

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Σύστημα Μηχανικής Μάθησης Για Διαχείριση Ανθρώπων Με Διαβήτη



-Πτυχιακή Εργασία –

2020 - 2021

Θωμάς Παραβαΐτης Π2014183

Επιβλέπων: -Θεμιστοκλής Έξαρχος-

Επιβλέπων

Θεμιστοκλής Έξαρχος, Επίκουρος Καθηγητής

Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Θεμιστοκλής Έξαρχος, Επίκουρος Καθηγητής

Ιόνιο Πανεπιστήμιο

Φοίβος Μυλωνάς, Αναπληρωτής Καθηγητής

Ιόνιο Πανεπιστήμιο

Κάτια Λήδα Κερμανίδου, Μόνιμη Επίκουρη Καθηγήτρια

Ιόνιο Πανεπιστήμιο

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	5
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....	6
1.1 Εισαγωγή στο πρόβλημα.....	6
1.2 Γενικά για τον σακχαρώδη διαβήτη.....	8
1.3 Σκοπός και στόχοι εργασίας.....	15
1.4 Βιβλιογραφική ανασκόπηση – έρευνα.....	15
1.5 Αποτελέσματα βιβλιογραφίας και συνεισφορά.....	38
ΚΕΦΑΛΑΙΟ 2: ΤΕΧΝΙΚΟ ΥΠΟΒΑΘΡΟ.....	39
2.1 WEKA.....	39
2.2 Android studio.....	39
2.3 Flutter.....	40
2.4 Dart.....	40
2.5 K-fold cross validation.....	41
ΚΕΦΑΛΑΙΟ 3: ΔΕΔΟΜΕΝΑ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ.....	43
3.1 Περιγραφή δεδομένων.....	43
3.2 Χρήση του WEKA στο σύστημα.....	52
3.3 Κώδικας – δέντρο WEKA.....	52
3.4 Πειράματα και αποτελέσματα έρευνας.....	56
3.5 Περιγραφή Εφαρμογής.....	60
3.6 Απαιτήσεις Προγράμματος.....	65

3.7 Επεκτασιμότητα.....	65
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	67

ΠΕΡΙΛΗΨΗ

Το αντικείμενο που πραγματεύεται αυτή η πτυχιακή εργασία είναι η δημιουργία ενός συστήματος machinelearning (μηχανικήμάθηση) χρησιμοποιώντας ταξινόμηση, βοηθητικό για τους ασθενείς που πάσχουν από διαβήτη έτσι ώστε να γίνεται καλύτερα η διάγνωση και η υποστήριξη τους αλλά και η μείωση του κόστους του νοσοκομείου, μειώνοντας τις επισκέψεις που θα χρειαστεί ο ασθενής και άρα το κόστος της ιατρικής περίθαλψης.

Ο διαβήτης είναι ένα από τα μεγαλύτερα ιατρικά προβλήματα στην σύγχρονη ανθρωπότητα καθώς σκοτώνει εκατομμύρια ανθρώπους κάθε χρόνο και δημιουργεί προβλήματα σε ακόμη περισσότερους ανθρώπους. Το γεγονός ότι ο διαβήτης δεν αντιμετωπίζεται, μόνο περιορίζεται τον κάνει ένα ακόμη πιο επικίνδυνο φαινόμενο το οποίο χρήζει άμεση ιατρική περίθαλψη με τρόπους όπως εξειδικευμένες δίαιτες, φάρμακα, συχνές ιατρικές εξετάσεις και πολλά άλλα. Συνεπώς δημιουργήθηκε αυτή η εφαρμογή έτσι ώστε να υποστηρίξει την ιατρική περίθαλψη μέσω μιας αυτοματοποιημένης μεθοδολογίας.

Η προτεινόμενη μεθοδολογία χρησιμοποιήθηκε για τον αυτόματο έλεγχο της αποδοτικότητας των θεραπειών που χορηγήθηκαν σε ασθενείς με διαβήτη χρησιμοποιώντας το εργαλείο WEKA. Αρχικά χρησιμοποιήθηκαν δεδομένα εισόδου – εξόδου που έχουν σχέση με το αν χορηγήθηκε Ινσουλίνη και μετορμίνη καθώς και στοιχεία που αφορούν τον ασθενή όπως αν επανεντάχθηκε στην κλινική ο ασθενής και μετά από πόσες μέρες, αν υπάρχουν τρέχον θεραπείες για τον διαβήτη, πόσες εξετάσεις γίνανε στον ασθενή, ηλικία, φύλλο, πόσος χρόνο έμεινε στο νοσοκομείο, πόσοι εργαστηριακοί έλεγχοι γίνανε, πόσοι εξωτερικές και εσωτερικές επισκέψεις γίνανε σε ιατρεία, επίπεδα γλυκόζης κ.α. με βάση τα οποία αναπτύχθηκε ένα σύστημα το οποίο ελέγχει αν θα χρειαστεί θεραπεία για τον διαβήτη ο ασθενής ή αν θα χρειαστεί να αλλάξει την θεραπεία που έχει πάρει από την κλινική στο μέλλον.

Το προτεινόμενο σύστημα και τα αποτελέσματα του προήλθαν από δοκιμές πάνω σε αληθινό dataset με ασθενείς από 130 νοσοκομεία με τον αριθμό τους να είναι γύρω στους 100000 ασθενείς. Τα αποτελέσματα έδειξαν ότι το σύστημα κάνει πρόβλεψη με περίπου 85%. Ο αλγόριθμος που χρησιμοποιήθηκε για το classification είναι ο J48(C48 γραμμένος σε Java).

1.ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή Στο Πρόβλημα

Στις μέρες μας, το πρόβλημα του σακχαρώδους διαβήτη είναι ένα πρόβλημα το οποίο απασχολεί πάνω από 170 εκατομμύρια ανθρώπους παγκοσμίως και έως το 2030 αναμένεται να διπλασιαστεί καθώς η συχνότητα της νόσου αυξάνει ραγδαία. Εκτιμάται ότι στην Ελλάδα το 5.9% του γενικού πληθυσμού πάσχει από σακχαρώδη διαβήτη. Ο διαβήτης εμφανίζεται συχνότερα στον ανεπτυγμένο κόσμο και φαίνεται ότι συνδέεται άμεσα με το λεγόμενο δυτικό τρόπο ζωής που περιλαμβάνει τη διαβίωση στα μεγάλα αστικά κέντρα, τον ανθυγιεινό τρόπο διατροφής και την καθιστική ζωή. Αποτελεί μια από τις κύριες αιτίες θανάτου παγκοσμίως.

Ο σακχαρώδης διαβήτης είναι μεταβολική ασθένεια η οποία χαρακτηρίζεται από αύξηση συγκέντρωσης του σακχάρου στο αίμα(Υπεργλυκαιμία) και διαταραχή του μεταβολισμού της γλυκόζης, είτε ως αποτέλεσμα ελαττωμένης έκκρισης ινσουλίνης είτε λόγω ελάττωσης της ευαισθησίας των κυττάρων του σώματος στην ινσουλίνη. Οι κύριοι τύποι σακχαρώδους διαβήτη είναι ο διαβήτης τύπου 1, ο διαβήτης τύπου 2 και ο διαβήτης της κύησης. Ο σακχαρώδης διαβήτης έχει χρόνια πορεία και μπορεί να προκαλέσει μια σειρά σοβαρών επιπλοκών, όπως καρδιαγγειακή νόσο, χρόνια νεφρική ανεπάρκεια, βλάβες του αμφιβληστροειδούς, βλάβες των νεύρων, στυτική δυσλειτουργία κ.α.

Ο διαβήτης είναι μια ασθένεια που χρειάζεται συνεχή αντιμετώπιση και αυτό ήταν το έναυσμα για το σύστημα που θα σας παρουσιάσω, το οποίο έχει σκοπό να βοηθήσει στην σωστή λήψη θεραπείας από τον ασθενή ή την βελτίωση της υπάρχουσας θεραπείας και κατά συνέπεια να μειωθούν οι επισκέψεις στο νοσοκομείο και άρα το κόστος των νοσοκομείων για ιατρική περίθαλψη.

Αναγνωρίζεται όλο και περισσότερο ότι η αντιμετώπιση της υπεργλυκαιμίας στον νοσοκομειακό ασθενή έχει σημαντική επίδραση στην έκβαση, τόσο ως προς τη νοσηρότητα όσο και τη θνησιμότητα. Αυτή η αναγνώριση οδήγησε στην ανάπτυξη τυποποιημένων πρωτοκόλλων στη ρύθμιση της μονάδας εντατικής θεραπείας με αυστηρούς στόχους γλυκόζης σε πολλά ιδρύματα. Ωστόσο, το ίδιο δεν μπορεί να ειπωθεί για τις περισσότερες εισαγωγές εσωτερικών ασθενών εκτός εντατικής θεραπείας. Αντίθετα, άλλα στοιχεία δείχνουν ότι η διαχείριση των εσωτερικών ασθενών είναι αυθαίρετη και συχνά οδηγεί είτε σε καθόλου θεραπεία είτε σε μεγάλες διακυμάνσεις της γλυκόζης όταν χρησιμοποιούνται παραδοσιακές

στρατηγικές διαχείρισης. Αν και τα δεδομένα είναι λίγα, πρόσφατες ελεγχόμενες δοκιμές έχουν δείξει ότι οι στρατηγικές εσωτερικών ασθενών που βασίζονται σε πρωτόκολλα μπορούν να είναι αποτελεσματικές και ασφαλείς. Ως εκ τούτου, συνιστάται τώρα η εφαρμογή πρωτοκόλλων στο νοσοκομείο. Ωστόσο, υπάρχουν λίγες εθνικές αξιολογήσεις της περίθαλψης του διαβήτη στον νοσοκομειακό ασθενή που θα μπορούσαν να χρησιμεύσουν ως βάση για την αλλαγή. Η παρούσα ανάλυση μιας μεγάλης κλινικής βάσης δεδομένων πραγματοποιήθηκε για την εξέταση ιστορικών προτύπων περίθαλψης διαβήτη σε ασθενείς με διαβήτη που εισήχθησαν σε νοσοκομείο των ΗΠΑ και για την ενημέρωση μελλοντικών κατευθύνσεων που θα μπορούσαν να οδηγήσουν σε βελτιώσεις στην ασφάλεια των ασθενών. Συγκεκριμένα, εξετάσαμε τη χρήση του HbA1c ως δείκτη προσοχής στη φροντίδα του διαβήτη σε μεγάλο αριθμό ατόμων που αναγνωρίζονται ότι έχουν διάγνωση σακχαρώδους διαβήτη. Υποθέτουμε ότι η μέτρηση του HbA1c σχετίζεται με μείωση των ποσοστών επανεισδοχής σε άτομα που εισάγονται στο νοσοκομείο.

Οι βάσεις δεδομένων κλινικών δεδομένων περιέχουν πολύτιμα αλλά ετερογενή και δύσκολα δεδομένα όσον αφορά τις ελλείπουσες τιμές, τις ελλειπείς ή ασυνεπείς εγγραφές και την υψηλή διάσταση που κατανοούνται όχι μόνο από τον αριθμό των χαρακτηριστικών αλλά και από την πολυπλοκότητά τους. Επιπλέον, η ανάλυση εξωτερικών δεδομένων είναι πιο δύσκολη από την ανάλυση των αποτελεσμάτων ενός προσεκτικά σχεδιασμένου πειράματος ή δοκιμής, επειδή δεν έχει καμία επίδραση στον τρόπο και τον τύπο των πληροφοριών που συλλέχθηκαν. Ωστόσο, είναι σημαντικό να χρησιμοποιήσετε αυτά τα τεράστια ποσά δεδομένων για να βρείτε νέες πληροφορίες / γνώσεις που πιθανώς δεν είναι διαθέσιμες πουθενά.

Συνεπώς γίνεται σαφές ότι η σωστή θεραπεία του ασθενή που πάσχει από διαβήτη (ιδιαίτερα τύπου 2) είναι πολύ σημαντική και θα πρέπει να χορηγείται η καταλληλότερη θεραπεία στον ασθενή. Πολλές φορές όμως η χορηγία της σωστής θεραπείας του διαβήτη είναι περίπλοκη υπόθεση καθώς σε πολλές περιπτώσεις, ιδιαίτερα σε διαβήτη τύπου 2 η ασθένεια εισβάλλει βαθμιαία και πολλές φορές ο ασθενής αισθάνεται υγιής.

Με αφορμή τα μεγάλα προβλήματα που υπάρχουν αλλά και για την σωστή θεραπεία των ασθενών του σακχαρώδους διαβήτη, γεννήθηκε η ιδέα για την δημιουργία ενός συστήματος που θα εξετάζει την συμπεριφορά ενός ασθενή και την θεραπεία του και θα προβλέπει εάν πρέπει να αλλάξει(η θα χρειαστεί να αλλάξει στο μέλλον) η θεραπεία, ή αν δεν έχει χορηγηθεί κάποια θεραπεία στον ασθενή αν θα χρειαστεί να του χορηγηθεί έτσι ώστε να βελτιωθεί η πρόληψη

αλλά και η αντιμετώπιση της ασθένειας και να αποφευχθούν μελλοντικά προβλήματα υγείας για τον ασθενή.

Το σύστημα πέρα από τα προβλήματα της ιατρικής φύσης που αφορούν τις εξετάσεις του ασθενή όπως γλυκόζη, Αιματολογικά τεστ A1C, ινσουλίνη(μέσα στον οργανισμό), ινσουλίνη(φάρμακο) και μετορμίνη, λαμβάνει υπόψη και παράγοντες που αφορούν θέματα όπως εσωτερικές επισκέψεις, επισκέψεις σε εξωτερικά ιατρεία, ηλικία, βάρος, φύλλο, επείγοντα περιστατικά κ.α. έτσι ώστε να παράγει αποτελεσματικότερες προβλέψεις.

1.2 Γενικά Για Τον Σακχαρώδη Διαβήτη

Ορισμός

Ο σακχαρώδης διαβήτης είναι μια κοινή και χρόνια πάθηση, που χαρακτηρίζεται από υψηλά επίπεδα γλυκόζης ή σακχάρου στο αίμα. Εκδηλώνεται όταν τα κύτταρα δεν αντιδρούν σωστά στην ινσουλίνη (μια ορμόνη που εκκρίνεται από το πάγκρεας) και όταν το πάγκρεας δεν μπορεί να παράγει περισσότερη ινσουλίνη ως αντίδραση. Δεν υπάρχει θεραπεία για τον διαβήτη αν και σε μερικές περιπτώσεις, η απώλεια βάρους και η διατήρησή του μπορούν να αναστρέψουν την πορεία της ασθένειας. Αν κάποιος δεν ακολουθήσει αγωγή ή την παραμελεί, ο διαβήτης μπορεί να προκαλέσει πολύ σοβαρές επιπλοκές μακροπρόθεσμα, όπως νεφρική ανεπάρκεια, ακρωτηριασμό και τύφλωση. Από την άλλη πλευρά, ο διαβήτης αυξάνει τον κίνδυνο εκδήλωσης καρδιακής προσβολής, εγκεφαλικού επεισοδίου και επιπλοκών που σχετίζονται με την κακή κυκλοφορία του αίματος.

Κατηγορίες Σακχαρώδη Διαβήτη

Ο σακχαρώδης διαβήτης είναι η πιο συχνή σοβαρή, μεταβολική νόσος στον άνθρωπο. Το χαρακτηριστικό του διαβήτη είναι τα αυξημένα επίπεδα του σακχάρου στο αίμα, αλλά η διαταραχή αυτή είναι μόλις μια από τον αριθμό των φυσιολογικών μεταβολών. Ο διαβήτης δεν είναι μια διαταραχή αλλά μπορεί να προκύψει ως αποτέλεσμα ενός αριθμού διαταραχών στη ρύθμιση της σύνθεσης, έκκρισης και δράσης της ινσουλίνης. Ο τύπος του διαβήτη που επηρεάζει πιο συχνά τα παιδιά είναι ο διαβήτης τύπου I ή ινσουλινοεξαρτώμενος σακχαρώδης διαβήτης (IDDM). Ο διαβήτης που ξεκινά γενικά στην ενήλικη ζωή και είναι ιδιαίτερα συχνός σε παχύσαρκα άτομα ονομάζεται τύπου II ή μη ινσουλινοεξαρτώμενος σακχαρώδης διαβήτης

(NIDDM). Επίσης ο σακχαρώδης διαβήτης που προσβάλλει τις έγκυες γυναίκες ονομάζεται σακχαρώδης διαβήτης κύησης. Ωστόσο οι δυο πιο συνηθισμένες μορφές διαβήτη, οι τύπου I και II έχουν πολλά κοινά χαρακτηριστικά, διαφέρουν όμως σε πολλά άλλα. Και οι δυο τύποι χαρακτηρίζονται από υψηλά επίπεδα σακχάρου στο αίμα, τα οποία προκαλούν μια σειρά από προβλήματα υγείας που επηρεάζουν την καρδιά, τα μάτια, τα νεφρά, τα νεύρα και τον εγκέφαλο (μακροπρόθεσμα προβλήματα). Παρότι υπάρχουν σημαντικές ομοιότητες ανάμεσα στις δύο παθήσεις, η συνηθισμένη διάγνωση και η αιτία του διαβήτη τύπου I διαφέρουν πολύ από αυτές του τύπου II.

Σακχαρώδης Διαβήτης τύπου I

Ο σακχαρώδης διαβήτης τύπου I ή ινσουλινοεξαρτώμενος σακχαρώδης διαβήτης (IDDM) παρατηρείται στο 5-10% του συνόλου των διαβητικών ασθενών και προκαλείται από αυτοάνοση εκλεκτική καταστροφή των β-κυττάρων στα νησίδια του Langerhans του παγκρέατος. Τα άλλα κύτταρα που υπάρχουν στα νησίδια δεν προσβάλλονται. Οι επιπτώσεις της απώλειας της ινσουλίνης, με τη διατήρηση της γλυκαγόνης, μπορούν να θεωρηθούν ως μια παρατεταμένη μορφή νηστείας ή υποσιτισμού. Ένα υγιές άτομο το οποίο είναι σε κατάσταση νηστείας για πολλές ημέρες συνεχίζει να εκκρίνει ινσουλίνη σε χαμηλό ρυθμό η οποία όμως είναι επαρκής για να υπάρχει ισορροπία με την δράση της γλυκαγόνης στην παραγωγή γλυκόζης και κετονοσωμάτων στο ήπαρ. Στον διαβήτη τύπου I, η ανεπάρκεια ινσουλίνης είναι σοβαρή και η παραγωγή γλυκόζης και κετονών από το ήπαρ παρατηρούνται σε ρυθμό που υπερβαίνει κατά πολύ το ρυθμό χρήσης τους. Ως αποτέλεσμα αυτού, η συγκέντρωση αυτών των υποστρωμάτων στο αίμα να είναι αυξημένη. Ακόμα και όταν οι συγκεντρώσεις της γλυκόζης φθάσουν να είναι 5 με 10 φορές πιο πάνω από το φυσιολογικό, δεν εκκρίνεται ινσουλίνη επειδή δεν υπάρχουν β κύτταρα. Τα αυξημένα επίπεδα γλυκόζης και κετονών παρέχουν ένα τεράστιο φορτίο ωσμωτικά δραστικών ουσιών στο νεφρό που προκαλεί ωσμωτική διούρηση. Επιπλέον, τα κετοξέα που παράγονται είναι μετρίως ισχυρά οργανικά οξέα και η αυξημένη παραγωγή τους προκαλεί σοβαρή μεταβολική οξέωση. Αν οι ασθενείς αυτοί δεν αντιμετωπιστούν με ινσουλίνη, η οξέωση και η αφυδάτωση οδηγούν σε θάνατο από διαβητική κετοξέωση. Η κλινική εκδήλωση μπορεί να είναι αρκετά ταχεία και ύπουλη και οι ασθενείς μπορεί να μεταπέσουν από μια κατάσταση πλήρως ασυμπτωματική σε ανεξέλεγκτη πολυουρία, πολυδιψία και πολυφαγία με απώλεια βάρους και θάμβους όρασης, σε μία

περίοδος ημερών έως εβδομάδων. Τα υπογλυκαιμικά φάρμακα είναι αναποτελεσματικά στα άτομα με σακχαρώδη διαβήτη τύπου I, καθώς για να δράσουν απαιτείται η ύπαρξη υπολειπόμενης λειτουργίας του β κυττάρου.

