοσης και κατανόησης του πόσο ακριβές ένα μοντέλο είναι. Ο πιο διαδεδομένος είναι το cross-validation. Η διαδικασία του είναι αρκετά απλή. Αυτό που κάνει είναι να χωρίζει τα δεδομένα σε 2 κατηγορίες (train, test). Το train αποτελείται από το 75-80% των δεδομένων και χρησιμοποιείται για να εξασκήσει τον classifier. Το test αποτελείται από το 20-25% των ίδιων δεδομένων και χρησιμοποιείται για να λάβουμε μετρικές απόδοσης. Αυτή η διαδικασία γίνεται πολλές φορές και υπολογίζεται ο μέσος όρος κάθε μετρικής. Αν το testing set είναι συνέχεια το ίδιο μπορεί η ανάλυση σου να έχει ρυθμιστεί τόσο πολύ σε ένα data set ώστε είναι πιθανό να μην μπορείς να αναλύσεις διαφορετικό set. Το Cross-validation βοηθάει συγκεκριμένα τέτοια περίπτωση.

### Precision, Recall, and Accuracy

Είναι μετρικές για να υπολογίσουμε την απόδοση ενός classifier.

- Το Precision μετράει πόσα κείμενα είχαν προβλεφθεί σωστά ως ανήκοντα σε μια δεδομένη κατηγορία από όλα τα κείμενα που είχαν προβλεφθεί (σωστά και εσφαλμένα) ως μέλη της κατηγορίας.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- Το Recall μετράει πόσα κείμενα είχαν προβλεφθεί σωστά ως ανήκοντα σε μια δεδομένη κατηγορία από όλα τα κείμενα που θα έπρεπε να είχαν προβλεφθεί ως ανήκοντα στην κατηγορία. Γνωρίζουμε επίσης ότι όσο περισσότερα δεδομένα θα διανεμούμε με τους ταξινομητές μας, τόσο καλύτερη θα είναι η ανάκληση.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- Το Accuracy μετράει πόσα κείμενα έχουν προβλεφθεί σωστά (που ανήκουν ή όχι σε μια δεδομένη κατηγορία) από όλα τα κείμενα του corpus.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- F1-score: αυτό είναι μόνο ο αρμονικός μέσος όρος ακρίβειας και ανάκλησης

$$f_1\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

### Confusion Matrix

Ο confusion matrix είναι μια σύνοψη των αποτελεσμάτων πρόβλεψης σε ένα πρόβλημα ταξινόμησης. Ο αριθμός σωστών και εσφαλμένων προβλέψεων συνοψίζεται με τις τιμές μετρήσεων και αναλύεται ανά κατηγορία. Παρουσιάζει τους τρόπους με τους οποίους το μοντέλο ταξινόμησης συγχέεται όταν κάνει προβλέψεις. Μας δίνει τη δυνατότητα να δούμε όχι μόνο τα σφάλματα που δημιουργούνται από έναν ταξινομητή αλλά κυρίως τα είδη των σφαλμάτων που γίνονται

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig 2.7

TP = True Positive (παρατήρηση θετική/πρόβλεψη θετική)  
 FP = False Positive (παρατήρηση αρνητική/πρόβλεψη θετική)  
 FN = False Negative (παρατήρηση θετική/πρόβλεψη αρνητική)  
 TN = True Negative (παρατήρηση αρνητική/πρόβλεψη αρνητική)

### 2.5.3 Ανάλυση συναισθήματος σε αγορές

#### Χρησιμότητα

- Ανάλυση κριτικών προϊόντων
- Σύγκριση συναισθήματος με παγκόσμιες αγορές
- Ανάλυση σε social media για real-time γεγονότα
- Ανάλυση κριτικών για ανεπιθύμητα σχόλια πελατών

#### Πλεονεκτήματα

- Νέες πηγές πληροφοριών
- Πρόσθεση της ποιοτικής διάστασης σε ήδη συγκεντρωμένες ποσοτικές γνώσεις.
- Παροχή πληροφοριών σε πραγματικό χρόνο και όχι εκ των υστέρων.
- Συμπλήρωση κενών όπου τα δημόσια δεδομένα είναι σπάνια στις αναδυόμενες αγορές, για παράδειγμα.

## 2.6 Εφαρμογές ανάλυσης μεγάλων δεδομένων

Σε μια ανταγωνιστική κατάσταση στην τρέχουσα εποχή, οι επιχειρήσεις χρησιμοποιούν επιχειρηματικά δεδομένα και εξωτερικές πληροφορίες για να υποστηρίξουν τακτικές και στρατηγικές αποφάσεις.

Η ικανότητα μιας επιχείρησης να λαμβάνει γρήγορες και ενημερωμένες αποφάσεις διαφοροποιείται από τους ανταγωνιστές σε ιδιαίτερα ανταγωνιστικές αγορές της τρέχουσας εποχής. Όπως περιγράφεται νωρίτερα, αναλύσεις πρόβλεψης και αναλύσεις κοινωνικών μέσων παρέχουν την ευκαιρία να γνώση της αγοράς από πρώτο χέρι. Οι καταναλωτές παρέχουν άμεση ανατροφοδότηση σχετικά με προϊόντα, υπηρεσίες ή ταινίες στα κοινωνικά μέσα ενημέρωσης. Αυτή είναι μια πολύτιμη πηγή για τις επιχειρήσεις να συλλέγουν πληροφορίες σχετικά με τα αισθήματα και τις απόψεις των καταναλωτών. Εκεί υπάρχουν αρκετοί οργανισμοί που αξιοποιούν την αξία των μεγάλων δεδομένων για τη βελτίωση της ικανοποίησης των πελατών, την παρακολούθηση διαδρομής πελατών για να αναλύσει και να αντιληφθούν πιθανή φθορά στην σχέση τους με τους πελάτες ή τις αποφάσεις αγοράς, να εντοπίσει τους κινδύνους της αλυσίδας εφοδιασμού, να συγκεντρώσει ανταγωνιστική νοημοσύνη ή για τη λήψη αποφάσεων.

Υπάρχουν πολλές περιπτώσεις - παραδείγματα μελέτης αποτελεσματικής χρήσης του Big Data Analytics όπως αναφέρονται και συνοψίζονται παρακάτω:

α) Με τη βοήθεια ιστορικών δεδομένων πωλήσεων στις περιοχές που επλήγησαν από τον τυφώνα, ο CIO της Walmart θα μπορούσε να προβλέψει υψηλότερο επίπεδο ζήτησης για ορισμένα προϊόντα ακριβώς μπροστά από τον τυφώνα Frances.

β) Τα στούντιο των κινηματογραφικών ταινιών έχουν δει εντυπωσιακή ακρίβεια από τον τρόπο με τον οποίο τα tweets που προέρχονται από την πρώτη προβολή μιας νέας ταινίας προβλέπουν την επιτυχία της καθώς και την επιτυχία του DVD και του Blu-Ray της.

γ) Όταν αντιμετωπίζονται μεγάλους όγκους δεδομένων της τάξης των εκατομμυρίων ή δισεκατομμυρίων, τα Big Data Analytics μπορούν να σας βοηθήσουν να ανακαλυφθούν πρότυπα και προβλήματα, όπως νέες μορφές καταγισμού των πελατών, επιχειρηματικές ευκαιρίες όπως νέα πεδία πελατών και προοπτικές πωλήσεων, κατανόηση της συμπεριφοράς των πελατών μέσω των κλικ κτλ.

δ) Τα μεγάλα δεδομένα χρησιμοποιούνται ήδη από τις επιχειρήσεις για την ανάπτυξη και απόκτηση πληροφοριών για την αγορά, από τις κυβερνήσεις για σχεδιασμό πολιτικών, από πολιτικούς για το σχεδιασμό πολιτικών εκστρατειών, από ιατρούς για έξυπνη διαχείριση της υγείας. Μερικά από τα αναδυόμενα ερευνητικά πεδία στον τομέα αυτό είναι το Big Data Analytics, Text Analytics, Analytics δικτύου και Mobile Analytics.

### **2.6.1 Ευκαιρίες και προκλήσεις ανάλυσης μεγάλων δεδομένων**

Τα μεγάλα δεδομένα χαρακτηρίζονται από μαζικό μέγεθος δείγματος και υψηλή διαστασιολόγηση.

Αναλύοντας αυτό το είδος δεδομένων με πολλές διαστάσεις, τεράστιο όγκο και ετερογένεια παρουσιάζουν ορισμένες ευκαιρίες ανταγωνιστικού πλεονεκτήματος καθώς και ορισμένες προκλήσεις όσον αφορά την αντιμετώπιση αυτών των μεγάλων και συνεχών συνόλων δεδομένων.

### **2.6.2 Ευκαιρίες**

Ο τεράστιος όγκος δεδομένων αποτελούσε ένα τεχνολογικό πρόβλημα μόλις πριν από μερικά χρόνια, τώρα παρουσιάζει ένα ευκαιρία. Τα Big Data παρέχουν πολλές ευκαιρίες και ανταγωνιστικά πλεονεκτήματα. Νωρίς ο μεταφορέας Amazon.com άρχισε να συλλέγει πληροφορίες πελατών, προτιμήσεις, ιστορικό αγοράς, ιστορικό αναζήτησης και σχόλια βιβλίων. Με βάση αυτά τα δεδομένα, παρέχει συστάσεις προϊόντων που παρακινούν τον πελάτη αγοράσει παρόμοιο ή σχετικό προϊόν, βελτιώνοντας τις πιθανότητες για επιπλέον αγορές από τον ίδιο πελάτη. Οι έμποροι λιανικής της επόμενης γενιάς θα μπορούν να

παρακολουθούν τη συμπεριφορά μεμονωμένων πελατών και να αναπτύσσουν μοντέλα για την πρόβλεψη ή τον επηρεασμό των υποψηφίων πελατών τους. Η Walmart κάνει χρήση του "Κοινωνικού Γονιδιώματος" που παρακολουθεί τις συνδέσεις μεταξύ ανθρώπων, προϊόντων, διάφορες μάρκες και άλλες συναφείς οντότητες. Το κοινωνικό γονιδίωμα χρησιμοποιείται για να κάνει συστήσει το προϊόν στους πελάτες όταν είναι στο διαδίκτυο ή στο κατάστημα. Τα μεγάλα δεδομένα παρέχουν πολλά πλεονεκτήματα σε σχέση με την παραδοσιακή μέθοδο συλλογής δεδομένων όπως τα δείγματα σχεδίασης.

### 2.6.3 Μεγάλο μέγεθος δείγματος για μεγάλα δεδομένα

Τα Big Data παρέχουν γιγαντιαία στατιστικά δείγματα, τα οποία ενισχύουν τα αποτελέσματα των αναλυτικών εργαλείων.

Σε μια δειγματοληψία, μια απόκλιση από ένα δείγμα μπορεί να ανήκει σε άλλο υποπληθυσμό. Αυτό μπορεί να μας οδηγήσει σε ένα μη έγκυρο μοντέλο καθώς και σε λανθασμένα αποτελέσματα εξαιτίας της έλλειψης συμπερίληψης όλων των παραμέτρων που εξηγούν τη μεταβολή μέσα στην ομάδα δείγματος.

Μέσα στα Big Data, εξετάζουμε ολόκληρο το σύνολο δεδομένων που περιλαμβάνει όλους τους υποπληθυσμούς. Αυτό μας βοηθάει για την καλύτερη κατανόηση της ετερογένειας στο σύνολο δεδομένων και για την καλύτερη κατανόηση των σχέσεων μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών σε ένα μοντέλο.

Επιπλέον, τα αναλυτικά εργαλεία είναι πλέον ικανά να χειριστούν μεγάλο όγκο δεδομένων με λογικό κόστος. Δεδομένου ότι τα δεδομένα αντιπροσωπεύουν ολόκληρο πληθυσμιακό μοντέλο που βασίζεται σε δεδομένα παρέχει ακριβείς πληροφορίες, οδηγώντας σε υψηλότερο επίπεδο ακρίβειας στις επιχειρηματικές αποφάσεις.

Τα μεγάλα δεδομένα περιλαμβάνουν τόσο δεδομένα συστημάτων επιχειρήσεων όσο και μη δομημένα δεδομένα, όπως π.χ. κοινωνικά μέσα ενημέρωσης που παρέχουν πληροφορίες σε πραγματικό χρόνο επί τόπου, όπως feedback από τους καταναλωτές, ζωντανές πληροφορίες για τον καιρό και την κυκλοφορία, οι οποίες συμπληρώνουν πολύ καλά τις πληροφορίες από τα επιχειρησιακά συστήματα.

### 2.6.4 Πρόοδος στην τεχνολογία

Τα Big Data Analytics απαιτούν αναζητήσεις σε μεγάλα σύνολα δεδομένων, αναλύοντας τα δεδομένα και αναζητώντας συσχετίσεις.

Οι προκλήσεις που παρουσιάζονται από τα μεγάλα σύνολα δεδομένων ώθησαν την ανάπτυξη της νέας υπολογιστικής υποδομής και των μεθόδων αποθήκευσης δεδομένων. Αποκορύφωση της επιστήμης των δεδομένων, των στατιστικών και των εφαρμοσμένων

Μαθηματικών οδηγούν σε καλύτερους αλγορίθμους βελτιστοποίησης που είναι κλιμακωτοί για να επεξεργαστούν μεγάλες ποσότητες και σύνολα δεδομένων με μεγάλες διαστάσεις.

Η τεχνολογία είναι έτοιμη με τις λύσεις να έρχονται γρηγορότερα από το ρυθμό δημιουργίας δεδομένων με καλύτερη, φθηνότερη ικανότητα αποθήκευσης και ικανότητα υπολογισμού υψηλών ταχυτήτων.

### **2.6.5 Σε πραγματικό χρόνο**

Μεγάλα δεδομένα, είτε προέρχονται από δομημένες είτε μη δομημένες πηγές, παρέχουν πληροφορίες σε πραγματικό χρόνο σε όσους έχουν την ικανότητα να αντλούν αξία από αυτή τη συγκεκριμένη πηγή δεδομένων. Υπάρχουν πολλά παραδείγματα όπου οι πληροφορίες σε πραγματικό χρόνο έχουν χρησιμοποιηθεί για την επίλυση προβλημάτων. Οι φυσικές καταστροφές έχουν τη δυνατότητα να διαταράξουν την κρίσιμη αλυσίδα εφοδιασμού. Ιδιαίτερα, ο αντίκτυπος σε χώρες όπως η Ινδία έχει μεγαλύτερη επίπτωση στο σύστημα εφοδιασμού.

Η Ινδία είναι κατακερματισμένη, η υποδομή είναι ανεπαρκής και αποτελείται από πολλούς μικρούς έως μεγάλους παίκτες. Δεδομένα που συλλέγονται από γεωγραφικές τοποθεσίες, καιρικές συνθήκες και ανάπτυξη φυσικών καταστροφών όπως καταιγίδες, πλημμύρες, σεισμούς κ.λπ. έχουν την κατεύθυνση εφαρμογής τους στην παρακολούθηση σε πραγματικό χρόνο της αλυσίδας προσφοράς προληπτικά. Η κυβέρνηση της Σιγκαπούρης χρησιμοποιεί τα δεδομένα γεωγραφικής κατανομής από κινητά τηλέφωνα για τη διαχείριση της κυκλοφορίας κατά τη διάρκεια της ώρα αιχμής. Αυτές οι πληροφορίες χρησιμοποιούνται για την πρόβλεψη ζήτησης σε πραγματικό χρόνο για υπηρεσίες μεταφοράς κατά τις ώρες αιχμής και για να εκτρέψουν τα ταξί στις περιοχές της πόλης. Οι πολίτες λαμβάνουν ενημερώσεις σε πραγματικό χρόνο σχετικά με την κυκλοφορία, τον καιρό μέσω των κοινωνικών μέσων ενημέρωσης και αναθεωρούν ανάλογα τα ταξιδιωτικά τους σχέδια. Οι Netizens παρέχουν σε πραγματικό χρόνο ενημερώσεις σχετικά με την ποικιλία των γεγονότων μέσω των κοινωνικών μέσων. Εταιρείες ηλεκτρονικού εμπορίου Amazon, Flipkart κάνουν συστάσεις προϊόντων που βασίζονται σε προηγούμενες αγορές και ιστορικό αναζήτησης, γεγονός που οδηγεί σε πωλήσεις πρόσθετων προϊόντων.

### **2.6.6 Προκλήσεις με τα Big Data**

Λόγω του τεράστιου μεγέθους και της μεγάλης διαστασιμότητας, τα Big Data δημιουργούν μερικές μοναδικές προκλήσεις.

Τα δεδομένα που συλλέχθηκαν σε τέτοια τεράστια ταχύτητα 'πνίγουν' τις περισσότερες οργανώσεις. Οι κυριότερες προκλήσεις σχετίζονται με το μέγεθος, την

ποιότητα συσσώρευσης θορύβου, την ψευδής συσχέτιση, την τυχαία ενδογενετικότητα και τα σφάλματα μέτρησης.

Ένα από τα σημαντικά γεγονότα που πρέπει να σημειωθούν είναι ότι τα μεγάλα δεδομένα οδηγούνται από τεράστια ποσά δεδομένων που παράγονται κάθε μέρα και αποθηκεύονται με φθηνότερο ρυθμό από ό,τι πριν. Για μια αποτελεσματική στατιστική διαδικασία είναι απαραίτητο να αντιμετωπιστούν τα εμπόδια μεγάλων δεδομένων όπως ο θόρυβος, η ετερογένεια, η συσχέτιση και η αναποτελεσματικότητα.

### **2.6.7 Μέγεθος και ποιότητα δεδομένων**

Η μεγαλύτερη πρόκληση για το χειρισμό μεγάλων δεδομένων είναι το μέγεθός τους. Οι παραδοσιακές βάσεις δεδομένων μπορούν να χρησιμοποιήσουν εύκολα Excel ή άλλα εργαλεία ETL σε πληροφοριακά συστήματα, ενώ τα μεγάλα δεδομένα απαιτούν μια εξειδικευμένη τεχνολογία όπως το Hadoop για τη διαχείριση δεδομένων.

Δεδομένου ότι τα μεγάλα δεδομένα παράγονται αρκετά γρήγορα, η μεταφορά και η αποθήκευση αυτών των τεράστιων δεδομένων απαιτεί προηγμένο σχεδιασμό και επενδύσεις σε υποδομές. Η δεύτερη πρόκληση σχετίζεται με την ποιότητα των δεδομένων που μπορούν να καθοριστούν με πληρότητα, ακρίβεια και έγκαιρη διαθεσιμότητα δεδομένων.

Σε μια παραδοσιακή διαδικασία δειγματοληψίας, ένα τυχαίο δείγμα δεδομένων επιλέγεται από ένα μεγάλο μέγεθος του πληθυσμού για τη συλλογή λεπτομερών πληροφοριών και έπειτα γίνεται στο δείγμα η στατιστική ανάλυση. Στην περίπτωση των μεγάλων δεδομένων, τα μεγάλα σύνολα δεδομένων απαιτούν εξελιγμένες στατιστικές και υπολογιστικές μεθόδους για ανάλυση. Παρατηρούνται και αναγνωρίζονται πολλά προβλήματα ποιότητας για τα μεγάλα δεδομένα στη διαχείριση της αλυσίδας εφοδιασμού και προτείνεται η διεπιστημονική έρευνα για την αντιμετώπιση προβλημάτων ποιότητας δεδομένων.

## **2.7 Μελέτη περίπτωσης**

### **Lesara**

#### **2.7.1 Περιγραφή εταιρείας**

Η εταιρεία Lesara, με έδρα το Βερολίνο, που ιδρύθηκε το 2013, είναι ο κορυφαίος παίκτης της Ευρωπαϊκής ευέλικτης αγοράς μόδας. Το 2017, συγκέντρωσαν 40 εκατομμύρια δολάρια για τη χρηματοδότηση στρατηγικών προς την ανάπτυξη και επέκτασή τους. Η



πλειοψηφία των επιτευγμάτων της εταιρείας επιτεύχθηκαν με τη βοήθεια μεγάλων δεδομένων.

