



ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ, ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

**Αξιοποίηση Μεγάλων Δεδομένων για Επιχειρηματική
Ανάλυση: Βελτιστοποίηση των Διαδικασιών Λήψης
Αποφάσεων**

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: ΠΑΝΑΓΙΩΤΗΣ ΠΑΠΑΔΟΠΟΥΛΟΣ

ΟΝΟΜΑΤΕΠΩΝΥΜΟ ΕΠΙΒΛΕΠΩΝΤΑ/ΠΟΥΣΑ: ΕΛΕΝΑ ΚΑΚΟΥΛΛΗ

ΙΑΝΟΥΑΡΙΟΣ 2025



ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ, ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

**Αξιοποίηση Μεγάλων Δεδομένων για Επιχειρηματική
Ανάλυση: Βελτιστοποίηση των Διαδικασιών Λήψης
Αποφάσεων**

ΟΝΟΜΑΤΕΠΩΝΥΜΟ: ΠΑΝΑΓΙΩΤΗΣ ΠΑΠΑΔΟΠΟΥΛΟΣ

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

ΕΛΕΝΑ ΚΑΚΟΥΛΛΗ,

ΕΛΕΥΘΕΡΙΟΣ ΖΑΧΑΡΙΟΥΔΑΚΗΣ,

ΣΑΛΩΜΗ ΕΥΡΙΠΙΔΟΥ

ΙΑΝΟΥΑΡΙΟΣ 2025

Πνευματικά δικαιώματα

Copyright © Παναγιώτης Παπαδόπουλος, 2025.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της Διπλωματικής Εργασίας από το Πανεπιστήμιο Νεάπολις δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Πανεπιστημίου.

Σελίδα Εγκυρότητας

Ονοματεπώνυμο Φοιτητή: Παναγιώτης Παπαδόπουλος

Τίτλος Διπλωματικής Εργασίας: Αξιοποίηση Μεγάλων Δεδομένων για Επιχειρηματική Ανάλυση:
Βελτιστοποίηση των Διαδικασιών Λήψης Αποφάσεων

Η παρούσα Διπλωματική Εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση εξ αποστάσεως μεταπτυχιακού τίτλου στο Πανεπιστήμιο Νεάπολις και εγκρίθηκε στις 3/2/2025 από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή:

Πρώτος επιβλέπων (Πανεπιστήμιο Νεάπολις Πάφος): ΕΛΕΝΑ ΚΑΚΟΥΛΛΗ

Μέλος Εξεταστικής Επιτροπής: ΕΛΕΥΘΕΡΙΟΣ ΖΑΧΑΡΙΟΥΔΑΚΗΣ

Μέλος Εξεταστικής Επιτροπής: ΣΑΛΩΜΗ ΕΥΡΙΠΙΔΟΥ

Ή ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

Ο Παναγιώτης Παπαδόπουλος, γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «Αξιοποίηση Μεγάλων Δεδομένων για Επιχειρηματική Ανάλυση: Βελτιστοποίηση των Διαδικασιών Λήψης Αποφάσεων», αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Ο Δηλών