Σακχαρώδης Διαβήτης τύπου II

Ο σακχαρώδης διαβήτης τύπου II ή μη ινσουλινοεξαρτώμενος σακχαρώδης διαβήτης (NIDDM) παρατηρείται στο 90-95% του συνόλου των διαβητικών ασθενών και έχει την δυνατότητα να συνεχίζει να παράγει ινσουλίνη. Πιο συγκεκριμένα τα β κύτταρα δεν είναι μόνο παρόντα αλλά είναι συχνά υπερπλαστικά και δεν ανταποκρίνονται φυσιολογικά στα αυξημένα επίπεδα γλυκόζης του πλάσματος με αντίστοιχη αύξηση στην έκκριση ινσουλίνης. Ωστόσο η μεταβολή στην έκκριση ινσουλίνης δεν είναι το μόνο πρόβλημα. Αν χορηγήσουμε παρόμοιες δόσεις ινσουλίνης στο ήπαρ, στο μυ και στο λιπώδη ιστό σε ένα άτομο με διαβήτη τύπου II και σε ένα υγιές άτομο, είναι δυνατό να διαπιστώσουμε ότι ο τύπου II διαβητικός έχει ανοχή στη δράση της ινσουλίνης. Έτσι τόσο ο μεταβολισμός της γλυκόζης σε απάντηση στην ινσουλίνη όπως και η έκκριση της ινσουλίνης δεν είναι φυσιολογικές στους ασθενείς με διαβήτη τύπου II. Στα αρχικά στάδια της νόσου οι ασθενείς με σακχαρώδη διαβήτη τύπου II δεν χρειάζονται τη λήψη εξωγενούς ινσουλίνης για διατήρηση της ευγλυκαιμίας. Οι περισσότεροι από τους εν λόγω ασθενείς είναι παχύσαρκοι και εμφανίζουν μεγάλη συγκέντρωση λίπους στην κοιλιακή χώρα (κοιλιακή παχυσαρκία). Σπάνια εμφανίζεται διαβητική κετοξέωση και αυτό συμβαίνει όταν υπάρχει καταπόνηση του οργανισμού (π.χ. λοίμωξη). Ο σακχαρώδης διαβήτης τύπου II συχνά παραμένει αδιάγνωστος για πολλά χρόνια διότι η υπεργλυκαιμία αναπτύσσεται βαθμιαία και στα αρχικά στάδια δεν είναι σοβαρή για να προκαλέσει τα συμπτώματα και σημεία που χαρακτηρίζουν το σακχαρώδη διαβήτη.

Συμπτώματα και Διάγνωση του Σακχαρώδη Διαβήτη

Οι πάσχοντες του σακχαρώδη διαβήτη μπορεί να είναι ασυμπτωματικοί. Το αυξημένο σάκχαρο (υπεργλυκαιμία) αποβάλλεται οσμωτικός δια των ούρων από τα νεφρά, γι' αυτό και προκαλείται αυξημένη διούρηση (πολυουρία), η οποία προς εξισορρόπηση δημιουργεί πολυδιψία. Το άτομο εμφανίζει αίσθημα πείνας, πολυφαγίας με σύγχρονο απώλεια βάρους, λόγω του ότι αδυνατεί να χρησιμοποιήσει ως πηγή ενέργειας τους προσλαμβανόμενους

υδατάνθρακες (λόγω απουσίας ινσουλίνης). Αντ' αυτών ο ασθενής καίει από τα αποθέματα των ιστών, δηλαδή λίπη και πρωτεΐνες. Συχνές είναι οι λοιμώξεις, οι δοθινώσεις και ο κνησμός γεννητικών οργάνων. Η διάγνωση είναι απαραίτητη και γίνεται με την εύρεση τιμής γλυκόζης πλάσματος (σε νηστικό άτομο) και με την θετική δοκιμασία ανοχής γλυκόζης.

Αντιμετώπιση Σακχαρώδη Διαβήτη

Η θεραπεία του διαβήτη έγκειται σε μια ελεγχόμενη διαίτα, γυμναστική και φάρμακα, εάν αυτά θεωρηθούν αναγκαία. Ο εντατικός έλεγχος του διαβήτη και η διατήρηση κανονικών επιπέδων γλυκόζης αίματος σε σχεδόν όλες τις στιγμές έχει βρεθεί ότι μειώνουν τον κίνδυνο αμφιβληστροειδοπάθειας, νεφροπάθειας και νευροπάθειας στο διαβήτη τύπου I και υπάρχουν κάποιες ενδείξεις ότι επίσης ελαττώνουν και την πιθανότητα για καρδιοπάθειες. Τα άτομα που πάσχουν από σακχαρώδη διαβήτη συχνά χρειάζονται κάποια χειρουργική αντιμετώπιση γιατί εμφανίζει διάφορες επιπλοκές.

Φαρμακευτική Αγωγή

Ο στόχος της θεραπείας στο διαβήτη είναι να ελευθερωθεί ο ασθενής από τα συμπτώματα, να έχει μια καλή ποιότητα ζωής, φυσιολογική συγκέντρωση γλυκόζης στο πλάσμα και μείωση του κινδύνου ανάπτυξης των επιπλοκών της νόσου. Υπάρχουν αρκετά φάρμακα που μπορούν να χρησιμοποιηθούν στη θεραπεία του διαβήτη: η ινσουλίνη και τα από του στόματος αντιδιαβητικά που περιλαμβάνουν τις σουλφονουλιδίες, τα διγουανίδια, οι θειαζολιδινεδιόνες και τους αναστολείς της α γλυκοσιδάσης.

- **Η ινσουλίνη:** η ινσουλίνη μπορεί να εξαχθεί είτε από βόειο είτε από χοίρειο πάγκρεας ή μπορεί να παραχθεί με τεχνολογία ανασυνδυασμένου DNA. Επειδή η ινσουλίνη έχει πολύ μικρό χρόνο ημίσειας ζωής (περίπου 10 λεπτά), η διατήρηση των φυσιολογικών επιπέδων γλυκόζης στο πλάσμα εξαρτάται από τη συνεχή παραγωγή της από το πάγκρεας. Σε αντίθεση με άλλα φάρμακα για τον διαβήτη, η ινσουλίνη είναι μια ορμόνη που μειώνει τη γλυκόζη που παράγεται με φυσικό τρόπο. Μειώνει τα επίπεδα σακχάρου στο αίμα πιο αποτελεσματικά από οποιοδήποτε άλλο διαθέσιμο φάρμακο για τον διαβήτη. Η αρχική αγωγή με ινσουλίνη μπορεί να έχει ως αποτέλεσμα μια περίοδο ύφεσης των συμπτωμάτων, στη διάρκεια της οποίας μπορεί να διακοπούν όλα

τα φάρμακα για τη μείωση του επιπέδου σακχάρου και να διατηρηθούν τα επίπεδα σακχάρου στο αίμα σε φυσιολογικές τιμές. Επιπλέον, αποδείχθηκε ότι η ινσουλίνη βελτιώνει τις τιμές των τριγλυκεριδίων και της HDL χοληστερόλη. Δεν υπάρχει ανώτατη αποτελεσματική δόση για την ινσουλίνη, γεγονός που σημαίνει ότι μπορεί να χορηγηθεί όση ποσότητα χρειάζεται για να υπάρχει η επιθυμητή αντίδραση. Ωστόσο, σχετίζεται επίσης με την αύξηση βάρους, όπως συμβαίνει και με πολλά φάρμακα για τον διαβήτη, καθώς επιτρέπει στα κύτταρα να απορροφούν το σάκχαρο που προηγουμένως απορριπτόταν μαζί με τα ούρα. Ο χρόνος έναρξης δράσης και η αποβολή της ινσουλίνης μπορούν να τροποποιηθούν με την προσθήκη διαφόρων ουσιών στις οποίες προσδένεται η ινσουλίνη, επηρεάζοντας έτσι την απελευθέρωσή της από το σημείο της ένεσης. Με την προσθήκη διαφόρων ουσιών στην ινσουλίνη αυτή μπορεί να ταξινομηθεί σε τέσσερις τύπους, ανάλογα με το χρόνο έναρξης δράσης μετά τη χορήγηση του με υποδόρια ένεση και τη διάρκεια της δράσης του.

- **Βραχείας δράσης ινσουλίνη:** παράγει αποτέλεσμα μετά από 10-30 λεπτά, ανάλογα με το ποια βραχείας δράσης ινσουλίνης χρησιμοποιείται, τη θέση της ένεσης και το βαθμό της φυσικής δραστηριότητας. Η μέγιστη δράση εκδηλώνεται περίπου 1-3 ώρες μετά την ένεση και ολοκληρώνεται μετά από 8 ώρες.
- **Ενδιάμεσης δράσης ινσουλίνης:** παράγει αποτέλεσμα μετά από περίπου 1-1,5 ώρες από τη χορήγηση. Η μέγιστη δράση της εκδηλώνεται περίπου 4-12 ώρες μετά τη χορήγηση και ολοκληρώνεται μετά από 24 ώρες.
- **Μακράς δράσης ινσουλίνης:** παράγει αποτέλεσμα 4 ώρες μετά τη χορήγηση. Η μέγιστη δράση της εκδηλώνεται 8-24 ώρες μετά τη χορήγηση και ολοκληρώνεται μετά από περίπου 28 ώρες.
- **Συνδυασμοί βραχείας και ενδιάμεσης δράσης ινσουλίνης:** παράγουν αποτελέσματα 30 λεπτά μετά τη χορήγηση. Η μέγιστη δράση εκδηλώνεται 4-8 ώρες μετά τη χορήγηση και ολοκληρώνεται μετά από περίπου 24 ώρες.
- **Σουλφονουλουρίες:** Οι σουλφονουλουρίες είναι η παλιότερη κατηγορία αντιδιαβητικών φαρμάκων που λαμβάνουν από το στόμα, ενεργοποιούν και διεγείρουν το πάγκρεας έτσι ώστε να παράγει περισσότερη ινσουλίνη. Η δράση τους συνίσταται στη διέγερση της έκκρισης της ινσουλίνης από τα β-κύτταρα λειτουργούν. Οι σουλφονουλουρίες, είναι υπεύθυνα για τη θεραπεία του σακχαρώδη διαβήτη τύπου II. Μπορούν να

χρησιμοποιηθούν είτε σε συνδυασμό με την ινσουλίνη ή άλλα υπογλυκαιμικά φάρμακα είτε μόνα τους. Προτιμώνται σε ασθενείς που τη στιγμή της διάγνωσης έχουν ηλικία μεγαλύτερη των 40 ή διάρκεια νόσου μικρότερη των 5 ετών, έχουν σωματικό βάρος λιγότερο του φυσιολογικού και επίπεδα γλυκόζης νηστείας μικρότερα των 180 mg/dl. Οι παρενέργειες αυτών των φαρμάκων είναι ένας βασικός παράγοντας. Ένας κίνδυνος μπορεί να είναι η υπογλυκαιμία, ιδιαίτερα σε ηλικιωμένα άτομα και σε όσους παίρνουν σουλφονυλουρίες παρατεταμένης δράσης. Το αλκοόλ ή η παράλειψη ενός γεύματος μπορεί να προκαλέσουν κάποια αντίδραση. Επίσης, ορισμένα άτομα που καταναλώνουν αλκοόλ, ενώ λαμβάνουν σουλφονυλουρίες, μπορεί να έχουν κοκκίνισμα στο πρόσωπο και ναυτία, στομαχικούς πόνους και ταχυπαλμίες. Ένα άλλο συχνό φαινόμενο για όσους παίρνουν σουλφονυλουρίες είναι η αύξηση του βάρους. Επίσης, ορισμένες μελέτες δείχνουν ότι συγκεκριμένες σουλφονυλουρίες ίσως να σχετίζονται με τις καρδιαγγειακές παθήσεις. Εξαιτίας αυτών των παρενεργειών, οι σουλφονυλουρίες αποτελούν τη δεύτερη επιλογή μετά τη μετφορμίνη στις περισσότερες περιπτώσεις.

- **Μετφορμίνη:** Η μετφορμίνη, το μόνο διγουανίδιο, χρησιμοποιείται ως μονοθεραπεία ή σε συνδυασμό με ινσουλίνη ή άλλα αντιδιαβητικά φάρμακα σε ασθενείς με σακχαρώδη διαβήτη τύπου II που δεν μπορούν να ανταποκριθούν στη χορήγηση των σουλφονυλουριών ως μονοθεραπεία. Η μετφορμίνη ελαττώνει τα επίπεδα γλυκόζης, μειώνοντας την ηπατική παραγωγή γλυκόζης και το μεταβολισμό του γλυκογόνου και βελτιώνοντας την αντίσταση στη δράση της ινσουλίνης, αυξάνοντας την ρυθμιζόμενη από την ινσουλίνη πρόσληψη γλυκόζης. Μειώνει τα επίπεδα των τριγλυκεριδίων και της ολικής χοληστερόλης, αυξάνοντας τα επίπεδα HDL.

Η μετφορμίνη έχει διάφορα πλεονεκτήματα σε σχέση με τις σουλφονυλουρίες έχει καλά αποτελέσματα στα άτομα στις κατηγορίες υπέρβαρου και μέσου βάρους. Μειώνει τα επίπεδα σακχάρου τόσο δραστικά όσο και οι σουλφονυλουρίες χωρίς να προκαλεί αύξηση βάρους, ενώ συνήθως δεν προκαλεί υπογλυκαιμία αν λαμβάνεται μόνη της. Από την άλλη πλευρά, η μετφορμίνη μπορεί να προκαλέσει γαστρεντερικές παρενέργειες, όπως ναυτία, μετεωρισμός και πολύ μαλακά κόπρανα ή διάρροια. Ωστόσο, αυτές οι παρενέργειες είναι σπάνιες, παρατηρούνται σε λιγότερο από το 10% των ατόμων που την λαμβάνουν. Αυτές οι παρενέργειες συνήθως ελαχιστοποιούνται, αν ο ασθενής ξεκινήσει με μια μικρή δόση την οποία όμως μπορεί να αποβεί

θανατηφόρα και που οφείλεται στα υψηλά επίπεδα γαλακτικού οξέος στο αίμα. Με δεδομένο αυτό τον κίνδυνο, τα άτομα με νεφρική ανεπάρκεια, παθήσεις του συκωτιού και προβλήματα του κυκλοφορικού δεν πρέπει να παίρνουν μετοφορμίνη.

- **Θειαζολιδινεδιόνες:** Οι θειαζολιδινεδιόνες (TZD), γνωστές και ως γλιταζόνες, μειώνουν την έλλειψη ανταπόκρισης στην ινσουλίνη και είναι πιο δραστικές σε συνδυασμό με άλλα αντιδιαβητικά φάρμακα. Αυξάνουν την ευαισθησία των μυών, του λίπους και του συκωτιού στην ινσουλίνη και αυξάνουν την δυνατότητα της να απορροφά θρεπτικές ουσίες από το αίμα. Οι συνηθισμένες παρενέργειες είναι η αύξηση του βάρους και η κατακράτηση υγρών. Ωστόσο, αυτά τα φάρμακα έχουν συσχετιστεί με διάφορα σοβαρά προβλήματα υγείας, με αποτέλεσμα η FDA και διεθνείς οργανισμοί ρύθμισης φαρμάκων να έχουν επιβάλλει περιορισμούς στη χρήση τους. Όσοι παίρνουν αυτά τα φάρμακα πρέπει να γνωρίζουν ότι όλες οι TZD μπορούν να προκαλέσουν κατακράτηση υγρών και αύξηση βάρους, καθώς επίσης και να αυξήσουν τον κίνδυνο καρδιακής ανεπάρκειας.
- **Οι αναστολείς της α-γλυκοσιδάσης:** οι αναστολείς της α-γλυκοσιδάσης (ακαρβόλη, μιγλιτόλη), μπορούν να χρησιμοποιηθούν είτε μόνοι τους είτε σε συνδυασμό με ινσουλίνη ή άλλα υπογλυκαιμικά φάρμακα για τη θεραπεία του σακχαρώδη διαβήτη τύπου II. Τα φάρμακα αυτά αναστέλλουν τις γλυκοσιδάσεις των λαχνών που βρίσκονται στις λάχνες του λεπτού εντέρου και διασπούν τους σύνθετους πολυσακχαρίτες και την συρκόζη προς απορροφήσιμους μονοσακχαρίτες. Με τον τρόπο αυτό καθυστερείται η πέψη των υδατανθράκων και η απορρόφηση της γλυκόζης, γεγονός που οδηγεί σε χαμηλότερες μεταγευματικές αιχμές γλυκόζης. Οι ανεπιθύμητες ενέργειες αφορούν το γαστρεντερικό σύστημα (μετεωρισμός, διάρροια, κοιλιακός πόνος) μπορούν να αντιμετωπιστούν με την χορήγηση χαμηλών δόσεων. Η χρήση αναστολείς της α-γλυκοσιδάσης με συνδυασμο ινσουλίνης ή άλλων υπογλυκαιμικών φαρμάκων μπορούν να προκαλέσουν υπογλυκαιμία

1.3 Σκοπός Και Στόχοι Εργασίας

Οι στόχοι όσον αφορά την κλινική που θα χρησιμοποιήσει το σύστημα είναι κυρίως για την βελτίωση της αποδοτικότητας και αποτελεσματικότητας της στην πρόληψη και της αντιμετώπιση της ασθένειας του σακχαρώδη διαβήτη τύπου 2, έτσι ώστε μέσω της τυποποίησης διαδικασιών και της τεχνητής νοημοσύνης να επιτυγχάνεται η καλύτερη περίθαλψη των ασθενών αλλά και η εξοικονόμηση οικονομικών πόρων μέσω της αποτελεσματικότερης θεραπείας και άρα την μείωση της πιθανότητας να ξαναχρειαστεί ιατρική περίθαλψη ο ασθενής και φυσικά η βελτίωση της φήμης της κλινικής μέσα από την καλύτερη αντιμετώπιση της ασθένειας.