Η ευέλικτη μόδα είναι μια νέα ιδέα που επικεντρώνεται στην πρόβλεψη της τωρινής και μελλοντικής ζήτησης, των απαιτούμενων ποσοτήτων και τάσεων στυλ μέσω μεγάλων αναλυτικών εργαλείων δεδομένων, όπως για παράδειγμα το πιο δημοφιλές λογισμικό ανοιχτού κώδικα το Hadoop. Είναι σημαντικό να αναφέρουμε ότι η συνταγή της Lesara για επιτυχία δεν περιλαμβάνει μόνο την ικανή χρήση μεγάλων δεδομένων · η εταιρεία έχει εξαιρετικά αποτελεσματικά συστήματα παραγωγής και διανομής, τα οποία τους επιτρέπουν να μειώσουν σημαντικά τα έξοδα. Μια άλλη απόφαση που άφησε τη Λέσαρα να σώσει ένα μεγάλο χρηματικό ποσό ήταν να εγκαταλείψουν τα παραδοσιακά φυσικά καταστήματα, υπέρ της πώλησης μέσω διαδικτύου.

### **2.7.2 Σκοπός των Big Data στην περίπτωση της Lesara**

Ωστόσο, η γρήγορη διανομή και οι ευέλικτοι κύκλοι παραγωγής θα ήταν άσκοποι χωρίς μεγάλα δεδομένα. Τα Big Data επιτρέπουν στη Lesara να κάνει ακριβείς προβλέψεις για τις τάσεις της μόδας και συνεπώς τις μελλοντικές απαιτήσεις των πελατών. Αυτή είναι μια σημαντική αλλαγή στην προσέγγιση της μόδας. Τέτοιες μεγάλες και επιτυχημένες εταιρείες λιανικής πώλησης όπως η H&M και η Inditex εξακολουθούν να είναι προσανατολισμένες κυρίως στις ήδη υπάρχοντες τάσεις, που καθορίζονται από τους παγκοσμίου φήμης σχεδιαστές, και αποκαλύπτονται σε fashion shows δύο φορές το χρόνο. Με έξυπνες αναλύσεις, εταιρείες όπως η Lesara είναι πάντα μερικά βήματα μπροστά από τους παραδοσιακούς λιανοπωλητές μόδας.

### **2.7.3 Οφέλη της Lesara από την χρήση των Big Data**

Ο Roman Kirsch, διευθύνων σύμβουλος της Lesara, περιγράφει το πλεονέκτημα της εταιρείας του έναντι των ανταγωνιστών: Βασιζόμενοι σε δεδομένα, είμαστε σε θέση να προσφέρουμε στους πελάτες τις ακριβείς τάσεις που επιδιώκουν και θέλουν να αγοράσουν, χρησιμοποιώντας μια πιο εξατομικευμένη εμπειρία αγοράς από οποιονδήποτε άλλο λιανοπωλητή βασισμένο σε παραδοσιακές, υποκειμενικές προβλέψεις τάσεων.

Οι προβλέψεις τάσεων θα μπορούσαν να γίνουν με πολλούς τρόπους, ένας από τα πιο συνηθισμένους είναι τα Google Trends - ένα δωρεάν εργαλείο που αναλύει διάφορες πτυχές της δημοτικότητας σε έναν ορισμένο όρο αναζήτησης. Το Google Trends είναι ένα μάλλον απλό εργαλείο, αλλά τα δεδομένα του θα μπορούσαν να συνδυάσουν πληροφορίες από άλλες πηγές, συμβάλλοντας στην τελική απόφαση.

Ο Kirsch προβλέπει επίσης ότι σε δέκα χρόνια η τεχνολογία θα κυβερνά την αγορά μόδας και οι τάσεις δεν πρόκειται να δημιουργηθούν στα εργαστήρια των σχεδιαστών αλλά στην τεχνολογία των βάσεων δεδομένων των εταιρειών.

#### 2.7.4 Τρία προβλήματα που λύνει η Lesara με την χρήση των Big Data

Η Lesara χρησιμοποιεί κυρίως μεγάλα δεδομένα για τρία πράγματα: για να ελαχιστοποιήσει την υπερβολική παραγωγή, να σπρώξει τη μαζική παραγωγή σε ένα εντελώς νέο επίπεδο και να μειώσει τον χρόνο της παραγωγής μεταξύ της έναρξης του σχεδιασμού και της πώλησης του προϊόντος.

Συνήθως οι λιανοπωλητές της μόδας παράγουν περισσότερα αντικείμενα από αυτά που χρειάζονται, λόγω του φόβου της μη ικανοποίησης της ζήτησης. Αυτό οδηγεί σε μεταφορά πολλών προϊόντων σε εκπτώσεις λιανικής πώλησης, με αποτέλεσμα την απώλεια χρημάτων. Τα μεγάλα δεδομένα επιτρέπουν στις εταιρείες να προβλέψουν τη ζήτηση με μεγαλύτερη ακρίβεια, αναλύοντας τις προηγούμενες εποχές και συνδυάζοντας τα αποτελέσματα με τα τρέχοντα δεδομένα πελατών. Έτσι, υπάρχει ελάχιστα έως καθόλου μη πωληθέντα αντικείμενα, πράγμα που σημαίνει χαμηλότερο κόστος παραγωγής, αποθέματος και μεταφοράς.

Σύμφωνα με διάφορες πηγές, η Lesara παράγει από 50.000 έως 100.000 νέα κομμάτια ενδυμάτων ετησίως, αριθμός ο οποίος είναι περισσότερο από πέντε φορές μεγαλύτερος από τους αριθμούς που εμφανίζονται από εταιρείες όπως οι H&M και τα Zara. Τα δεδομένα βοηθούν στην γρήγορη κατανόηση των στοιχείων που αναπτύσσονται σε ζήτηση και στη συνέχεια προσαρμόζουν τον κύκλο παραγωγής για να ικανοποιήσουν αυτή τη ζήτηση. Ο κόσμος της μόδας εξακολουθεί να ζει από την φθινοπωρινή / χειμωνιάτικη μέχρι την ανοιξιάτικη / καλοκαιρινή συλλογή, αλλά αυτό το σχέδιο σύντομα θα γίνει ένα πράγμα του παρελθόντος. Η ευέλικτη μόδα με την βοήθεια των μεγάλων δεδομένων μειώνει δραστικά τον χρόνο μεταξύ της αρχικής ιδέας και της αποστολής του τελικού προϊόντος στον πελάτη. Ο Roman Kirsch ισχυρίζεται ότι με τη χρήση των μεγάλων δεδομένων από διαφορετικές πηγές, είναι σε θέση να σχεδιάσουν, να κατασκευάσουν και να ξεκινήσουν αποστολή σε πελάτες σε 10 ημέρες.

#### 2.7.5 Πηγές της Lesara

Αναφερόμενος στις πηγές, ο Kirsch υπογραμμίζει τη σημασία της επέκτασης του συνόλου δεδομένων (Data Pool). Η Lesara χρησιμοποιεί τις τάσεις της Google, τα κοινωνικά μέσα, τους μπλόγκερ μόδας, τις δικές της αναλύσεις ιστού και συστήματα διαχείρισης αποθεμάτων. Οι περισσότερες εταιρείες χρησιμοποιούν αρχεία cookie για τη συλλογή και αποθήκευση πληροφοριών από ιστότοπους και πλατφόρμες κοινωνικών μέσων. Η Lesara

σίγουρα διανέμει ένα σημαντικό μέρος των δαπανών της για την αγορά των cookies πρώτου μέρους από πολλές ιστοσελίδες, που συνδέονται με τον κόσμο της μόδας.

Ο Kirsch εξηγεί ότι "όσο περισσότερα δεδομένα και κλίμακα η Lesara έχουν, τόσο περισσότερο εξελιγμένοι και ακριβείς οι αλγόριθμοι που γίνονται σε όλα τα σχετικά πεδία εφαρμογής - από την κατανόηση και την πρόβλεψη των απαιτήσεων των καταναλωτών σε πραγματικό χρόνο, στη βελτίωση της αποτελεσματικότητας της εμπορίας.

Ωστόσο περισσότερα δεδομένα ενδέχεται να θέσουν το πρόβλημα του "θορύβου". Για να είναι σε θέση να χειριστεί όλα αυτά τα δεδομένα, η Lesara θα χρειαζόταν ένα μεγάλο και αξιόπιστο σύστημα αποθήκευσης.

### 2.7.6 Σύνοψη

Συνολικά, η Lesara θα μπορούσε να είναι υπεύθυνη για την επανάσταση της λιανικής πώλησης ειδών ένδυσης με την εισαγωγή μιας μεγάλης προσέγγισης της μόδας με την χρήση βάσεων δεδομένων. Η εταιρεία δεν εφαρμόζει μόνο τέτοιες πολυτομεακές πρακτικές, όπως η ακριβής εκτίμηση της ζήτησης για να εξαλείψει την υπερπαραγωγή αλλά επίσης προβλέπει και επιτυχώς τις τάσεις της μόδας μπροστά από τον ανταγωνισμό. Καθώς ο ρυθμός της ζωής αυξάνεται, η ευέλικτη μόδα με τα συνεχή νέα της σχέδια και τις χαμηλές τιμές της είναι πιθανό να αντικαταστήσει σύντομα την παραδοσιακή εποχιακή μόδα.

## ΚΕΦΑΛΑΙΟ 3ο : Εξόρυξη γνώσης (data mining) και μεγάλων δεδομένων

### 3.1 Ορισμός Εξόρυξης γνώσης και Δεδομένων.

Η εξόρυξη γνώσης από δεδομένα (data mining) ή πιο απλά η εξόρυξη γνώσης είναι μια νέα δυναμική τεχνολογία που βοηθάει τις επιχειρήσεις να εστιάσουν στην σημαντική πληροφορία που βρίσκεται μέσα στις αποθήκες δεδομένων τους (data warehouses). Οι τεχνικές της είναι σε θέση να αναζητήσουν και να βρουν γρήγορα και λεπτομερειακά βάσεις δεδομένων για την αναζήτηση κρυμμένων προτύπων (patterns). Έτσι λοιπόν μπορούμε να πούμε ότι η εξόρυξη γνώσης είναι μια διαδικασία εξαγωγής κρυμμένης πληροφορίας από μεγάλες βάσεις δεδομένων. «Εξόρυξη δεδομένων είναι η διαδικασία εξαγωγής υπονοούμενης και εν πολλοίς άγνωστης αλλά ενδεχομένως χρήσιμης γνώσης υπό την μορφή συσχετίσεων προτύπων και τάσεων, μέσω της εξέτασης ανάλυσης και επεξεργασίας βάσεων δεδομένων, συνδυάζοντας και χρησιμοποιώντας τεχνικές από την μηχανική μάθηση, την αναγνώριση προτύπων, την στατιστική, τις βάσεις δεδομένων και την οπτικοποίηση.»

Παρά το γεγονός ότι υπάρχει μια γενικότερη συμφωνία ότι ο στόχος της εξόρυξης δεδομένων είναι η ανακάλυψη νέας και χρήσιμης πληροφορίας σε βάσεις δεδομένων, τα μέσα για την επίτευξη του στόχου αυτού ποικίλουν σε πολύ υψηλό βαθμό. Η εξόρυξη γνώσης περιλαμβάνει ένα ευρύ πεδίο υπολογιστικών μεθόδων που μεταξύ άλλων περιλαμβάνουν, την στατιστική ανάλυση (statistical analysis), τα δένδρα αποφάσεων (decision trees), τα νευρωνικά δίκτυα (neural networks), την εξαγωγή κανόνων (rule induction) και την γραφική οπτικοποίηση (graphic visualization). Τέτοιες μέθοδοι χρησιμοποιούνται για την εύρεση συσχετίσεων, προτύπων και δομών σε μεγάλες και διαρκώς αυξανόμενες βάσεις δεδομένων. Ειδικά η εύρεση εργαλείων είναι ένα ιδιαίτερα σημαντικό εξαγόμενο της εξόρυξης δεδομένων μέσω σχέσεων μεταξύ των χαρακτηριστικών των βάσεων δεδομένων.

### 3.2 Εξόρυξη δεδομένων και ανεύρεση γνώσης.

Η εξόρυξη γνώσης βοηθά τις σύγχρονες εταιρείες να εστιάζουν στα πιο σημαντικά στοιχεία από τις αποθήκες δεδομένων τους. Με άλλα λόγια είναι η διαδικασία εφαρμογής μεθόδων ανάλυσης σε μεγάλο όγκο δεδομένων. Ο χρήστης των εργαλείων εξόρυξης μπορεί να προβλέψει μελλοντικές συμπεριφορές και συνήθειες, ώστε οι εταιρίες να παίρνουν επιτυχημένες αποφάσεις. Συνειδητοποιούμε ότι οι τεχνικές εξόρυξης γνώσης αναπτύσσονται

γρήγορα, δίχως αλλαγές στην υποδομή και με μοναδικό στόχο την αξιοποίηση των επεξεργασμένων δεδομένων.

Στη διεθνή βιβλιογραφία υπάρχει μια γενικότερη σύγχυση ανάμεσα στους όρους «Εξόρυξη Γνώσης» (Data mining) και «Ανεύρεση γνώσης στις βάσεις δεδομένων» (Knowledge discovery in data bases, KDD). Σε πολλές περιπτώσεις αξίζει να σημειωθεί ότι οι δύο αυτοί όροι ταυτίζονται, ενώ στην πραγματικότητα η εξόρυξη δεδομένων αποτελεί τμήμα της ανεύρεσης γνώσης, συγκροτώντας το πυρήνα αυτής (Zaiane, 1999). Προκειμένου λοιπόν να κατανοηθεί καλύτερα η εξόρυξη δεδομένων, θα γίνει μια σύντομη αναφορά στη διαδικασία της ανεύρεσης γνώσης.

Η ανεύρεση γνώσης είναι μια επαναληπτική διαδικασία που αποτελείται από μια σειρά βημάτων, τα οποία οδηγούν από τη συλλογή των δεδομένων στην ανακάλυψη και εξαγωγή χρήσιμης πληροφορίας από αυτά.

Τα βήματα από τα οποία αποτελείται η διαδικασία ανεύρεσης γνώσης είναι τα ακόλουθα:

1. Καθαρισμός δεδομένων (Data cleaning): Στο βήμα αυτό, αφαιρούνται από τη βάση δεδομένων αυτά που παράγουν θόρυβο, δηλαδή όλα εκείνα τα στοιχεία που μπορούν να επηρεάσουν ή και να διαστρεβλώσουν το αποτέλεσμα.

2. Ενσωμάτωση δεδομένων (Data integration): Σε αυτό το βήμα τα δεδομένα που έχουν συλλεχθεί, πολλές φορές ανομοιογενή και από πολλές διαφορετικές πηγές, ενσωματώνονται σε μια κοινή βάση δεδομένων.

3. Επιλογή δεδομένων (Data selection): Από όλα εκείνα τα δεδομένα που έχουμε στη διάθεση μας, επιλέγονται προσεκτικά εκείνα που είναι σχετικά και χρήσιμα για την ανάλυση που θα ακολουθήσει.

4. Τροποποίηση δεδομένων (Data transformation): Τα δεδομένα που έχουμε επιλέξει δέχονται τις απαραίτητες τροποποιήσεις έτσι ώστε η μορφή τους να είναι κατάλληλη για την διαδικασία της εξόρυξης.

5. Εξόρυξη δεδομένων (Data mining): Είναι το σημαντικότερο από τα βήματα της διαδικασίας και αυτό γιατί στο συγκεκριμένο στάδιο, ποικίλες εξελιγμένες τεχνικές χρησιμοποιούνται για την εξαγωγή δυνητικά χρήσιμων προτύπων.

6. Αξιολόγηση προτύπων (Pattern evaluation): Στο βήμα αυτό αναγνωρίζονται χρήσιμα πρότυπα που αναπαριστούν γνώση, βάσει συγκεκριμένων μέτρων αξιολόγησης (evaluation measures).

7. Αναπαράσταση γνώσης (Knowledge representation): Στο τελικό αυτό στάδιο, η γνώση που έχει ανακαλυφθεί παρουσιάζεται στον χρήστη, βοηθώντας τον έτσι να κατανοήσει και να ερμηνεύσει τα αποτελέσματα της εξόρυξης δεδομένων.

Πολλές φορές κάποια από τα παραπάνω βήματα μπορούν να συνδυαστούν μεταξύ τους για το καλύτερο δυνατό αποτέλεσμα. Για παράδειγμα, τα βήματα του καθαρισμού και της ενσωμάτωσης των δεδομένων, μπορούν να υλοποιηθούν μαζί με στόχο την δημιουργία μια αποθήκης δεδομένων. Με την ίδια λογική μπορούν να συνδυαστούν και τα βήματα της επιλογής και τροποποίησης των δεδομένων.

Από τα παραπάνω λοιπόν συμπεραίνουμε ότι η εξόρυξη δεδομένων είναι μια διαδικασία-κλειδί για την ανεύρεση γνώσης. Παρόλα αυτά, δεν καταλαμβάνει παρά μόνο ένα μικρό μέρος της όλης προσπάθειας, δεδομένου της πολυπλοκότητας της. Σε αυτό το σημείο αξίζει να σημειωθεί ότι ο χρήστης, εκμεταλλευόμενος την επαναληπτική μορφή της διαδικασίας ανεύρεσης γνώσης, έχει την δυνατότητα να τροποποιήσει τα μέτρα αξιολόγησης, να τελειοποιήσει την διαδικασία της εξόρυξης, να επιλέξει νέα δεδομένα, να τροποποιήσει περαιτέρω τα ήδη υπάρχοντα ή να ενσωματώσει στη βάση νέα από καινούργιες πηγές, με τελικό στόχο την εξαγωγή διαφορετικών και ακόμη πιο κατάλληλων αποτελεσμάτων.

### 3.2.1 Εξόρυξη δεδομένων σε Supervised Data

Η supervised μάθηση είναι ένα task μηχανικής μάθησης για την εκμάθηση μιας συνάρτησης που χαρτογραφεί μια είσοδο σε μια έξοδο βασισμένη σε παραδείγματα ζευγών εισόδου-εξόδου. Συγκεντρώνει μια συνάρτηση από τα supervised data κατάρτισης που αποτελούνται από ένα σύνολο εκπαιδευτικών παραδειγμάτων. Ένας supervised αλγόριθμος εκμάθησης αναλύει τα δεδομένα εκπαίδευσης και παράγει μια συναγόμενη συνάρτηση, η οποία μπορεί να χρησιμοποιηθεί για χαρτογράφηση νέων παραδειγμάτων. Ένα βέλτιστο σενάριο θα επιτρέψει στον αλγόριθμο να προσδιορίσει σωστά τις ετικέτες της κλάσης για αθέατες περιπτώσεις. Αυτό απαιτεί τον αλγόριθμο μάθησης να γενικεύεται από τα δεδομένα εκπαίδευσης σε αόρατες καταστάσεις με έναν "λογικό" τρόπο.