ΠΕΡΙΕΧΟΜΕΝΑ

Κατάλογος Εικόνων.....	5
Κατάλογος Πινάκων	5
Περίληψη	6
1. Εισαγωγή.....	7
1.1 Ορισμός και Εξέλιξη των Μεγάλων Δεδομένων.....	7
1.2 Σημασία των Μεγάλων Δεδομένων στη Σύγχρονη Επιχειρηματική Ανάλυση.....	9
1.3 Σκοπός και στόχοι της Πτυχιακής Εργασίας.....	10
1.4 Ερευνητικά Ερωτήματα.....	11
2. Μεθοδολογία	12
3. Θεωρητικό Υπόβαθρο	15
3.1 Τα Χαρακτηριστικά των Μεγάλων Δεδομένων	15
3.2 Τεχνολογίες Μεγάλων Δεδομένων.....	16
3.2.1 Υποδομές και Πλατφόρμες: Hadoop, Spark.....	17
3.2.2 Εργαλεία Αποθήκευσης και Επεξεργασίας Δεδομένων: NoSQL, Data Lakes	22
3.3 Επεξεργασία Δεδομένων για Επιχειρηματική Ανάλυση	25
3.3.1 Εξόρυξη δεδομένων και Ανάλυση Δεδομένων	26
3.3.2 Προγνωστικά Μοντέλα και Στατιστικές Τεχνικές	27
3.3.3 Μηχανική Μάθηση και Μεγάλα Δεδομένα	33
3.4 Η Επιχειρηματική Ανάλυση (Business Analytics) και τα Μεγάλα Δεδομένα.....	35
3.4.1 Κατηγορίες Επιχειρηματικής Ανάλυσης: Περιγραφική, Διαγνωστική, Προγνωστική, Προδιαγραφική.....	35
3.4.2 Ο ρόλος της Επιχειρηματικής Ανάλυσης στη Λήψη Αποφάσεων.....	38
4. Η Σημασία των Μεγάλων Δεδομένων στη Λήψη Αποφάσεων.....	40
4.1 Πώς τα Μεγάλα Δεδομένα Βελτιστοποιούν τη Διαδικασία Λήψης Αποφάσεων	40
4.2 Προκλήσεις και Προβλήματα στην Εφαρμογή των Μεγάλων Δεδομένων στην Διαδικασία Λήψης Αποφάσεων	42
4.3 Παραδείγματα Χρήσης των Μεγάλων Δεδομένων σε Επιχειρηματικό Πλαίσιο Εφοδιαστική Αλυσίδα, Εξυπηρέτηση Πελατών, Ανάλυση Αγοράς.....	44
4.3.1 Χρήση των Μεγάλων Δεδομένων σε Ανάλυση Αγοράς.....	44
4.3.2 Χρήση των Μεγάλων Δεδομένων σε Εφοδιαστική Αλυσίδα	45
4.3.3 Χρήση των Μεγάλων Δεδομένων σε Εξυπηρέτηση Πελατών	46
5.Βελτιστοποίηση Διαδικασιών Λήψης Αποφάσεων με Μεγάλα Δεδομένα	48
5.1 Βελτιστοποίηση Συλλογής και Οργάνωσης Μεγάλων Δεδομένων	48
5.2 Τεχνικές Βελτιστοποίησης Ανάλυσης και Επεξεργασίας Δεδομένων	55
5.3 Παραδείγματα οργανισμών που λαμβάνουν αποφάσεις με γνώμονα τα δεδομένα.....	61

6. Συμπεράσματα.....	67
Βιβλιογραφία.....	71

Κατάλογος Εικόνων

Εικόνα 1: Διάγραμμα ροής PRISMA.....	14
--------------------------------------	----

Κατάλογος Πινάκων

Πίνακας 1: Συγκριτικός Πίνακας: Apache Hadoop vs Apache Spark.....	20
Πίνακας 2: Συγκριτικός πίνακας βάσης δεδομένων και λίμνης δεδομένων.....	24
Πίνακας 3: Συγκριτικός πίνακας περιγραφικής ανάλυσης, προγνωστικής ανάλυσης και προδιαγραφικής ανάλυσης.....	37
Πίνακας 4: Πίνακας Τεχνολογιών Βελτιστοποίησης Συλλογής και Οργάνωσης Μεγάλων Δεδομένων.....	51
Πίνακας 5: Συγκριτικός Πίνακας των Τεχνικών Βελτιστοποίησης Ανάλυσης και Επεξεργασίας Δεδομένων.....	58
Πίνακας 6: Παραδείγματα οργανισμών που λαμβάνουν αποφάσεις με γνώμονα τα δεδομένα	64