Ακόμη ένας πολύ σημαντικός στόχος είναι η συνεχής προσοχή της κατάστασης του ασθενή από τον ίδιο τον ασθενή ώστε να παρατηρεί και να προβλέπει μέσω του συστήματος εάν θα πρέπει να προσέχει ή ακόμη και να επισκεφθεί τον γιατρό που παρακολουθεί την υγεία του.

Σκοπός αυτής της εργασίας είναι η εκμάθηση ενός νέου εργαλείου machine learning, σε ένα διαφορετικό περιβάλλον εργασίας και και με διαφορετικού είδους δεδομένα . Επίσης η χρήση Δένδρων απόφασης για την δημιουργία συστημάτων τα οποία θα μπορούν να συνδεθούν με μία γλώσσα προγραμματισμού για την παραγωγή προγραμμάτων τεχνητής νοημοσύνης.

1.4 Βιβλιογραφική Ανασκόπηση - Έρευνα

Στο άρθρο[1] γίνεται η κατασκευή ενός συστήματος machine learning όπου βοηθά στον περιορισμό του διαβήτη τύπου 1 και την πρόληψή του ελέγχοντας τα επίπεδα γλυκόζης του ασθενή. Η έρευνα που αναλύεται έχει ως πρωταρχικό ρόλο να δημιουργήσει και να αναλύσει το σύστημα με το οποίο μέσω της καλύτερης διάγνωσης θα μειώσει το κόστος που θα δαπανά το κράτος της Αμερικής για ανθρώπους με προβλήματα διαβήτη τύπου 1. Εδώ βλέπουμε και ομοιότητες με το σύστημα που δημιουργήθηκε για την εργασία καθώς και τα δύο συστήματα αποσκοπούν στην μείωση εξόδων των κλινικών ωστόσο το ένα απευθύνεται σε διαβήτη τύπου 1 ενώ το άλλο για διαβήτη τύπου 2. Το σύστημα για την δημιουργία του ανάλυσε ένα πολύ μεγάλο όγκο δεδομένων αφού αναλύθηκαν δεδομένα από διάφορες πηγές τα οποία πάρθηκαν από διάφορα επιστημονικά άρθρα που είναι δημοσιευμένα. Το εύρος της

συγκεκριμένης έρευνας αφορά κυρίως τεχνικές πρόβλεψης που εστιάζουν σε διαφορετικές κατηγορίες αλγορίθμων μηχανικής μάθησης, τεχνητών νευρωνικών δικτύων, μηχανών φορέα υποστήριξης, Bayesian νευρωνικών δικτύων, αποφάσεις δέντρων και άλλα. Το σύστημα βλέπει τα επίπεδα γλυκόζης του ασθενή καθώς και τα επίπεδα γλυκόζης που είχε στο παρελθόν και στέλνει τις κατάλληλες ειδοποιήσεις που αφορούν την υπογλυκαιμία και την υπεργλυκαιμία. Επίσης άλλα δεδομένα που λαμβάνει υπόψη η έρευνα είναι τα επίπεδα ινσουλίνης στο αίμα, τα επίπεδα δραστηριότητας του ασθενή καθώς και η διατροφή του. Ωστόσο επηρεάζεται και από άλλους παράγοντες όπως ο δείκτης μάζας σώματος του ατόμου, τα επίπεδα άγχους του ασθενή, ο χρόνος ύπνου του, η παρουσία κάποιας ασθένειας, εάν είναι καπνιστής, ιατρικές αγωγές και φάρμακα που μπορεί να λαμβάνει, αλκοολισμός και αλλεργίες. Η μέθοδος της έρευνας είναι να δημιουργήσει στρατηγικές μοντελοποίησης και στρατηγικές machine learning πάνω σε συστήματα πρόβλεψης γλυκόζης στο αίμα για διαβήτη τύπου 1. Οι πληροφορίες που πάρθηκαν από τα άρθρα έγινε βάση κάποιων κατηγοριών όπως ηλικία ασθενών, τύπος εισαγωγής δεδομένων, μορφή και μέγεθος δεδομένων, προεπεξεργασία δεδομένων, κλάση μηχανικής μάθησης, αλγόριθμος που χρησιμοποιήθηκε για εκμάθηση(training), τρόπος τεχνικής επικύρωσης, μετρήσεις απόδοσης αλλά και κριτήρια αξιολόγησης. Όπως φαίνεται το άρθρο που εξετάζεται σε σχέση με το κυρίως θέμα είναι εντελώς διαφορετικού χαρακτήρα καθώς η έρευνα που εξετάζεται εδώ αφορά μετρήσεις και ανάλυση χρησιμοποιώντας μηχανική μάθηση και είναι ξεκάθαρα ερευνητικού χαρακτήρα, ενώ το κυρίως θέμα είναι πάνω στο πρακτικό κομμάτι και αφορά μετρήσεις μηχανικής μάθησης για μελλοντικές προβλέψεις. Τα δεδομένα όπου χρησιμοποιούν είναι παρόμοια καθώς γίνονται μετρήσεις όπως ινσουλίνης στο αίμα αλλά υπάρχουν αρκετές διαφορές καθώς η έρευνα του άρθρου αφορά ανθρώπους με διαβήτη τύπου 1 ενώ το κυρίως θέμα αφορά ανθρώπους με διαβήτη τύπου 2, συνεπώς τα datasets είναι διαφορετικά. Ωστόσο και τα δύο θέματα δίνουν ιδιαίτερη έμφαση σε παράγοντες που μπορεί να επηρεάζονται τα δεδομένα του dataset όπως το σωματικό βάρος, ηλικία κ.τ.λ. Το άρθρο το οποίο αναλύθηκε χρησιμοποίησε πολλούς διαφορετικούς αλγορίθμους μηχανικής μάθησης καθώς είναι ερευνητικού χαρακτήρα σε αντίθεση με το κυρίως θέμα που χρησιμοποιήθηκε ο αλγόριθμος j48 του εργαλείου weka και η δημιουργία της εφαρμογής έγινε χρησιμοποιώντας δέντρα αποφάσεων. Ο όγκος των δεδομένων τα οποία χρησιμοποιήθηκαν είναι και στις 2 περιπτώσεις μεγάλος καθώς στο άρθρο οι συλλογή τους έγινε από μια μεγάλη επιλογή μέσα από ιστοσελίδες ερευνητικού περιεχομένου όπως IEEE Xplore, google Scholar, ScienceDirect, PubMed κ.α. , ενώ στην έρευνα του κυρίως θέματος χρειάστηκε ένας μεγάλος όγκος δεδομένων για να γίνει σωστή

εκμάθηση(training) στον αλγόριθμο για να πάρουμε αξιόπιστα αποτελέσματα για το σύστημα το οποίο κατασκευάστηκε.

Στο άρθρο[2] γίνεται η ανάλυση ενός συστήματος μηχανικής μάθησης για την πρόβλεψη και της ανάλυση του διαβήτη χρησιμοποιώντας τεχνικές που έγιναν μέσω της γλώσσας προγραμματισμού R και τα εργαλεία της. Το Dataset που επιλέχθηκε αποτελείται από γυναίκες ηλικίας τουλάχιστον 21 ετών ,το οποίο περιέχει τουλάχιστον 768 στοιχεία στους πίνακες του και κατηγοριοποιήθηκαν ως διαβητική γυναίκα η όχι διαβητική. Τα στοιχεία που περάστηκαν εντός των πινάκων είναι πόσες φορές υπήρξε έγκυος η γυναίκα, συγκέντρωση γλυκόζης, διαστολική αρτηριακή πίεση, διπλώσεις δέρματος τρικέφαλου, ινσουλίνη στο αίμα, δείκτης μάζας σώματος, γενεαλογικό δέντρο διαβήτη και τέλος της ηλικία της. Στη διαδικασία της προεπεξεργασίας των δεδομένων εφαρμόστηκε επιλογή χαρακτηριστικών, αφαίρεση ακραίων τιμών καθώς και χρήση του KNN για τις τιμές που λείπουν. Για την πρόβλεψη εάν κάποιος είναι διαβητικός η όχι χρησιμοποιήθηκαν πέντε αλγόριθμοι, οι οποίοι είναι οι εξής, Linear kernel(RBF), Kernel support vector machine(SVM), K-Nearest neighbor(KNN), Artificial neural network(ANN) και τέλος ο multifactor dimensionality reduction(MDR). Ένα ακόμη εξίσου σημαντικό στοιχείο που πρέπει να αναφερθεί είναι ότι στην διαδικασία της προεπεξεργασίας εφαρμόστηκε ο αλγόριθμος Boruta wrapper έτσι ώστε να βρεθούν τα χρήσιμα στοιχεία που χρειάζονται οι αλγόριθμοι που χρησιμοποιούνται. Ο τρόπος με τον οποίο δουλεψήκανε οι αλγόριθμοι είναι με 70% training data και τα υπόλοιπα 30% χρησιμοποιήθηκαν για να γίνουν τα κατάλληλα test. Η διαδικασία που χρησιμοποιήθηκε είναι ten-fold cross validation.

Τα αποτελέσματα της έρευνας είναι :

Linear Kernel SVM □ 84%

Radial Basis Kernel SVM □ 89%

K-NN □ 88%

ANN □ 86%

MDN □ 83%

Όπως φαίνεται από τα παραπάνω αποτελέσματα το σύστημα το οποίο αναπτύχθηκε στην παρακάτω έρευνα είναι αρκετά αποτελεσματικό και τέτοιου είδους συστήματα είναι απαραίτητα πλέον για τον έλεγχο του διαβήτη. Το άρθρο περιέχει αρκετές ομοιότητες με το

κυρίως θέμα της εργασίας καθώς και οι δύο έρευνες έχουν ως στόχο να προβλέψουν αν κάποιος έχει διαβήτη, ωστόσο το σύστημα που αναπτύχθηκε στο κυρίως θέμα προχωρά ένα βήμα παρακάτω δίνοντας την δυνατότητα στον ασθενή να ελέγξει και επιπλέον επιπλοκές που μπορεί να έχει στο μέλλον. Τα δύο συστήματα δημιουργήθηκαν χρησιμοποιώντας δύο εντελώς διαφορετικούς αλγορίθμους καθώς το σύστημα του άρθρου έγινε ελέγχοντας και δοκιμάζοντας πέντε διαφορετικούς αλγορίθμους ενώ το σύστημα της εργασίας δημιουργήθηκε χρησιμοποιώντας τον αλγόριθμο Naïve Bayes, K nearest neighbor, J48 και αξιοποιώντας τα δέντρα αποφάσεων του αλγορίθμου. Ωστόσο και τα δύο συστήματα χρησιμοποιούν cross-form validation με 70% training και 30% testing. Τα αποτελέσματα των δύο ερευνών ήταν αρκετά καλά και αξιόπιστα και στις 2 περιπτώσεις, όπου έχουμε στο σύστημα του άρθρου [2] 89% ενώ στο σύστημα της εργασίας 85%. Τα εργαλεία που χρησιμοποιήθηκαν διαφέρουν καθώς το σύστημα που αναλύθηκε στο άρθρο υλοποιήθηκε με τα εργαλεία της R ενώ το σύστημα της εργασίας μέσω του εργαλείου WEKA. Τέλος αξίζει να αναφερθεί ότι και τα δύο συστήματα στοχεύουν σε ανθρώπους με διαβήτη, παρόλα αυτά το άρθρο αναλύει ένα σύστημα γενικής πρόβλεψης διαβήτη και κυρίως σε γυναίκες ενώ το σύστημα της εργασίας απευθύνεται σε ανθρώπους με διαβήτη τύπου 2 και ανεξαρτήτως φύλου.

Το άρθρο[3] προτείνει ένα σύστημα μηχανικής μάθησης το οποίο θα διαβάζει την γλυκόζη και την πίεση του ασθενή και θα ειδοποιεί τον γιατρό σε περίπτωση που υπάρχουν ανησυχητικά αποτελέσματα. Το σύστημα θα παίρνει κάποιες αποφάσεις μέσω ενός συστήματος μηχανικής μάθησης και αποφάσεις δένδρων και θα προβλέπει εάν ο ασθενής έχει υπέρταση και διαβήτη τύπου 2. Ο αλγόριθμος ο οποίος φαίνεται να είχε τα καλύτερα αποτελέσματα είναι ο SVM ο οποίος και χρησιμοποιήθηκε στο σύστημα. Στόχοι αυτού του συστήματος είναι να βελτιώσουν την υγεία πάνω στα προβλήματα υπέρτασης, να αναπτυχθεί η εφαρμογή ως μια Desktop εφαρμογή για να μπορεί ο χρήστης να την έχει στον προσωπικό του υπολογιστή, να γίνει η ανάλυση ενός τέτοιου προβλήματος με αυτοματοποιημένο τρόπο και να μπορούν να δουλέψουν οι γιατροί με το σύστημα σε συνδυασμό για βελτιωθεί η υγεία του ασθενή. Το σύστημα κατηγοριοποιεί τα αποτελέσματα της πίεσης ανάλογα με τον βαθμό επικινδυνότητας τους και χρησιμοποιεί αυτά τα αποτελέσματα για να προβλέψει αν ο ασθενής είναι διαβητικός ή όχι. Το Dataset που χρησιμοποιήθηκε για να κατασκευαστεί το σύστημα περιέχει τα εξής δεδομένα που εξετάζονται όπως, πόσες φορές υπήρξε έγκυος ο ασθενής(εάν είναι γυναίκα), τα επίπεδα γλυκόζης, το γενεαλογικό δένδρο του ασθενή, πίεση, πάχος του δέρματος, ινσουλίνη, δείκτης

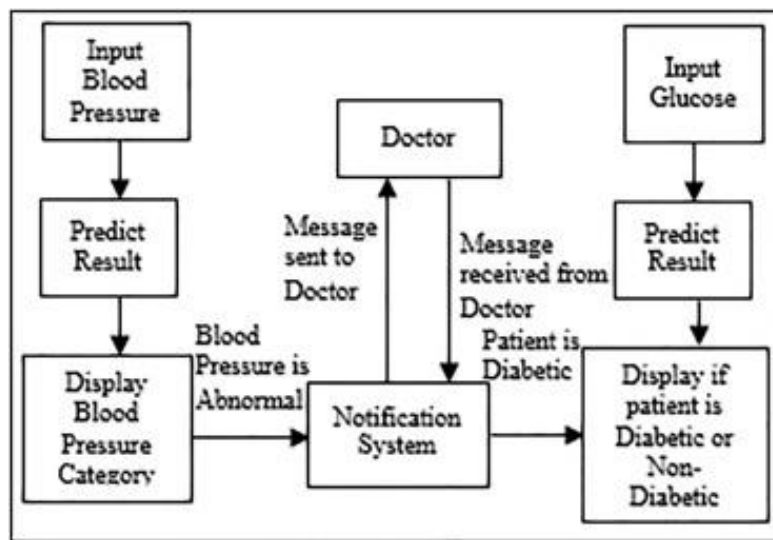
μάζας σώματος, ηλικία και το αν είναι διαβητικός η όχι ο ασθενής το οποίο παρουσιάζεται με τους αριθμούς 0 η 1, δηλαδή όχι η ναι αντίστοιχα. Το dataset που χρησιμοποιήθηκε πρώτα υπέστη προεπεξεργασία καθώς κάποιες τιμές στο dataset ξεκινούσαν με 0 ως τιμή τους πράγμα που είναι αδύνατο καθώς η ινσουλίνη παράδειγμα δεν μπορεί να πάρει αυτή την τιμή. Τέλος για την σωστή προεπεξεργασία έγινε καθάρισμα των δεδομένων αφαιρώντας όλα τα δεδομένα που έχουν μη εφικτές τιμές η λάθος δεδομένα. Το σύστημα το οποίο εξετάζει αυτό το άρθρο χρησιμοποιεί ως αλγόριθμο πρόβλεψης τον αλγόριθμο SVM ο οποίος τρέχει με 75% accuracy. Το σύστημα σχεδιάστηκε για να λειτουργεί ως Desktop εφαρμογή με graphical interface όπου θα παρουσιάζεται γραφικά η απεικόνιση των αποτελεσμάτων και θα εμφανίζονται ειδοποιητήρια μηνύματα στον χρήστη του προγράμματος ανάλογα με την κατάσταση του.

Το σύστημα για να βρει την κατηγορία της πίεσης, δηλαδή πόσο επικίνδυνη είναι λειτουργεί με βάση τον παρακάτω κανόνα

- Επικίνδυνα χαμηλή : Εάν η μεγάλη πίεση ή πίεση συστολής είναι λιγότερο του 40 και η μικρή πίεση η αλλιώς πίεση διαστολής μικρότερο του 20.
- Χαμηλή : Εάν η πίεση συστολής είναι μεταξύ του 40 και 90 και η πίεση διαστολής είναι μεταξύ του 20 και του 60.
- Φυσιολογική : Εάν η πίεση συστολής είναι μεταξύ του 90 και 140 και πίεση διαστολής είναι μεταξύ του 60 και του 90.
- Μεγάλη : : Εάν η πίεση συστολής είναι μεταξύ του 140 και 180 και η πίεση διαστολής είναι μεταξύ του 90 και του 120.
- Επικίνδυνα Υψηλή : Εάν η πίεση συστολής είναι πάνω του 180 και η πίεση διαστολής είναι μεγαλύτερη του 120.

Το σύστημα για να δοκιμάσει και να τρέξει τον αλγόριθμο χρησιμοποιεί 80% training και 20% για testing χρησιμοποιώντας ten-cross validation. Το σύστημα εκτός από SVM που έφερε τα καλύτερα αποτελέσματα δοκιμάστηκε και KNN, Decision tree κ.α.

1.1.Αρχιτεκτονική Συστήματος Άρθρου 2



Όπως φαίνεται από το άρθρο η θεματολογία με το κυρίως θέμα είναι αρκετά κοντά η μια με την άλλη καθώς στοχεύουν σε ανθρώπους με διαβήτη τύπου 2, ωστόσο τα αποτελέσματα είναι διαφορετικά αφού τα αποτελέσματα του συστήματος του άρθρου αφορά το αν κάποιος είναι διαβητικός ή όχι ενώ το σύστημα της εργασίας ελέγχει αν κάποιος είναι διαβητικός και στη συνέχεια ελέγχει και αν πρέπει να παρακολουθεί την θεραπεία του σύμφωνα με τα στοιχεία που θα εισάγει στο σύστημα. Ωστόσο το σύστημα του άρθρου δεν παρακολουθεί την θεραπεία αλλά ελέγχοντας τα στοιχεία στέλνει τα κατάλληλα μηνύματα στον γιατρό όταν αυτό θεωρηθεί κατάλληλο. Συνεπώς θα μπορούσαμε να πούμε ότι η δυνατότητα αυτή θα μπορούσε να είναι μια μελλοντική επέκταση και της εφαρμογής της εργασίας. Τα δεδομένα του dataset έχουν αρκετές διαφορές αλλά υπάρχουν και κάποιες ομοιότητες που χρησιμοποιήθηκαν στα δέντρα απόφασης όπως η ινσουλίνη και αυτό οφείλεται κυρίως στο ότι και τα δύο συστήματα στοχεύουν σε ανθρώπους με διαβήτη τύπου 2. Ο αλγόριθμος ο οποίος χρησιμοποιήθηκε είναι διαφορετικός καθώς στο σύστημα του άρθρου χρησιμοποιήθηκε ο SVM με accuracy 75% ενώ στο σύστημα της εργασίας χρησιμοποιήθηκε ο j48 με accuracy 85%, ενώ και στις δύο περιπτώσεις χρησιμοποιήθηκε K-Cross Validation με K=10.