#### Supervised Technics and Algorithms

Κατηγοριοποίηση (Classification) : είναι μια διαδικασία στην οποία οι ιδέες και τα αντικείμενα αναγνωρίζονται, διαφοροποιούνται και κατανοούνται. (Random Forest, Naïve Bayes, Logistic Regression) (Fig 2.3)

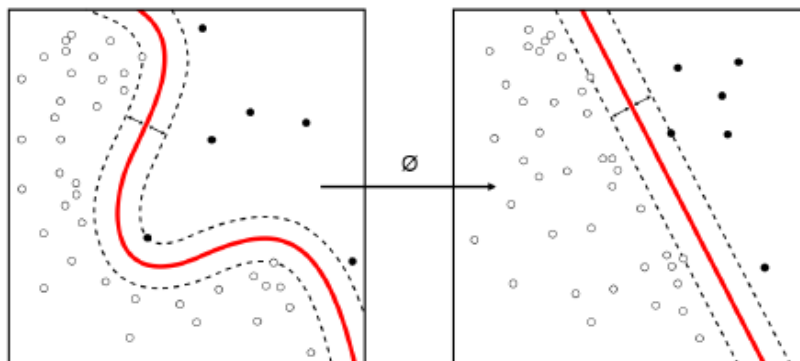


Fig 2.3

Οπισθοδρόμηση (Regression) : Υπολογίζει την εξαρτώμενη προσδοκία της εξαρτημένης μεταβλητής λαμβάνοντας υπόψη τις ανεξάρτητες μεταβλητές – δηλαδή τη μέση τιμή της εξαρτημένης μεταβλητής όταν οι ανεξάρτητες μεταβλητές είναι σταθερές. (Linear-Polynomial Regression) (Fig 2.4)

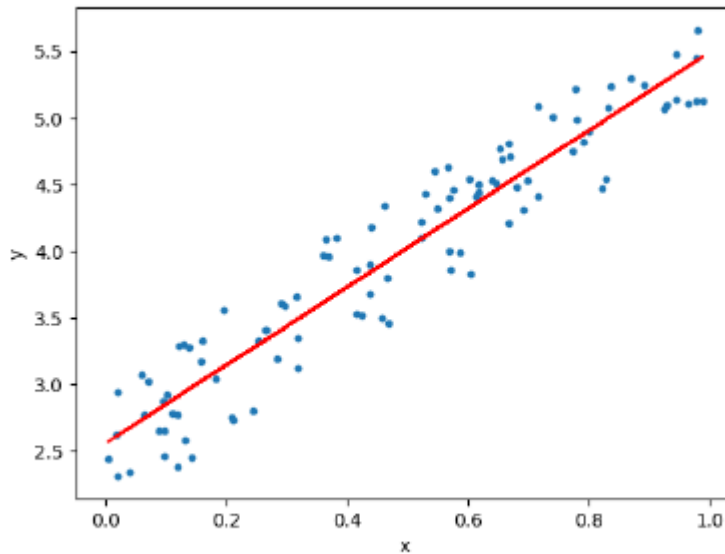


Fig 2.4

### 3.2.2 Εξόρυξη δεδομένων σε Supervised Data

Η unsupervised μάθηση είναι ένας τύπος αυτό-οργανωμένης εκμάθησης Hebbian που βοηθάει να βρούμε παλιότερα άγνωστα μοτίβα σε σύνολο δεδομένων χωρίς προϋπάρχουσες ετικέτες. Είναι επίσης γνωστή ως αυτό-οργάνωση και επιτρέπει πυκνότητες πιθανότητας μοντελοποίησης συγκεκριμένων εισροών.

#### Unsupervised Technics and Algorithms

Συσσωμάτωση (Clustering) : Είναι η διαδικασία ομαδοποίησης ενός συνόλου αντικειμένων με τέτοιο τρόπο ώστε τα αντικείμενα της ίδιας ομάδας (που ονομάζεται σύμπλεγμα) να είναι πιο παρόμοια μεταξύ τους (σε κάποια έννοια) με αυτά που ανήκουν σε άλλες ομάδες (συστάδες). (K-means, DBSCAN) (Fig 2.5)

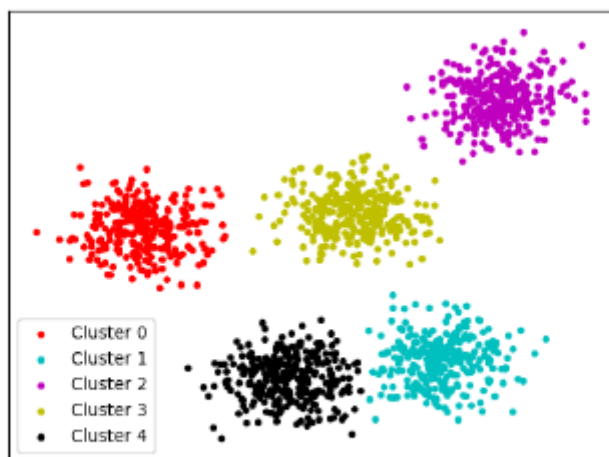


Fig 2.5

Νευρωνικά Δίκτυα (Neural Networks) : Είναι συστήματα υπολογιστών που εμπνέονται από, αλλά όχι ταυτόσημα, βιολογικά νευρωνικά δίκτυα που αποτελούν ζωικό εγκέφαλο. Αυτά τα συστήματα «μαθαίνουν» να εκτελούν εργασίες εξετάζοντας παραδείγματα, γενικά χωρίς να προγραμματίζονται με συγκεκριμένους κανόνες. (Hebbian Learning, Autoencoders) (Fig 2.6)

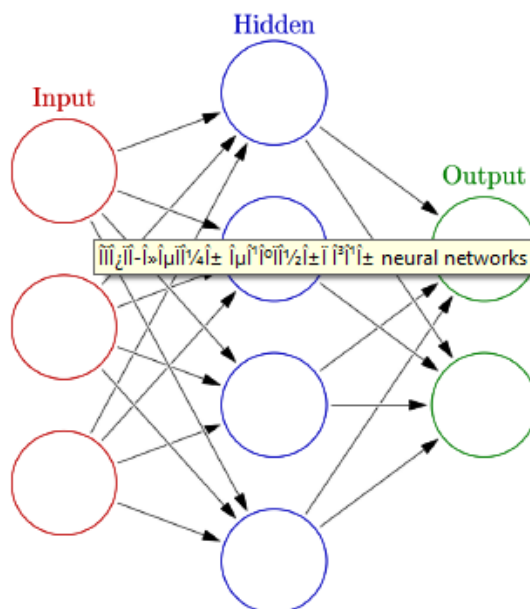


Fig 2.6



### 3.3 Θεωρητικό υπόβαθρο

#### Ανάλυση σελίδων πωλήσεων

#### Amazon

Είναι μια αμερικανική πολυεθνική εταιρεία τεχνολογίας με έδρα το Σιάτλ της Ουάσιγκτον, η οποία επικεντρώνεται στο ηλεκτρονικό εμπόριο, στο cloud computing, στην ψηφιακή ροή και στην τεχνητή νοημοσύνη. Θεωρείται μία από τις εταιρείες τεχνολογίας Big Four μαζί με το Google, την Apple και το Facebook. Η Amazon είναι γνωστή για τη διαταραχή των καθιερωμένων βιομηχανιών μέσω της τεχνολογικής καινοτομίας και της μαζικής κλίμακας. Είναι η μεγαλύτερη αγορά ηλεκτρονικού εμπορίου παγκοσμίως, AI assistant provider και πλατφόρμα υπολογιστικού cloud, όπως μετράται από τα έσοδα και την κεφαλαιοποίηση της αγοράς. Η Amazon είναι η μεγαλύτερη εταιρεία διαδικτύου από έσοδα στον κόσμο. Είναι ο δεύτερος μεγαλύτερος ιδιωτικός εργοδότης στις Ηνωμένες Πολιτείες και μία από τις πιο πολύτιμες εταιρείες στον κόσμο. Είναι η δεύτερη μεγαλύτερη εταιρεία τεχνολογίας από τα έσοδα. Όλα αυτά τα χρόνια, στα πλαίσια της αναζήτησης ενός ανοιχτού, ευέλικτου και δομημένου συστήματος, πρόσθεσε λειτουργίες όπως Verified Purchase, ψήφους και κριτικές.

#### Alibaba

Είναι μια κινεζική πολυεθνική εταιρεία χαρτοφυλακίου που ειδικεύεται στο ηλεκτρονικό εμπόριο, το λιανικό εμπόριο, το διαδίκτυο και την τεχνολογία. Η εταιρεία, που ιδρύθηκε στις 4 Απριλίου 1999 στο Hangzhou, Zhejiang, παρέχει υπηρεσίες διαδικτυακών πωλήσεων από καταναλωτές προς καταναλωτές, επιχειρήσεις προς καταναλωτές και επιχειρηματικές συναλλαγές, καθώς και ηλεκτρονικές υπηρεσίες πληρωμών, μηχανές αναζήτησης αγορών και υπηρεσίες cloud computing. Διαθέτει και εκμεταλλεύεται μια ποικιλία επιχειρήσεων σε όλο τον κόσμο σε πολλούς τομείς και ονομάζεται ως μία από τις πιο θαυμαστές εταιρείες του κόσμου από την Fortune.

#### eBay

Είναι μια αμερικανική πολυεθνική εταιρεία ηλεκτρονικού εμπορίου που εδρεύει στο San Jose της Καλιφόρνια και διευκολύνει τις πωλήσεις μεταξύ καταναλωτών μέσω των ιστοσελίδων τους. Το eBay ιδρύθηκε από τον Pierre Omidyar το φθινόπωρο του 1995 και

έγινε μια αξιοσημείωτη ιστορία επιτυχίας της φούσκας dot-com. Το eBay είναι μια επιχείρηση πολλών δισεκατομμυρίων δολαρίων με δραστηριότητες σε περίπου 30 χώρες, από το 2011. Η εταιρεία διαχειρίζεται τον ιστότοπο του eBay, έναν ηλεκτρονικό ιστότοπο δημοπρασιών και αγορών, στον οποίο οι άνθρωποι και οι επιχειρήσεις αγοράζουν και πωλούν μια μεγάλη ποικιλία αγαθών και υπηρεσιών παγκοσμίως. Ο ιστότοπος είναι ελεύθερος να χρησιμοποιηθεί για τους αγοραστές, αλλά οι πωλητές χρεώνουν τα τέλη για την καταχώριση στοιχείων.

## Walmart

Είναι μια αμερικανική πολυεθνική εταιρία λιανικής που λειτουργεί μια αλυσίδα υπερκαταστημάτων, εκπτώτικων πολυκαταστημάτων και παντοπωλείων με έδρα το Bentonville του Αρκάνσας. Η εταιρεία ιδρύθηκε από τον Sam Walton το 1962 και ενσωματώθηκε στις 31 Οκτωβρίου 1969. Διαθέτει επίσης και λειτουργεί τις λιανικές αποθήκες του Sam's Club. Από τις 31 Ιουλίου 2019, η Walmart διαθέτει 11.389 καταστήματα και κλαμπ σε 27 χώρες, που λειτουργούν με 55 διαφορετικά ονόματα. Η εταιρεία λειτουργεί με την επωνυμία Walmart στις Ηνωμένες Πολιτείες και τον Καναδά, ως Walmart de México y Centroamérica στο Μεξικό και την Κεντρική Αμερική, ως Asda στο Ηνωμένο Βασίλειο και ως ο όμιλος Seiyu στην Ιαπωνία. Έχει επιχειρήσεις που ανήκουν εξ ολοκλήρου στην Αργεντινή, τη Χιλή, τον Καναδά και τη Νότια Αφρική. Από τον Αύγουστο του 2018, η Walmart κατέχει μόνο μειοψηφική συμμετοχή στην Walmart Brasil, η οποία μετονομάστηκε σε Grupo Big τον Αύγουστο του 2019, με το 20% των μετοχών της εταιρείας ενώ η εταιρεία ιδιωτικών μετοχών Advent International κατέχει το 80% της ιδιοκτησίας της.

### **3.4 Στόχοι της εξόρυξης δεδομένων**

Οι μέθοδοι εξόρυξης γνώσης στοχεύουν στην ανακάλυψη στοιχείων που θα είναι χρήσιμα για τους οργανισμούς και τις επιχειρήσεις. Πληροφορίες για τυποποιημένες μορφές όπως για παράδειγμα, ότι υπάρχουν πελάτες που θα ψωνίσουν περισσότερο από δύο φορές σε περίοδο εκπτώσεων ή προσφορών, ή είναι πιθανό να αγοράσουν τουλάχιστον μια φορά κατά την διάρκεια των εορταστικών ημερών, Πάσχα και Χριστουγέννων, είτε για συσχετίσεις όπως όταν ένας πελάτης αγοράζει dvd player τότε πιθανότατα να αγοράσει και κάποια άλλη ηλεκτρονική συσκευή, μπορεί να αποτελέσουν καθοριστικούς παράγοντες για την λήψη αποφάσεων όσον αφορά τη λειτουργία μιας εμπορικής επιχείρησης. Αυτό συμβαίνει επειδή μπορεί να ληφθούν αποφάσεις σχετικά με το ωράριο, το ύψος και τη διάρκεια των εκπτώσεων, ακόμη και για την τοποθέτηση των προϊόντων μέσα στα καταστήματα.

Παράλληλα τέτοιου είδους πληροφορίες χρησιμοποιούνται για τον προγραμματισμό χρήσης πρόσθετων αποθηκευτικών χώρων ή και για τον σχεδιασμό διαφορετικών στρατηγικών μάρκετινγκ. Τα στελέχη της επιχείρησης, που είναι υπεύθυνα για την λήψη των αποφάσεων εκμεταλλεύονται τις δυνατότητες της εξόρυξης γνώσης και μετατρέπουν τις

γνώσεις σε επιτυχή αποτελέσματα. Παρακάτω περιγράφονται και αναλύονται οι στόχοι της εξόρυξης δεδομένων.

Η εξόρυξη δεδομένων έχει λοιπόν σαν βασικούς της στόχους την εφαρμογή τεχνικών πρόβλεψης και συμπεριφοράς τάσεων (prediction), την αναγνώριση, την περιγραφή (description) σε μεγάλες βάσεις δεδομένων, καθώς επίσης την ταξινόμηση και την βελτιστοποίηση των πόρων της.

Ειδικότερα:

- **Πρόβλεψη:** Περιλαμβάνει την χρήση μερικών μεταβλητών ή χαρακτηριστικών μιας βάσης δεδομένων για την πρόβλεψη άγνωστων ή μελλοντικών τιμών χρήσιμων μεταβλητών. Με άλλα λόγια, οι διαδικασίες πρόβλεψης της εξόρυξης δεδομένων (predictive data mining tasks), προσπαθούν να κάνουν εκτιμήσεις βγάζοντας συμπεράσματα από τα διαθέσιμα δεδομένα. Η προσπάθεια πρόβλεψης μελλοντικών συμπεριφορών έχει ως στόχο να ληφθούν αποφάσεις που να μεγιστοποιούν το κέρδος και να προλαμβάνουν δυσάρεστες καταστάσεις. Τα αποτελέσματα της εξόρυξης μπορεί να είναι πληροφορίες σχετικές με το ύψος των πωλήσεων ενός καταστήματος για μια συγκεκριμένη χρονική περίοδο, αλλά και αν το κλείσιμο μιας γραμμής παραγωγής θα είχε θετική επίδραση στις πωλήσεις. Συγχρόνως σε επιστημονικό επίπεδο, η μελέτη παλαιότερων σεισμικών φαινομένων ίσως να οδηγούσε στην πρόβλεψη σεισμικής δραστηριότητας.

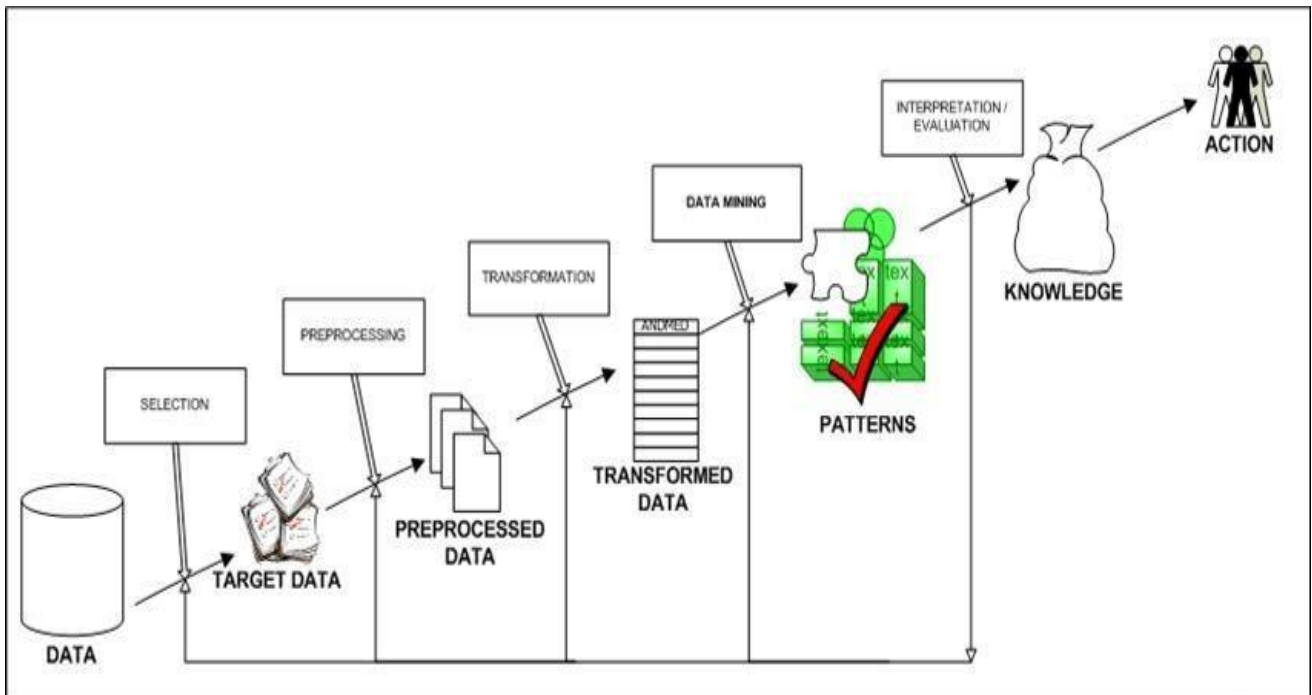
- **Αναγνώριση:** Σε αυτή τη φάση οι τυποποιημένες μορφές των δεδομένων χρησιμοποιούνται για να δείξουν την ύπαρξη μιας δραστηριότητας ή ενός γεγονότος.

- **Περιγραφή:** Είναι η διαδικασία η οποία επικεντρώνεται στην ανακάλυψη προτύπων και αναπαριστά τα δεδομένα μιας πολύπλοκης βάσης δεδομένων με όσο το δυνατό πιο κατανοητό και αξιοποιήσιμο τρόπο. Με άλλα λόγια, οι περιγραφικές διαδικασίες της εξόρυξης δεδομένων (descriptive data mining tasks) περιγράφουν τις γενικές ιδιότητες των υπαρχόντων διαθέσιμων δεδομένων.

- **Ταξινόμηση:** Σε αυτό το στάδιο έχουμε διαχωρισμό των στοιχείων, με αποτέλεσμα να προκύπτουν διαφορετικές κατηγορίες ή κλάσεις. Για παράδειγμα, οι πελάτες ενός σούπερ μάρκετ είναι δυνατόν να χωριστούν σε παρορμητικούς, πιστούς ή αλλιώς όπως θα λέγαμε κανονικούς, σπάνιους και σε φίλους των εκπτώσεων και προσφορών. Κατά την ανάλυση των πωλήσεων αυτή η κατηγοριοποίηση χρησιμοποιείται για να ληφθούν αποφάσεις, ώστε να προσελκυστούν περισσότεροι πελάτες ανεξαρτήτως κατηγορίας.

- **Βελτιστοποίηση:** Μεταξύ των άλλων σκοπός της εξόρυξης γνώσης είναι η βέλτιστη χρήση κάποιων πόρων κάτω από περιορισμούς. Τέτοιοι πόροι μπορεί να είναι ο χρόνος, ο χώρος, το χρήμα και η μεγιστοποίηση κάποιων μεγεθών, όπως είναι τα κέρδη είτε οι

πωλήσεις. Σε αυτή την περίπτωση η εξόρυξη γνώσης έχει κοινά σημεία με την επιχειρησιακή έρευνα.



Εικόνα 3.1: Στάδια της διαδικασίας εξόρυξης γνώσης.

### 3.5 Που εφαρμόζεται η εξόρυξη γνώσης.

Τα παραδείγματα εξόρυξης γνώσης από δεδομένα ποικίλουν ανάλογα με τον τομέα στον οποίο εφαρμόζονται. Στη σημερινή εποχή όπου τα δεδομένα υπάρχουν σχεδόν παντού και τις περισσότερες φορές βρίσκονται σε ηλεκτρονική μορφή, η σωστή ανάλυση τους οδηγεί πάντα στην ανάδειξη και οργάνωση της πληροφορίας, η γνώση της οποίας είναι ο σημαντικότερος παράγοντας για την εύρεση μιας στρατηγικής και την ορθολογική λήψη αποφάσεων.

Ο χρηματοοικονομικός τομέας, ο τομέας των τηλεπικοινωνιών, της υγείας και της εκπαίδευσης, ο δημόσιος τομέας καθώς επίσης και αυτός της βιομηχανίας και της έρευνας, αποτελούν ίσως το μεγαλύτερο δείγμα εφαρμογών των τεχνολογιών εξόρυξης γνώσης από δεδομένα.