Περίληψη

Τα μεγάλα δεδομένα αναφέρονται στον όγκο δομημένων και μη δομημένων στοιχείων που παράγουν οι άνθρωποι και οι μηχανές και σε αυτά μπορεί να περιλαμβάνονται οι αναρτήσεις στα κοινωνικά δίκτυα, τα δεδομένα αισθητήρων, οι οικονομικές συναλλαγές και κάθε είδος κατάσταση που παράγει στοιχεία. Τα μεγάλα δεδομένα είναι σημαντικά για τις επιχειρήσεις, καθώς παρέχουν ανεκτίμητες γνώσεις και πληροφορίες που ενισχύουν σημαντικά τις διαδικασίες λήψης αποφάσεων. Επηρεάζουν διάφορες πτυχές μιας επιχείρησης, από τις στρατηγικές δέσμευσης πελατών έως τη διαχείριση αποθεμάτων. Η μελέτη των τεχνικών βελτιστοποίησης έδειξε ότι η διαχείριση και η επεξεργασία των μεγάλων δεδομένων από τους οργανισμούς έχει συντελέσει στην αποδοτικότερη λειτουργία των οργανισμούς σε πολλούς τομείς. Τεχνικές όπως η κατάτμηση δεδομένων είναι ζωτικής σημασίας για μεγάλης κλίμακας κατανομημένα συστήματα, απαιτώντας προσεκτική σχεδίαση για αποφυγή ανισορροπίας. Επίσης η διαχείριση ευρετηρίων και μεταδεδομένων είναι αποτελεσματική για συχνές ερωτήσεις, αλλά απαιτεί τακτική συντήρηση για να διατηρηθεί επεκτάσιμη αλλά και η επεξεργασία στη μνήμη είναι ιδανική για αναλύσεις σε πραγματικό χρόνο, αν και περιορίζεται από τους διαθέσιμους πόρους μνήμης. Η ανάλυση μεγάλων δεδομένων επέτρεψε στους οργανισμούς να χειρίζονται τεράστια σύνολα δεδομένων και να αποκαλύπτουν πολύτιμες πληροφορίες που προηγουμένως ήταν ανέφικτες λόγω υπολογιστικών περιορισμών. Η ανάλυση δεδομένων είναι η διαδικασία εξέτασης συνόλων δεδομένων για την εξαγωγή συμπερασμάτων σχετικά με τις πληροφορίες που περιέχουν και ο πρωταρχικός στόχος της ανάλυσης δεδομένων είναι η υποστήριξη της λήψης αποφάσεων με την παροχή αξιοποιήσιμων πληροφοριών. Οι τρεις κύριοι τύποι μοντέλων ανάλυσης δεδομένων είναι η περιγραφική, η προγνωστική και η συντακτική ανάλυση, καθένας από τους οποίους εξυπηρετεί έναν μοναδικό σκοπό και παρέχει διαφορετικές γνώσεις. Εργαλεία όπως το Apache Hadoop και το Apache Spark εστιάζουν στην επεξεργασία μεγάλων δεδομένων, ενώ τεχνολογίες όπως το Amazon Lambda και το Google Cloud Functions προσανατολίζονται σε serverless υπολογιστικά μοντέλα. Οι προαναφερθείσες μεθοδολογίες και τεχνικές συντελούν στην βελτιστοποίηση της λήψης τεκμηριωμένων και επαληθευμένων αποφάσεων με στόχο την ανάπτυξη της επιχείρησης.

Λέξεις Κλειδιά: Μεγάλα Δεδομένα, Περιγραφική Ανάλυση, Προγνωστική Ανάλυση, □ Προδιαγραφική Ανάλυση, Εξόρυξη Δεδομένων, Apache Hadoop, Apache Spark, Apache Kafka