Η έρευνα του άρθρου[4] επικεντρώνεται στην ανάλυση του διαβήτη και πως μπορούμε να βελτιώσουμε την περίθαλψη του για νέους και ηλικιωμένους ασθενείς χρησιμοποιώντας το εργαλείο ODM της Oracle. Στο σύστημα οι τιμές περαστήκανε ως $p(o)$ για ηλικιωμένους και ως $p(y)$ για άτομα νεότερης ηλικίας. Η έρευνα αυτή προσπαθεί μέσα από της εξόρυξη δεδομένων να καθορίσει την καλύτερη αγωγή που πρέπει να πάρει κάποιος για την

αντιμετώπιση του διαβήτη ανάλογα με την ηλικία του. Το dataset αφορά ασθενείς από της Σαουδική Αραβία και αποτελείται από ανθρώπους όλων των ηλικιών. Τα δεδομένα τα οποία χρησιμοποιήθηκαν για την δημιουργία του συστήματος είναι ο αριθμός φαρμάκων που χορηγείται στον ασθενή, η διατροφή του, το βάρος του ασθενή, εάν ο ασθενής είναι καπνιστής, εάν ασκείται ο ασθενής, καθώς και εάν του έχει χορηγηθεί ινσουλίνη. Για τα δεδομένα εφαρμόστηκε ο αλγόριθμος predictive regression έτσι ώστε να βρεθεί ποια θεραπεία είναι η καταλληλότερη για κάθε ηλικία. Για την συλλογή δεδομένων χρησιμοποιήθηκε το εργαλείο της Oracle, το Oracle Data Miner(ODM).

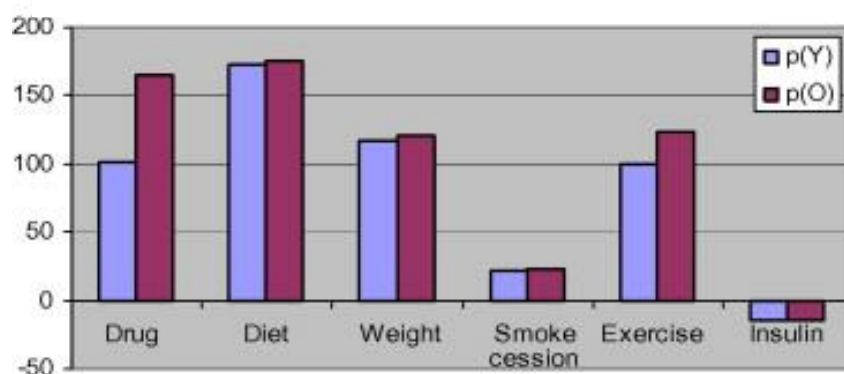
Η συλλογή δεδομένων έγινε μέσα από έξι στάδια:

- Στάδιο 1: Συλλογή δεδομένων και προεπεξεργασίας, έτσι ώστε να συμπληρωθούν κενές τιμές και την διόρθωση λαθών στα δεδομένα καθώς και μη εφικτές τιμές που μπορεί να υπήρχαν.
- Στάδιο 2: Προετοιμασία αρχείου δεδομένων: Το εργαλείο της Oracle(ODM) δεν μπορούσε να επεξεργαστεί την αρχική μορφή .xls και .mdb καθώς έπρεπε να γίνει μετατροπή σε κατάλληλη μορφή αρχείου για το εργαλείο της Oracle.
- Στάδιο 3: Ανάλυση δεδομένων: Σε αυτό το στάδιο γίνεται η ανάλυση δεδομένων με έναν από τους γνωστούς αλγορίθμους ανάλυσης δεδομένων. Στο συγκεκριμένο σύστημα χρησιμοποιήθηκε ο αλγόριθμος SVM.
- Στάδιο 4: Αποτελέσματα δεδομένων: Σε αυτό το στάδιο θα φανούν τα τελικά αποτελέσματα του αλγορίθμου που πραγματοποιήθηκαν στο προηγούμενο στάδιο.
- Στάδιο 5: Πρόβλεψη βάση δεδομένων και ανασκόπηση δεδομένων: Σε αυτό το στάδιο παίρνουμε χρήσιμες πληροφορίες από τα αποτελέσματα και εξάγουμε δεδομένα και νέες τεχνικές από τα αποτελέσματα. Επίσης έχουμε γραφικές απεικονίσεις σύμφωνα με τα αποτελέσματα.
- Στάδιο 6: Παραγωγή: Σε αυτό το στάδιο γίνονται οι προβλέψεις βάση της επεξεργασίας των δεδομένων που περαστήκανε στο σύστημα.

Τα αποτελέσματα της έρευνας δείχνουν ότι οι άνθρωποι μεγαλύτερης ηλικίας $p(o) > 44$ είναι προτιμότερο να ακολουθούν την ιατρική οδό και να λαμβάνουν ιατρικές αγωγές, ενώ τα άτομα μικρότερης ηλικίας $p(y) \leq 44$ θα πρέπει να ακολουθούν άλλους τρόπους όπως καλύτερη διατροφή, γυμναστική κ.α. Η έρευνα απευθύνεται σε ανθρώπους με διαβήτη τύπου

και διαβήτη τύπου 2. Μέσω των αποτελεσμάτων αντλήθηκαν και άλλες πληροφορίες όπως ότι πρέπει και στις δύο κατηγορίες να γίνεται προσοχή στην διατροφή τους, όπως και ότι θα πρέπει και στις δύο κατηγορίες να μειωθεί η δόση ζάχαρης που καταναλώνεται. Αλλά και άλλες χρήσιμες πληροφορίες εκτός από τις παραπάνω άρα αυτό μας δίνει το συμπέρασμα ότι η διαδικασία του Data mining μπορεί να βοηθήσει της πρόοδο όχι μόνο μέσα από το σύστημα το οποίο δημιουργείται αλλά και από την σκοπιά των επιπλέον πληροφοριών που αντλούμε. Μέσα από την έρευνα μπορούμε να πούμε ότι αυτά που υπερισχύουν και πρέπει να προσεχθούν για τους νέους είναι η σωστή διατροφή, τα περιττά κιλά, η αποφυγή αχρείαστων κιλών, το κάπνισμα καθώς και η ινσουλίνη. Στις μεγαλύτερες ηλικίες θα πρέπει να προσεχθούν η διατροφή, η αγωγή φαρμάκων που θα τους χορηγηθεί, η σωματική άσκηση, το βάρος του ασθενή, το κάπνισμα και τέλος η ινσουλίνη.

1.2 Σύγκριση δεδομένων για την πρόβλεψη του διαβήτη σύμφωνα με το άρθρο 4



Τα δύο συστήματα έχουν αρκετές ομοιότητες ως προς την έρευνα και το περιεχόμενό τους αφού και τα δύο απευθύνονται σε άτομα με διαβήτη και πως θα βελτιωθεί η πρόσληψη του αλλά και την συλλογή σημαντικών πληροφοριών μέσα από την έρευνα πάνω τους. Μια μεγάλη διαφορά ανάμεσα στις δύο έρευνες είναι ότι το άρθρο απευθύνεται σε ανθρώπους με διαβήτη και των δύο ειδών ενώ το θέμα της εργασίας αφορά και απευθύνεται σε ανθρώπους με διαβήτη τύπου 2. Η αλγόριθμοι που χρησιμοποιήθηκαν στις δύο περιπτώσεις είναι ο SVM για το σύστημα που αναπτύχθηκε ώστε να αντληθούν οι πληροφορίες από το άρθρο χρησιμοποιώντας το εργαλείο της Oracle ODM, ενώ στο κυρίως θέμα χρησιμοποιήθηκε ο J48 του εργαλείου WEKA. Ο τρόπος με τον οποίο αναπτύχθηκαν τα δύο συστήματα είναι διαφορετικός καθώς στο άρθρο μας περιγράφει ότι η ανάπτυξη έγινε με μια σειρά σταδίων όπου σε κάθε στάδιο

αντλούνταν πληροφορίες ώστε να γίνει η συλλογή δεδομένων ενώ στην εργασία ο τρόπος ανάπτυξης είναι μέσω δέντρων απόφασης όπου για κάθε επιλογή ακολουθήθηκε και ένα μονοπάτι με διαφορετικό αποτέλεσμα σύμφωνα με τα δεδομένα του χρήστη.

Το άρθρο[5] ξεκινά κάνοντας μια πολύ ενδιαφέρον παρατήρηση επισημαίνοντας ότι στον κόσμο της μηχανικής μάθησης οι περισσότερες έρευνες υποθέτουν ότι τα δεδομένα θα είναι ξεχωριστά το ένα με το άλλο και το ίδιο μοιραζόμενα μέσα στο σύνολο. Αυτό σύμφωνα με το άρθρο τα κάνει αναξιόπιστα για μακροχρόνιες προβλέψεις και χρήση. Σε αυτή την έρευνα δημιουργήθηκε μία κατασκευή η οποία εντάσσει τα τυχαία αποτελέσματα του GLMM σε ένα σύστημα μηχανικής μάθησης ικανό να προβλέψει διαφορετικών ειδών δεδομένα. Ο αλγόριθμος που χρησιμοποιήθηκε είναι ο MEm1 για να γίνει η σωστή πρόβλεψη μακροχρόνιων αλλαγών στον γλυκαιμικό έλεγχο ο οποίος μετρήθηκε με βάση την ουσία A1c μέσα σε ενήλικους με διαβήτη τύπου 2. Τα αποτελέσματα της έρευνας δείχνουν ότι ο MEm1 είναι ανταγωνιστικός σε συνδυασμό με τον παραδοσιακό GLMM και πως δίνει καλύτερα αποτελέσματα από άλλα μοντέλα μηχανικής μάθησης τα οποία δεν λαμβάνουν υπόψη τους τον τυχαίο παράγοντα στα δεδομένα. Συνεπώς μπορούμε να πούμε ότι η έρευνα αυτή μας δείχνει το πόσο σημαντικό είναι οι ερευνητές να λαμβάνουν υπόψη τους τον τυχαίο παράγοντα που μπορεί να υπάρχει στα δεδομένα. Το σύστημα το οποίο αναπτύχθηκε για να ολοκληρώσει τις έρευνες του έκανε δοκιμές πάνω σε ασθενείς τέσσερις φορές. Τα χαρακτηριστικά του ασθενή που λήφθηκαν υπόψη είναι φύλλο, καταγωγή, νοσήματα που μπορεί να έχει ο ασθενής, γεωγραφική περιοχή στην οποία κατοικεί, το τρέχον HbA1c καθώς και αν γίνεται χρήση κάποιας αγωγής. Για την καλύτερη χρήση του συστήματος ο ερευνητής έκανε δοκιμές και σε άλλα dataset διαφορετικών θεμάτων χρησιμοποιώντας το σύστημα έτσι ώστε να διαπιστώσει εάν είναι αποτελεσματικό. Για την αποτελεσματικότερη και ακριβέστερη καταγραφή των αποτελεσμάτων αποκλειστήκανε από το dataset όλοι οι ασθενείς οι οποίοι κάνουν χρήση ινσουλίνης 12 μέρες πριν την δειγματοληψία, ασθενής με διαβητική κετοξέωση, διαβητικό κόμμα, υπεργλυκαιμική υπερομοριακή κατάσταση, μεγάλο πρόβλημα υπογλυκαιμίας και η χρήση αντιδιαβητικών φαρμάκων το τελευταίο έτος. Για να προβλέψει το σύστημα αν ένας ασθενής θα χάσει τον γλυκαιμικό έλεγχο, χρησιμοποιεί πληροφορίες από τις τρέχον αλλά και παλιές επισκέψεις καθώς και πληροφορίες σχετικά με τις υποθέσεις για το αν θα υπάρξει το συγκεκριμένο πρόβλημα στο μέλλον. Ο αλγόριθμος χρησιμοποιήθηκε με 63% training και το υπόλοιπο ως testing χρησιμοποιώντας K-cross validation.

Όπως φαίνεται το σύστημα το οποίο αναπτύχθηκε στην έρευνα σε σχέση με το σύστημα της εργασίας έχουν αρκετές ομοιότητες αλλά και πολλές διαφορές. Η κύρια ομοιότητα τους είναι ότι απευθύνονται και τα δύο συστήματα σε ανθρώπου με διαβήτη και συγκεκριμένα με διαβήτη τύπου 2. Ακόμη για την σωστή αξιολόγηση του δοκιμάστηκε τρέχοντας διαφορετικούς αλγορίθμους και χρησιμοποιήθηκαν και στις δύο περιπτώσεις δέντρα αποφάσεις. Ωστόσο στο σύστημα του άρθρου τα αποτελέσματα προέρχονται από τέσσερις διαφορετικές μετρήσεις που παρουσιάζουν μελλοντικά αποτελέσματα ενώ στο σύστημα της εργασίας έχουμε ένα αποτέλεσμα, δηλαδή εάν κάποιος πρέπει να προσέξει καθώς βρίσκεται σε κίνδυνο σύμφωνα με τα δεδομένα που έδωσε στο σύστημα ή όχι. Ο αλγόριθμος που χρησιμοποιήθηκε για την δημιουργία του συστήματος είναι ο MEMl και ο GLMM, σε αντίθεση με τον αλγόριθμο της εργασίας όπου δημιουργήθηκε χρησιμοποιώντας τον j48 του εργαλείου WEKA. Οι αλγόριθμοι και ο τρόπος λειτουργίας τους είναι εντελώς διαφορετική καθώς ο MEMl και GLMM λειτουργούν με δεδομένα τα οποία είναι τυχαία κατανεμημένα σε αντίθεση με τον j48 ο οποίος τα κατανέμει όμοια μεταξύ τους. Τέλος πρέπει να αναφερθεί ότι και στις δύο περιπτώσεις έγινε k-cross validation με $k = 10$ αλλά με διαφορετικά ποσοστά σε training και testing των δεδομένων.

Στο άρθρο[6] αυτό προσπαθεί ο ερευνητής να δείξει πόσο σημαντικός είναι ο φαινότυπος των υπερτρυγκαιμιδίων σε συνδυασμό με ανθρωπομετρήσεις για την πρόβλεψη χρήσιμων δεδομένων για τον διαβήτη τύπου 2. Για την έρευνα αυτή έγινε μέτρηση της γλυκόζης πλάσματος σε συνδυασμό με τα τριγλυκερίδια για την μέτρηση ανθρώπινων δεδομένων σχετικά με τον διαβήτη. Ο αλγόριθμος ο οποίος χρησιμοποιήθηκε είναι ο logistic binary regression για να γίνει μέτρηση μεταξύ των ανθρώπων χωρίς πρόβλημα διαβητικό, έναντι αυτών που πάσχουν από διαβήτη τύπου 2. Για ακόμη πιο αξιόπιστα αποτελέσματα χρησιμοποιήθηκαν ακόμη 2 αλγόριθμοι, ο αλγόριθμος Naïve bayes αλλά και ο αλγόριθμος Logistic Regression για να βρεθούν οι δυνατότητες πρόβλεψης διαφόρων φαινοτύπων. Όλες οι προβλέψεις γίνανε χρησιμοποιώντας K-cross validation. Ένα ενδιαφέρον αποτέλεσμα είναι ότι η σχέση μεταξύ υπερτρυγκαιμιδίων και διαβήτη τύπου 2 είναι μεγαλύτερη από ότι τριγλυκεριδίων με διαβήτη τύπου 2. Για να γίνουν οι απαραίτητες μετρήσεις για την δημιουργία του συστήματος συλλέχθηκαν 11937 διαφορετικά στοιχεία ασθενών από διάφορες χώρες. Όλοι οι ασθενείς που συλλέχθηκαν για να δημιουργηθεί η έρευνα τους είχε ζητηθεί να μην φάνε τις τελευταίες 8 ώρες πριν να γίνουν οι απαραίτητες μετρήσεις τριγλυκεριδίων και πλάσματος γλυκόζης αλλά και κάποιες επιπλέον αιματολογικές εξετάσεις. Τα εργαλεία τα

οποία χρησιμοποιήθηκαν για να δημιουργηθεί το σύστημα είναι το SPSS 19 για windows αλλά και το εργαλείο WEKA για ανάλυση δεδομένων και την εξόρυξη τους. Η μεθοδολογία που ακολουθήθηκε είναι η εξής, στατιστική ανάλυση του συσχετισμού μεταξύ της υπεργλυκαιμίας, προσωπικών μετρήσεων και διαβήτη τύπου 2, καθώς και ανάλυση των προβλέψεων των προσωπικών μετρήσεων και φαινοτύπων. Αυτό έγινε χρησιμοποιώντας δύο αλγόριθμους. Για την στατιστική εξέταση των δεδομένων χρησιμοποιήθηκε binary logistic regression και για την σύγκριση των δυνατοτήτων πρόβλεψης όλων των φαινοτύπων για την διάγνωση του διαβήτη τύπου 2 χρησιμοποιήθηκε ο αλγόριθμος Naïve bayes αλλά και Logistic Regression για να αποκτήσουμε πιο αξιόπιστα αποτελέσματα. Όλοι οι αλγόριθμοι τρέξαν με 10-fold cross validation. Τα αποτελέσματα της έρευνας είναι ότι για τους άνδρες η υπεργλυκαιμία έχει την μεγαλύτερη σχέση με τον διαβήτη τύπου 2 ενώ για τις γυναίκες δεν βγαίνει μια σαφή ένδειξη για το τι έχει την μεγαλύτερη σχέση με διαβήτη τύπου 2 και άρα μπορεί να προέρχεται από οποιοδήποτε χαρακτηριστικό. Μέσω ερευνών ακόμη βλέπουμε ότι η δυνατότητα πρόβλεψης για τους άντρες ήταν καλύτερη χρησιμοποιώντας υπερτρυγλυκαιμικά χαρακτηριστικά ενώ τα λιγότερα ακριβές ήταν οι μετρήσεις που έγιναν με βάση τα τριγλυκερίδια. Στις γυναίκες ο καλύτερος παράγοντας πρόβλεψης φαίνεται είναι και εκεί τα υπερτρυγλυκαιμικά. Συνεπώς η έρευνα μας δείχνει ότι τα υπερτρυγλυκαιμικά είναι ο πιο σημαντικός παράγοντας για να κάνει κάποιος συσχέτιση με τον διαβήτη τύπου 2.

