



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS**

**INTERDISCIPLINARY POSTGRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"**

MASTER THESIS

Data Analytics in Chronic Disease Self-Management

Statistical and Machine Learning Methodologies
for Knowledge Discovery based on Quantified Self Data

Aikaterini V. Georgountzou

Supervisor: **Dr. Anastasia Krithara**, Post-Doctoral Research Associate,
Software and Knowledge Engineering Laboratory, Institute of
Informatics and Telecommunications, NCSR "Demokritos"

ATHENS

FEBRUARY 2019



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Η Ανάλυση Δεδομένων στην
Αυτοδιαχείριση Χρόνιων Παθήσεων**

Μεθοδολογίες Στατιστικής και Μηχανικής Μάθησης για Εξόρυξη Γνώσεων
από Προσωπικά Δεδομένα Υγείας και Τρόπου Ζωής

Αικατερίνη Β. Γεωργούντζου

Επιβλέπουσα: **Δρ. Αναστασία Κριθαρά**, Μεταδιδακτορική Ερευνήτρια,
Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού, Ινστιτούτο
Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ «Δημόκριτος»

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2019

MASTER THESIS

Data Analytics in Chronic Disease Self-Management
Statistical and Machine Learning Methodologies
for Knowledge Discovery based on Quantified Self Data

Aikaterini V. Georgountzou

S.N.: ΠΙΒ0160

SUPERVISOR: **Dr. Anastasia Krithara**, Post-Doctoral Research Associate,
Software and Knowledge Engineering Laboratory, Institute of
Informatics and Telecommunications, NCSR “Demokritos”

**EXAMINATION
COMMITTEE:**

Dr. Anastasia Krithara, Post-Doctoral Research Associate,
Software and Knowledge Engineering Laboratory, Institute of
Informatics and Telecommunications, NCSR “Demokritos”

Dr. Ilias Maglogiannis, Associate Professor, Director of the
Computational Biomedicine Laboratory, Department of
Digital Systems, University of Piraeus

Dr. Vangelis Karkaletsis, Research Director of the Institute
of Informatics and Telecommunications, Head of the
Software and Knowledge Engineering Laboratory, NCSR
"Demokritos"

February 2019

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Η Ανάλυση Δεδομένων στην Αυτοδιαχείριση Χρόνιων Παθήσεων:
Μεθοδολογίες Στατιστικής και Μηχανικής Μάθησης
για Εξόρυξη Γνώσεων από Προσωπικά Δεδομένα Υγείας και Τρόπου Ζωής

Αικατερίνη Β. Γεωργούντζου

A.M.: ΠΙΒ0160

ΕΠΙΒΛΕΠΟΥΣΑ: **Δρ. Αναστασία Κριθαρά**, Μεταδιδακτορική Ερευνήτρια,
Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού, Ινστιτούτο
Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ «Δημόκριτος»

**ΕΞΕΤΑΣΤΙΚΗ
ΕΠΙΤΡΟΠΗ:**

Δρ. Αναστασία Κριθαρά, Μεταδιδακτορική Ερευνήτρια,
Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού,
Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών, ΕΚΕΦΕ
«Δημόκριτος»

Δρ. Ηλίας Μαγκλογιάννης, Αναπληρωτής Καθηγητής,
Διευθυντής Εργαστηρίου Υπολογιστικής Βιοϊατρικής, Τμήμα
Ψηφιακών Συστημάτων, Πανεπιστήμιο Πειραιώς

Δρ. Βαγγέλης Καρκλέτσης, Διευθυντής Έρευνας,
Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών, Υπεύθυνος
Εργαστηρίου Τεχνολογίας Γνώσεων και Λογισμικού, ΕΚΕΦΕ
«Δημόκριτος»

Φεβρουάριος 2019

ABSTRACT

Chronic diseases management is one of the greatest challenges of modern healthcare systems. Given the fact that non-communicable diseases are responsible for more than 70% of deaths worldwide [1], the constant monitoring of a patient's health condition has become vital need and, hence, the era of mobile health starts to rise. At the same time, the idea of self-managing personal aspects of life, and not only, through the prism of new technologies, the so-called Quantified Self, gains ground rapidly. Nowadays, sensors constitute an integral part of the daily life and monitor almost every aspect of it, gathering enormous quantities of data. The challenge is how to control the data that derive from the combination of electronic health services with wearable sensor technologies and broaden the horizons of scientific research [2]. At this point, data analytics assumes its decisive role. Patients using such technologies gain the capability to record and process their vital signs, fitness activities, everyday habits, or even feelings [3]. The resulting data constitute the gemstone for statistical and machine learning techniques to be performed so that knowledge discovery can take place and, as a consequence, identify the risk factors in patients' health and provide personalized medical follow-up and immediate feedback to avoid emergent situations.

This graduate thesis proposes a data analytics solution that will examine patients' consistency in their measurement schedule and study the interaction among the different daily measurements, with the scope of determining how these factors can influence the monitoring of their health. Studies generalized on a demographic level, including sex, age and geolocation, will also take place so that statistical significant differences can be identified in the medical values and, thus, appropriate recommendations can be derived per population group. Aiming at improving and personalizing the medical monitoring of chronic health conditions, the proposed solution can circumvent the challenges of electronic health systems and provide benefits for the involved patients, such as enhancement of their welfare, early detection of dangerous situations, assumption of further targeted monitoring, motivation to engage in self-caring activities and follow treatment and, last, modeling of their behavior to improve self-care and enjoy a better quality of life.

SUBJECT AREA: Data Analytics

KEYWORDS: chronic diseases management, quantified self, mobile health, statistical data analysis, machine learning

ΠΕΡΙΛΗΨΗ

Η διαχείριση των χρόνιων παθήσεων συνιστά μια από τις σημαντικότερες προκλήσεις των σύγχρονων συστημάτων υγείας. Η επιτακτική ανάγκη της συνεχούς διαχείρισης των νοσημάτων αυτών, που συνιστούν αιτία θανάτου για περισσότερο από το 70% του πληθυσμού παγκοσμίως [1], ήταν ένας από τους λόγους που οδήγησαν τον τομέα της ηλεκτρονικής υγείας να γνωρίσει ραγδαία ανάπτυξη. Παράλληλα, η ιδέα της αυτοδιαχείρισης προσωπικών δεδομένων υγείας και τρόπου ζωής, υπό το πρίσμα των νέων τεχνολογιών, κερδίζει έδαφος πολύ γρήγορα. Στις μέρες μας, οι αισθητήρες συνιστούν αναπόσπαστο κομμάτι της καθημερινότητας και συλλέγουν τεράστιες ποσότητες δεδομένων, ελέγχοντας κάθε πτυχή αυτής. Η πρόκληση, λοιπόν, είναι πώς θα καταφέρουμε να διαχειριστούμε όλα αυτά τα δεδομένα που προκύπτουν από το συνδυασμό των υπηρεσιών ηλεκτρονικής υγείας με τις τεχνολογίες φορητών αισθητήρων και κυρίως πώς θα τα ερμηνεύσουμε, ώστε να διευρύνουμε τους ορίζοντες της επιστημονικής έρευνας [2]. Στο σημείο αυτό, ο τομέας της ανάλυσης δεδομένων καλείται να αναλάβει καθοριστικό ρόλο. Οι ασθενείς που χρησιμοποιούν τέτοιες τεχνολογίες, αποκτούν τη δυνατότητα να καταγράψουν και να επεξεργαστούν τα βιοσήματά τους, τις αθλητικές τους δραστηριότητες, τις καθημερινές συνήθειές τους ή ακόμα και τα συναισθήματά τους [3]. Τα δεδομένα που προκύπτουν συνιστούν τον πολύτιμο λίθο της στατιστικής και των τεχνικών μηχανικής μάθησης, η εφαρμογή των οποίων θα οδηγήσει σε εξόρυξη γνώσεων σχετικά με τους παράγοντες αυξημένης επικινδυνότητας για την υγεία ενός ασθενούς και θα παράσχει τη δυνατότητα εξατομικευμένης ιατρικής παρακολούθησης και άμεσης ενημέρωσης για αποφυγή επειγόντων περιστατικών.

Η παρούσα διπλωματική εργασία προτείνει μια μεθοδολογία ανάλυσης δεδομένων που θα εξετάσει τη συνέπεια των ασθενών στο πρόγραμμα λήψης των μετρήσεών τους και θα μελετήσει την αλληλεπίδραση μεταξύ των διαφορετικών ημερήσιων μετρήσεων, με σκοπό τον προσδιορισμό του τρόπου με τον οποίο αυτοί οι παράγοντες μπορούν να επηρεάσουν την παρακολούθηση της υγείας των ασθενών. Παράλληλα, θα πραγματοποιηθούν μελέτες που γενικεύονται σε δημογραφικό επίπεδο, συμπεριλαμβανομένου του φύλου, της ηλικίας και της γεωγραφικής κατανομής, έτσι ώστε να εντοπιστούν οι στατιστικά σημαντικές διαφορές στις ιατρικές τιμές ανα πληθυσμιακή ομάδα και να εξαχθούν πιο στοχευμένα, κατάλληλα συμπεράσματα. Στοχεύοντας στη βελτίωση και εξατομίκευση της ιατρικής παρακολούθησης χρόνιων καταστάσεων υγείας, η προτεινόμενη λύση δύναται να αντιμετωπίσει τις προκλήσεις των ηλεκτρονικών υπηρεσιών υγείας, παρέχοντας στους ασθενείς τη δυνατότητα έγκαιρου εντοπισμού επικίνδυνων καταστάσεων, ενίσχυση της ευημερίας τους, κινητοποίηση για συμμόρφωση στο πρόγραμμα λήψης των μετρήσεών τους αλλά και την εξειδικευμένη θεραπευτική τους αγωγή, δέσμευση για άσκηση και, τέλος, μοντελοποίηση της συμπεριφοράς τους με σκοπό τη βελτίωση της φροντίδας του εαυτού τους και την απόκτηση μιας καλύτερης ποιότητας ζωής.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Ανάλυση Δεδομένων

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: διαχείριση χρόνιων νοσημάτων, προσωπικά δεδομένα υγείας και τρόπου ζωής, κινητή υγεία, στατιστική ανάλυση δεδομένων, μηχανική μάθηση