Στον χρηματοοικονομικό τομέα η παρουσία αρκετών τραπεζικών και ασφαλιστικών ιδρυμάτων σε συνδυασμό με το υπάρχον οικονομικό κλίμα έχει αυξήσει κατά πολύ τον ανταγωνισμό μεταξύ των επιχειρήσεων. Πολλά από τα επιχειρηματικά ζητήματα του κλάδου, όπως η προσέλκυση νέων πελατών που θα αποφέρουν κέρδος, η προώθηση και πώληση

επιπρόσθετων προϊόντων ή παροχή υπηρεσιών, η διατήρηση πελατών, ο προσδιορισμός οικονομικού δόλου, η ανάλυση του πιστωτικού κινδύνου ενδεχόμενων πελατών, μπορούν να αντιμετωπιστούν με τα κατάλληλα εργαλεία της εξόρυξης γνώσης.

Στον κλάδο των **τηλεπικοινωνιών** το διαρκώς μεταβαλλόμενο επιχειρηματικό περιβάλλον με τον αστείρευτο ανταγωνισμό, αναγκάζουν τις εταιρείες να προβούν στην αναζήτηση νέων τρόπων ενίσχυσης της θέσης τους έναντι άλλων. Έχοντας γίνει πλήρως αντιληπτό από τη πλευρά τους το συγκριτικό πλεονέκτημα που τους προσφέρει η εξόρυξη γνώσης από δεδομένα, εφαρμόζουν μεταξύ άλλων τεχνικές έγκαιρης πρόβλεψης διακοπής υπηρεσιών από πελάτες, κατηγοριοποίηση των απαιτήσεων τους, ομαδοποίηση των συνηθειών τους και όλα αυτά με τελικό σκοπό την συγκράτηση των πελατών που ήδη έχουν αλλά και την προσέλκυση νέων.

Ο τομείς του λιανικού εμπορίου είναι ένας άλλος κλάδος ιδιαίτερα ανταγωνιστικός, όπου οι εφαρμογές εξόρυξης γνώσης βρίσκουν μεγάλη ανταπόκριση. Οι συνεχείς αλλαγές των καταναλωτικών προτιμήσεων και οι τεράστιοι όγκοι δεδομένων πωλήσεων, κρύβουν πολύτιμα στοιχεία εκ των οποίων ελάχιστα μπορούν να αξιοποιηθούν από τα συμβατικά συστήματα ανάλυσης πληροφορίας. Αντιθέτως οι εφαρμογές εξόρυξης γνώσης δίνουν μια νέα διάσταση στην παλαιότερη και βασικότερη επιχειρηματική διαδικασία που είχε σαν αρχή: «αναλύοντας ότι έγινε στο παρελθόν και κατανοώντας τα αποτελέσματα μπορούμε να γίνουμε αποτελεσματικότεροι στο μέλλον.» Επιπλέον, οι εφαρμογές εξόρυξης γνώσης κάνουν εφικτή μια προσωποποιημένη σχέση με κάθε ένα πελάτη χωριστά, κάτι που εξασφαλίζει την διαχρονική σχέση και την μεγιστοποίηση του κέρδους ανά πελάτη.

Οι επαγγελματίες στο χώρο της υγείας, πάντα αντιμετωπίζουν την ανάγκη να συλλέγουν, να αποθηκεύουν και να αναλύουν μεγάλες ποσότητες δεδομένων που μπορεί να περιλαμβάνουν καρτέλες ασθενών, δοκιμές νέων φαρμάκων, εξάρσεις ασθενειών και πολλά άλλα.

Από όλα τα παραπάνω, συμπεραίνουμε ότι η εξόρυξη γνώσης είναι ένα απαραίτητο εργαλείο σε πολλούς τομείς της σύγχρονης κοινωνίας. Η ραγδαία αύξηση του όγκου δεδομένων έχει καταστήσει σαφές ότι παλιές παραδοσιακές τεχνικές και μέθοδοι, δε μπορούν πλέον να βοηθήσουν στην ανάλυση και οργάνωση της πληροφορίας. Πολύ περισσότερο, δε μπορούν να φέρουν στην επιφάνεια γνώση που τα δεδομένα περιέχουν καλά κρυμμένη και η οποία απαιτεί εφαρμογή ειδικών για να αποκαλυφθεί.

### **3.5.1 Εξόρυξη δεδομένων και οικονομία.**

Ένας από τους τομείς που εφαρμόζεται κατά κόρον η εξόρυξη δεδομένων είναι αυτός της οικονομίας. Τα οικονομικά δεδομένα συλλέγονται κυρίως από τράπεζες, σουπερμάρκετ και από άλλους οικονομικούς οργανισμούς. Τα δεδομένα αυτά συνήθως είναι αξιόπιστα, ολοκληρωμένα, έχουν υψηλή ποιότητα και απαιτούν συστηματική μέθοδο για την ανάλυση τους. Η συνεισφορά της εξόρυξης δεδομένων στην επιστήμη της οικονομίας συναντάται στην συλλογή, κατανόηση και βελτίωση των δεδομένων, στην δημιουργία και εκτίμηση ενός μοντέλου και στην ανάπτυξη αυτού.

Η σωστή ανάλυση των οικονομικών δεδομένων διευκολύνει στο να παρθούν καλύτερες αποφάσεις ενεργώντας σύμφωνα με την ανάλυση της αγοράς. Τα εργαλεία και οι τεχνικές της εξόρυξης δεδομένων βοηθούν στο να αναλύσουμε τα οικονομικά δεδομένα και είναι τέτοια η συμβολή τους έτσι ώστε για παράδειγμα, τα οικονομικά ινστιτούτα να αναγνωρίζουν τις απάτες από παραποιημένα δεδομένα από τις διάφορες βάσεις δεδομένων και από το ιστορικό συναλλαγών που έγιναν από τους πελάτες. Οι τεχνικές οπτικοποίησης βοηθούν στην παρουσίαση δεδομένων με διαφορετικές μορφές, όπως γράφοι που βασίζονται σε συγκεκριμένα γνωρίσματα. Παραδείγματος χάρη προβάλλοντας τα δεδομένα από διάφορες οπτικές γωνίες, μία τράπεζα δύναται να διακρίνει τους πελάτες που έχουν επιχειρήσει παράνομες πράξεις και μετά μια λεπτομερή έρευνα αυτών των ύποπτων περιπτώσεων βοηθάει στην εξιχνίαση των απατών και των εγκλημάτων.

### 3.5.2 Εφαρμογές εξόρυξης γνώσης σε επιστημονικά πεδία.

Στο σημείο αυτό θα αναφέρουμε συνοπτικά ορισμένες εφαρμογές των μεθόδων της εξόρυξης γνώσης οι οποίες έχουν συλλεχθεί από διάφορα επιστημονικά πεδία.

- **Μάρκετινγκ μέσω mail (Direct mail marketing):** Το Body Shop International δοκιμάζει τεχνικές εξόρυξης γνώσης ώστε να καταφέρει να αυξήσει την αποτελεσματικότητα των παραγγελιών του μέσω email. Τα διοικητικά στελέχη ενδιαφέρονται στο να μειώσουν το κόστος αποστολής διαφημιστικών καταλόγων, εστιάζοντας μόνο σε πελάτες οι οποίοι θεωρούνται ως κερδοφόροι.

- **Κατηγοριοποίηση μέσω μάνατζμεντ και έλεγχος για κάτι νέο (Category management and inventory control):** Η εταιρεία Rubbermaid χρησιμοποιεί μεθόδους εξόρυξης γνώσης για να κατηγοριοποιεί το στυλ της στρατηγικής που χρησιμοποιεί ανάλογα με την αγορά στην οποία εστιάζεται, στο να αγοράζει και να πουλά τα οικονομικά αγαθά που παράγει. Ανάλυση καλαθιού της νοικοκυράς (Market basket analysis): Τα στελέχη της J.Crew Group συνδυάζουν click system analysis μέσα από το επίσημο site τους σε συνδυασμό με την μέθοδο point-of-sale (POS) στις λιανικές τους πωλήσεις. Με αυτό τον τρόπο θέλουν να δούνε τι ρούχα, τι υποδήματα και άλλα αξεσουάρ αγοράζονται μαζί. Τα δεδομένα μετά θα αναλυθούν και έπειτα θα σταλθούν κατάλογοι με οικονομικά αγαθά και προσφορές των σε on-line αγοραστές.

- **Κατανόηση του προφίλ του κάθε πελάτη (Customer relationship management):** Σύμφωνα με διάφορες εταιρείες λιανικής, άλλες από τον τραπεζικό τομέα είναι δυνατό χρησιμοποιώντας τεχνικές εξόρυξης γνώσης, όπως κανόνες ταξινόμησης ή και ομαδοποίηση να κατατάσσουμε τους πελάτες ανάλογα με προσωπικά τους στοιχεία και την προηγούμενη τους συμπεριφορά προς την εταιρεία που μας ενδιαφέρει σε καλούς, μέτριους και κακούς. Έτσι δίνεται η δυνατότητα να επιλέγουν το αγοραστικό κοινό με το οποίο συναλλάσσονται και στο οποίο εστιάζουν την πολιτική της εταιρείας.

- **Αστρονομία (Astronomy):** Έχει αναπτυχθεί ένα σύστημα με την ονομασία SKICAT από το JPL/Caltech και χρησιμοποιείται από τους αστρονόμους στο να αναγνωρίζουν αυτόματα

τους διάφορους γαλαξίες και αστεροειδείς σε μία μεγάλη κλίμακα η οποία περιέχει διάφορα αστρονομικά μεγέθη.

- **Βιολογία (Biology):** Διάφορα συστήματα έχουν αναπτυχθεί για να εξάγουν κανόνες που αφορούν την δομή των οργανισμών, την ανάλυση του DNA, καθώς και την δυνατότητα για εύρεση φαρμάκων για την καταπολέμηση ασθενειών.

- **Παγκόσμιο μοντέλο κλιματολογικών συνθηκών (Global climate modeling):** Διάφορα συστήματα έχουν υιοθετηθεί τα οποία επιτρέπουν την ανάλυση κλιματολογικών συνθηκών δίνοντας έτσι την δυνατότητα προβλέψεων κλιματολογικών φαινομένων όπως οι κυκλώνες, οι καταιγίδες, οι καύσωνες και άλλα πολλά.

- **Η μέθοδος εξόρυξης από δεδομένα σε οικονομικές εφαρμογές (Data mining for financial applications):** Πολλές φορές τεχνικές της μεθόδου εξόρυξης από δεδομένα όπως τα νευρωνικά δίκτυα και τα δένδρα αποφάσεων μπορούν να χρησιμοποιηθούν από οικονομικούς αναλυτές για την λήψη στρατηγικών αποφάσεων στο ανάλογο οικονομικό πεδίο που ενδιαφέρει κάθε φορά. Φυσικά χρειάζεται και το ανάλογο υπόβαθρο από ιστορικά δεδομένα ώστε να είναι δυνατή η ανάλυση που θα γίνει.

### 3.6 Αποτυγχάνουμε να αντλήσουμε γνώση από τα big data, συνεχίζουμε την εξόρυξη

Ένας στους τρεις οργανισμούς που επιχειρούν να εξάγουν ειδικές γνώσεις από την ανάλυση όλων των δεδομένων τους αποτυγχάνουν, δείχνει νέα έρευνα της Hewlett Packard. Η ανάλυση των επονομαζόμενων big data αποδεικνύεται δύσκολη υπόθεση, εντούτοις έξι στους δέκα οργανισμούς δηλώνουν πρόθυμοι να δεσμεύσουν το 10% του προϋπολογισμού τους για την καινοτομία για να εξορύξουν πολύτιμη πληροφορία από δομημένα και αδόμητα δεδομένα.

Ότι κι αν σημαίνει big data, η νεότερη έρευνα της Hewlett Packard τον Μάιο του 2013 επαναλαμβάνει ότι η ανάλυση της διαρκούς ροής δεδομένων δομημένων και αδόμητων παραμένει πρόκληση για τις επιχειρήσεις. Ο αυξανόμενος όγκος, η ποικιλομορφία αλλά και η ευπάθεια των δεδομένων που ρέουν από το εσωτερικό αλλά και το περιβάλλον κάθε επιχείρησης αποτελούν την αιτία της αποτυχίας των πρωτοβουλιών big data. Περισσότεροι από έναν στους τρεις οργανισμούς που έχει επιχειρήσει την ανάλυση έχει αποτύχει, δείχνει η έρευνα που πραγματοποιήθηκε για λογαριασμό της Hewlett Packard υπό τον τίτλο "Big Data and Cloud" από την Coleman Parkers Research Ltd., τον Μάιο του 2013.

Σε προγενέστερη παγκόσμια έρευνα που εκδόθηκε τον Απρίλιο και διενεργήθηκε επίσης για λογαριασμό της HP, περισσότερα από ένα στα δύο στελέχη επιχειρήσεων ανέφεραν ότι οι οργανισμοί τους δεν είναι εξοπλισμένοι με τις σωστές λύσεις για να αντλήσουν ειδικές γνώσεις από τα Big Data. Επιπλέον, δεν διαθέτουν την τεχνογνωσία καθώς και τη συνεκτική στρατηγική για να συγκεντρώσουν όλα τα στοιχεία και στη συνέχεια

να ενσωματώσουν νέα αλλά και παλιά δεδομένα. Παρά τις συνεχείς αποτυχίες, το 60% των εταιρειών που συμμετείχαν στην νεότερη έρευνα δηλώνουν πρόθυμες να δεσμεύσουν για το big data το 10% του προϋπολογισμού τους για την καινοτομία. Αυτό δείχνει ότι έχουν πειστεί ότι τα επονομαζόμενα big data κρύβουν πλούσιες γνώσεις και η ανάλυσή τους μπορεί να προσφέρει αποτελέσματα σε πραγματικό χρόνο. Με την εξαγωγή ειδικών γνώσεων που είναι κρυμμένες μέσα στα big data οι επιχειρήσεις μπορούν να εξορθολογήσουν τις βασικές οργανωτικές διαδικασίες όπως, οι προσφορές, οι προμήθειες, η εφοδιαστική αλυσίδα και οι λειτουργίες απογραφής, υποστηρίζουν εταιρείες όπως η Hewlett Packard, η IBM και άλλες εταιρείες που παρέχουν hardware, software και υπηρεσίες.

«Τα big data επιτρέπουν στους οργανισμούς να επωφεληθούν από το σύνολο των πληροφοριών τους, τόσο των εσωτερικών όσο και των εξωτερικών, σε πραγματικό χρόνο. Παράγουν εξαιρετικά γρήγορη διαδικασία λήψης αποφάσεων και ως αποτέλεσμα, μοναδικούς και καινοτόμους τρόπους για την προσφορά υπηρεσιών στους πελάτες και την κοινωνία», δήλωσε ο George Kadifa, επικεφαλής του τμήματος software στην HP κατά την ανακοίνωση του HAVEn. Το HAVEn αποτελεί την νεότερη πρόταση της HP για την ανάλυση big data, μια πλατφόρμα big data analytics, η οποία αξιοποιεί το λογισμικό, το hardware και τις υπηρεσίες analytics της HP (συνδυάζει τις δοκιμασμένες τεχνολογίες των HP Autonomy, HP Vertica, HP Arc Sight και HP Operations Management). Η Hewlett Packard υποστηρίζει πως με αυτή την πλατφόρμα οι επιχειρήσεις μπορούν να αντλήσουν αξία 56 από το 100% των πληροφοριών, συμπεριλαμβανομένων των δομημένων, των ημι-δομημένων και των μη-δομημένων δεδομένων.

Περνώντας από τις υψηλού επιπέδου έννοιες των big data στην πραγματικότητα, δηλαδή στην αξιοποίησή τους και τη δράση με βάση την πληροφορία, οι ειδικές γνώσεις που προκύπτουν από την ανάλυση τους μπορούν να βοηθήσουν μια επιχείρηση να βελτιώσει:

- Το cross-channel marketing σε πραγματικό χρόνο
- Την εμπειρία του πελάτη και την πρόβλεψη της ζήτησης

Άλλες περιπτώσεις χρήσης περιλαμβάνουν:

- Την υποστήριξη των προσπαθειών των πελατών να εντοπίσουν και να προλάβουν την απάτη,
- Να διασφαλίσουν τη συμμόρφωση σε πραγματικό χρόνο
- Να χρησιμοποιήσουν τα κοινωνικά δίκτυα για να διαχειριστούν τους κινδύνους και τη φήμη της μάρκας τους.

Είναι γεγονός ότι ο όρος big data έχει πολλούς ορισμούς, όπως άλλωστε και η πρακτική αξιοποίησή της ανάλυσης των πάντοτε αδιευκρίνιστων δεδομένων.





*Εικόνα 3.2: Η διαδικασία της εξόρυξης δεδομένων.*

Η εξόρυξη δεδομένων είναι ένα ισχυρό εργαλείο Επιχειρηματικής Ευφυΐας για την ανακάλυψη γνώσης και παράλληλα είναι μια διαδικασία διαχείρισης γνώσης γιατί περιλαμβάνει ανθρώπινη γνώση. Συνεπώς η εξόρυξη δεδομένων συνδέει την Επιχειρηματική Ευφυΐα με τη διαχείριση γνώσης. Η Επιχειρηματική Ευφυΐα και η Διαχείριση Γνώσης πρέπει να ενοποιηθούν προκειμένου να προωθηθεί η οργανωτική εκμάθηση και η αποτελεσματική λήψη αποφάσεων και παράλληλα θα πρέπει να υπολογισθεί η αποτελεσματικότητα της Επιχειρηματικής Ευφυΐας βάσει της βελτίωσης της γνώσης ενός οργανισμού.

## ΚΕΦΑΛΑΙΟ 4ο : Κανονισμοί - Νομοθεσία GDPR

### 4.1 Κανονισμοί

(6) Οι ραγδαίες τεχνολογικές εξελίξεις και η παγκοσμιοποίηση δημιούργησαν νέες προκλήσεις για την προστασία των δεδομένων προσωπικού χαρακτήρα. Η κλίμακα της συλλογής και της ανταλλαγής δεδομένων προσωπικού χαρακτήρα αυξήθηκε σημαντικά. Η τεχνολογία επιτρέπει τόσο σε ιδιωτικές επιχειρήσεις όσο και σε δημόσιες αρχές να κάνουν χρήση δεδομένων προσωπικού χαρακτήρα σε πρωτοφανή κλίμακα για την επιδίωξη των δραστηριοτήτων τους. Τα φυσικά πρόσωπα ολοένα και περισσότερο δημοσιοποιούν προσωπικές πληροφορίες και τις καθιστούν διαθέσιμες σε παγκόσμιο επίπεδο. Η τεχνολογία έχει αλλάξει τόσο την οικονομία όσο και την κοινωνική ζωή και θα πρέπει να διευκολύνει περαιτέρω την ελεύθερη κυκλοφορία δεδομένων προσωπικού χαρακτήρα εντός της Ένωσης και τη διαβίβαση σε τρίτες χώρες και διεθνείς οργανισμούς, διασφαλίζοντας παράλληλα υψηλό επίπεδο προστασίας των δεδομένων προσωπικού χαρακτήρα.

(13) Για να διασφαλιστεί συνεκτικό επίπεδο προστασίας των φυσικών προσώπων σε ολόκληρη την Ένωση και προς αποφυγή αποκλίσεων που εμποδίζουν την ελεύθερη κυκλοφορία των δεδομένων προσωπικού χαρακτήρα στην εσωτερική αγορά, απαιτείται κανονισμός ο οποίος θα κατοχυρώνει την ασφάλεια δικαίου και τη διαφάνεια για τους οικονομικούς παράγοντες, περιλαμβανομένων των πολύ μικρών, των μικρών και των μεσαίων επιχειρήσεων, και θα προβλέπει για τα φυσικά πρόσωπα σε όλα τα κράτη μέλη το ίδιο επίπεδο νομικά εκτελεστών δικαιωμάτων και υποχρεώσεων, καθώς και ευθυνών για τους υπευθύνους επεξεργασίας και τους εκτελούντες την επεξεργασία, ώστε να διασφαλιστεί η συνεκτική παρακολούθηση της επεξεργασίας δεδομένων προσωπικού χαρακτήρα, καθώς και οι ισοδύναμες κυρώσεις σε όλα τα κράτη μέλη και η αποτελεσματική συνεργασία μεταξύ των εποπτικών αρχών των διάφορων κρατών μελών. Για την ομαλή λειτουργία της εσωτερικής αγοράς, η ελεύθερη κυκλοφορία δεδομένων προσωπικού χαρακτήρα εντός της Ένωσης δεν πρέπει να περιορίζεται, ούτε να απαγορεύεται για λόγους που σχετίζονται με την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα.