1.3 Ανάλυση των συσχετίσεων μεταξύ διαβήτη τύπου 2 και ανθρωπομετρικών ενδείξεων για Άνδρες

	p	OR		p	OR ^a
HW phenotype	<0.001	2.01 (1.68-2.41)		<0.0001	2.07 (1.72-2.49)
Weight	<0.001	1.31 (1.22-1.40)		<0.0001	1.56 (1.45-1.68)
BMI	<0.001	1.39 (1.29-1.49)		<0.0001	1.53 (1.42-1.65)
NeckC	<0.001	1.49 (1.39-1.60)		<0.0001	1.61 (1.49-1.73)
ChestC	<0.001	1.51 (1.41-1.62)		<0.0001	1.60 (1.49-1.73)
RibC	<0.001	1.61 (1.50-1.73)		<0.0001	1.64 (1.52-1.77)
WaistC (WC)	<0.001	1.59 (1.48-1.71)		<0.0001	1.60 (1.48-1.72)
HipC	<0.001	1.19 (1.11-1.28)		<0.0001	1.30 (1.21-1.40)
Neck_Hip	<0.001	1.35 (1.26-1.44)		<0.0001	1.33 (1.24-1.43)
Rib_Hip	<0.001	1.63 (1.52-1.76)		<0.0001	1.60 (1.48-1.73)
Waist_Hip (WHR)	<0.001	1.73 (1.60-1.86)		<0.0001	1.66 (1.54-1.80)
Forehead_Waist	<0.001	0.60 (0.56-0.65)		<0.0001	0.62 (0.58-0.68)
Forehead_Rib	<0.001	0.59 (0.55-0.64)		<0.0001	0.61 (0.56-0.66)
Forehead_Neck	<0.001	0.65 (0.60-0.69)		<0.0001	0.64 (0.59-0.69)
WHR	<0.001	1.36 (1.27-1.46)		<0.0001	1.57 (1.46-1.69)
TG	<0.001	1.34 (1.26-1.44)		<0.0001	1.38 (1.28-1.47)

* Results were obtained after adjustment for age and region (investigation site).

1.4 Ανάλυση των συσχετίσεων μεταξύ διαβήτη τύπου 2 και ανθρωπομετρικών ενδείξεων για Γυναίκες

Index	Unadjusted		Adjusted	
	p	OR	p	OR*
HW phenotype	<0.001	2.83 (2.44-3.28)	<0.001	2.09 (1.79-2.45)
Weight	<0.001	1.33 (1.24-1.42)	<0.001	1.46 (1.36-1.57)
BMI	<0.001	1.56 (1.46-1.67)	<0.001	1.50 (1.39-1.61)
NeckC	<0.001	1.66 (1.55-1.78)	<0.001	1.63 (1.52-1.76)
ChestC	<0.001	1.85 (1.72-1.98)	<0.001	1.68 (1.56-1.81)
RibC	<0.001	1.95 (1.81-2.09)	<0.001	1.70 (1.57-1.84)
WaistC (WC)	<0.001	1.81 (1.68-1.94)	<0.001	1.54 (1.43-1.67)
HipC	<0.001	1.21 (1.13-1.30)	<0.001	1.21 (1.13-1.30)
Neck_Hip	<0.001	1.42 (1.33-1.52)	<0.001	1.36 (1.27-1.47)
Rib_Hip	<0.001	2.23 (2.07-2.40)	<0.001	1.91 (1.75-2.08)
Waist_Hip (WHR)	<0.001	1.99 (1.85-2.14)	<0.001	1.65 (1.52-1.80)
Forehead_Waist	<0.001	0.49 (0.45-0.53)	<0.001	0.60 (0.55-0.65)
Forehead_Rib	<0.001	0.44 (0.41-0.48)	<0.001	0.53 (0.48-0.58)
Forehead_Neck	<0.001	0.51 (0.48-0.55)	<0.001	0.57 (0.52-0.61)
WHR	<0.001	1.46 (1.36-1.56)	<0.001	1.49 (1.39-1.60)
TG	<0.001	1.50 (1.41-1.59)	<0.001	1.37 (1.29-1.46)

* Results were obtained after adjusting for age and region (investigation site).

Τα δύο συστήματα που αναπτύχθηκαν απευθύνονται στους ίδιους ασθενείς, δηλαδή ασθενείς με διαβήτη τύπου 2, ενώ και στις 2 περιπτώσεις χρησιμοποιήθηκε το εργαλείο WEKA ωστόσο για διαφορετικές λειτουργίες καθώς στο άρθρο χρησιμοποιήθηκε για εξόρυξη δεδομένων ώστε να αντληθούν σημαντικά στοιχεία για την σχέση διαβήτη τύπου 2 και διάφορων άλλων μετρήσεων όπως τριγλυκερίδια και τριγλυκαιμικά ενώ στην εργασία χρησιμοποιήθηκε για να γίνει classification και πρόβλεψη των αποτελεσμάτων της εφαρμογής που αναπτύχθηκε. Επίσης αξίζει να σημειωθεί ότι στο σύστημα που περιγράφει το άρθρο χρησιμοποιήθηκε και δεύτερο εργαλείο το SPSS της Microsoft. Μολονότι φαίνεται να χρησιμοποιούνται τα ίδια εργαλεία οι αλγόριθμοι που χρησιμοποιούνται είναι διαφορετικοί καθώς στην εργασία χρησιμοποιήθηκε ο αλγόριθμος j48 ενώ στο άρθρο που περιγράφεται χρησιμοποιήθηκε ένας συνδυασμός αλγορίθμων για να γίνει ο κατάλληλος συσχετισμός και να βρεθούν τα κατάλληλα αποτελέσματα. Τα δύο συστήματα τρέξαν με K-cross validation, k=10 με ένα αρκετά μεγάλο Dataset.

1.5 Σύγκριση των αποτελεσμάτων σε Balanced δεδομένα

Gender	Class	Logistic regression					Naïve Bayes classifier				
		AUC	Sensitivity	1-specificity	Precision	F-measure	AUC	Sensitivity	1-specificity	Precision	F-measure
Female	Normal	0.937	0.937	0.183	0.836	0.884	0.937	0.944	0.185	0.837	0.887
	High FPG		0.817	0.063	0.928	0.869		0.815	0.056	0.936	0.872
Male	Normal	0.724	0.584	0.256	0.696	0.635	0.713	0.586	0.27	0.685	0.631
	High FPG		0.744	0.416	0.641	0.689		0.73	0.414	0.638	0.681

The results were evaluated by tenfold cross validation. AUC: area under the receiver operating characteristic curve.

Στο άρθρο[7] παρουσιάζεται μια έρευνα στην οποία έγινε προσπάθεια πρόβλεψης γλυκόζης ανθρώπων που απείχαν από φαγητό τουλάχιστον 8 ώρες έτσι ώστε να γίνει η διάγνωση του διαβήτη τύπου ανάμεσα σε ενήλικους Κορεάτες. Το Dataset αποτελείται από 4870 ασθενής. Βασιζόμενο σε 37 ανθρωπομετρικά στοιχεία, έγινε σύγκριση της γλυκόζης χωρίς να έχει φάει ο ασθενής χρησιμοποιώντας εξατομικευμένες μετρήσεις έναντι ενωμένων μετρήσεων χρησιμοποιώντας δύο αλγορίθμους μηχανικής μάθησης. Το σύστημα που αναπτύχθηκε μας δείχνει ότι οι προβλέψεις γλυκόζης χωρίς να έχει φάει ο ασθενής χρησιμοποιώντας ανθρωπομετρικές μετρήσεις είναι αισθητά ανώτερες από εξατομικευμένες μετρήσεις και στις γυναίκες αλλά και στους άνδρες. Το dataset το οποίο χρησιμοποιήθηκε πρώτα πέρασε από φιλτράρισμα κάνοντας όλα του τα attributes σε balanced. Μερικές από τις ανθρωπομετρήσεις που χρησιμοποιήθηκαν στο Dataset είναι περιφέρεια μετώπου, περιφέρεια λαιμού, περιφέρεια στήθους, περιφέρεια λεκάνης, απόσταση λαιμού προς μέτωπο, απόσταση λαιμού προς στήθος και πολλά άλλα, εκτός από τις μετρήσεις αυτές όμως χρησιμοποιήθηκαν και κάποια βασικά στοιχεία όπως ηλικία, ύψος, βάρος και δείκτης μάζας. Βλέποντας τα είδη ανθρωπομετρικών στοιχείων που χρησιμοποιήθηκαν φαίνεται γιατί στο σύστημα μετρήθηκαν ξεχωριστά άντρες με γυναίκες καθώς η σωματική μορφολογία είναι εντελώς διαφορετική. Όλες οι αναλύσεις γίνανε χρησιμοποιώντας τα εργαλείο SPSS και το εργαλείο WEKA. Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι ο αλγόριθμος Logistic regression αλλά και ο αλγόριθμος Naïve Bayes. Ένα ακόμη πολύ σημαντικό στοιχείο της έρευνας είναι ότι για να γίνουν οι μετρήσεις έγινε oversampling του dataset. Όλοι οι αλγόριθμοι χρησιμοποιούν tenfold cross validation. Τα αποτελέσματα τα οποία μετρήθηκαν είναι πως οι γυναίκες έχουν καλύτερα ποσοστά με accuracy 0.47 χρησιμοποιώντας logistic regression και 0.69 χρησιμοποιώντας naïve Bayes. Στους άντρες έχουμε 0.49 χρησιμοποιώντας logistic regression και 0.67 χρησιμοποιώντας Naïve Bayes. Στις αναλύσεις στις οποίες έγινε ο συνδυασμός δεδομένων φαίνεται μεγάλη διαφορά καθώς έχουμε για γυναίκες 0.74 και 0.73 για logistic regression και Naïve Bayes αντίστοιχα. Στους άντρες έχουμε 0.68 και για τους δύο αλγορίθμους.

Το θέμα της έρευνας που αναλύθηκε όπως φαίνεται είναι εντελώς διαφορετικό από το θέμα της εργασίας ωστόσο απευθύνεται σε ίδιους ανθρώπους με διαβήτη τύπου 2, το θέμα της εργασίας είναι πιο πρακτικό δημιουργώντας μια εφαρμογή ενώ το θέμα της έρευνας του άρθρου είναι περισσότερο ερευνητικό κάνοντας μετρήσεις για να βγάλει χρήσιμα συμπεράσματα. Τα συστήματα που αναπτύχθηκαν γίνανε με τον εργαλείο WEKA και τα δύο, ενώ στο άρθρο αναφέρεται ότι χρησιμοποιήθηκε και το εργαλείο SPSS. Οι αλγόριθμοι που

χρησιμοποιήθηκαν είναι εντελώς διαφορετικοί με το σύστημα του άρθρου να χρησιμοποιεί 2 αλγόριθμους, τον αλγόριθμο Logistic Regression και τον αλγόριθμο Naïve Bayes, ενώ στην εργασία χρησιμοποιήθηκε ο αλγόριθμος J48. Ο λόγος που χρησιμοποιήθηκαν δύο αλγόριθμοι στο άρθρο είναι για λόγους δοκιμών και σωστή άντληση στοιχείων. Το accuracy της εργασίας είναι αρκετά καλύτερο αλλά δεδομένου της διαφορετικότητας των θεμάτων είναι αναμενόμενο. Και στις δύο περιπτώσεις έγινε το σύστημα χρησιμοποιώντας tenfold cross validation με την διαφορά ότι το dataset του άρθρου πρώτα επεξεργάστηκε χρησιμοποιώντας διάφορα φίλτρα.

Το άρθρο[8] παραθέτει ένα σύστημα το οποίο κάνει εξόρυξη δεδομένων και προβλέπει τον σακχαρώδη διαβήτη βασιζόμενος σε χρόνιες ασθένειες του ασθενή, το σύστημα έγινε με τα εργαλεία της R και η έρευνα αφορά κατοίκους της βορειοδυτικής Νιγηρίας. Τα ποσοστά Accuracy του συστήματος είναι 97%. Η συλλογή των δεδομένων έγινε μέσω ερωτηματολογίων και συνεντεύξεων με ανθρώπους οι οποίοι πάσχουν από σακχαρώδη διαβήτη ή άλλες χρόνιες ασθένειες. Τα δεδομένα που χρησιμοποιήθηκαν είναι τα εξής, τύπος διαβήτη του ασθενή, ηλικία του ασθενή, επίπεδα γλυκόζης του ασθενή, διαστολική πίεση του ασθενή, ο δείκτης μάζας σώματος του ασθενή, συμπτώματα που αφορούν προβλήματα στο νεφρό, προβλήματα που αφορούν καρδιακά προβλήματα και συμπτώματα που αφορούν προβλήματα στα μάτια. Τα δεδομένα είναι είτε binary δηλαδή με απαντήσεις που αντιπροσωπεύονται ως 0 ή 1 ή μπορεί να είναι numeric που μπορεί να είναι αριθμητικές απαντήσεις.

Ο αλγόριθμος ο οποίος χρησιμοποιήθηκε είναι ο Logistic Regression ο οποίος είχε τα εξής αποτελέσματα:

1.6 Αποτελέσματα Αλγορίθμου

Variables	Coefficients	Std error	Z values	P values
constant	-48.12040	21.61366	-2.226	0.0260*
AGE	0.30854	0.19877	1.552	0.0246*
GLU	6.52985	2.70314	2415	0.0157*
BMI	-0.13504	0.09778	-1.381	0.1673
DBP	-0.06805	0.17965	-0.379	0.7048
KDYP	7.97545	4.15712	1.919	0.0550
HRTP	3.28995	3.06506	1.073	0.2831
EYEP	-1.68871	2.02458	-0.834	0.4042

Όπως φαίνεται η έρευνα του άρθρου έχει αρκετές διαφορές με την εργασία που παρατέθηκε, αρχικά θα πρέπει να αναφερθεί ότι τα δύο συστήματα φτιάχνονται σε δύο εντελώς διαφορετικά εργαλεία, με το σύστημα του άρθρου να φτιάχνετε χρησιμοποιώντας τα εργαλεία της R ενώ το σύστημα της εργασίας δημιουργήθηκε με το εργαλείο WEKA. Τα θέματα των δύο ερευνών είναι εντελώς διαφορετικά καθώς το ένα απευθύνεται σε εξόρυξη δεδομένων και πρόβλεψη με βάση των μακροχρόνιων ασθενειών του ασθενή αλλά και του σακχαρώδους διαβήτη ενώ από την άλλη το σύστημα της εργασίας απευθύνεται σε πιο συνηθισμένα δεδομένα για να κάνει τις προβλέψεις του. Οι συλλογή των δεδομένων στο άρθρο έγινε με ερωτηματολόγια, συλλογή από διάφορες όμοιες έρευνες αλλά και συνεντεύξεις ενώ στην εργασία τα δεδομένα που συλλέχθηκαν είναι από έρευνα διαφορετικού συστήματος με όμως τις ίδιες ανάγκες σε δεδομένα.

Σε αυτό το paper[9] προτείνεται ένα νέο σύστημα για σαφέστερη διάγνωση του διαβήτη και καλύτερη προεπεξεργασίας των δεδομένων. Το σύστημα χρησιμοποιεί διαφορετικές απόψεις για τα ασαφή συμπεράσματα, των συλλογιστικών οντολογιών και μια ασαφή αναλυτική διαδικασία ιεραρχίας για να φτιάξουμε μια καλύτερη κατασκευή. Το σύστημα προσπαθεί να αποδείξει ακόμη ότι τα ασαφή συστήματα έχουν καλύτερα αποτελέσματα από τα συστήματα μηχανικής μάθησης. Το κύριο πρόβλημα ενός τέτοιου συστήματος είναι αρχικά η δυσκολία στην ερμηνευση του καθώς έχουν γίνει αρκετές αποτυχημένες προσπάθειες για να δημιουργηθούν τέτοιου είδους συστήματα. Ο τρόπος κατασκευής γίνεται χρησιμοποιώντας έναν ειδικό ώστε να εισάγει την γνώση απευθείας μέσα στο σύστημα. Το πρόβλημα σε αυτές τις περιπτώσεις είναι ότι μπορεί να είναι πολλά τα δεδομένα ή να μην υπάρχει ο ειδικός. Ο δεύτερος τρόπος είναι να χρησιμοποιηθεί ένα σύστημα για την εισαγωγή γνώσης και δεδομένων. Συνοψίζοντας θα πρέπει να πούμε ότι για την σωστή κατασκευή ενός ασαφούς συστήματος θα πρέπει να ψάξουμε για στοχευσιμότητα αποτελεσμάτων, ταχύτητα και αποτελεσματικότητα. Ένα ακόμη πρόβλημα είναι να διατυπωθούν σωστά τα λεξιλογικά δεδομένα. Η δημιουργία ενός τέτοιου συστήματος μόνο με αριθμητικά δεδομένα είναι αδύνατη και η σωστή αναπαράσταση λεξιλογικών δεδομένων έχει μεγάλη σημασία για την δημιουργία ενός τέτοιου συστήματος. Τα συστήματα σχετικά με το θέμα θα πρέπει να παρέχουν στον χρήστη τις μέγιστες δυνατότητες αυτοματοποίησης πληροφοριών. Μια εναλλακτική σε αυτό είναι να συνδυαστούν οι δυνατότητες της λογικής των δεδομένων με τις ιατρικές οντολογίες. Στο σύστημα που δημιουργήθηκε, φτιαχτήκανε διαφορετικοί αλγόριθμοι για διαφορετικές λειτουργίες λόγω της πολυπλοκότητας του

συστήματος. Τα βήματα του είναι τα εξής, αρχικά γίνεται η κατανομή, στη συνέχεια γίνεται η διέγερση κανόνων, έπειτα γίνεται η απλούστευση και τέλος γίνεται η εκτέλεση. Στη συνέχεια θα χτιστεί το συνδυασμένο ιεραρχικό σύστημα και θα δοκιμαστεί. Έπειτα επεκτείνεται το μοντέλο που προκύπτει ενσωματώνοντας την λογική οντολογίας στην διαδικασία ασαφούς συλλογισμού με βάση έναν επιλεγμένο αλγόριθμο σημασιολογικής ομοιότητας. Το σύστημα που προκύπτει είναι πιο ακριβές, διαισθητικό, διαλειτουργικό, δυναμικό, ερμηνεύσιμο και σημασιολογικά έξυπνο. Η προεπεξεργασία των δεδομένων συμπεριλάμβανε την εύρεση τιμών που λείπουν, εντοπισμό τιμών που δεν ήταν σωστές, κωδικοποίηση κατηγορηματικών ή ονομαστικών χαρακτηριστικών και σημασιολογική κωδικοποίηση ιατρικών εννοιών σύμφωνα με την τυπική ορολογία. Το σύστημα δημιουργήθηκε χρησιμοποιώντας τα JFML και το Jena JAVA APIs. Το σύστημα αποτελείται από 39 attributes και το αποτέλεσμα είναι σε binary μορφή με δύο αποτελέσματα, διαβητικός ή όχι. Ο τρόπος με τον οποίο έγινε το testing είναι με 10 fold cross validation και HFP. Επίσης πολύ σημαντικό είναι το γεγονός ότι για την λειτουργία και την κατασκευή του συστήματος υπήρχε πάντα η γνώμη και η καθοδήγηση από έναν ειδικό στον τομέα. Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι οι Naïve bayes, SVM, logistic regression, KNN με $k = 3$, C45, ANN καθώς και ο Random Forest. Το σύστημα έχει ένα interface το οποίο επιτρέπει την ενημέρωση και διαγραφή και αναζήτηση ασθενών καθώς και συγκεκριμένης διάγνωσης και επίσης υπάρχει authentication για τους ασθενείς. Το backend της εφαρμογής έγινε πάνω σε Microsoft access.

