



ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

Extracting Top-K Insights from Multi-Dimensional Data

ΤΟΥ

Γεώργιου Κυνηγόπουλου - 3093

Εκπόνηση πτυχιακής ως μέρος του
Προπτυχιακού Τίτλου Σπουδών

στη
Σχολή Θετικών Επιστημών
Τμήμα Πληροφορικής

Επιβλέπων Καθηγητής: Αναστάσιος Γούναρης

25/06/2021

Δήλωση Συγγραφικής Ιδιότητας

Εγώ, ο Γεώργιος Κυνηγόπουλος, δηλώνω ότι αυτή η πτυχιακή εργασία με τίτλο, Extracting Top-K Insights from Multi-Dimensional Data, και η δουλειά που παρουσιάζεται σε αυτή είναι δικά μου. Επιβεβαιώνω ότι:

- Αυτή η δουλειά πραγματοποιήθηκε ολοκληρωτικά ή κυρίως κατά την υποψηφιότητά μου για τίτλο προπτυχιακών σπουδών σε αυτό το πανεπιστήμιο.
- Όπου οποιοδήποτε μέρος αυτής της πτυχιακής εργασίας έχει προηγουμένως κατατεθεί για την απόκτηση πτυχίου ή άλλου τίτλου σε αυτό ή άλλο πανεπιστήμιο, αυτό διατυπώνεται ξεκάθαρα.
- Όπου έχω συμβουλευτεί την δημοσιευμένη δουλειά τρίτων, αυτό αποδίδεται ορθώς.
- Όπου έχω παραθέσει από δουλειά τρίτων, η πηγή δίνεται πάντα. Με εξαίρεση αυτές τις παραθέσεις, αυτή η πτυχιακή εργασία είναι εξ ολοκλήρου προσωπική μου δουλειά.
- Έχω παραθέσει όλες τις κύριες πηγές βοήθειας.
- Όπου αυτή η πτυχιακή εργασία είναι βασισμένη σε συνεργατική δουλειά δική μου και τρίτων, έχω καταστήσει ξεκάθαρο ποια κομμάτια έχουν πραγματοποιηθεί από άλλους και πώς συνέβαλα εγώ.

Υπογραφή:

Ημερομηνία:

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

Σύνοψη

Σχολή Θετικών Επιστημών

Τμήμα Πληροφορικής

Προπτυχιακός Τίτλος Σπουδών

του Γεώργιου Κυνηγόπουλου - 3093

Με την εκθετική αύξηση της ποσότητας πληροφορίας που μας κατακλύζει, η εξαγωγή insights (συμπερασμάτων) από πολυδιάστατα δεδομένα καθίσταται όλο και πιο δύσκολη, απαιτώντας αρκετή εμπειρία στον τομέα της ανάλυσης δεδομένων. Σε αυτή την εργασία, αναλύεται ο τρόπος για να αυτοματοποιηθεί αυτή η διεργασία, επιτρέποντας ακόμα και σε μη-έμπειρους χρήστες να εξάγουν εύκολα συμπεράσματα σε γρήγορο χρόνο. Επιπλέον, προτείνονται πρακτικές χρήσεις για το συγκεκριμένο λογισμικό, επιπρόσθετες επεκτάσεις που θα μπορούσαν να εφαρμοστούν και παρουσιάζονται αποτελέσματα με χρόνους εκτέλεσης.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή, κ. Αναστάσιο Γούναρη για την εξαιρετική επικοινωνία που υπήρχε, για την καθοδήγησή του και για την επίλυση όλων των αποριών κατά την διάρκεια αυτής της πτυχιακής εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω την οικογένειά μου, η οποία με στήριξε για την ολοκλήρωση των σπουδών μου.

Περιεχόμενα

Δήλωση Συγγραφικής Ιδιότητας	i
Σύνοψη	ii
Ευχαριστίες	iii
1 Εισαγωγή	1
1.1 Σκοπός της εργασίας	1
1.2 Δομή της εργασίας	2

Κεφάλαιο 1

Εισαγωγή

Η ηλεκτρονική αναλυτική επεξεργασία, ή αλλιώς OLAP, επιτρέπει την οργάνωση και συντήρηση μεγάλων βάσεων δεδομένων. Επιτρέπουν ταχύτατες πράξεις για την μείωση διαστάσεων και την σύμπτυξη πληροφορίας. Ωστόσο, αυτό απαιτεί ορισμένη εμπειρία και αναπόφευκτη ώρα εργασίας για να βρεθούν ουσιώδη αποτελέσματα, τα οποία πολλές φορές αφήνονται στην κρίση του χρήστη.

1.1 Σκοπός της εργασίας

Σκοπός της εργασίας είναι να αυτοματοποιηθεί κατά μεγάλο ποσοστό η διεργασία για την εξαγωγή συμπερασμάτων. Ο χρήστης δεν χρειάζεται να κάνει πράξεις, βέβαια πρέπει να ορίσει ορισμένες μεταβλητές όπως τον αριθμό των insights (συμπερασμάτων) που αναζητά, τις διαστάσεις που τον ενδιαφέρουν κτλ. Ο τρόπος για να υλοποιηθεί αυτή η διεργασία είναι:

- **Υπολογισμός όλων των δυνατών συνδυασμών extractor (εξαγωγέων).**
Οι extractors είναι παρόμοιοι με τις συναρτήσεις. Δέχονται μια πληθώρα μεταβλητών και επιστρέφουν μια τιμή πίσω.
- **Υπολογισμός όλων των δυνατών υποχώρων.**
Υποχώρος θεωρείται ένας χώρος που ανήκει στο σύνολο των δεδομένων που επεξεργαζόμαστε.
- **Εξαγωγή ενός συνόλου τιμών.**
Κάθε έγκυρο ζεύγος extractor με υποχώρο δίνει ένα συγκεκριμένο σύνολο τιμών.

- Υπολογισμός σκορ από αυτό το σύνολο.

Ορίζεται μια συνάρτηση όπου επιστρέφει ένα σκορ αναλόγως τον τύπο του insight που αναζητείται.

Στην συνέχεια της εργασίας θα αναλυθούν διεξοδικά τα βήματα ένα προς ένα.

1.2 Δομή της εργασίας

Για να γίνει πλήρως αντιληπτό το αντικείμενο που αναλύεται, η δομή της εργασίας ακολουθεί αυτόν τον απλό σκελετό:

1. Αναφέρονται με πλήρη σαφήνεια όλοι οι ορισμοί και το υπόβαθρο που πρέπει να υπάρχει (Κεφάλαιο 2)