1. Εισαγωγή

1.1 Ορισμός και Εξέλιξη των Μεγάλων Δεδομένων

Τα μεγάλα δεδομένα αναφέρονται σε μεγάλα και ποικίλα σύνολα πληροφοριών που μεταβάλλονται με συνεχώς αυξανόμενους ρυθμούς. Ο όρος περιλαμβάνει τον όγκο της πληροφορίας, την ταχύτητα με την οποία δημιουργούνται και συλλέγονται, καθώς και την ποικιλία και το εύρος των δεδομένων. Μπορεί να είναι είτε δομημένα για παράδειγμα αριθμητικά με συγκεκριμένη μορφοποίηση, αλλά μπορεί να είναι και αδόμητα. Πηγές των μεγάλων δεδομένων μπορούν να αποτελέσουν τα κοινωνικά δίκτυα, ιστότοποι, προσωπικές ηλεκτρονικές συσκευές αλλά και ερωτηματολόγια. Άλλοτε συλλέγονται με τη συγκατάθεση του χρήστη και άλλοτε όχι, εγείροντας συχνά ανησυχίες για την προστασία της ιδιωτικής ζωής και συνήθως αποθηκεύονται ηλεκτρονικά και αναλύονται με τη χρήση λογισμικού ειδικά σχεδιασμένου για να χειρίζεται μεγάλα, πολύπλοκα σύνολα δεδομένων (Gandomi & Haider, 2015).

Τα μεγάλα δεδομένα αναφέρονται στον όγκο δομημένων και μη δομημένων στοιχείων που παράγουν οι άνθρωποι και οι μηχανές και σε αυτά μπορεί να περιλαμβάνονται οι αναρτήσεις στα κοινωνικά δίκτυα, τα δεδομένα αισθητήρων, οι οικονομικές συναλλαγές και κάθε είδος κατάσταση που παράγει στοιχεία. Η επεξεργασία τους συνεπάγεται την εξαγωγή συμπερασμάτων από αυτή την ευρεία συλλογή τους τα οποία μπορούν να βοηθήσουν τους οργανισμούς να γίνουν πιο αποτελεσματικοί, να καινοτομήσουν ταχύτερα, και εν δυνάμει να αξιοποιήσει τις πληροφορίες που εξάγονται από αυτά με σκοπό να αυξήσει την κερδοφορία του. Τα μεγάλα δεδομένα χρειάζονται ειδική διαχείριση από τους οργανισμούς αφού τα σύνολα αυτά είναι τόσο μεγάλα και πολύπλοκα σε όγκο, ταχύτητα και ποικιλία, ώστε τα παραδοσιακά συστήματα διαχείρισης τους δεν μπορούν να τα αποθηκεύσουν, να τα επεξεργαστούν και να τα αναλύσουν (Marr, 2021).

Τα μεγάλα δεδομένα δεν αφορούν μόνο τον όγκο των δεδομένων, αλλά και το τι κάνουν οι οργανισμοί με αυτά. Μπορούν να χρησιμοποιηθούν για την ανάλυση πληροφοριών που οδηγούν σε καλύτερες αποφάσεις και στρατηγικές επιχειρηματικές κινήσεις. Ο πρωταρχικός στόχος της ανάλυσης μεγάλων δεδομένων είναι να βοηθήσει τους οργανισμούς να λαμβάνουν πιο τεκμηριωμένες επιχειρηματικές αποφάσεις, επιτρέποντας στους επιστήμονες δεδομένων και άλλους χρήστες να αναλύουν τεράστιους όγκους δεδομένων συναλλαγών και άλλων πηγών δεδομένων που τα παραδοσιακά επιχειρηματικά συστήματα ενδέχεται να μην είναι σε θέση να διαχειριστούν. Μπορούν να αναλυθούν με εργαλεία λογισμικού που χρησιμοποιούνται συνήθως ως μέρος προηγμένων κλάδων ανάλυσης, όπως η προγνωστική ανάλυση, η εξόρυξη δεδομένων, η ανάλυση κειμένου και η στατιστική ανάλυση (Manyika et al., 2011).