Όπως φαίνεται από το paper παρόλο που η θεματολογία είναι ίδια μεταξύ του paper και του θέματος της εργασίας, δηλαδή και τα δύο προσπαθούν μέσω του συστήματος που αναπτύχθηκε να προβλέψουν αν κάποιος είναι διαβητικός ή όχι(με το σύστημα της εργασίας να ελέγχει και επιπλέον πληροφορίες για μελλοντικά προβλήματα που μπορεί να έχει ο ασθενής)μέσω binary αποτελέσματος, λειτουργούν με εντελώς διαφορετικό τρόπο. Αρχικά μία από τις κυριότερες διαφορές είναι ότι το σύστημα του paper χρησιμοποιεί backend για λειτουργίες όπως authentication και user management σε αντίθεση με το σύστημα της εργασίας που δουλεύει με το front end κομμάτι και δεν κρατάει δεδομένα, αν και υπάρχει ως μελλοντική επέκταση για επιπλέον λειτουργίες, επίσης μια ακόμη πολύ σημαντική διαφορά είναι ότι το σύστημα του paper χρησιμοποιεί πολλούς διαφορετικούς αλγορίθμους για την λειτουργία διαφορετικών επιλογών σε αντίθεση με τον αλγόριθμο της εργασίας που δουλεύει με δέντρα απόφασης χρησιμοποιώντας ένα αλγόριθμο. Ακόμη μια πολύ μεγάλη διαφορά είναι ότι για την ανάπτυξη του συστήματος του paper χρησιμοποιήθηκαν 60 ασθενείς σε αντίθεση με το

σύστημα της εργασίας στην οποία έγιναν εκτεταμένες δοκιμές πάνω σε 100000 περίπου διαφορετικούς ασθενείς. Τα συστήματα αναπτύξανε χρησιμοποιώντας διαφορετικά εργαλεία ωστόσο και τα δύο για training και testing χρησιμοποιούν το k-fold cross-validation με $k = 10$. Το σύστημα της εργασίας είναι ένα σύστημα μηχανικής μάθησης το οποίο κάνει κάποιες έρευνες σύμφωνα με τα στοιχεία του ασθενή ενώ το σύστημα του paper είναι ένα ασαφές σύστημα(fuzzy system) δηλαδή προσπαθεί να δώσει αποτελέσματα παίρνοντας ασαφή δεδομένα για να πιάσει μεγαλύτερες πτυχές του προβλήματος που εξετάζει. Τα δύο συστήματα έχουν μεγάλη διαφορά στα attributes που περιέχεται μέσα στην εφαρμογή που κατασκευάστηκε με το app της εφαρμογής της εργασίας να λειτουργεί με 13 attributes, ενώ του paper με 39 attributes. Συνεπώς αυτό που μπορούμε να πούμε είναι ότι τα δύο συστήματα εξετάζουν όμοια θέματα ωστόσο με εντελώς διαφορετικό τρόπο και λογική.

Στο paper[10] αυτό προτείνεται ένα σύστημα πρόγνωσης διαβήτη τύπου 2 και υπέρτασης βασισμένο στα δεδομένα του χρήστη που αφορούν κινδύνους στην υγεία του. Για την κατασκευή του χρησιμοποιήθηκε ο iForest, αφαιρώντας μη προσιτά δεδομένα, και ο SMOTETomek κάνοντας ίση κατανομή των δεδομένων. Για την δημιουργία του συστήματος χρησιμοποιήθηκαν τέσσερα διαφορετικά Dataset για να βρεθεί το καλύτερο μοντέλο δεδομένων. Για την κατασκευή του χρησιμοποιήθηκε ένας συνδυασμός αλγορίθμων. Για την κατασκευή του συστήματος χρησιμοποιήθηκαν τέσσερα διαφορετικά Datasets τα οποία περιέχουν δεδομένα όπως χοληστερόλη, γλυκόζη, Τοποθεσία ασθενή, ηλικία, φύλλο, ύψος, βάρος, διάφορες μετρήσεις πίεσης αίματος, δείκτης μάζας σώματος κ.α. Το class του συστήματος ήταν η γλυκοζυλιωμένη αιμοσφαιρίνη, της οποίας το αποτέλεσμα δείχνει εάν κάποιος είναι διαβητικός ή όχι ως εξής, σε περίπτωση που είναι <7 τότε η διάγνωση δείχνει ότι ο ασθενής είναι αρνητικός ενώ εάν είναι ≥ 7 τότε η διάγνωση δείχνει ότι πάσχει από διαβήτη τύπου 2. Ο τρόπος με τον οποίο χρησιμοποιήθηκαν τα τέσσερα Datasets που χρησιμοποιήθηκαν είναι παίρνοντας για class που αναφέρθηκε το 1^ο dataset, χρησιμοποιώντας το 3^ο και 4^ο Dataset για να προβλέπει εάν έχει υπέρταση ο ασθενής και το 4^ο για να γίνει η σύνδεση μεταξύ της υπέρτασης και του αποτελέσματος του διαβήτη τύπου 2. Ο τρόπος πρόβλεψης γίνεται χρησιμοποιώντας δύο επίπεδα, στο πρώτο επίπεδο χρησιμοποιείται ο αλγόριθμος MLP, SVM, Decision Tree ενώ στο δεύτερο επίπεδο χρησιμοποιείται Logistic Regression. Ο τρόπος με τον οποίον έγινε η χρήση πολλών αλγορίθμων είναι με το λογισμικό Scikit-learn και Mlxtend της Python. Για το σωστό training και testing των δεδομένων έγινε 10-fold cross validation. Το accuracy του συστήματος είναι στο 97%.

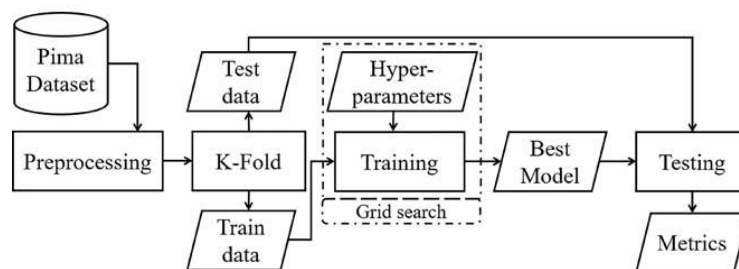
Το σύστημα που αναπτύσσεται σε αυτό το paper[11] είναι ένα σύστημα με το οποίο θα μπορούμε να κάνουμε πρόβλεψη για μακροχρόνια προβλήματα διαβήτη τύπου 2 σε διάστημα 7-8 ετών. Το σύστημα προβλέπει εάν κάποιος θα έχει προβλήματα διαβήτη τύπου 2 εντός 7 – 8 ετών με ποσοστό 96%. Το σύστημα για να δημιουργηθεί συλλέχθηκε ένα Dataset από 3226 ασθενής και εξετάστηκαν δεδομένα όπως ηλικία, εθνικότητα, δείκτης μάζας σώματος, γλυκόζη, καθώς και μετρήσεις γλυκόζης μετά από 30, 60 και 120 λεπτά, επίπεδα ινσουλίνης και οι μετρήσεις ινσουλίνης σε 30,60 και 120 λεπτά. Για προεπεξεργασία αλλάχθηκε η μορφή του dataset καθώς η αρχική μορφή ήταν imbalanced έχοντας μόλις το 11% των ασθενών θετικούς σε διαβήτη. Για να γίνει αυτό δημιουργήθηκαν 7 συνολικά πίνακες όπου κάθε πίνακας περιείχε τον ίδιο αριθμό ασθενών που διαγνώστηκαν θετικά και αρνητικά στον διαβήτη. Επίσης για την φάση της προεπεξεργασίας έγινε και συμπλήρωση κάποιων στοιχείων τα οποία ήταν κενά. Για το training και testing του συστήματος χρησιμοποιήθηκε 10-fold cross-validation ενώ το τελικό αποτέλεσμα αποτυπώνεται σε binary μορφή (0 – 1) όπου 0 είναι αν ο χρήστης δεν βρέθηκε θετικός και 1 εάν ο χρήστης είναι θετικός στον διαβήτη. Για την ανάπτυξη του συστήματος χρησιμοποιήθηκαν οι αλγόριθμοι SVM όπου ενεργοποιήθηκαν και οι επιλογές C και gamma του αλγορίθμου και ο δεύτερος αλγόριθμος ο οποίος χρησιμοποιήθηκε είναι ο Naïve Bayes. Μια αρκετά έξυπνη μέθοδος που ακολουθήθηκε είναι η εξής, πάρθηκαν μερικές από τις αποφάσεις οι οποίες ήταν <<αδύναμες>> και ενώθηκαν ώστε να δημιουργήσουν μια αξιόπιστη, αυτή η διαδικασία αποδείχθηκε αρκετά ικανοποιητική καθώς αποφεύγεται το overfitting αλλά και είναι πιο τυχαία τα δεδομένα που εξετάζονται και άρα πιο αξιόπιστα σε αληθινές περιστάσεις.

Συγκρίνοντας το paper με τη εργασία η κυριότερη διαφορά που βλέπουμε είναι ότι στοχεύουν σε διαφορετικά χρονικά πλαίσια, με το σύστημα που περιγράφει το paper να στοχεύει σε απόσταση 7 – 8 χρόνια μετά ενώ το σύστημα της εργασίας δίνει αποτελέσματα άμεσα σύμφωνα με κάποια στοιχεία που εισάγει ο χρήστης, ωστόσο και τα δύο συστήματα αφορούν ανθρώπους με διαβήτη τύπου 2 με διαφορετικά όμως αποτελέσματα έρευνας. Τα δύο συστήματα είναι αρκετά αποτελεσματικά με αρκετά καλό ποσοστό accuracy ενώ τα dataset αποτελούνται από μεγάλο αριθμό για ασθενών. Οι αλγόριθμοι οι οποίοι δουλεψτηκαν είναι οι εξής, στο paper το σύστημα που περιγράφηκε χρησιμοποιεί τους αλγορίθμους Naïve Bayes και SVM κάνοντας επεξεργασία κυρίως για τα δεδομένα καθώς υπάρχει πρόβλημα στην διαφορά ασθενών με διαβήτη και όχι, ενώ στο σύστημα της εργασίας δουλέψαμε με δέντρα απόφασης χρησιμοποιώντας τον αλγόριθμο j48 εφαρμόζοντας κάποια φίλτρα για καλύτερο

accuracy, και στις δύο περιπτώσεις το training και testing έγινε με 10-fold cross-validation. Τα αποτελέσματα και στις δύο περιπτώσεις διατυπώθηκαν ως binary παίρνοντας τιμές 0 και 1 για κάθε πρόβλεψη.

Σε αυτό το Paper[12] περιγράφεται ένα σύστημα μηχανικής μάθησης όπου γίνεται η διάγνωση του διαβήτη εφαρμόζοντας ένα συνδυασμό αλγορίθμων και τεχνικών προεπεξεργασίας. Οι τεχνικές που χρησιμοποιήθηκαν είναι η απόρριψη μη έγκυρων δεδομένων, συμπλήρωση τιμών που λείπουν, επιλογή στοιχείων, τυποποίηση δεδομένων και για το training και testing χρησιμοποιήθηκε ο k-fold cross-validation. Οι αλγόριθμοι οι οποίοι χρησιμοποιήθηκαν είναι K-nearest neighbour(K-NN), RF, DT, Naïve Bayes, AB, ο XGBoost καθώς και ο Multilayer Perception(MLP). Τα δεδομένα στα οποία έγιναν οι μετρήσεις είναι 768 γυναίκες από τις περιοχές Phoenix και Arizona. Από τους 768, οι 268 είναι διαβητικοί και οι 500 έχουν διαγνωστεί αρνητικά. Τα δεδομένα τα οποία χρησιμοποιήθηκαν είναι εγκυμοσύνη, γλυκόζη, πίεση, πάχος δέρματος τρικέφαλου, ινσουλίνη, Δείκτης μάζας σώματος, ηλικία και γενεαλογικά προβλήματα. Ο τρόπος με τον οποίο δημιουργήθηκε το σύστημα είναι μέσω της γλώσσας python και των εργαλείων της. Τα αποτελέσματα τα οποία μετρηθήκανε στο confusion Matrix είναι True Positive, False Positive, True Negative και False Negative.

1.7 Διάγραμμα του συστήματος που περιγράφεται στο άρθρο



Όπως φαίνεται το paper και η εργασία ασχολούνται με το ίδιο θέμα τον διαβήτη ωστόσο υπάρχει διαφορά στο θέμα πρόβλεψης, αφού το σύστημα που περιγράφηκε στο paper ασχολείται με την απλή πρόβλεψη του διαβήτη ενώ το σύστημα της εργασίας προβλέπει επιπλέον πληροφορίες σχετικά με την κατάσταση του ασθενή. Επίσης μια μεγάλη διαφορά που πρέπει να αναφερθεί είναι η προεπεξεργασία που δέχτηκε το σύστημα που περιγράφηκε στο paper, καθώς είναι ένα πολύ σημαντικό κομμάτι της όλης έρευνας θέλοντας να δείξει την σημαντικότητα της προεπεξεργασίας σε ένα σύστημα. Όσον αφορά τα Dataset, το paper μελέτησε μόνο γυναικείο πληθυσμό σχετικά με τον διαβήτη.

Το άρθρο[13] αυτό έχει ως στόχο να περιγράψει την υβριδικότητα που μπορούν να πάρουν τα συστήματα που επικεντρώνονται στην ανωμαλία της γλυκόζης στο αίμα και την ανίχνευση της μαζί με τις γλυκαιμικές διαφορετικοποιήσεις, την υπεργλυκαιμία και την υπογλυκαιμία στον διαβήτη τύπου 1 μέσα από ένα σύστημα υποστήριξης αποφάσεων. Στο σύστημα που αναπτύχθηκε σε αυτήν την έρευνα χρησιμοποιήθηκαν αλγόριθμοι όπως Artificial neural networks(ANN), Decision Trees(DTs), Support Vector Machine(SVM), Evolutionary algorithms κ.α. Για την εύρεση των δεδομένων συλλέχθηκαν συνολικά 496 papers ώστε να δημιουργηθεί η συγκεκριμένη έρευνα, ωστόσο μετά από ξεκαθάρισμα των paper, χρησιμοποιήθηκαν συνολικά 47. Μαζεύοντας δεδομένα και πληροφορίες από τα papers χρησιμοποιήθηκαν τα εξής στοιχεία για να δημιουργηθεί ο αλγόριθμος τα οποία είναι γλυκόζη, ινσουλίνη, παλμοί καρδιάς, δίαιτα, γυμναστική, ώρες ύπνου κ.α. Το κυριότερο από τα χαρακτηριστικά που χρησιμοποιήθηκαν για να τρέξουν οι αλγόριθμοι είναι οι παλμοί καρδιάς, καθώς ήταν ο πυρήνας του συστήματος που αναπτύχθηκε. Σχετικά με την γυμναστική και την δραστηριότητα του ασθενή, χρησιμοποιούνται μια σειρά αισθητήρων στον ασθενή. Στο άρθρο χρησιμοποιήθηκε για κάθε paper όπου μαζευτήκανε δεδομένα διαφορετικό αλγόριθμο για να πάρει αποτελέσματα ώστε να κατασκευαστεί το σύστημα υποστήριξης αποφάσεων.

Τα δύο συστήματα τα οποία συγκρίνουμε όπως φαίνεται έχουν αρκετές διαφορές με την κυριότερη ότι στην εργασία δημιουργείται ένα σύστημα μηχανικής μάθησης για την πρόβλεψη χρήσιμων πληροφοριών σχετικά με τον διαβήτη, ενώ στο σύστημα του άρθρου δημιουργήθηκε ένα σύστημα υποστήριξης αποφάσεων το οποίο ασχολείται με διαβήτη τύπου 1 σε αντίθεση με την εργασία που ασχολείται με διαβήτη τύπου 2. Στο άρθρο έγινε ένας συνδυασμός αλγορίθμων όπου ο κάθε αλγόριθμος αξιολογούσε συγκεκριμένα από τα 47 paper όπου χρησιμοποιήθηκαν για την κατασκευή του συστήματος σε αντίθεση από την εργασία που χρησιμοποιήθηκε ένα paper για να αντληθούν τα κατάλληλα στοιχεία όπου χρησιμοποιήθηκαν για την κατασκευή του συστήματος χρησιμοποιώντας τον αλγόριθμο j48 μέσω δέντρα αποφάσεων. Στο άρθρο χρησιμοποιήθηκε η γλώσσα Python και τα εργαλεία της ενώ στην εργασία χρησιμοποιήθηκε το εργαλείο WEKA.

Το άρθρο[14] έχει ως σκοπό να δημιουργήσει μια έρευνα ώστε να γίνει περιγραφή σχετικά με την αντίδραση των ανθρώπων στα συστήματα υποστήριξης αποφάσεων μέσα σε δύο κύκλους, από Δεκέμβριο του 2013 έως Φεβρουάριο 2014 και από Αύγουστο 2014 έως Νοέμβριο 2014. Τα αποτελέσματα πάρθηκαν μέσα από ερωτηματολόγια. Τα αποτελέσματα έδειξαν ότι οι γιατροί είναι θετικοί με την χρήση των συστημάτων υποστήριξης αποφάσεων

στην λειτουργία της ιατρικής και πως πιστεύουν ότι βοηθούν στην αποτελεσματικότερη λειτουργία της ιατρικής με την χρήση τους. Για δημιουργία της έρευνας χρησιμοποιήθηκε ένα σύστημα υποστήριξης απόφασης σχετικά με τον διαβήτη στο οποίο ήταν εγγεγραμμένοι χρήστες σχετικά με την έρευνα και από τους οποίους πήραμε και τα αποτελέσματα. Τα αποτελέσματα ήταν από ασθενής και γιατρούς και τα δεδομένα τους στάλθηκαν ως ανώνυμα. Οι ασθενείς οι οποίοι χρησιμοποιήθηκαν είναι ασθενείς οι οποίοι ανάμεσα στον 1^ο και στο 2^ο κύκλο είχαν βελτίωση της κατάστασης τους. Οι ερωτήσεις που κυρίως ρωτήθηκαν στους ασθενείς είναι σχετικά με την αλληλεπίδραση με τους γιατρούς, χρήση φαρμάκων καθώς και γενικής ικανοποίησης. Για τους γιατρούς δημιουργήθηκαν δύο ερωτηματολόγια τα οποία μετά την απάντηση τους χρησιμοποιήθηκαν ως μεταβλητές για τον αλγόριθμο multivariable linear regression, ώστε να γίνει η ανάλυση τους. Για την καλύτερη ανασκόπηση της έρευνας σχετικά με τους γιατρούς χρησιμοποιήθηκαν δεδομένα όπως αριθμός των <<κλικ>> στα στοιχεία ενός ασθενή και πόση ώρα κάποιος γιατρός κοιτούσε τα στοιχεία του ασθενή. Για την ανάλυση του συστήματος πολύ μεγάλη βαρύτητά είχαν και οι περιπτώσεις στην οποία οι ασθενείς έχουν κάποιου είδους μηνύματος από το σύστημα. Ένα πολύ ενδιαφέρον αποτέλεσμα που πήραμε από την ανάλυση του συστήματος είναι ότι από την στιγμή που ο χρήστης έλαβε μήνυμα από το σύστημα συνήθως είχε αλλαγή της κατάστασης του γλυκαιμικού ελέγχου. Επίσης πολύ ενδιαφέρον είναι ότι υπάρχει αλλαγή και στην χοληστερόλη, στην πίεση(συστολική και διαστολική) καθώς και στην αναλογία λευκωματίνης / κρεατινίνης των ούρων μετά από 9 – 15 μήνες. Από τους 2072 ασθενείς που ήταν εγγεγραμμένοι στο σύστημα συλλέχθηκαν οι απαντήσεις των 359 από τον 1^ο και 2^ο κύκλο. Το ποσοστό ευχαριστημένων χρηστών από την χρήση συστήματος υποστήριξης αποφάσεων για τον διαβήτη ήταν μεγαλύτερο του 90%. Από την μεριά των γιατρών, το μεγαλύτερο ποσοστό απάντησε θετικά η ουδέτερα στα συστήματα υποστήριξης αποφάσεων, ωστόσο παρόλο που υπάρχει θετική ροή απαντήσεων, οι αναφορές έδειξαν μικρή χρήση των συστημάτων.

Από ότι φαίνεται το άρθρο και η εργασία είναι δύο τελείως διαφορετικά θέματα με το άρθρο να προσπαθεί να δημιουργήσει μια έρευνα σχετικά με το πόσο βοηθούν τα συστήματα υποστήριξης αποφάσεων στην καθημερινότητα αλλά και το πόσο αποδεκτά είναι από το κοινό, ενώ η εργασία δημιουργεί ένα σύστημα. Ωστόσο και στις δύο περιπτώσεις μελετώνται συστήματα σχετικά με τον διαβήτη με το άρθρο να προσπαθεί να δώσει χρήσιμες πληροφορίες στους δημιουργούς των συστημάτων μέσω των ερωτηματολογίων της έρευνας. Τα δεδομένα που χρησιμοποιήθηκαν για τις δύο περιπτώσεις είναι διαφορετικά καθώς στην εργασία είναι

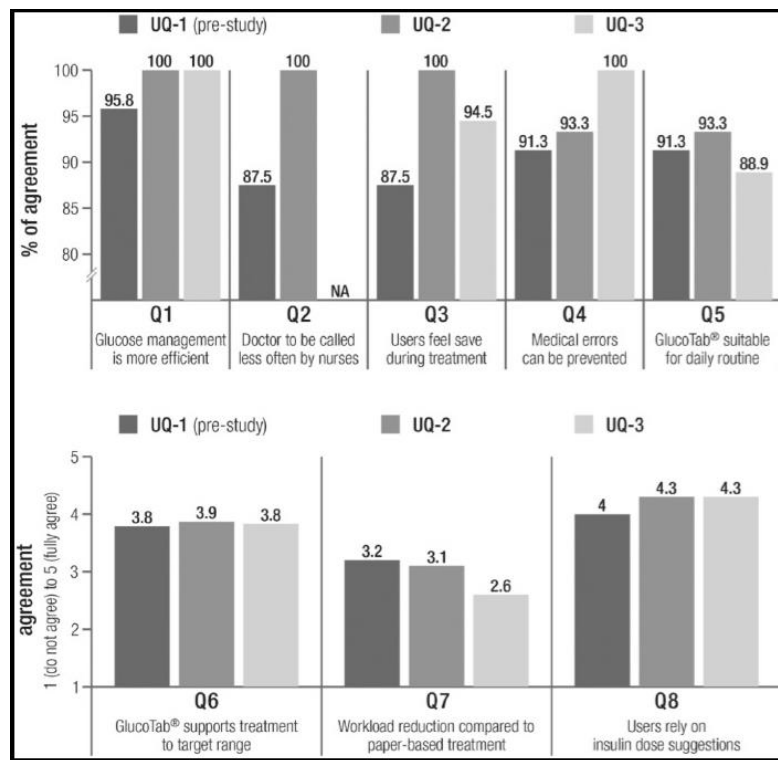
περισσότερο εργαστηριακά με δεδομένα όπως ινσουλίνη, μετφορμίνη κ.α., ενώ στο άρθρο σαν δεδομένα έχουμε δεδομένα όπως πόση ώρα χρησιμοποιούνται τέτοιου είδους συστήματα.

Το άρθρο[15] αυτό έχει ως στόχο να δημιουργήσει ένα σύστημα με το οποίο θα αποτρέπει την υπογλυκαιμία σε άτομα με διαβήτη τύπου 2. Τα αποτελέσματα του συστήματος ήταν αρκετά ενθαρρυντικά, ιδίως μετά την λειτουργία του όπου δόθηκαν ερωτηματολόγια σε κάποιους από τους ανώνυμους χρήστες με τις απαντήσεις να δείχνουν ότι ήταν αρκετά ικανοποιημένοι από το σύστημα. Το σύστημα στοχεύει κυρίως σε χρήστες οι οποίοι έχουν χορηγηθεί ινσουλίνη καθώς είναι η κύρια πηγή παρενεργειών όπως η υπογλυκαιμία σε ανθρώπους με διαβήτη. Το σύστημα χρησιμοποιήθηκε για την κατασκευή ενός mobile app το οποίο δόθηκε στους γιατρούς και τους νοσηλευτές το οποίο βοηθά στην καλύτερη ιατρική περίθαλψη των ασθενών με διαβήτη τύπου 2. Μια από τις κύριες λειτουργίες του συστήματος είναι η παροχή συστάσεων δόσης ινσουλίνης για θεραπεία με ινσουλίνη σε ασθενείς με Διαβήτη τύπου 2. Περιλαμβάνει τις ακόλουθες λειτουργίες που βοηθούν τους γιατρούς και τους νοσηλευτές:

- (1) εισαγωγή δεδομένων στο app,
- (2) εισαγωγή φαρμάκων με υποστήριξη λήψης αποφάσεων δόσης ινσουλίνης για γιατρούς,
- (3) υποστήριξη χορήγησης φαρμάκων συμπεριλαμβανομένου υπολογισμού δόσης ινσουλίνης για νοσοκόμες και
- (4) διαχείριση ροής εργασίας για γιατρούς και νοσηλευτές.

Το κινητό σύστημα βοηθά στην οργάνωση της ροής εργασίας θεραπείας, συμπεριλαμβανομένης της προβολής για ανοιχτές εργασίες, διευκολύνοντας την τεκμηρίωση και παρέχοντας οπτικοποίηση των τιμών της γλυκόζης, των γευμάτων και των δόσεων ινσουλίνης. Από τα δεδομένα τα οποία συλλέχθηκαν αφαιρέθηκαν ασθενείς με σημαντικά μειωμένη νεφρική λειτουργία, ασθενής με μόνιμες βλάβες, ασθενής με υποψία αλλεργίας στην ινσουλίνη, έγκυος γυναίκες, ασθενής που πάσχουν από διαβήτη τύπου 1 και ασθενής που έχουν ενταχθεί σε κάποια κλινική μελέτη τους τελευταίους 3 μήνες. Τα δεδομένα τα οποία χρησιμοποιήθηκαν είναι δεδομένα όπως, γλυκόζη, δίαιτα ασθενή, ευαισθησία στην ινσουλίνη, προτεινόμενη δόση ινσουλίνης, φύλλο, ηλικία, δείκτης μάζας σώματος, βάρος, διάρκεια διαβήτη, κ.α. Το μέγεθος του Dataset αποτελείται από 30 ασθενείς μεγαλύτεροι των 18 χρόνων.

1.8 Ερωτηματολόγιο Άρθρου



Το άρθρο που περιγράφηκε όπως και η εργασία ασχολείται με ασθενείς που πάσχουν διαβήτη τύπου 2 ωστόσο στους ασθενείς που εξετάζονται στο άρθρο δίνεται ιδιαίτερη βαρύτητα στην ινσουλίνη και στις παρενέργειες της όπως η υπογλυκαιμία. Επίσης μια σημαντική διαφορά μεταξύ των δύο ερευνών είναι ότι στο άρθρο υπάρχει και η ερευνητική προσέγγιση καθώς υπάρχουν τα ερωτηματολόγια τα οποία δίνονται στους γιατρούς και στους νοσηλευτές. Τα Dataset των δύο άρθρων είχαν δύο σημαντικές διαφορές καθώς στο άρθρο επικεντρώθηκε η έρευνα σε λίγα άτομα, συγκεκριμένα 30, με μεγάλο όμως όγκο δεδομένων ενώ στην εργασία επικεντρώθηκε σε μεγάλο όγκο δεδομένων για καλύτερο testing και πιο στοχευμένα στοιχεία για τους ασθενείς. Τα δύο συστήματα ενσωματώθηκαν και σε mobile app με γραφικό περιβάλλον για τους ασθενείς και τους γιατρούς.

1.5 Αποτελέσματα Βιβλιογραφίας και Συνεισφορά

Όπως φαίνεται παραπάνω από την βιβλιογραφία μπορούμε να συμπεράνουμε ότι η μηχανική μάθηση και όλος ο κλάδος της πληροφορικής μπορεί να συμβάλλει καθοριστικά στην καλύτερη περίθαλψη του διαβήτη τύπου 1 και τύπου 2. Αυτό το οποίο βλέπουμε είναι ότι τα συστήματα τα οποία δημιουργούνται πάνω στον κλάδο έχουν πολύ μεγάλη αποτελεσματικότητα με πολύ καλά ποσοστά πρόβλεψης πλησιάζοντας ποσοστά 90% τουλάχιστον στις περισσότερες περιπτώσεις. Συνεπώς θα πρέπει να προωθούνται τέτοιου είδους συστήματα στις κλινικές έτσι ώστε να χρησιμοποιούνται ως δεύτερη γνώμη για τους γιατρούς. Ακόμη ένα πολύ ενθαρρυντικό στοιχείο που αντλήθηκε από τις βιβλιογραφίες είναι το γεγονός πώς οι χρήστες αλλά και οι γιατροί είναι αρκετά δεκτικοί στην χρήση τέτοιου είδους συστημάτων, που μας δείχνει ότι θα είναι και ευκολότερη η αφομοίωση τέτοιου είδους συστημάτων στον κλάδο της ιατρικής. Αυτό ίσως οφείλεται και στην πλέον προστασία προσωπικών δεδομένων των ασθενών σε τέτοιου είδους εφαρμογών το οποίο αυξάνει την εμπιστοσύνη τους προς τα πληροφοριακά συστήματα. Ένα πολύ κρίσιμο στοιχείο το οποίο αντλούμε από την ανάλυση της βιβλιογραφίας είναι πολύ σημαντικό για την δημιουργία ενός τέτοιου συστήματος να χρησιμοποιούνται διαφορετικοί συνδυασμοί αλγορίθμων μέχρι την τελική κατασκευή καθώς καλύπτονται περισσότερες περιπτώσεις και παίρνουμε διάφορα αποτελέσματα για τον τρόπο με τον οποίο θα πρέπει να κατασκευαστεί το σύστημα. Ακόμη βλέπουμε πολλά επιπλέον στοιχεία όπως με ποιον τρόπο θα πρέπει να γίνει το φιλτράρισμα των στοιχείων που χρησιμοποιούνται. Ένα ακόμη πολύ σημαντικό συμπέρασμα που θα μπορούσαμε να πάρουμε είναι μέσω των στοιχείων που χρησιμοποιήθηκε στην βιβλιογραφία, καθώς βλέπουμε ότι όλες οι βιβλιογραφίες έχουν παρόμοια δεδομένα προς ανάλυση και χρήση για την κατασκευή του συστήματος που μας δείχνει ότι οι περισσότεροι άνθρωποι με διαβήτη πάσχουν από παρόμοια περιστατικά και δεδομένα. Τέλος ένα ακόμη αξιοσημείωτο συμπέρασμα που πήραμε από την βιβλιογραφία είναι η πρόβλεψη του διαβήτη αλλά και η σωστή περίθαλψη του είναι μέγιστης σημασίας καθώς ένα πρόβλημα διαβήτη το οποίο δεν ελέγχεται και δεν λαμβάνει την κατάλληλη περίθαλψη μπορεί να προκαλέσει πολλά προβλήματα η ακόμη και θάνατο στον ασθενή και μέσω των συστημάτων της βιβλιογραφίας βλέπουμε ότι μπορούμε να περιορίσουμε τέτοιου είδους καταστάσεις μέσα από την έγκαιρη πρόβλεψη αλλά και τα αποτελέσματα που λαμβάνουμε για την κατάσταση του ασθενή.

2. ΤΕΧΝΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 WEKA

Το WEKA είναι ένας δημοφιλής λογισμικός για Machine Learning (Μηχανική Μάθηση) γραμμένο σε Java, που αναπτύχθηκε στο Πανεπιστήμιο του Waikato της Νέας Ζηλανδίας. Είναι ελεύθερο λογισμικό (Open Source) και περιέχει μια μεγάλη συλλογή από εργαλεία οπτικοποίησης (Visualization) και αλγορίθμους για την ανάλυση δεδομένων και την προγνωστική μοντελοποίηση, μαζί με γραφικές διεπαφές χρήστη για εύκολη πρόσβαση σε αυτές τις λειτουργίες.

Η αρχική μη Java έκδοση του WEKA ήταν ένα Tcl/tk front-end για μοντελοποίηση αλγορίθμων που εφαρμόζονται σε άλλες γλώσσες προγραμματισμού, παρέχοντας δυνατότητες pre-processing (προεπεξεργασία) δεδομένων σε γλώσσα C, και ένα σύστημα βασισμένο σε Makefile για την πραγματοποίηση πειραμάτων μηχανικής μάθησης. Αυτή η αρχική έκδοση είχε σχεδιαστεί ως ένα εργαλείο για την ανάλυση δεδομένων από γεωργικούς τομείς, αλλά η πιο πρόσφατη πλήρης έκδοση βασισμένη σε Java (WEKA 3), η ανάπτυξη της οποίας άρχισε το 1997, έχει πλέον πολλούς τομείς εφαρμογής, κυρίως εκπαιδευτικούς σκοπούς και έρευνες.

2.2 Android Studio

Το **Android Studio** είναι ένα ολοκληρωμένο προγραμματιστικό περιβάλλον (IDE) για ανάπτυξη εφαρμογών στην πλατφόρμα Android. Ανακοινώθηκε στις 16 Μαΐου 2013 στο συνέδριο Google I/O από την Google Product Manager, Katherine Chou. Το Android Studio είναι διαθέσιμο ελεύθερα με την άδεια Apache License 2.0.

Το Android Studio ήταν διαθέσιμο σε πρώιμο στάδιο για προεπισκόπηση ξεκινώντας από την έκδοση 0.1 τον Μάιο του 2013, έπειτα ξεκίνησε το δοκιμαστικό στάδιο από την έκδοση 0.8 που βγήκε τον Ιούνιο του 2014¹. Η πρώτη σταθερή έκδοση βγήκε το Δεκέμβριο του 2014, με την έκδοση 1.0.

Βασισμένο στο λογισμικό της JetBrains' IntelliJ IDEA, το Android Studio σχεδιάστηκε αποκλειστικά για προγραμματισμό Android. Είναι διαθέσιμο για Windows Mac OS και Linux,

και αντικατέστησε τα Eclipse Android Development Tools (ADT) ως το κύριο IDE της Google για ανάπτυξη εφαρμογών Android.

2.3 Flutter

Το Flutter είναι ένα λογισμικό open source το οποίο δημιουργήθηκε από την google. Χρησιμοποιείται για την ανάπτυξη εφαρμογών για Android, iOS, Windows, Mac, Linux αλλά και για Web εφαρμογές καθώς και το νέο λογισμικό της google, Fuchsia. Το Flutter φτιάχτηκε με την λογική να μπορεί ένας Developer να δημιουργεί App για οποιαδήποτε συσκευή με ένα μόνο Codebase.

Η πρώτη έκδοση του Flutter βγήκε με το όνομα sky και έτρεχε σε λειτουργικό σύστημα Android. Παρουσιάστηκε πρώτη φορά στην σύνοδο προγραμματιστών γλώσσας Dart του 2015 με την υπόσχεση ενός framework που θα τρέχει στα 120 frames το δευτερολεπτο. Η πλέον επίσημη έκδοση του Flutter είναι η 1.17.0 που τρέχει μαζί με την γλώσσα Dart έκδοσης 2.8

Τα κύρια συστατικά του Flutter περιλαμβάνουν

- Την Πλατφόρμα Dart
- Την Flutter Engine
- Την επίσημη βιβλιοθήκη του Flutter
- Καθώς και ειδικά διαμορφωμένα Widgets για το Flutter

2.4 Dart

Η Dart είναι μία client - optimized, client – side φτιαγμένη για να δημιουργεί συστήματα και εφαρμογές σε πολλαπλές πλατφόρμες. Δημιουργήθηκε από την Google και οι κυριότερες χρήσεις της μέχρι στιγμής είναι για την δημιουργία mobile εφαρμογών, Desktop εφαρμογών, προγραμματισμός σε Server και Web Apps.

Η Dart είναι μια αντικειμενοστραφής γλώσσα προγραμματισμού με συντακτικό παρόμοιο με την γλώσσα C. Η Dart μπορεί να κάνει Compile σε native code ή ακόμη και σε Javascript.

Η Dart παρουσιάστηκε σε συνέδριο στην Δανία το 2011. Το Project ξεκίνησε από τους Lars Bak και Kasper Lund. Η σταθερή έκδοση Dart 1.0 βγήκε στο κοινό το 2015. Τον Αύγουστο το 2018 βγήκε η νεότερη της έκδοση Dart 2.0 όπου βελτίωσε αισθητά την Dart.

Η νέα πρόσφατη έκδοση 2.6 της Dart συνοδεύεται από μια νέα επέκταση που λέγεται dart2native. Η δυνατότητα αυτή επιτρέπει την Dart να κάνει Native Compile σε Linux, macOS και Windows. Η Dart έχει «συνδεθεί» με την Flutter καθώς είναι το εργαλείο που την ανέδειξε σαν γλώσσα προγραμματισμού και έχει κάθε χρόνο άνοδο λόγω της υποστήριξής του.