(15) Προκειμένου να αποτραπεί σοβαρός κίνδυνος καταστρατήγησης, η προστασία των φυσικών προσώπων θα πρέπει να είναι τεχνολογικά ουδέτερη και να μην εξαρτάται από τις χρησιμοποιούμενες τεχνικές. Η προστασία των φυσικών προσώπων θα πρέπει να εφαρμόζεται τόσο στην επεξεργασία δεδομένων προσωπικού χαρακτήρα με αυτοματοποιημένα μέσα, όσο και στη χειροκίνητη επεξεργασία, εάν τα δεδομένα προσωπικού χαρακτήρα περιέχονται ή προορίζονται να περιληφθούν σε σύστημα αρχειοθέτησης

(24) Η επεξεργασία δεδομένων προσωπικού χαρακτήρα προσώπων που βρίσκονται στην Ένωση από υπεύθυνο επεξεργασίας ή εκτελούντα την επεξεργασία μη εγκατεστημένο στην Ένωση θα πρέπει επίσης να διέπεται από τον παρόντα κανονισμό, εφόσον αφορά την παρακολούθηση της συμπεριφοράς των εν λόγω υποκειμένων των δεδομένων στον βαθμό που η συμπεριφορά τους λαμβάνει χώρα εντός της Ένωσης. Για τον καθορισμό του κατά πόσον μια δραστηριότητα επεξεργασίας μπορεί να θεωρηθεί ότι παρακολουθεί τη συμπεριφορά υποκειμένου των δεδομένων, θα πρέπει να εξακριβωθεί κατά πόσον φυσικά πρόσωπα παρακολουθούνται στο Διαδίκτυο, συμπεριλαμβανομένης της δυνητικής μετέπειτα χρήσης τεχνικών επεξεργασίας δεδομένων προσωπικού χαρακτήρα οι οποίες συνίστανται στη διαμόρφωση του «προφίλ» ενός φυσικού προσώπου, ιδίως με σκοπό να ληφθούν αποφάσεις που το αφορούν ή να αναλυθούν ή να προβλεφθούν οι προσωπικές προτιμήσεις, οι συμπεριφορές και οι νοοτροπίες του.

(30) Τα φυσικά πρόσωπα μπορεί να συνδέονται με επιγραμματικά αναγνωριστικά στοιχεία ταυτότητας, τα οποία παρέχονται από τις συσκευές, τις εφαρμογές, τα εργαλεία και τα πρωτόκολλά τους, όπως διεύθυνσης διαδικτυακού πρωτοκόλλου, αναγνωριστικά cookies ή άλλα αναγνωριστικά στοιχεία όπως ετικέτες αναγνώρισης μέσω ραδιοσυχνοτήτων. Αυτά μπορεί να αφήνουν ίχνη τα οποία, ιδίως όταν συνδυαστούν με μοναδικά αναγνωριστικά στοιχεία ταυτότητας και άλλες πληροφορίες που λαμβάνουν οι εξυπηρετητές, μπορούν να χρησιμοποιηθούν για να δημιουργηθεί το προφίλ των φυσικών προσώπων και να αναγνωριστεί η ταυτότητά τους.

(31) Οι δημόσιες αρχές στις οποίες κοινολογούνται τα δεδομένα προσωπικού χαρακτήρα σύμφωνα με νομική υποχρέωση για την άσκηση των επίσημων καθηκόντων τους, όπως φορολογικές και τελωνειακές αρχές, μονάδες οικονομικής έρευνας, ανεξάρτητες διοικητικές αρχές ή αρχές χρηματοπιστωτικών αγορών που είναι αρμόδιες για τη ρύθμιση και την εποπτεία των αγορών κινητών αξιών δεν θα πρέπει να θεωρηθούν αποδέκτες, εάν λαμβάνουν δεδομένα προσωπικού χαρακτήρα που είναι απαραίτητα για τη διενέργεια ειδικής έρευνας για το γενικό συμφέρον, σύμφωνα με το δίκαιο της Ένωσης ή κράτους μέλους. Τα αιτήματα κοινολόγησης που αποστέλλονται από δημόσιες αρχές θα πρέπει να είναι πάντα γραπτά, αιτιολογημένα και σύμφωνα με την περίπτωση και δεν θα πρέπει να αφορούν το σύνολο ενός συστήματος αρχειοθέτησης ή να οδηγούν στη διασύνδεση των συστημάτων αρχειοθέτησης. Η επεξεργασία δεδομένων προσωπικού χαρακτήρα από τις εν λόγω δημόσιες αρχές θα πρέπει να συμμορφώνεται προς τους ισχύοντες κανόνες προστασίας των δεδομένων ανάλογα με τους σκοπούς της επεξεργασίας.

(32) Η συγκατάθεση θα πρέπει να παρέχεται με σαφή θετική ενέργεια η οποία να συνιστά ελεύθερη, συγκεκριμένη, ρητή και εν πλήρη επίγνωση ένδειξη της συμφωνίας του υποκειμένου των δεδομένων υπέρ της επεξεργασίας των δεδομένων που το αφορούν, για παράδειγμα με γραπτή δήλωση, μεταξύ άλλων με ηλεκτρονικά μέσα, ή με προφορική δήλωση. Αυτό θα μπορούσε να περιλαμβάνει τη συμπλήρωση ενός τετραγωνιδίου κατά την επίσκεψη σε διαδικτυακή ιστοσελίδα, την επιλογή των επιθυμητών τεχνικών ρυθμίσεων για υπηρεσίες της κοινωνίας των πληροφοριών ή μια δήλωση ή συμπεριφορά που δηλώνει

σαφώς, στο συγκεκριμένο πλαίσιο, ότι το υποκείμενο των δεδομένων αποδέχεται την πρόταση επεξεργασίας των οικείων δεδομένων προσωπικού χαρακτήρα. Επομένως, η σιωπή, τα προσυμπληρωμένα τετραγωνίδια ή η αδράνεια δεν θα πρέπει να εκλαμβάνονται ως συγκατάθεση. Η συγκατάθεση θα πρέπει να καλύπτει το σύνολο των δραστηριοτήτων επεξεργασίας που διενεργείται για τον ίδιο σκοπό ή για τους ίδιους σκοπούς. Όταν η επεξεργασία έχει πολλαπλούς σκοπούς, θα πρέπει να δίνεται συγκατάθεση για όλους αυτούς τους σκοπούς. Εάν η συγκατάθεση του υποκειμένου των δεδομένων πρόκειται να δοθεί κατόπιν αιτήματος με ηλεκτρονικά μέσα, το αίτημα πρέπει να είναι σαφές, περιεκτικό και να μην διαταράσσει αδικαιολόγητα τη χρήση της υπηρεσίας για την οποία παρέχεται.

(33) Συχνά, δεν είναι δυνατόν να προσδιορίζεται πλήρως ο σκοπός της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα για σκοπούς επιστημονικής έρευνας κατά τον χρόνο συλλογής των δεδομένων. Ως εκ τούτου, τα υποκείμενα των δεδομένων θα πρέπει να μπορούν να δώσουν τη συγκατάθεσή τους για ορισμένους τομείς της επιστημονικής έρευνας, όταν ακολουθούνται τα αναγνωρισμένα πρότυπα δεοντολογίας για την επιστημονική έρευνα. Τα υποκείμενα των δεδομένων θα πρέπει να έχουν τη δυνατότητα να παρέχουν τη συναίνεσή τους μόνο σε ορισμένους τομείς της έρευνας ή μόνο σε μέρη προγραμμάτων έρευνας, στον βαθμό που επιτρέπεται από τον επιδιωκόμενο σκοπό.

(38) Τα παιδιά απαιτούν άδίκη προστασία όσον αφορά τα δεδομένα τους προσωπικού χαρακτήρα, καθώς τα παιδιά μπορεί να έχουν μικρότερη επίγνωση των σχετικών κινδύνων, συνεπειών και εγγυήσεων και των δικαιωμάτων τους σε σχέση με την επεξεργασία δεδομένων προσωπικού χαρακτήρα. Αυτή η άδίκη προστασία θα πρέπει να ισχύει ιδίως στη χρήση των δεδομένων προσωπικού χαρακτήρα με σκοπό την εμπορία ή τη δημιουργία προφίλ προσωπικότητας ή προφίλ χρήστη και τη συλλογή δεδομένων προσωπικού χαρακτήρα όσον αφορά παιδιά κατά τη χρήση υπηρεσιών που προσφέρονται άμεσα σε ένα παιδί. Η συγκατάθεση του γονέα ή κηδεμόνα δεν θα πρέπει να είναι απαραίτητη σε συνάρτηση με υπηρεσίες πρόληψης ή παροχής συμβουλών που προσφέρονται άμεσα σε ένα παιδί.

(39) Κάθε επεξεργασία δεδομένων προσωπικού χαρακτήρα θα πρέπει να είναι σύννομη και δίκαιη. Θα πρέπει να είναι σαφές για τα φυσικά πρόσωπα ότι δεδομένα προσωπικού χαρακτήρα που τα αφορούν συλλέγονται, χρησιμοποιούνται, λαμβάνονται υπόψη ή υποβάλλονται κατ' άλλο τρόπο σε επεξεργασία, καθώς και σε ποιο βαθμό τα δεδομένα προσωπικού χαρακτήρα υποβάλλονται ή θα υποβληθούν σε επεξεργασία. Η αρχή αυτή απαιτεί κάθε πληροφορία και ανακοίνωση σχετικά με την επεξεργασία των εν λόγω δεδομένων προσωπικού χαρακτήρα να είναι εύκολα προσβάσιμη και κατανοητή και να χρησιμοποιεί σαφή και απλή γλώσσα. Αυτή η αρχή αφορά ιδίως την ενημέρωση των υποκειμένων των δεδομένων σχετικά με την ταυτότητα του υπευθύνου επεξεργασίας και τους σκοπούς της επεξεργασίας και την περαιτέρω ενημέρωση ώστε να διασφαλιστεί δίκαιη και διαφανής επεξεργασία σε σχέση με τα εν λόγω φυσικά πρόσωπα και το δικαίωμά τους να λαμβάνουν επιβεβαίωση και να επιτυγχάνουν ανακοίνωση των σχετικών με αυτά δεδομένων προσωπικού χαρακτήρα που υπόκεινται σε επεξεργασία. Θα πρέπει να γνωστοποιείται στα

φυσικά πρόσωπα η ύπαρξη κινδύνων, κανόνων, εγγυήσεων και δικαιωμάτων σε σχέση με την επεξεργασία δεδομένων προσωπικού χαρακτήρα και πώς να ασκούν τα δικαιώματά τους σε σχέση με την επεξεργασία αυτή. Ιδίως, οι συγκεκριμένοι σκοποί της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα θα πρέπει να είναι σαφείς, νόμιμοι και προσδιορισμένοι κατά τον χρόνο συλλογής των δεδομένων προσωπικού χαρακτήρα. Τα δεδομένα προσωπικού χαρακτήρα θα πρέπει να είναι επαρκή και συναφή και να περιορίζονται στα αναγκαία για τους σκοπούς της επεξεργασίας τους. Αυτό απαιτεί ειδικότερα να διασφαλίζεται ότι το διάστημα αποθήκευσης των δεδομένων προσωπικού χαρακτήρα να περιορίζεται στο ελάχιστο δυνατό. Τα δεδομένα προσωπικού χαρακτήρα θα πρέπει να υποβάλλονται σε επεξεργασία μόνο εάν ο σκοπός της επεξεργασίας δεν μπορεί να επιτευχθεί με άλλα μέσα. Για να διασφαλιστεί ότι τα δεδομένα προσωπικού χαρακτήρα δεν διατηρούνται περισσότερο από όσο είναι αναγκαίο, ο υπεύθυνος επεξεργασίας θα πρέπει να όριζα προθεσμίες για τη διαγραφή τους ή για την περιοδική επανεξέτασή τους. Θα πρέπει να λαμβάνεται κάθε εύλογο μέτρο, ώστε να διασφαλίζεται ότι τα δεδομένα προσωπικού χαρακτήρα που δεν είναι ακριβή διορθώνονται ή διαγράφονται. Τα δεδομένα προσωπικού χαρακτήρα θα πρέπει να υφίστανται επεξεργασία κατά τρόπο που να διασφαλίζει την ενδεδειγμένη προστασία και εμπιστευτικότητα των δεδομένων προσωπικού χαρακτήρα, μεταξύ άλλων και για να αποτρέπεται κάθε ανεξουσιοδότητη πρόσβαση σε αυτά τα δεδομένα προσωπικού χαρακτήρα και στον εξοπλισμό που χρησιμοποιείται για την επεξεργασία τους ή η χρήση αυτών των δεδομένων προσωπικού χαρακτήρα και του εν λόγω εξοπλισμού.

(42) Όταν η επεξεργασία βασίζεται στη συναίνεση του υποκειμένου των δεδομένων, ο υπεύθυνος επεξεργασίας θα πρέπει να είναι σε θέση να αποδείξει ότι το υποκείμενο των δεδομένων συγκατατέθηκε στη πράξη επεξεργασίας. Ειδικότερα, στο πλαίσιο έγγραφης δήλωσης για άλλο θέμα, θα πρέπει να παρέχονται εγγυήσεις που να διασφαλίζουν ότι το υποκείμενο των δεδομένων γνωρίζει αυτό το γεγονός και σε ποιο βαθμό έχει συγκατατεθεί. Σύμφωνα με την οδηγία 93/13/ΕΟΚ του Συμβουλίου ('), θα πρέπει να παρέχεται δήλωση συγκατάθεσης, διατυπωμένη εκ των προτέρων από τον υπεύθυνο επεξεργασίας σε κατανοητή και εύκολα προσβάσιμη μορφή, με σαφή και απλή διατύπωση, χωρίς καταχρηστικές ρήτρες. Για να θεωρηθεί η συγκατάθεση εν επι γνώση, το υποκείμενο των δεδομένων θα πρέπει να γνωρίζει τουλάχιστον την ταυτότητα του υπευθύνου επεξεργασίας και τους σκοπούς της επεξεργασίας για την οποία προορίζονται τα δεδομένα προσωπικού χαρακτήρα. Η συγκατάθεση δεν θα πρέπει να θεωρείται ότι δόθηκε ελεύθερα αν το υποκείμενο των δεδομένων δεν έχει αληθινή ή ελεύθερη επιλογή ή δεν είναι σε θέση να αρνηθεί ή να αποσύρει τη συγκατάθεσή του χωρίς να ζημιωθεί.

(60) Οι αρχές της δίκαιης και διαφανούς επεξεργασίας απαιτούν να ενημερώνεται το υποκείμενο των δεδομένων για την ύπαρξη της πράξης επεξεργασίας και τους σκοπούς της. Ο υπεύθυνος επεξεργασίας θα πρέπει να παρέχει στο υποκείμενο των δεδομένων κάθε περαιτέρω πληροφορία που είναι αναγκαία για τη διασφάλιση δίκαιης και διαφανούς επεξεργασίας, λαμβάνοντας υπόψη τις ειδικές συνθήκες και το πλαίσιο εντός του οποίου πραγματοποιείται η επεξεργασία των δεδομένων προσωπικού χαρακτήρα. Περαιτέρω, το υποκείμενο των δεδομένων θα πρέπει να ενημερώνεται αν καταρτίζεται το προφίλ του και ποιες συνέπειες έχει αυτό. Εάν τα δεδομένα προσωπικού χαρακτήρα παρέχονται από το

υποκείμενο των δεδομένων, το υποκείμενο των δεδομένων θα πρέπει να ενημερώνεται επίσης για το κατά πόσον υποχρεούται να παράσχει τα δεδομένα προσωπικού χαρακτήρα και για τις συνέπειες, όταν δεν παρέχει τα εν λόγω δεδομένα. Οι πληροφορίες αυτές μπορούν να παρέχονται σε συνδυασμό με τυποποιημένα εικονίδια προκειμένου να δίνεται με ευδιάκριτο, κατανοητό και ευανάγνωστο τρόπο μια ουσιαστικής επισκόπηση της σκοπούμενης επεξεργασίας. Εάν τα εικονίδια διατίθενται ηλεκτρονικά, θα πρέπει να είναι μηχανικώς αναγνώσιμα.

(61) Οι πληροφορίες σε σχέση με την επεξεργασία των δεδομένων προσωπικού χαρακτήρα που σχετίζονται με το υποκείμενο των δεδομένων θα πρέπει να του παρέχονται κατά τη συλλογή από το υποκείμενο των δεδομένων ή, εάν τα δεδομένα προσωπικού χαρακτήρα λαμβάνονται από άλλη πηγή, εντός εύλογης προθεσμίας, ανάλογα με τις συνθήκες κάθε περίπτωσης. Εάν τα δεδομένα προσωπικού χαρακτήρα επιτρέπεται να κοινοποιηθούν σε άλλον αποδέκτη, το υποκείμενο των δεδομένων θα πρέπει να ενημερώνεται, όταν τα δεδομένα προσωπικού χαρακτήρα κοινολογούνται για πρώτη φορά στον αποδέκτη. Όταν ο υπεύθυνος επεξεργασίας προτίθεται να επεξεργαστεί τα δεδομένα προσωπικού χαρακτήρα για σκοπό άλλο από εκείνον για τον οποίο συλλέχθηκαν, ο υπεύθυνος επεξεργασίας θα πρέπει να παρέχει στο υποκείμενο των δεδομένων, πριν από την εν λόγω περαιτέρω επεξεργασία, πληροφορίες για τον σκοπό αυτόν και άλλες αναγκαίες πληροφορίες. Όταν η προέλευση των δεδομένων προσωπικού χαρακτήρα δεν μπορεί να γνωστοποιηθεί στο υποκείμενο των δεδομένων διότι έχουν χρησιμοποιηθεί διάφορες πηγές, θα πρέπει να παρέχονται γενικές πληροφορίες.

(65) Ένα υποκείμενο των δεδομένων θα πρέπει να έχει το δικαίωμα να ζητεί τη διόρθωση των δεδομένων προσωπικού χαρακτήρα που το αφορούν, καθώς και το «δικαίωμα στη λήθη», εάν η διατήρηση των εν λόγω δεδομένων παραβιάζει τον παρόντα κανονισμό ή το δίκαιο της Ένωσης ή κράτους μέλους στο οποίο υποκαίτε ο υπεύθυνος επεξεργασίας. Ιδίως, το υποκείμενο των δεδομένων θα πρέπει να έχει το δικαίωμα να ζητεί τη διαγραφή και την παύση της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα που το αφορούν, εάν τα δεδομένα προσωπικού χαρακτήρα δεν είναι πλέον απαραίτητα σε σχέση με τους σκοπούς για τους οποίους συλλέγονται ή υποβάλλονται κατ' άλλο τρόπο σε επεξεργασία, εάν το υποκείμενο των δεδομένων αποσύρει τη συγκατάθεσή του για την επεξεργασία ή εάν αντιτάσσεται στην επεξεργασία δεδομένων προσωπικού χαρακτήρα που το αφορούν ή εάν η επεξεργασία των δεδομένων προσωπικού χαρακτήρα που το αφορούν δεν είναι σύμφωνη προς τον παρόντα κανονισμό κατ' άλλο τρόπο. Το δικαίωμα αυτό έχει ιδίως σημασία όταν το υποκείμενο των δεδομένων παρέσχε τη συγκατάθεσή του ως παιδί, όταν δεν είχε πλήρη επίγνωση των κινδύνων που ενέχει η επεξεργασία, και θέλει αργότερα να αφαιρέσει τα συγκεκριμένα δεδομένα προσωπικού χαρακτήρα, κυρίως από το Διαδίκτυο. Το υποκείμενο των δεδομένων θα πρέπει να μπορεί να ασκήσει το εν λόγω δικαίωμα παρά το γεγονός ότι δεν είναι πλέον παιδί. Ωστόσο, η περαιτέρω διατήρηση των δεδομένων προσωπικού χαρακτήρα θα πρέπει να είναι σύνομη όταν είναι αναγκαία για την άσκηση του δικαιώματος ελευθερίας της έκφρασης και ενημέρωσης, για τη συμμόρφωση με νομική υποχρέωση, για την εκπλήρωση καθήκοντος που εκτελείται προς το δημόσιο συμφέρον ή κατά την άσκηση δημόσιας εξουσίας που έχει ανατεθεί στον υπεύθυνο επεξεργασίας, για λόγους δημόσιου

συμφέροντος στον τομέα της δημόσιας υγείας, για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, για σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς, ή για τη θεμελίωση, άσκηση ή υποστήριξη νομικών αξιώσεων.

(66) Για να ενισχυθεί το δικαίωμα στη λήθη στο επιγραμμικό περιβάλλον, το δικαίωμα διαγραφής θα πρέπει επίσης να επεκταθεί με τέτοιο τρόπο ώστε ο υπεύθυνος επεξεργασίας ο οποίος δημοσιοποίησε τα δεδομένα προσωπικού χαρακτήρα να υποχρεούται να ενημερώνει τους υπευθύνους επεξεργασίας που επεξεργάζονται τα εν λόγω δεδομένα προσωπικού χαρακτήρα ώστε να διαγράψουν οποιουδήποτε συνδέσμους ή αντίγραφα ή αναπαραγωγή των εν λόγω δεδομένων προσωπικού χαρακτήρα. Όταν το πράττει, ο εν λόγω υπεύθυνος επεξεργασίας θα πρέπει να λαμβάνει εύλογα μέτρα, λαμβάνοντας υπόψη τη διαθέσιμη τεχνολογία και τα μέσα που έχει στη διάθεσή του ο υπεύθυνος επεξεργασίας, μεταξύ άλλων και τεχνικά μέτρα, ώστε να ενημερωθούν οι υπεύθυνοι επεξεργασίας που επεξεργάζονται τα δεδομένα προσωπικού χαρακτήρα σχετικά με το αίτημα του υποκαμένου των δεδομένων.

(70) Όταν δεδομένα προσωπικού χαρακτήρα υφίστανται επεξεργασία για σκοπούς της απευθείας εμπορικής προώθησης, το υποκείμενο των δεδομένων θα πρέπει να έχει το δικαίωμα να ανατεθεί στην εν λόγω επεξεργασία, συμπεριλαμβανομένης της κατάρτισης προφίλ στον βαθμό που αυτή συνδέεται με την εν λόγω απευθείας εμπορική προώθηση, είτε πρόκειται για αρχική είτε για περαιτέρω επεξεργασία, ανά πάσα στιγμή και χωρίς χρέωση. Το δικαίωμα αυτό θα πρέπει να περιέρχεται ρητά ως γνώσιν του υποκαμένου των δεδομένων και να παρουσιάζεται σαφώς και ξεχωριστά από κάθε άλλη πληροφορία.

(71) Το υποκείμενο των δεδομένων θα πρέπει να έχει το δικαίωμα να μην υπόκειται σε απόφαση, η οποία μπορεί να περιλαμβάνει κάποιο μέτρο, με την οποία αξιολογούνται προσωπικές πτυχές που το αφορούν, λαμβανόμενη αποκλειστικά βάσει αυτοματοποιημένης επεξεργασίας και η οποία παράγει έννομα αποτελέσματα έναντι του προσώπου αυτού ή το επηρεάζει σημαντικά κατά ανάλογο τρόπο, όπως η αυτόματη άρνηση επιγραμμικής αίτησης πίστωσης ή πρακτικές ηλεκτρονικών προσλήψεων χωρίς ανθρώπινη παρέμβαση. Η επεξεργασία αυτή περιλαμβάνει την «κατάρτιση προφίλ» που αποτελείται από οποιαδήποτε μορφή αυτοματοποιημένης επεξεργασίας δεδομένων προσωπικού χαρακτήρα για την αξιολόγηση προσωπικών πτυχών σχετικά με ένα φυσικό πρόσωπο, ιδίως την ανάλυση ή την πρόβλεψη πτυχών που αφορούν τις επίδοσης στην εργασία, την οικονομική κατάσταση, την υγεία, τις προσωπικές προτιμήσεις ή συμφέροντα, την αξιοπιστία ή τη συμπεριφορά, τη θέση ή κίνηση του υποκειμένου των δεδομένων, στον βαθμό που παράγει νομικά αποτελέσματα έναντι του προσώπου αυτού ή το επηρεάζει σημαντικά κατά ανάλογο τρόπο. Ωστόσο, η λήψη απόφασης που βασίζεται σε αυτήν την επεξεργασία, συμπεριλαμβανομένης της κατάρτισης προφίλ, θα πρέπει να επιτρέπεται όταν προβλέπεται ρητά από το δίκαιο της Ένωσης ή κράτους μέλους, στο οποίο υποκαίτε ο υπεύθυνος επεξεργασίας, μεταξύ άλλων για σκοπούς παρακολούθησης και πρόληψης της απάτης και της φοροδιαφυγής σύμφωνα με τους κανονισμούς, τα πρότυπα και τις συστήσεις των θεσμικών οργάνων της Ένωσης ή των εθνικών οργάνων εποπτείας και προκειμένου να διασφαλιστεί η ασφάλεια και η αξιοπιστία της υπηρεσίας που παρέχει ο υπεύθυνος επεξεργασίας, ή όταν είναι αναγκαία για τη σύναψη

ή την εκτέλεση σύμβασης μεταξύ υποκειμένου των δεδομένων και υπευθύνου επεξεργασίας ή όταν το υποκείμενο των δεδομένων παρέσχε τη ρητή συγκατάθεσή του. Σε κάθε περίπτωση, η επεξεργασία αυτή θα πρέπει να υπόκειται σε κατάλληλες εγγυήσεις, οι οποίες θα πρέπει να περιλαμβάνουν ειδική ενημέρωση του υποκειμένου των δεδομένων και το δικαίωμα εξασφάλισης ανθρώπινης παρέμβασης, το δικαίωμα διατύπωσης της άποψής του, το δικαίωμα να λάβει αιτιολόγηση της απόφασης που ελήφθη στο πλαίσιο της εν λόγω εκτίμησης και το δικαίωμα αμφισβήτησης της απόφασης. Το εν λόγω μέτρο θα πρέπει να μην αφορά παιδί. Προκειμένου να διασφαλισθεί δίκαιη και διαφανής επεξεργασία σε σχέση με το υποκείμενο των δεδομένων, λαμβανομένων υπόψη των ειδικών συνθηκών και του πλαισίου εντός του οποίου πραγματοποιείται η επεξεργασία των δεδομένων προσωπικού χαρακτήρα, ο υπεύθυνος επεξεργασίας θα πρέπει να χρησιμοποιεί κατάλληλες μαθηματικές ή στατιστικές διαδικασίες για την κατάρτιση του προφίλ, να εφαρμόζει τεχνικά και οργανωτικά μέτρα, ώστε να διορθώνονται οι παράγοντες που οδηγούν σε ανακρίβειες σε δεδομένα προσωπικού χαρακτήρα και να ελαχιστοποιείται ο κίνδυνος σφαλμάτων, να καθιστά ασφαλή τα δεδομένα προσωπικού χαρακτήρα κατά τρόπο που να λαμβάνει υπόψη τους πιθανούς κινδύνους που συνδέονται με τα συμφέροντα και τα δικαιώματα του υποκειμένου των δεδομένων και κατά τρόπο που να προλαμβάνει, μεταξύ άλλων, τα αποτελέσματα διακρίσεων σε βάρος φυσικών προσώπων βάση της φυλετικής ή εθνοτικής καταγωγής, των πολιτικών φρονιμάτων, της θρησκείας ή των πεποιθήσεων, της συμμετοχής σε συνδικαλιστικές οργανώσεις, της γενετικής κατάστασης ή της κατάστασης της υγείας ή του γενετήσιου προσανατολισμού, ή μέτρων ισοδύναμου αποτελέσματος. Η αυτοματοποιημένη λήψη αποφάσεων και κατάρτιση προφίλ που βασίζονται σε ειδικές κατηγορίες δεδομένων προσωπικού χαρακτήρα θα πρέπει να επιτρέπονται μόνο υπό ειδικές προϋποθέσεις.

(78) Η προστασία των δικαιωμάτων και των ελευθεριών των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα απαιτεί τη λήψη κατάλληλων τεχνικών και οργανωτικών μέτρων ώστε να διασφαλίζεται ότι τηρούνται οι απαιτήσεις του παρόντος κανονισμού. Προκειμένου να μπορεί να απόδειξη συμμόρφωση προς τον παρόντα κανονισμό, ο υπεύθυνος επεξεργασίας θα πρέπει να θεσπίζει εσωτερικές πολιτικές και να εφαρμόζει μέτρα τα οποία ανταποκρίνονται ειδικότερα στις αρχές της προστασίας των δεδομένων ήδη από τον σχεδιασμό και εξ ορισμού. Τέτοια μέτρα θα μπορούσαν να περιλαμβάνουν, μεταξύ άλλων, την ελαχιστοποίηση της επεξεργασίας δεδομένων προσωπικού χαρακτήρα, την ψευδωνυμοποίηση δεδομένων προσωπικού χαρακτήρα το συντομότερο δυνατόν, τη διαφάνεια όσον αφορά τις λειτουργίες και την επεξεργασία των δεδομένων προσωπικού χαρακτήρα, ώστε να μπορεί το υποκείμενο των δεδομένων να παρακολουθεί την επεξεργασία δεδομένων και να είναι σε θέση ο υπεύθυνος επεξεργασίας να δημιουργεί και να βελτιώνει τα χαρακτηριστικά ασφάλειας. Κατά την ανάπτυξη, τον σχεδιασμό, την επιλογή και τη χρήση εφαρμογών, υπηρεσιών και προϊόντων που βασίζονται στην επεξεργασία δεδομένων προσωπικού χαρακτήρα ή επεξεργάζονται δεδομένα προσωπικού χαρακτήρα για την εκπλήρωση του έργου τους, οι παραγωγοί προϊόντων, υπηρεσιών και εφαρμογών θα πρέπει να ενθαρρύνονται να λαμβάνουν υπόψη τους το δικαίωμα προστασίας των δεδομένων, κατά την ανάπτυξη και τον σχεδιασμό τέτοιων προϊόντων, υπηρεσιών και εφαρμογών, ώστε, λαμβανομένων υπόψη των τελευταίων εξελίξεων, να διασφαλίζεται ότι οι υπεύθυνοι επεξεργασίας και οι εκτελούντες την επεξεργασία θα είναι σε θέση να εκπληρώνουν τις υποχρέωσής τους όσον αφορά την



προστασία των δεδομένων. Οι αρχές της προστασίας των δεδομένων ήδη από τον σχεδιασμό και εξ ορισμού θα πρέπει επίσης να λαμβάνονται υπόψη στο πλαίσιο των δημόσιων διαγωνισμών.

(110) Ένας όμιλος επιχειρήσεων, όπως επίσης και ένας όμιλος εταιρειών που ασκούν κοινή οικονομική δραστηριότητα, θα πρέπει να μπορεί να κάνει χρήση εγκεκριμένων δεσμευτικών εταιρικών κανόνων για τις διεθνείς διαβιβάσεις του από την Ένωση σε οργανισμούς εντός του ίδιου ομίλου επιχειρήσεων ή ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα, εφόσον οι εν λόγω εταιρικοί κανόνες περιλαμβάνουν όλες τις βασικές αρχές και δικαιώματα τα οποία τυγχάνουν δικαστικής προστασίας, ώστε να διασφαλίζονται κατάλληλες εγγυήσεις για διαβιβάσεις ή κατηγορίες διαβιβάσεων δεδομένων προσωπικού χαρακτήρα.

(156) Η επεξεργασία δεδομένων προσωπικού χαρακτήρα για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς θα πρέπει να υπόκειται σε κατάλληλες εγγυήσεις για τα δικαιώματα και τις ελευθερίες του υποκειμένου των δεδομένων σύμφωνα με τον παρόντα κανονισμό. Οι εν λόγω εγγυήσεις θα πρέπει να διασφαλίζουν ότι έχουν θεσπιστεί τα τεχνικά και οργανωτικά μέτρα που εγγυώνται, ειδικότερα, την αρχή της ελαχιστοποίησης των δεδομένων. Η περαιτέρω επεξεργασία δεδομένων προσωπικού χαρακτήρα για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς πραγματοποιείται όταν ο υπεύθυνος της επεξεργασίας έχει εκτιμήσει κατά πόσο είναι εφικτό να εκπληρωθούν οι σκοποί αυτοί μέσω της επεξεργασίας δεδομένων τα οποία δεν επιτρέπουν ή δεν επιτρέπουν πλέον την ταυτοποίηση των υποκειμένων των δεδομένων, υπό την προϋπόθεση ότι υπάρχουν κατάλληλες εγγυήσεις (όπως, για παράδειγμα, η ψευδωνυμοποίηση των δεδομένων). Τα κράτη μέλη θα πρέπει να προβλέπουν κατάλληλες διασφαλίσεις σχετικά με την επεξεργασία δεδομένων προσωπικού χαρακτήρα για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς. Θα πρέπει να επιτρέπεται στα κράτη μέλη να παρέχουν, υπό συγκεκριμένες προϋποθέσεις και με δέουσες εγγυήσεις για τα υποκείμενα των δεδομένων, προδιαγραφές και παρέκκλισης όσον αφορά τις απαιτήσεις πληροφόρησης και τα δικαιώματα διόρθωσης και διαγραφής, το δικαίωμα στη λήθη, το δικαίωμα περιορισμού της επεξεργασίας, το δικαίωμα στη φορητότητα των δεδομένων και το δικαίωμα ανάταξης κατά την επεξεργασία δεδομένων προσωπικού χαρακτήρα για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς. Οι εν λόγω προϋποθέσεις και εγγυήσεις ενδέχεται να συνεπάγονται άδικές διαδικασίες, ώστε τα υποκείμενα των δεδομένων να ασκούν τα δικαιώματα αυτά, εφόσον είναι σκόπιμο για τους σκοπούς που επιδιώκονται με τη συγκεκριμένη επεξεργασία, παράλληλα με τεχνικά και οργανωτικά μέτρα που αποσκοπούν στην ελαχιστοποίηση της επεξεργασίας δεδομένων προσωπικού χαρακτήρα σύμφωνα με τις αρχές της αναλογικότητας και της αναγκαιότητας. Η επεξεργασία δεδομένων προσωπικού χαρακτήρα για επιστημονικούς σκοπούς θα πρέπει να συμμορφώνεται επίσης με άλλες σχετικές νομοθεσίες, όπως αυτή για τις κλινικές δοκιμές.

(159) Όταν δεδομένα προσωπικού χαρακτήρα υφίστανται επεξεργασία για σκοπούς επιστημονικής έρευνας, ο παρών κανονισμός θα πρέπει να ισχύει και για την επεξεργασία αυτή. Για τους σκοπούς του παρόντος κανονισμού, η επεξεργασία δεδομένων προσωπικού χαρακτήρα για σκοπούς επιστημονικής έρευνας θα πρέπει να ερμηνεύεται διασταλτικά, δηλαδή να περιλαμβάνει παραδείγματος χάριν τεχνολογική ανάπτυξη και επίδειξη, βασική έρευνα, εφαρμοσμένη έρευνα και ιδιωτικά χρηματοδοτούμενη έρευνα. Επιπλέον, θα πρέπει να λαμβάνει υπόψη τον στόχο της Ένωσης δυνάμει του άρθρου 179 παράγραφος 1 ΣΛΕΕ για την επίτευξη ενός ευρωπαϊκού χώρου έρευνας. Στους σκοπούς επιστημονικής έρευνας θα πρέπει να περιλαμβάνονται και μελέτες που πραγματοποιούνται για το δημόσιο συμφέρον στον τομέα της δημόσιας υγείας. Για να ληφθούν υπόψη οι ιδιαιτερότητες της επεξεργασίας δεδομένων προσωπικού χαρακτήρα για σκοπούς επιστημονικής έρευνας, θα πρέπει να ισχύουν ειδικοί όροι ιδίως όσον αφορά τη δημοσίευση ή με άλλο τρόπο δημοσιοποίηση δεδομένων προσωπικού χαρακτήρα στο πλαίσιο των σκοπών επιστημονικής έρευνας. Εάν το αποτέλεσμα της επιστημονικής έρευνας ειδικότερα στον τομέα της υγείας αιτιολογεί τη λήψη περαιτέρω μέτρων προς το συμφέρον του υποκειμένου των δεδομένων, ισχύουν οι γενικοί κανόνες του παρόντος κανονισμού όσον αφορά τα μέτρα αυτά.

(160) Όταν δεδομένα προσωπικού χαρακτήρα υφίστανται επεξεργασία για σκοπούς ιστορικής έρευνας, ο παρών κανονισμός θα πρέπει επίσης να ισχύει και για την επεξεργασία αυτή. Συμπεριλαμβάνονται εν προκειμένω η ιστορική έρευνα και η έρευνα για γενεαλογικούς σκοπούς, έχοντας κατά νου ότι ο παρών κανονισμός δεν θα πρέπει να ισχύει για τους θανόντες.

(162) Όταν δεδομένα προσωπικού χαρακτήρα υφίστανται επεξεργασία για στατιστικούς σκοπούς, ο παρών κανονισμός θα πρέπει να ισχύει και για την επεξεργασία αυτή. Το ενωσιακό δίκαιο ή το δίκαιο των κρατών μελών θα πρέπει, εντός των ορίων του παρόντος κανονισμού, να καθορίζει το στατιστικό περιεχόμενο, τον έλεγχο της πρόσβασης, τις προδιαγραφές για την επεξεργασία δεδομένων προσωπικού χαρακτήρα για στατιστικούς σκοπούς και κατάλληλα μέτρα που διασφαλίζουν τα δικαιώματα και τις ελευθερίες του υποκειμένου των δεδομένων και που αποσκοπούν στην εξασφάλιση του στατιστικού απορρήτου. Ο όρος «στατιστικοί σκοποί» σημαίνει κάθε πράξη συλλογής και την επεξεργασία δεδομένων προσωπικού χαρακτήρα που είναι αναγκαία για την πραγματοποίηση στατιστικών ερευνών ή για την παραγωγή στατιστικών αποτελεσμάτων. Τα εν λόγω στατιστικά αποτελέσματα μπορούν να χρησιμοποιηθούν περαιτέρω για διάφορους σκοπούς, μεταξύ άλλων και για σκοπούς επιστημονικής έρευνας. Ο στατιστικός σκοπός συνεπάγεται ότι το αποτέλεσμα της επεξεργασίας για στατιστικούς σκοπούς δεν είναι δεδομένα προσωπικού χαρακτήρα αλλά συγκεντρωτικά δεδομένα και ότι το αποτέλεσμα αυτό ή τα δεδομένα προσωπικού χαρακτήρα δεν χρησιμοποιούνται προς υποστήριξη μέτρων ή αποφάσεων που αφορούν συγκεκριμένο φυσικό πρόσωπο.

(163) Θα πρέπει να προστατεύονται οι εμπιστευτικές πληροφορίες που συλλέγουν οι ενωσιακές και εθνικές στατιστικές υπηρεσίες για την κατάρτιση επίσημων ευρωπαϊκών και εθνικών στατιστικών. Οι ευρωπαϊκές στατιστικές θα πρέπει να αναπτύσσονται, να

καταρτίζονται και να διαδίδονται σύμφωνα με τις στατιστικές αρχές που θεσπίζονται στο άρθρο 338 παράγραφος 2 ΣΛΕΕ, ενώ οι εθνικές στατιστικές θα πρέπει επίσης να συμμορφώνονται με το δίκαιο των κρατών μελών. Ο κανονισμός (ΕΚ) αριθ. 223/2009 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου (2) παρέχει περαιτέρω διευκρινίσεις περί του στατιστικού απορρήτου για τις ευρωπαϊκές στατιστικές.

## 4.1 ΑΡΘΡΑ

### Άρθρο 4

#### Ορισμοί

Για τους σκοπούς του παρόντος κανονισμού νοούνται ως:

- 1) «δεδομένα προσωπικού χαρακτήρα»: κάθε πληροφορία που αφορά ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο («υποκείμενο των δεδομένων»): το ταυτοποιήσιμο φυσικό πρόσωπο είναι εκείνο του οποίου η ταυτότητα μπορεί να εξακριβωθεί, άμεσα ή έμμεσα, ιδίως μέσω αναφοράς σε αναγνωριστικό στοιχείο ταυτότητας, όπως όνομα, σε αριθμό ταυτότητας, σε δεδομένα θέσης, σε επιγραμμικό αναγνωριστικό ταυτότητας ή σε έναν ή περισσότερους παράγοντες που προσιδιάζουν στη σωματική, φυσιολογική, γενετική, ψυχολογική, οικονομική, πολιτιστική ή κοινωνική ταυτότητα του εν λόγω φυσικού προσώπου,
- 2) «επεξεργασία»: κάθε πράξη ή σειρά πράξεων που πραγματοποιείται με ή χωρίς τη χρήση αυτοματοποιημένων μέσων, σε δεδομένα προσωπικού χαρακτήρα ή σε σύνολα δεδομένων προσωπικού χαρακτήρα, όπως η συλλογή, η καταχώριση, η οργάνωση, η διάρθρωση, η αποθήκευση, η προσαρμογή ή η μεταβολή, η ανάκτηση, η αναζήτηση πληροφοριών, η χρήση, η κοινολόγηση με διαβίβαση, η διάδοση ή κάθε άλλη μορφή διάθεσης, η συσχέτιση ή ο συνδυασμός, ο περιορισμός, η διαγραφή ή η καταστροφή,
- 3) «περιορισμός της επεξεργασίας»: η επισήμανση αποθηκευμένων δεδομένων προσωπικού χαρακτήρα με στόχο τον περιορισμό της επεξεργασίας τους στο μέλλον,
- 4) «κατάρτιση προφίλ»: οποιαδήποτε μορφή αυτοματοποιημένης επεξεργασίας δεδομένων προσωπικού χαρακτήρα που συνίσταται στη χρήση δεδομένων προσωπικού χαρακτήρα για την αξιολόγηση ορισμένων προσωπικών πτυχών ενός φυσικού προσώπου, ιδίως για την ανάλυση ή την πρόβλεψη πτυχών που αφορούν την απόδοση στην εργασία, την οικονομική κατάσταση, την υγεία, τις προσωπικές προτιμήσεις, τα ενδιαφέροντα, την αξιοπιστία, τη συμπεριφορά, τη θέση ή τις μετακινήσεις του εν λόγω φυσικού προσώπου,
- 5) «ψευδωνυμοποίηση»: η επεξεργασία δεδομένων προσωπικού χαρακτήρα κατά τρόπο ώστε τα δεδομένα να μην μπορούν πλέον να αποδοθούν σε συγκεκριμένο υποκείμενο των δεδομένων χωρίς τη χρήση συμπληρωματικών πληροφοριών, εφόσον οι εν λόγω συμπληρωματικές πληροφορίες διατηρούνται χωριστά και υπόκεινται σε τεχνικά και

οργανωτικά μέτρα προκειμένου να διασφαλιστεί ότι δεν μπορούν να αποδοθούν σε ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο,

6) «σύστημα αρχειοθέτησης»: κάθε διαρθρωμένο σύνολο δεδομένων προσωπικού χαρακτήρα τα οποία είναι προσβάσιμα με γνώμονα συγκεκριμένα κριτήρια, είτε το σύνολο αυτό είναι συγκεντρωμένο είτε αποκεντρωμένο είτε κατανεμημένο σε λειτουργική ή γεωγραφική βάση,

7) «υπεύθυνος επεξεργασίας»: το φυσικό ή νομικό πρόσωπο, η δημόσια αρχή, η υπηρεσία ή άλλος φορέας που, μόνα ή από κοινού με άλλα, καθορίζουν τους σκοπούς και τον τρόπο της επεξεργασίας δεδομένων προσωπικού χαρακτήρα· όταν οι σκοποί και ο τρόπος της επεξεργασίας αυτής καθορίζονται από το δίκαιο της Ένωσης ή το δίκαιο κράτους μέλους, ο υπεύθυνος επεξεργασίας ή τα άδικα κριτήρια για τον διορισμό του μπορούν να προβλέπονται από το δίκαιο της Ένωσης ή το δίκαιο κράτους μέλους,

8) «εκτελών την επεξεργασία»: το φυσικό ή νομικό πρόσωπο, η δημόσια αρχή, η υπηρεσία ή άλλος φορέας που επεξεργάζεται δεδομένα προσωπικού χαρακτήρα για λογαριασμό του υπευθύνου της επεξεργασίας,

9) «αποδέκτης»: το φυσικό ή νομικό πρόσωπο, η δημόσια αρχή, η υπηρεσία ή άλλος φορέας, στα οποία κοινολογούνται τα δεδομένα προσωπικού χαρακτήρα, είτε πρόκειται για τρίτον είτε όχι. Ωστόσο, οι δημόσιες αρχές που ενδέχεται να λάβουν δεδομένα προσωπικού χαρακτήρα στο πλαίσιο συγκεκριμένης έρευνας σύμφωνα με το δίκαιο της Ένωσης ή κράτους μέλους δεν θεωρούνται ως αποδέκτες· η επεξεργασία των δεδομένων αυτών από τις εν λόγω δημόσιες αρχές πραγματοποιείται σύμφωνα με τους ισχύσαντες κανόνες προστασίας των δεδομένων ανάλογα με τους σκοπούς της επεξεργασίας,

10) «τρίτος»: οποιοδήποτε φυσικό ή νομικό πρόσωπο, δημόσια αρχή, υπηρεσία ή φορέας, με εξαίρεση το υποκείμενο των δεδομένων, τον υπεύθυνο επεξεργασίας, τον εκτελούντα την επεξεργασία και τα πρόσωπα τα οποία, υπό την άμεση εποπτεία του υπευθύνου επεξεργασίας ή του εκτελούντος την επεξεργασία, είναι εξουσιοδοτημένα να επεξεργάζονται τα δεδομένα προσωπικού χαρακτήρα,

11) «συγκατάθεση» του υποκειμένου των δεδομένων: κάθε ένδειξη βουλήσεως, ελεύθερη, συγκεκριμένη, ρητή και εν πλήρη επίγνωση, με την οποία το υποκείμενο των δεδομένων εκδηλώνει ότι συμφωνεί, με δήλωση ή με σαφή θετική ενέργεια, να αποτελέσουν αντικείμενο επεξεργασίας τα δεδομένα προσωπικού χαρακτήρα που το αφορούν,

12) «παραβίαση δεδομένων προσωπικού χαρακτήρα»: η παραβίαση της ασφάλειας που οδηγεί σε τυχαία ή παράνομη καταστροφή, απώλεια, μεταβολή, άνευ άδειας κοινολόγηση ή πρόσβαση δεδομένων προσωπικού χαρακτήρα που διαβιβάστηκαν, αποθηκεύτηκαν ή υποβλήθηκαν κατ' άλλο τρόπο σε επεξεργασία,

13) «γενετικά δεδομένα»: τα δεδομένα προσωπικού χαρακτήρα που αφορούν τα γενετικά χαρακτηριστικά φυσικού προσώπου που κληρονομήθηκαν ή αποκτήθηκαν, όπως προκύπτουν, ιδίως, από ανάλυση βιολογικού δείγματος του εν λόγω φυσικού προσώπου και τα οποία παρέχουν μοναδικές πληροφορίες σχετικά με την φυσιολογία ή την υγεία του εν λόγω φυσικού προσώπου,

14) «βιομετρικά δεδομένα»: δεδομένα προσωπικού χαρακτήρα τα οποία προκύπτουν από ειδική τεχνική επεξεργασία συνδεδεμένη με φυσικά, βιολογικά ή συμπεριφορικά

χαρακτηριστικά φυσικού προσώπου και τα οποία επιτρέπουν ή επιβεβαιώνουν την αδιαμφισβήτητη ταυτοποίηση του εν λόγω φυσικού προσώπου, όπως εικόνες προσώπου ή δακτυλοσκοπικά δεδομένα,

15) «δεδομένα που αφορούν την υγεία»: δεδομένα προσωπικού χαρακτήρα τα οποία σχετίζονται με τη σωματική ή ψυχική υγεία ενός φυσικού προσώπου, περιλαμβανόμενης της παροχής υπηρεσιών υγειονομικής φροντίδας, και τα οποία αποκαλύπτουν πληροφορίες σχετικά με την κατάσταση της υγείας του,

## Άρθρο 5

Αρχές που διέπουν την επεξεργασία δεδομένων προσωπικού χαρακτήρα

1. Τα δεδομένα προσωπικού χαρακτήρα:

α) υποβάλλονται σε σύννομη και θεμιτή επεξεργασία με διαφανή τρόπο σε σχέση με το υποκείμενο των δεδομένων («νομιμότητα, αντικειμενικότητα και διαφάνεια»),

β) συλλέγονται για καθορισμένους, ρητούς και νόμιμους σκοπούς και δεν υποβάλλονται σε περαιτέρω επεξεργασία κατά τρόπο ασύμβατο προς τους σκοπούς αυτούς· η περαιτέρω επεξεργασία για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον ή σκοπούς επιστημονικής ή ιστορικής έρευνας ή στατιστικούς σκοπούς δεν θεωρείται ασύμβατη με τους αρχικούς σκοπούς σύμφωνα με το άρθρο 89 παράγραφος 1 («περιορισμός του σκοπού»),

ε) διατηρούνται υπό μορφή που επιτρέπει την ταυτοποίηση των υποκειμένων των δεδομένων μόνο για το διάστημα που απαιτείται για τους σκοπούς της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα· τα δεδομένα προσωπικού χαρακτήρα μπορούν να αποθηκεύονται για μεγαλύτερα διαστήματα, εφόσον τα δεδομένα προσωπικού χαρακτήρα θα υποβάλλονται σε επεξεργασία μόνο για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, για σκοπούς επιστημονικής ή ιστορικής έρευνας ή για στατιστικούς σκοπούς, σύμφωνα με το άρθρο 89 παράγραφος 1 και εφόσον εφαρμόζονται τα κατάλληλα τεχνικά και οργανωτικά μέτρα που απαιτεί ο παρών κανονισμός για τη διασφάλιση των δικαιωμάτων και ελευθεριών του υποκειμένου των δεδομένων («περιορισμός της περιόδου αποθήκευσης»).

## Άρθρο 6

Νομιμότητα της επεξεργασίας

1. Η επεξεργασία είναι σύννομη μόνο εάν και εφόσον ισχύει τουλάχιστον μία από τις ακόλουθες προϋποθέσεις:

α) το υποκείμενο των δεδομένων έχει συναινέσει στην επεξεργασία των δεδομένων προσωπικού χαρακτήρα του για έναν ή περισσότερους συγκεκριμένους σκοπούς,

β) η επεξεργασία είναι απαραίτητη για την εκτέλεση σύμβασης της οποίας το υποκείμενο των δεδομένων είναι συμβαλλόμενο μέρος ή για να ληφθούν μέτρα κατ' αίτηση του υποκειμένου των δεδομένων πριν από τη σύναψη σύμβασης,

γ) η επεξεργασία είναι απαραίτητη για τη συμμόρφωση με έννομη υποχρέωση του υπευθύνου επεξεργασίας,

δ) η επεξεργασία είναι απαραίτητη για τη διαφύλαξη ζωτικού συμφέροντος του υποκαμένου των δεδομένων ή άλλου φυσικού προσώπου,

ε) η επεξεργασία είναι απαραίτητη για την εκπλήρωση καθήκοντος που εκτελείται προς το δημόσιο συμφέρον ή κατά την άσκηση δημόσιας εξουσίας που έχει ανατεθεί στον υπεύθυνο επεξεργασίας,

στ) η επεξεργασία είναι απαραίτητη για τους σκοπούς των έννομων συμφερόντων που επιδιώκει ο υπεύθυνος επεξεργασίας ή τρίτος, εκτός εάν έναντι των συμφερόντων αυτών υπερισχύει το συμφέρον ή τα θεμελιώδη δικαιώματα και οι ελευθερίες του υποκειμένου των δεδομένων που επιβάλλουν την προστασία των δεδομένων προσωπικού χαρακτήρα, ιδίως εάν το υποκείμενο των δεδομένων είναι παιδί

## Άρθρο 9

### Επεξεργασία ειδικών κατηγοριών δεδομένων προσωπικού χαρακτήρα

1. Απαγορεύεται η επεξεργασία δεδομένων προσωπικού χαρακτήρα που αποκαλύπτουν τη φυλετική ή εθνοτική καταγωγή, τα πολιτικά φρονήματα, τις θρησκευτικές ή φιλοσοφικές πεποιθήσεις ή τη συμμετοχή σε συνδικαλιστική οργάνωση, καθώς και η επεξεργασία γενετικών δεδομένων, βιομετρικών δεδομένων με σκοπό την αδιαμφισβήτητη ταυτοποίηση προσώπου, δεδομένων που αφορούν την υγεία ή δεδομένων που αφορούν τη σεξουαλική ζωή φυσικού προσώπου ή τον γενετήσιο προανατολισμό.

## Άρθρο 13

### Ενημέρωση και πρόσβαση σε δεδομένα προσωπικού χαρακτήρα

Πληροφορίες που παρέχονται εάν τα δεδομένα προσωπικού χαρακτήρα συλλέγονται από το υποκείμενο των δεδομένων

1. Όταν δεδομένα προσωπικού χαρακτήρα που αφορούν υποκείμενο των δεδομένων συλλέγονται από το υποκείμενο των δεδομένων, ο υπεύθυνος επεξεργασίας, κατά τη λήψη των δεδομένων προσωπικού χαρακτήρα, παρέχει στο υποκείμενο των δεδομένων όλες τις ακόλουθες πληροφορίες:

α) την ταυτότητα και τα στοιχεία επικοινωνίας του υπευθύνου επεξεργασίας και, κατά περίπτωση, του εκπροσώπου του υπευθύνου επεξεργασίας,

β) τα στοιχεία επικοινωνίας του υπευθύνου προστασίας δεδομένων, κατά περίπτωση,

γ) τους σκοπούς της επεξεργασίας για τους οποίους προορίζονται τα δεδομένα προσωπικού χαρακτήρα, καθώς και τη νομική βάση για την επεξεργασία,

δ) εάν η επεξεργασία βασίζεται στο άρθρο 6 παράγραφος 1 στοιχείο στ), τα έννομα συμφέροντα που επιδιώκονται από τον υπεύθυνο επεξεργασίας ή από τρίτο,

ε) τους αποδέκτες ή τις κατηγορίες αποδεκτών των δεδομένων προσωπικού χαρακτήρα, εάν υπάρχουν,

στ) κατά περίπτωση, την πρόθεση του υπευθύνου επεξεργασίας να διαβιβάσει δεδομένα προσωπικού χαρακτήρα σε τρίτη χώρα ή διεθνή οργανισμό και την ύπαρξη ή την απουσία απόφασης επάρκειας της Επιτροπής ή, όταν πρόκειται για τις διαβιβάσεις που αναφέρονται στο άρθρο 46 ή 47 ή στο άρθρο 49 παράγραφος 1 δεύτερο εδάφιο, αναφορά στις ενδεδειγμένες ή κατάλληλες εγγυήσεις και τα μέσα για να αποκτηθεί αντίγραφό τους ή στο πού διατέθηκαν.

2. Εκτός από τις πληροφορίες που αναφέρονται στην παράγραφο 1, ο υπεύθυνος επεξεργασίας, κατά τη λήψη των δεδομένων προσωπικού χαρακτήρα, παρέχει στο υποκείμενο των δεδομένων τις εξής επιπλέον πληροφορίες που είναι αναγκαίες για την εξασφάλιση θεμιτής και διαφανούς επεξεργασίας:

α) το χρονικό διάστημα για το οποίο θα αποθηκευτούν τα δεδομένα προσωπικού χαρακτήρα ή, όταν αυτό είναι αδύνατο, τα κριτήρια που καθορίζουν το εν λόγω διάστημα,

β) την ύπαρξη δικαιώματος υποβολής αιτήματος στον υπεύθυνο επεξεργασίας για πρόσβαση και διόρθωση ή διαγραφή των δεδομένων προσωπικού χαρακτήρα ή περιορισμό της επεξεργασίας που αφορούν το υποκείμενο των δεδομένων ή δικαιώματος ανάταξης στην επεξεργασία, καθώς και δικαιώματος στη φορητότητα των δεδομένων,

γ) όταν η επεξεργασία βασίζεται στο άρθρο 6 παράγραφος 1 στοιχείο α) ή στο άρθρο 9 παράγραφος 2 στοιχείο α), την ύπαρξη του δικαιώματος να ανακαλέσει τη συγκατάθεσή του οποτεδήποτε, χωρίς να θίγει η νομιμότητα της επεξεργασίας που βασίστηκε στη συγκατάθεση πριν από την ανάκλησή της,

δ) το δικαίωμα υποβολής καταγγελίας σε εποπτική αρχή,

ε) κατά πόσο η παροχή δεδομένων προσωπικού χαρακτήρα αποτελεί νομική ή συμβατική υποχρέωση ή απαίτηση για τη σύναψη σύμβασης, καθώς και κατά πόσο το υποκείμενο των δεδομένων υποχρεούται να παρέχει τα δεδομένα προσωπικού χαρακτήρα και ποιες ενδεχόμενες συνέπειες θα είχε η μη παροχή των δεδομένων αυτών,

στ) την ύπαρξη αυτοματοποιημένης λήψης αποφάσεων, συμπεριλαμβανομένης της κατάρτισης προφίλ, που αναφέρεται στο άρθρο 22 παράγραφοι 1 και 4 και, τουλάχιστον στις περιπτώσεις αυτές, σημαντικές πληροφορίες σχετικά με τη λογική που ακολουθείται, καθώς και τη σημασία και τις προβλεπόμενες συνέπειες της εν λόγω επεξεργασίας για το υποκείμενο των δεδομένων.

3. Όταν ο υπεύθυνος επεξεργασίας προτίθεται να επεξεργαστεί περαιτέρω τα δεδομένα προσωπικού χαρακτήρα για άλλο σκοπό από εκείνο για τον οποίο τα δεδομένα προσωπικού χαρακτήρα συλλέχθηκαν, ο υπεύθυνος επεξεργασίας παρέχει στο υποκείμενο των δεδομένων, πριν από την εν λόγω περαιτέρω επεξεργασία, πληροφορίες για τον σκοπό αυτόν και άλλες τυχόν αναγκαίες πληροφορίες, όπως αναφέρεται στην παράγραφο 2.

4. Οι παράγραφοι 1, 2 και 3 δεν εφαρμόζονται, όταν και εφόσον το υποκείμενο των δεδομένων έχει ήδη τις πληροφορίες.

## Άρθρο 17

### Δικαίωμα διαγραφής («δικαίωμα στη λήθη»)

1. Το υποκείμενο των δεδομένων έχει το δικαίωμα να ζητήσει από τον υπεύθυνο επεξεργασίας τη διαγραφή δεδομένων προσωπικού χαρακτήρα που το αφορούν χωρίς αδικαιολόγητη καθυστέρηση και ο υπεύθυνος επεξεργασίας υποχρεούται να διαγράψει δεδομένα προσωπικού χαρακτήρα χωρίς αδικαιολόγητη καθυστέρηση, εάν ισχύει ένας από τους ακόλουθους λόγους:

α) τα δεδομένα προσωπικού χαρακτήρα δεν είναι πλέον απαραίτητα σε σχέση με τους σκοπούς για τους οποίους συλλέχθηκαν ή υποβλήθηκαν κατ' άλλο τρόπο σε επεξεργασία,

β) το υποκείμενο των δεδομένων ανακαλεί τη συγκατάθεση επί της οποίας βασίζεται η επεξεργασία σύμφωνα με το άρθρο 6 παράγραφος 1 στοιχείο α) ή το άρθρο 9 παράγραφος 2 στοιχείο α) και δεν υπάρχει άλλη νομική βάση για την επεξεργασία,

γ) το υποκείμενο των δεδομένων ανατίθεται στην επεξεργασία σύμφωνα με το άρθρο 21 παράγραφος 1 και δεν υπάρχουν επιτακτικοί και νόμιμοι λόγοι για την επεξεργασία ή το υποκείμενο των δεδομένων ανατίθεται στην επεξεργασία σύμφωνα με το άρθρο 21 παράγραφος 2,

δ) τα δεδομένα προσωπικού χαρακτήρα υποβλήθηκαν σε επεξεργασία παράνομα,

ε) τα δεδομένα προσωπικού χαρακτήρα πρέπει να διαγραφούν, ώστε να τηρηθεί νομική υποχρέωση βάσει του ενωσιακού δικαίου ή του δικαίου κράτους μέλους, στην οποία υπόκειται ο υπεύθυνος επεξεργασίας,

στ) τα δεδομένα προσωπικού χαρακτήρα έχουν συλλεχθεί σε σχέση με την προσφορά υπηρεσιών της κοινωνίας των πληροφοριών που αναφέρονται στο άρθρο 8 παράγραφος 1.