Η ανάλυση μεγάλων δεδομένων είναι η διαδικασία συλλογής, οργάνωσης και ανάλυσης μεγάλων συνόλων δεδομένων για την ανακάλυψη μοτίβων και άλλων χρήσιμων πληροφοριών. Μπορεί να βοηθήσει τους οργανισμούς να κατανοήσουν τις πληροφορίες που περιέχονται στα δεδομένα και θα βοηθήσει επίσης στον εντοπισμό των δεδομένων που είναι πιο σημαντικά για την επιχείρηση και τις μελλοντικές επιχειρηματικές αποφάσεις. Οι αναλυτές μεγάλων δεδομένων χρησιμοποιούν την ανάλυση μεγάλων δεδομένων για να κατευθύνουν την επιχείρηση και να λαμβάνουν αποφάσεις. Υπάρχουν διάφορα βήματα που εμπλέκονται στην ανάλυση δεδομένων, τα οποία περιλαμβάνουν τη συλλογή δεδομένων, την επεξεργασία δεδομένων, τον καθαρισμό δεδομένων, την ανάλυση δεδομένων, την οπτικοποίηση και την ερμηνεία. Οι τεχνολογίες μεγάλων δεδομένων είναι σημαντικές για την παροχή ακριβέστερης ανάλυσης, η οποία μπορεί να οδηγήσει σε πιο συγκεκριμένη λήψη αποφάσεων με αποτέλεσμα μεγαλύτερη επιχειρησιακή αποτελεσματικότητα, μείωση του κόστους και μείωση των κινδύνων για την επιχείρηση. Για να αξιοποιηθούν οι δυνατότητες των μεγάλων δεδομένων, θα χρειαστεί μια υποδομή που μπορεί να διαχειρίζεται και να επεξεργάζεται τεράστιους όγκους δομημένων και μη δομημένων δεδομένων σε πραγματικό χρόνο και μπορεί να προστατεύει το απόρρητο και την ασφάλεια των δεδομένων. Υπάρχουν διάφορες τεχνολογίες στην αγορά από διάφορους προμηθευτές, όπως η Amazon, η IBM, η Microsoft κ.λπ., για τη διαχείριση των μεγάλων δεδομένων. Ενώ η επιλογή αυτών των τεχνολογιών μπορεί να εξαρτάται από τις συγκεκριμένες ανάγκες ενός οργανισμού, υπάρχουν ορισμένες τεχνολογίες που χρησιμοποιούνται ευρέως. Αυτές περιλαμβάνουν: Hadoop, πλατφόρμες μεγάλων δεδομένων όπως η Microsoft HDInsight, βάσεις δεδομένων NoSQL όπως η MongoDB και πλαίσια επεξεργασίας δεδομένων υψηλής ταχύτητας όπως η Spark.

Κάθε ένας από τους τομείς έχει διαφορετικές χρήσεις για τις πληροφορίες που συλλέγει. Για παράδειγμα, στην υγειονομική περίθαλψη, τα μεγάλα δεδομένα χρησιμοποιούνται για την πρόβλεψη επιδημιών, τη θεραπεία ασθενειών, τη βελτίωση της ποιότητας ζωής και την αποφυγή θανάτων που μπορούν να αποφευχθούν. Στον τραπεζικό τομέα, τα μεγάλα δεδομένα χρησιμοποιούνται για τον εντοπισμό δόλιων συναλλαγών. Στη γεωργία, οι αγρότες τα χρησιμοποιούν για να λαμβάνουν καλύτερες αποφάσεις σχετικά με το ποιες καλλιέργειες θα φυτέψουν, πώς θα μεγιστοποιήσουν την απόδοση και πώς θα μειώσουν τις ασθένειες των καλλιεργειών. Στην υγειονομική περίθαλψη, χρησιμοποιούνται για την πρόβλεψη επιδημιών, τη θεραπεία ασθενειών, τη βελτίωση της ποιότητας ζωής και την αποφυγή θανάτων που μπορούν να αποφευχθούν. Καθώς ο κόσμος γίνεται όλο και πιο ευαίσθητοποιημένος σε θέματα υγείας, η χρήση τους στην υγειονομική περίθαλψη πρόκειται να αυξηθεί. Η προγνωστική ανάλυση πρόκειται να αλλάξει τα δεδομένα για τον κλάδο της υγειονομικής περίθαλψης (Gandomi et al., 2015).