2.5 K – fold Cross Validation

Το k - fold Cross Validation είναι μια τεχνική για την αξιολόγηση προγνωστικών μοντέλων με διαίρεση του αρχικού δείγματος σε ένα training set για την εκπαίδευση του μοντέλου και ένα σύνολο δοκιμών για την αξιολόγησή του. Σε k-fold cross-validation, το αρχικό δείγμα χωρίζεται τυχαία σε υπο-δείγματα k ίσου μεγέθους. Από τα υπο-δείγματα k, ένα μεμονωμένο δείγμα διατηρείται ως δεδομένα επικύρωσης για τη δοκιμή του μοντέλου και τα υπόλοιπα υπο-δείγματα k-1 χρησιμοποιούνται ως δεδομένα εκπαίδευσης. Η διαδικασία στη συνέχεια επαναλαμβάνεται k φορές με κάθε ένα από τα υπο-δείγματα k να χρησιμοποιείται ακριβώς μία φορά ως δεδομένα επικύρωσης. Τα αποτελέσματα k από τις πτυχώσεις μπορούν στη συνέχεια να υπολογίζονται κατά μέσο όρο (ή να συνδυάζονται διαφορετικά) για να παράγουν μία μόνο εκτίμηση. Το πλεονέκτημα αυτής της μεθόδου είναι ότι όλες οι παρατηρήσεις χρησιμοποιούνται τόσο για την εκπαίδευση όσο και για την επικύρωση, και κάθε παρατήρηση χρησιμοποιείται για επικύρωση ακριβώς μία φορά. Για προβλήματα ταξινόμησης, κάποιος συνήθως χρησιμοποιεί στρωματοποιημένη διασταυρούμενη επικύρωση k-fold, στην οποία οι πτυχές επιλέγονται έτσι ώστε κάθε πτυχή να περιέχει περίπου τις ίδιες αναλογίες ετικετών κλάσης. Η διαδικασία επαναλαμβάνεται n φορές, αποδίδοντας n τυχαίες κατατμήσεις του αρχικού δείγματος. Τα αποτελέσματα n υπολογίζονται και πάλι κατά μέσο όρο (ή συνδυάζονται διαφορετικά) για να παράγουν μία μόνο εκτίμηση.

2.1 Παράδειγμα λειτουργίας ενός K fold cross validation με $K = 5$ 

3. ΔΕΔΟΜΕΝΑ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ

3.1 Περιγραφή Δεδομένων

Gender □ Το Gender παίρνει 2 τιμές, και οι τιμές είναι Άνδρας ή Γυναίκα και προσδιορίζει το φύλο του ανθρώπου που μελετάμε. Οι τιμές που θα πάρει αυτό το attribute είναι nominal. Ο διαβήτης τύπου 2 απαντάται συχνότερα σε άνδρες, ειδικά σε ηλικίες 35-54 ετών, όπου οι άνδρες έχουν διπλάσιες πιθανότητες να αναπτύξουν διαβήτη, με έναρξη σε πολύ χαμηλότερο μέσο Δείκτη μάζας σώματος. Τα περισσότερα διαβητικά συμπτώματα είναι τα ίδια σε άνδρες και γυναίκες. Αυτά τα γενικά συμπτώματα περιλαμβάνουν συνεχή δίψα, συνεχή ούρηση, κόπωση, ζάλη και απώλεια βάρους. Ωστόσο, τα συμπτώματα που παρατηρούνται ιδιαίτερα στους άνδρες είναι η απώλεια μυϊκής μάζας και τα γενετικά προβλήματα. Επιπλέον, οι γυναίκες συχνά εμφανίζουν συμπτώματα όπως λοιμώξεις ζυμομυκήτων, μολύνσεις ουροποιητικού συστήματος και σύνδρομο πολυκυστικών ωοθηκών. Εάν δεν αντιμετωπιστεί σωστά, ο διαβήτης μπορεί να οδηγήσει σε πολλές σοβαρές επιπλοκές στην υγεία. Αυτά περιλαμβάνουν ακρωτηριασμό, νευροπάθεια, αμφιβληστροειδοπάθεια, καρδιαγγειακές παθήσεις και νεφρική νόσο. Το 45% των ανδρών με διαβήτη αναπτύσσουν επίσης στυτική δυσλειτουργία λόγω βλάβης των νεύρων, των μυών και των αιμοφόρων αγγείων. Ωστόσο, οι γυναίκες έχουν πολύ περισσότερες πιθανότητες καρδιακής νόσου, νεφρικής νόσου και κατάθλιψη. Συνολικά, αυτό το καθιστά πολύ πιο απειλητικό για τη ζωή των γυναικών σε σύγκριση με τους άνδρες. Ένα επιπλέον ζήτημα που οι γυναίκες με διαβήτη έχουν είναι η εμμηνόπαυση. Ο συνδυασμός του διαβήτη με αυτήν την αλλαγή ορμονών μπορεί να οδηγήσει σε περαιτέρω αύξηση της γλυκόζης στο αίμα, αύξηση βάρους και προβλήματα με τον ύπνο. Αυτό μπορεί επομένως να οδηγήσει σε περαιτέρω ανάπτυξη σοβαρών επιπλοκών, επιδεινώνοντας περαιτέρω προηγούμενα ζητήματα υγείας. Συνολικά, τα αρσενικά μπορούν να αναπτύξουν διαβήτη με χαμηλότερο Δείκτη Μάζας Σώματος, με επιπλέον επιπλοκές όπως η στυτική δυσλειτουργία και η απώλεια μυϊκής μάζας. Ένας πιθανός λόγος για αυτό είναι η απώλεια της τεστοστερόνης σε μεταγενέστερη ζωή στους άνδρες. Ωστόσο, οι γυναίκες με διαβήτη αντιμετωπίζουν σοβαρότερες συνέπειες, όπως πιθανή καρδιακή νόσο. Επομένως, ο διαβήτης επηρεάζει κάθε φύλο με διαφορετικό τρόπο και μπορεί να οδηγήσει σε σοβαρές επιπλοκές στην υγεία που αλλάζουν τη ζωή. Ακόμη να αναφερθεί ότι το ποσοστό εμφάνισης στους άνδρες είναι διπλάσιο από τις γυναίκες.

Age □ Το attribute age προσδιορίζει την ηλικία του ανθρώπου που μελετάμε και παίρνει τιμές από 0 έως 100 με τον εξής τρόπο, οι τιμές που τρέχει ο αλγόριθμος είναι [0 – 10), [10 – 20), [20 – 30), [30 – 40), [40 – 50), [50 – 60), [60 – 70), [70 – 80), [80 – 90), [90- 100). Οι μεσήλικες και ηλικιωμένοι ενήλικες εξακολουθούν να διατρέχουν τον υψηλότερο κίνδυνο εμφάνισης διαβήτη τύπου 2 Σύμφωνα με την Εθνική Έκθεση Στατιστικών Διαβήτη του CDC 2017, υπήρχαν περίπου 1,5 εκατομμύρια νέες συνολικές περιπτώσεις διαβήτη μεταξύ των ενηλίκων το 2015. Το 2015, οι ενήλικες ηλικίας 45 έως 64 ετών ήταν η πιο διαγνωσμένη ηλικιακή ομάδα για διαβήτη. Νέες περιπτώσεις διαβήτη τύπου 1 και τύπου 2 σε άτομα ηλικίας 18 ετών και άνω διανεμήθηκαν ως εξής:

ηλικίας 18 έως 44 ετών: 355.000 νέες περιπτώσεις

ηλικίας 45 έως 64: 809.000 νέες περιπτώσεις

ηλικίας 65 ετών και άνω: 366.000 νέες περιπτώσεις

Time In Hospital □ Το Attribute αυτό στέλνει μέσα στο αλγόριθμο ένα ακέραιο αριθμό ο οποίος δηλώνει τις μέρες που έμεινε ο ασθενής από την ημέρα που εντάχθηκε στο νοσοκομείο / κλινική έως την ημέρα που πήρε εξιτήριο. Για να είναι αξιόπιστο αυτό το δεδομένο όταν θα εισαχθεί ο αριθμός στο σύστημα θα πρέπει να έχουν συμβεί τα εξής.

Κατά την διάρκεια της παραμονής στο νοσοκομείο ή ιατρείο θα πρέπει ο ασθενής να ελέγχετε συχνά από το προσωπικό του νοσοκομείου. Πρέπει να ελέγχετε τακτικά το σάκχαρό του ασθενή. Αυτό μπορεί να γίνεται κάθε ώρα. Το επίπεδο οξυγόνου στο αίμα θα πρέπει να παρακολουθείται από έναν αισθητήρα που είναι προσαρτημένος στο δάχτυλό ή στο λοβό του αυτιού του ασθενή. Μπορεί να τοποθετηθεί ένας μικρός σωλήνας (καθετήρα) μέσα στην ουροδόχο κύστη μέσω της ουρήθρας για να γίνει μέτρηση των ούρων από την ουροδόχο κύστη. Εξετάσεις αίματος (Αιμοσφαιρίνη A1c) για τον έλεγχο του μέσου σακχάρου στο αίμα τους τελευταίους 3 μήνες (δεδομένο το οποίο θα χρειαστεί και παρακάτω). Εξέταση αίματος για έλεγχο μολύνσεων. Εξέταση αίματος για να ελεγχθεί εάν το σώμα παράγει ινσουλίνη ώστε να διαπιστωθεί με σιγουριά εάν πρόκειται για διαβήτη τύπου 2 και άρα είναι αξιόπιστα τα αποτελέσματα του συστήματος που εξετάζουμε καθώς αφορούν ασθενείς με διαβήτη τύπου 2. Αίμα, ούρα ή άλλες εξετάσεις για την παρακολούθηση της λειτουργίας των οργάνων του ασθενή καθώς και δοκιμές ούρων για έλεγχο βακτηρίων στα ούρα. Ακτινογραφίες: Εικόνες του εσωτερικού του θώρακα για έλεγχο λοίμωξης

Insulin1 □ Το attribute αυτό περιγράφεται με 2 τιμές, up το οποίο σημαίνει ότι είναι ανεβασμένα τα επίπεδα ινσουλίνης στο αίμα και η δεύτερη τιμή είναι !up δηλαδή ότι είναι σε φυσιολογικά επίπεδα ή ότι είναι κατεβασμένο. Η ινσουλίνη είναι μια ορμόνη που βοηθά στη μετακίνηση του σακχάρου στο αίμα, γνωστό ως γλυκόζη, από την κυκλοφορία του αίματος στα κύτταρα σας. Η γλυκόζη προέρχεται από τις τροφές που τρώει και πίνει ένας άνθρωπος. Είναι η κύρια πηγή ενέργειας του σώματός. Η ινσουλίνη παίζει βασικό ρόλο στη διατήρηση της γλυκόζης στα σωστά επίπεδα. Εάν τα επίπεδα γλυκόζης είναι πολύ υψηλά ή πολύ χαμηλά, μπορεί να προκαλέσει σοβαρά προβλήματα υγείας. Τα επίπεδα γλυκόζης που δεν είναι φυσιολογικά είναι γνωστά ως:

Υπεργλυκαιμία, επίπεδα γλυκόζης στο αίμα που είναι πολύ υψηλά. Συμβαίνει όταν το σώμα δεν παράγει αρκετή ινσουλίνη. Εάν δεν υπάρχει αρκετή ινσουλίνη, η γλυκόζη δεν μπορεί να εισέλθει στα κύτταρα. Παραμένει στην κυκλοφορία του αίματος αντ' αυτού.

Υπογλυκαιμία, επίπεδα γλυκόζης στο αίμα που είναι πολύ χαμηλά. Εάν το σώμα στέλνει υπερβολική ινσουλίνη στο αίμα, πάρα πολύ γλυκόζη θα πάει στα κύτταρα. Αυτό αφήνει λιγότερο στην κυκλοφορία του αίματος.

Ο διαβήτης είναι η πιο κοινή αιτία ανώμαλων επιπέδων γλυκόζης.

Η αντίσταση στην ινσουλίνη συχνά αναπτύσσεται πριν από τον διαβήτη τύπου 2. Αρχικά, η αντίσταση στην ινσουλίνη αναγκάζει το σώμα να παράγει επιπλέον ινσουλίνη, για να αναπληρώσει την αναποτελεσματική ινσουλίνη. Η επιπλέον ινσουλίνη στην κυκλοφορία του αίματος μπορεί να προκαλέσει υπογλυκαιμία. Αλλά η αντίσταση στην ινσουλίνη τείνει να χειροτερεύει με την πάροδο του χρόνου. Τελικά, μειώνει την ικανότητα του σώματός να παράγει ινσουλίνη. Καθώς τα επίπεδα ινσουλίνης μειώνονται, τα επίπεδα σακχάρου στο αίμα αυξάνονται. Εάν τα επίπεδα δεν επανέλθουν στο φυσιολογικό, αυξάνεται ο κίνδυνος για διαβήτη τύπου 2. Συνεπώς μπορούμε να δούμε γιατί είναι τόσο σημαντικό αυτό το δεδομένο για το σύστημα που αναπτύχθηκε.

Insulin2 □ Το attribute αυτό παίρνει 2 τιμές οι οποίες είναι no το οποίο σημαίνει ότι ο ασθενής δεν παίρνει φαρμακευτική αγωγή ινσουλίνης και η δεύτερη τιμή είναι !no η οποία δηλώνει ότι ο ασθενής παίρνει φαρμακευτική αγωγή πάνω στην ινσουλίνη.

Ο καλός έλεγχος του σακχάρου στο αίμα είναι σημαντικός επειδή βοηθά να αποφύγει ο ασθενής επιπλοκές όπως απώλεια όρασης, νευρική και νεφρική βλάβη και καρδιακές

παθήσεις. Στο παρελθόν, ο γιατρός μπορεί να χορηγούσε ινσουλίνη μόνο εάν τα στοματικά φάρμακα όπως η μετφορμίνη(την οποία θα δούμε παρακάτω) και ο τρόπος ζωής αλλάζουν όπως η διατροφή και η άσκηση δεν μειώνουν αρκετά τα επίπεδα σακχάρου στο αίμα. Η ινσουλίνη μπορεί να βοηθήσει στον καλύτερο έλεγχο του διαβήτη. Μπορεί επίσης να βοηθήσει το πάγκρεας να κάνει τη δική του ινσουλίνη περισσότερη. Το πλεονέκτημα της ινσουλίνης είναι ότι μειώνει το σάκχαρο στο αίμα καλύτερα από τα στοματικά φάρμακα. Η ινσουλίνη επιτρέπει επίσης πιο ευέλικτη δοσολογία. Ο γιατρός μπορεί να τροποποιήσει τη δόση για να επιτύχει τους ακριβείς στόχους του σακχάρου στο αίμα. Ο γιατρός θα ελέγξει τον έλεγχο του σακχάρου στο αίμα με μια δοκιμή A1c. Αυτή η εξέταση αίματος μετράει το μέσο επίπεδο σακχάρου στο αίμα κατά τους τελευταίους 2 ή 3 μήνες. Ο στόχος A1c για κάποιον με διαβήτη τύπου 2 είναι μικρότερος από 7%.

Μπορεί να πάρει κάποιος ασθενής ινσουλίνη για διαβήτη τύπου 2 επειδή:

- Πρέπει να ελέγξει το σάκχαρο στο αίμα για μικρό χρονικό διάστημα. Μπορεί να χρειαστεί ινσουλίνη για να αντιμετωπίσει προσωρινά υψηλά επίπεδα σακχάρου στο αίμα, λόγω:
 - Χειρουργικής επέμβαση
 - Μια ιατρική κατάσταση έκτακτης ανάγκης όπως καρδιακή προσβολή
 - Εγκυμοσύνη
 - Στεροειδή φάρμακα
 - Διαβητική κετοξέωση - όταν το σώμα δεν έχει αρκετή ινσουλίνη και καίει λίπος για καύσιμο αντί για ζάχαρη

Metformin □ Το attribute metformin στέλνει 4 τιμές μέσα στον αλγόριθμο, αυτές οι τιμές είναι yes όπου δηλώνει ότι ο ασθενής λαμβάνει κάποια αγωγή μετφορμίνης, no το οποίο δηλώνει ότι ο ασθενής δεν λαμβάνει κάποια αγωγή μετφορμίνης, steady το οποίο δηλώνει ότι ο ασθενής παίρνει σταθερή ποσότητα μετφορμίνης από την στιγμή που του χορηγήθηκε και !steady το οποίο δηλώνει το αντίθετο, δηλαδή ότι ο ασθενής άλλαξε την ποσότητα μετφορμίνης που λαμβάνει από την ώρα που ξεκίνησε την αγωγή.

Η μετφορμίνη είναι ένα φάρμακο που χρησιμοποιείται για τη θεραπεία του διαβήτη τύπου 2 και για την πρόληψη του διαβήτη τύπου 2 εάν διατρέχει ο ασθενής υψηλό κίνδυνο για

να αναπτύξει στο μέλλον. Η μετφορμίνη χρησιμοποιείται κατά τη θεραπεία του συνδρόμου πολυκυστικών ωοθηκών (PCOS), αν και δεν είναι επίσημα εγκεκριμένη για PCOS. Το PCOS είναι μια κατάσταση που επηρεάζει τον τρόπο λειτουργίας των ωοθηκών. Η μετφορμίνη μειώνει τα επίπεδα σακχάρου στο αίμα βελτιώνοντας τον τρόπο με τον οποίο το σώμα σας χειρίζεται την ινσουλίνη. Συνήθως συνταγογραφείται για διαβήτη όταν η δίαιτα και η άσκηση μόνοι τους δεν ήταν αρκετά για να ελέγχουν τα επίπεδα σακχάρου στο αίμα. Για γυναίκες με PCOS, η μετφορμίνη μειώνει τα επίπεδα ινσουλίνης και σακχάρου στο αίμα και μπορεί επίσης να διεγείρει την ωορρηξία. Η μετφορμίνη διατίθεται με ιατρική συνταγή ως δισκία και ως υγρό που πίνετε.

Num medications □ Το συγκεκριμένο attribute δηλώνει πόσες διαφορετικές ουσίες παίρνει από τα φάρμακα του ο ασθενής(συμπεριλαμβάνονται και τα φάρμακα που του χορηγήθηκαν κατά την παραμονή του στο νοσοκομείο / κλινική) στέλνοντας στον αλγόριθμο ένα ακέραιο νούμερο.

Σε αυτό το δεδομένο μπορούν να είναι πολλά και διαφορετικά είδη φαρμάκων αλλά τα πιο συνηθισμένα που θα πρέπει να προσέξει και οπωσδήποτε να συμπεριλάβει ως φάρμακα που χρησιμοποιεί μέσα στο σύστημα είναι τα εξής:

- Ινσουλίνη
- Μετφορμίνη
- Ακαρβόζη
- Μιγλιτόλη
- Γλιπιζίδη
- Γλυβουρίδη
- Πιογλιταζόνη
- Ρεπαγλινίδη
- Ροσιγλιταζόνη
- Ντοπαμίνη
- Αλογλιπτίνη

- Λιναγλιπτίνη
- