



Τεχνολογικό
Πανεπιστήμιο
Κύπρου

Σχολή Μηχανικής και
Τεχνολογίας

Πτυχιακή εργασία

**Ευφυής επεξεργασία και ανάλυση δεδομένων μεγάλου όγκου:
Συλλογή και επεξεργασία δεδομένων μεγάλης συχνότητας και
εύρους σε πραγματικό χρόνο για τον εντοπισμό ανωμαλιών**

Χριστόφορος Βασιλείου

Λεμεσός, Μάιος 2018

ΤΕΧΝΟΛΟΓΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΗΣ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πτυχιακή εργασία

Ευφυής επεξεργασία και ανάλυση δεδομένων μεγάλου όγκου :
Συλλογή και επεξεργασία δεδομένων μεγάλης συχνότητας και
εύρους σε πραγματικό χρόνο για τον εντοπισμό ανωμαλιών

του

Χριστόφορου Βασιλείου

Επιβλέπων Καθηγητής

Δρ. Ανδρέας Ανδρέου

Λεμεσός, Μάιος 2018

Πνευματικά δικαιώματα

Copyright © Χριστόφορος Βασιλείου, 2018

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής του Τεχνολογικού Πανεπιστημίου Κύπρου δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα της πτυχιακής εργασίας καθηγητή Δρα. Ανδρέα Ανδρέου του Τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, για την εμπιστοσύνη που μας έδειξε αναθέτοντάς μας τη συγκεκριμένη εργασία, για την καθοδήγηση του κατά την υλοποίησή της και τη βοήθεια που μας παρείχε. Θέλω επίσης να ευχαριστήσω το συμφοιτητή μου Παναγιώτη Αντωνίου για την υπέροχη συνεργασία που είχαμε για την πραγματοποίηση της παρούσας πτυχιακής εργασίας. Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα την οικογένεια μου για τη συμπαράστασή της, την αγάπη της και τη στήριξή που μου επέδειξε καθ' όλη τη διάρκεια αυτών των τεσσάρων χρόνων της φοίτησής μου.

ΠΕΡΙΛΗΨΗ

Η εργασία επικεντρώνεται σε τρόπους συλλογής και ανάλυσης Μεγάλων Δεδομένων σε μια συνεχώς μεταβαλλόμενη εποχή, καθώς και την υποβοήθηση της διαδικασίας λήψης αποφάσεων βασισμένων στα δεδομένα. Η παραγωγή τεράστιου όγκου δεδομένων σε επιχειρησιακό επίπεδο αλλά και γενικότερα, οδήγησε στην ανάγκη αξιοποίησης αυτής της πληροφορίας. Απώτερος σκοπός της παρούσας εργασίας είναι η ανάδειξη της σημαντικότητας της χρήσης καινοτόμων τεχνολογιών για διαχείριση και ανάλυση μεγάλων όγκων δεδομένων και πώς μπορούν να συμβάλουν σε τομείς όπως της αυτοκινητοβιομηχανίας. Μελετήθηκαν περιπτώσεις ανάλυσης σε πληροφορίες πραγματικού χρόνου οι οποίες είχαν προέλευση την πλατφόρμα κοινωνικής δικτύωσης Twitter, αλλά και μεγάλων συνόλων δεδομένων από τομείς της αυτοκίνησης. Για την πραγματοποίηση των πειραμάτων χρησιμοποιήθηκε η καινοτόμα τεχνολογία του Apache Spark. Αποτελέσματα της έρευνας ήταν η σκιαγράφηση της επιρροής των επιμέρους τμημάτων των οχημάτων στη συνολική ομαλή τους λειτουργία. Τα ευρήματα αναδεικνύουν πως οι αναλυτικές τεχνικές που εφαρμόζονται στα Μεγάλα Δεδομένα της αυτοκίνησης, έχουν τεράστια αξία για την προγνωστική συντήρηση των οχημάτων, την αποσαφήνιση των σφαλμάτων, καθώς και των αιτιών που τα προκαλούν. Σε τελικό στάδιο της εργασίας, τα δεδομένα που επεξεργάστηκαν και αναλύθηκαν μετατράπηκαν σε κατάλληλη μορφή που να επιτρέπει τη γραφική τους απεικόνιση, η οποία αποτελεί διαφορετική πτυχιακή εργασία συμφοιτητή μου. Εν κατακλείδι, αποτυπώθηκαν τα πλεονεκτήματα της ύπαρξης μιας πλατφόρμας που θα προσφέρει υπηρεσίες “ Software as a service ” (SaaS) σε επιχειρήσεις και οργανισμούς παντός είδους.

Λέξεις κλειδιά: Apache Spark, Ανάλυση Δεδομένων, Μεγάλα Δεδομένα, Spark Streaming, Ροή Δεδομένων, Pyspark, Επιχειρηματική Ευφυΐα, Αυτοκινητοβιομηχανία.

ABSTRACT

The aim of this thesis is collecting and analysing Big Data in an era that is constantly changing. It is also very helpful to the process of taking decisions relying on the data. The huge production of data in business and other uses in general, has lead to the need of using and exploiting this information. The ulterior aim of this task is to produce and highlight the importance of using innovative technology to handle and analyse huge quantities of data and how this would be able to contribute to such sections as the car manufacture. Cases of analysing information of real time coming from social networking platforms such as Twitter have been studied. Datasets from automotive data have been also studied. The innovative Apache Spark technology has been used to carry out these experiments. The results of this research were sketching the role of the individual parts of the vehicles to their normal function in general. The findings showed that the analytic technics applied in huge data of the automotive, are valuable for the predictive maintenance of the vehicles, the clarification of errors and the causes of them. As a conclusion of the thesis, the data that has been worked out and analyzed, has converted into a proper form that allows the graphical appearance that is another task taken by a fellow student. Finally, the advantages of the creation of a platform that will be able to provide services to business and other organizations have been imprinted.

Keywords: Apache Spark, Data Analysis, Bid Data, Spark Streaming, Streaming Data, Pyspark, Business Intelligence, Automotive.