2. Όταν ο υπεύθυνος επεξεργασίας έχει δημοσιοποιήσει τα δεδομένα προσωπικού χαρακτήρα και υποχρεούται σύμφωνα με την παράγραφο 1 να διαγράψει τα δεδομένα προσωπικού χαρακτήρα, ο υπεύθυνος επεξεργασίας, λαμβάνοντας υπόψη τη διαθέσιμη τεχνολογία και το κόστος εφαρμογής, λαμβάνει εύλογα μέτρα, συμπεριλαμβανομένων των τεχνικών μέτρων, για να ενημερώσει τους υπευθύνους επεξεργασίας που επεξεργάζονται τα δεδομένα προσωπικού χαρακτήρα, ότι το υποκείμενο των δεδομένων



ζήτησε τη διαγραφή από αυτούς τους υπευθύνους επεξεργασίας τυχόν συνδέσμων με τα δεδομένα αυτά ή αντιγράφων ή αναπαραγωγών των εν λόγω δεδομένων προσωπικού χαρακτήρα.

3. Οι παράγραφοι 1 και 2 δεν εφαρμόζονται στον βαθμό που η επεξεργασία είναι απαραίτητη:

α) για την άσκηση του δικαιώματος ελευθερίας της έκφρασης και του δικαιώματος στην ενημέρωση,

β) για την τήρηση νομικής υποχρέωσης που επιβάλλει την επεξεργασία βάση του δικαιού της Ένωσης ή του δικαιού κράτους μέλους στο οποίο υπάγεται ο υπεύθυνος επεξεργασίας ή για την εκπλήρωση καθήκοντος που εκτελείται προς το δημόσιο συμφέρον ή κατά την άσκηση δημόσιας εξουσίας που έχει ανατεθεί στον υπεύθυνο της επεξεργασίας,

γ) για λόγους δημόσιου συμφέροντος στον τομέα της δημόσιας υγείας σύμφωνα με το άρθρο 9 παράγραφος 2 στοιχεία η) και 9), καθώς και το άρθρο 9 παράγραφος 3,

δ) για σκοπούς αρχειοθέτησης προς το δημόσιο συμφέρον, για σκοπούς επιστημονικής ή ιστορικής έρευνας ή για στατιστικούς σκοπούς σύμφωνα με το άρθρο 89 παράγραφος 1, εφόσον το δικαίωμα που αναφέρεται στην παράγραφο 1 είναι πιθανόν να καταστήσει αδύνατη ή να εμπόδισα σε μεγάλο βαθμό την επίτευξη σκοπών της εν λόγω επεξεργασίας, ή

ε) για τη θεμελίωση, άσκηση ή υποστήριξη νομικών αξιώσεων.

#### Άρθρο 18

##### Δικαίωμα περιορισμού της επεξεργασίας

β) η επεξεργασία είναι παράνομη και το υποκείμενο των δεδομένων αντιτάσσεται στη διαγραφή των δεδομένων προσωπικού χαρακτήρα και ζητεί, αντ' αυτής, τον περιορισμό της χρήσης τους

#### Άρθρο 21

##### Δικαίωμα εναντίωσης

2. Εάν δεδομένα προσωπικού χαρακτήρα υποβάλλονται σε επεξεργασία για σκοπούς απευθείας εμπορικής προώθησης, το υποκείμενο των δεδομένων δικαιούται να αναταχθεί ανά πάσα στιγμή στην επεξεργασία των δεδομένων προσωπικού χαρακτήρα που το αφορούν για την εν λόγω εμπορική προώθηση, περιλαμβανομένης της κατάρτισης προφίλ, εάν σχετίζεται με αυτήν την απευθείας εμπορική προώθηση.

3. Στις περιπτώσεις που αναφέρονται στην παράγραφο 2 στοιχεία α) και γ), ο υπεύθυνος επεξεργασίας των δεδομένων εφαρμόζει κατάλληλα μέτρα για την προστασία των δικαιωμάτων, των ελευθεριών και των έννομων συμφερόντων του υποκαμένου των δεδομένων, τουλάχιστον του δικαιώματος εξασφάλισης ανθρώπινης παρέμβασης από την πλευρά του υπευθύνου επεξεργασίας, έκφρασης άποψης και αμφισβήτησης της απόφασης.

#### Άρθρο 34

Ανακοίνωση παραβίασης δεδομένων προσωπικού χαρακτήρα στο υποκείμενο των δεδομένων

1. Όταν η παραβίαση δεδομένων προσωπικού χαρακτήρα ενδέχεται να θέσει σε υψηλό κίνδυνο τα δικαιώματα και τις ελευθερίες των φυσικών προσώπων, ο υπεύθυνος επεξεργασίας ανακοινώνει αμελλητί την παραβίαση των δεδομένων προσωπικού χαρακτήρα στο υποκείμενο των δεδομένων.

#### Άρθρο 44

Γενικές αρχές για διαβιβάσεις

Κάθε διαβίβαση δεδομένων προσωπικού χαρακτήρα τα οποία υποβάλλονται σε επεξεργασία ή προορίζονται να υποβληθούν σε επεξεργασία μετά από τη διαβίβασή τους σε τρίτη χώρα ή διεθνή οργανισμό πραγματοποιείται μόνο εάν, με την επιφύλαξη των λοιπών διατάξεων του παρόντος κανονισμού, οι προϋποθέσεις που θεσπίζονται στο παρόν κεφάλαιο τηρούνται από τον υπεύθυνο επεξεργασίας και τον εκτελούντα την επεξεργασία, μεταξύ άλλων για περαιτέρω διαβίβασης δεδομένων προσωπικού χαρακτήρα από την τρίτη χώρα ή τον διεθνή οργανισμό σε άλλη τρίτη χώρα ή άλλο διεθνή οργανισμό. Όλες οι διατάξεις του παρόντος κεφαλαίου εφαρμόζονται με σκοπό να διασφαλίζεται ότι το επίπεδο προστασίας των φυσικών προσώπων που εγγυάται ο παρών κανονισμός δεν υπονομεύεται.

#### Άρθρο 47

Δεσμευτικοί εταιρικοί κανόνες

1. Η αρμόδια εποπτική αρχή εγκρίνει δεσμευτικούς εταιρικούς κανόνες σύμφωνα με τον μηχανισμό συνεκτικότητας που προβλέπεται στο άρθρο 63, υπό τον όρο ότι:

α) είναι νομικά δεσμευτικοί και εφαρμόζονται σε κάθε οικείο μέλος και επιβάλλονται από κάθε οικείο μέλος του ομίλου επιχειρήσεων, ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα, περιλαμβανομένων των υπαλλήλων τους,

β) απονέμουν ρητώς εκτελεστά δικαιώματα στα υποκείμενα των δεδομένων όσον αφορά την επεξεργασία των δεδομένων προσωπικού χαρακτήρα που τα αφορούν και

γ) πληρούν τις απαιτήσεις που προβλέπονται στην παράγραφο 2.

2. Οι δεσμευτικοί εταιρικοί κανόνες που αναφέρονται στην παράγραφο 1 διευκρινίζουν τουλάχιστον:

α) τη δομή και τα στοιχεία επικοινωνίας του ομίλου επιχρίσεων, ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα και κάθε μέλους του,

β) τις διαβιβάσας δεδομένων ή το σύνολο των διαβιβάσεων δεδομένων, συμπεριλαμβανομένων των κατηγοριών δεδομένων προσωπικού χαρακτήρα, τον τύπο επεξεργασίας και τους σκοπούς της, τον τύπο των υποκειμένων των δεδομένων που επηρεάζονται και τον καθορισμό της εν λόγω τρίτης χώρας ή τρίτων χωρών,

γ) τη νομικά δεσμευτική φύση τους, τόσο εσωτερικά όσο και εξωτερικά,

δ) την εφαρμογή των γενικών αρχών προστασίας δεδομένων, ιδίως τον περιορισμό του σκοπού, την ελαχιστοποίηση των δεδομένων, τις περιορισμένες περιόδους αποθήκευσης, την ποιότητα των δεδομένων, την προστασία των δεδομένων ήδη από τον σχεδιασμό και εξ ορισμού, τη νομική βάση για την επεξεργασία, την επεξεργασία αδικών κατηγοριών δεδομένων προσωπικού χαρακτήρα, τα μέτρα διασφάλισης της ασφάλειας των δεδομένων, καθώς και την εφαρμογή των απαιτήσεων σχετικά με περαιτέρω διαβιβάσας σε φορείς που δεν δεσμεύονται από τους δεσμευτικούς εταιρικούς κανόνες,

ε) τα δικαιώματα των υποκαμένων των δεδομένων όσον αφορά την επεξεργασία και τα μέσα άσκησης των εν λόγω δικαιωμάτων, περιλαμβανομένων του δικαιώματος μη υπαγωγής σε απόφασής που λαμβάνονται αποκλειστικά βάση αυτοματοποιημένης επεξεργασίας, συμπεριλαμβανομένης της κατάρτισης προφίλ σύμφωνα με το άρθρο 22, του δικαιώματος υποβολής καταγγελίας ενώπιον της αρμόδιας εποπτικής αρχής και ενώπιον των αρμόδιων δικαστηρίων των κρατών μελών σύμφωνα με το άρθρο 79, καθώς και της εξασφάλισης επανόρθωσης και, όπου απαιτείται, αποζημίωσης για παράβαση των δεσμευτικών εταιρικών κανόνων,

στ) την αποδοχή ευθύνης από τον υπεύθυνο επεξεργασίας ή τον εκτελούντα την επεξεργασία που είναι εγκατεστημένος στο έδαφος κράτους μέλους για τυχόν παραβάσεις των δεσμευτικών εταιρικών κανόνων από οποιοδήποτε οικείο μέλος που δεν είναι εγκατεστημένο στην Ένωση· ο υπεύθυνος επεξεργασίας ή ο εκτελών την επεξεργασία εξαιρείται από την ευθύνη αυτή, εν όλω ή εν μέρη, μόνο αποδεικνύοντας ότι το εν λόγω μέλος δεν ευθύνεται για το γενεσιουργό γεγονός της ζημίας,

ζ) τον τρόπο παροχής της ενημέρωσης σχετικά με τους δεσμευτικούς εταιρικούς κανόνες προς τα υποκείμενα των δεδομένων, ιδίως τις διατάξεις που αναφέρονται στα στοιχεία δ), ε) και στ) της παρούσας παραγράφου, επιπλέον των άρθρων 13 και 14,

η) τα καθήκοντα κάθε υπευθύνου προστασίας δεδομένων που ορίζεται σύμφωνα με το άρθρο 37 ή οποιοσδήποτε προσώπου ή οντότητας επιφορτισμένων με την παρακολούθηση της συμμόρφωσης προς τους δεσμευτικούς εταιρικούς κανόνες εντός του ομίλου επιχειρήσεων, ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα, καθώς και με την παρακολούθηση της κατάρτισης και του χαρισμού των καταγγελιών,

θ) τις διαδικασίες καταγγελίας,

ι) τους μηχανισμούς εντός του ομίλου επιχειρήσεων, ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα για τον έλεγχο της συμμόρφωσης προς τους δεσμευτικούς εταιρικούς κανόνες. Οι εν λόγω μηχανισμοί περιλαμβάνουν ελέγχους για την προστασία των δεδομένων και μεθόδους διασφάλισης διορθωτικών δράσεων για την προστασία των δικαιωμάτων του υποκαμένου των δεδομένων. Τα αποτελέσματα του ελέγχου αυτού πρέπει να ανακοινώνονται στο πρόσωπο ή την οντότητα που αναφέρονται στο στοιχείο η) και στο διοικητικό συμβούλιο της ελέγχου σας επιχείρησης του ομίλου επιχειρήσεων ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα, ενώ επίσης πρέπει να παρέχονται κατόπιν αιτήματος στην αρμόδια εποπτική αρχή,

ια) τους μηχανισμούς αναφοράς και καταχώρισης αλλαγών στους κανόνες και αναφοράς των εν λόγω αλλαγών στην εποπτική αρχή

ιβ) τον μηχανισμό συνεργασίας με την εποπτική αρχή, ώστε να διασφαλίζεται η συμμόρφωση κάθε μέλους του ομίλου επιχειρήσεων, ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα, ιδίως θέτοντας στη διάθεση της εποπτικής αρχής τα αποτελέσματα των ελέγχων των μέτρων που αναφέρονται στο στοιχείο ι),

ιγ) τους μηχανισμούς αναφοράς στην αρμόδια εποπτική αρχή κάθε νομικής απαίτησης στην οποία ένα μέλος του ομίλου επιχειρήσεων, ή του ομίλου εταιρειών που ασκούν κοινή οικονομική δραστηριότητα υπόκειται σε τρίτη χώρα και η οποία ενδέχεται να έχει σημαντικές αρνητικές επιπτώσεις στις εγγυήσεις που παρέχονται από τους δεσμευτικούς εταιρικούς κανόνες και

ιδ) την κατάλληλη εκπαίδευση στην προστασία δεδομένων του προσωπικού που έχει μόνιμη ή τακτική πρόσβαση σε δεδομένα προσωπικού χαρακτήρα.

3. Η Επιτροπή μπορεί να προσδιορίσει τον μορφότυπο και τις διαδικασίες για την ανταλλαγή πληροφοριών μεταξύ υπευθύνων επεξεργασίας, εκτελούντων την επεξεργασία και εποπτικών αρχών για τους δεσμευτικούς εταιρικούς κανόνες κατά την έννοια του παρόντος άρθρου. Οι εν λόγω εκτελεστικές πράξεις εκδίδονται σύμφωνα με τη διαδικασία εξέτασης που αναφέρεται στο άρθρο 93 παράγραφος 2

# Βιβλιογραφικές Αναφορές

## Κεφάλαιο 1:

TDWI RESEARCH, BIG DATA ANALYTICS. Αναρτήθηκε τέταρτο τρίμηνο του 2011, από τον Philip Russom. Co-sponsored by SAS.

<https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf>

WISHWORKS, understanding the 3 v's of Big Data.

<https://www.whishworks.com/blog/big-data/understanding-the-3-vs-of-big-data-volume-velocity-and-variety>

All about big data applications, Deepak Kumar Sahu , εκδόθηκε 28 Ιουνίου 2018.

<https://www.digitalvidya.com/blog/big-data-applications/>

The guardian e-newspaper. Title: Big data and open data: what's what and why does it matter?

<https://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government>

International science Council

<https://council.science/publications/open-data-in-a-big-data-world/>

Big data In context 2018, institute for Information, Telecommunication and Media Law University of Münster

Big data and data protection by ICO ( Information Commissioner's Office )

Katalyst Data Management

<https://www.katalystdm.com/hadoop-de-hadoop-cloud-storage-processing-analytics-big-data/>

Big Data Storage. Εκδόθηκε από τους Martin Strohbach, Jeorg Daubert, Herman Ravkin, and Mario Lischka

[https://www.researchgate.net/publication/299617100\\_Big\\_Data\\_Storage](https://www.researchgate.net/publication/299617100_Big_Data_Storage)

Big data made simple.

Using big data to understand customers' behavior and increase the business profits.

<https://bigdata-madesimple.com/using-big-data-to-understand-customers-behavior-and-increase-the-business-profits/>

## Κεφάλαιο 2:

Big Data in Retail – How to win with Predictive Analytics

Ankit Khattri - PROJECT LEAD, ES-Tier II ERP-SCE-GIS | December 19, 2017

Co-author: Gaurang Singh

<https://www.hcltech.com/blogs/big-data-retail-how-win-predictive-analytics>

Role of big data and predictive analytics | Shirish JebleSneha KumariYogesh PatilYogesh Patil

November 2016

[https://www.researchgate.net/publication/309809606\\_Role\\_of\\_big\\_data\\_and\\_predictive\\_analytics](https://www.researchgate.net/publication/309809606_Role_of_big_data_and_predictive_analytics)

Big Data consumer analytics and the transformation of marketing | Author links open overlay

ISunil Erellesha, Nobuyuki Fukawa, Linda Swaynea

February 2016

<https://www.sciencedirect.com/science/article/abs/pii/S0148296315002842>

Predictive Analytics. What it is and why it matters.

[https://www.sas.com/en\\_ae/insights/analytics/predictive-analytics.html](https://www.sas.com/en_ae/insights/analytics/predictive-analytics.html)

Why Do Supply Chain Managers Need Predictive Analytics? | Author: Jasmine Morgan

<https://datafloq.com/read/amp/supply-chain-managers-need-predictive-analytics/5799?fbclid=IwAR2ykU89DeWdWkp4o9oUqIH0UpDsZtEZ5FOognGFe3nsUtzRLsZ6WABXmhc>

Social Media Analytics and Intelligence

Publisher: IEEE | Authors, Daniel Zeng

Chinese Academy of Sciences and University of Arizona, Hsinchun Chen, Robert Lusch, Shu-Hsing Li

<https://ieeexplore.ieee.org/abstract/document/5678581>

Sentiment Analysis | Monkey Learn

<https://monkeylearn.com/sentiment-analysis/>

Sentiment Analysis: Concept, Analysis and Applications

Shashank Gupta, Jan 7, 2018

<https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>

Understanding Confusion Matrix | Sarang Narkhede May 9, 2018 ·

<https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>

Confusion Matrix in Machine Learning | Geeks for Geeks

<https://www.geeksforgeeks.org/confusion-matrix-machine-learning/>

Role of big data and predictive analytics | Research Gate

[https://www.researchgate.net/publication/309809606\\_Role\\_of\\_big\\_data\\_and\\_predictive\\_analytics](https://www.researchgate.net/publication/309809606_Role_of_big_data_and_predictive_analytics)

Data science Central | How Machines Are Learning From Customers And Predicting Human Behavior

Posted by Hemant Warudkar on April 23, 2018

<https://www.datasciencecentral.com/profiles/blogs/how-machines-are-learning-from-customers-and-predicting-human?fbclid=IwAR3vsoRfpVvLVFa01ujzAZsW2m0oXloc1cA3DoNZ1JgHTyIVjGdEBfNOJoc>

### **Κεφάλαιο 3:**

Traffic Flow Prediction With Big Data: A Deep Learning Approach

Publisher: IEEE

<https://ieeexplore.ieee.org/document/6894591>

Information Security in Big Data: Privacy and Data Mining

Publisher: IEEE

<https://ieeexplore.ieee.org/document/6919256>

Expert system | Machine Learning for Big Data Analytics 28 March 2017

<https://expertsystem.com/machine-learning-big-data-analytics/>

Machine learning + Consumer Behavior = great customer experience

<https://analyticsindiamag.com/why-machine-learning-consumer-behaviour-great-customer-experience/>

Tech.co | 3 Ways Big Data and Machine Learning Affect Consumer Behavior

<https://tech.co/news/3-ways-big-data-and-machine-learning-are-affecting-consumer-behavior-2016-10>

Centric Digital July 1, 2016

By Centric Digital

On Big Data Visualization

About: Financial Services Government Healthcare Retail & Apparel

<https://centricdigital.com/blog/big-data-visualization/how-can-predictive-analytics-improve-business/>

Predictive analytics For Businesses: What It Is And How It Impacts You | Author, VANHISHIKHA BHARGAVA

<https://www.experfy.com/blog/predictive-analytics-for-businesses-what-it-is-and-how-it-impacts-you>

## **Κεφάλαιο 4:**

ΚΑΝΟΝΙΣΜΟΣ (ΕΕ) 2016/679 ΤΟΥ ΕΥΡΩΠΑΪΚΟΥ ΚΟΙΝΟΒΟΥΛΙΟΥ ΚΑΙ ΤΟΥ ΣΥΜΒΟΥΛΙΟΥ

της 27ης Απριλίου 2016 για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων



αυτών και την κατάργηση της οδηγίας 95/46/ΕΚ (Γενικός Κανονισμός για την Προστασία Δεδομένων)

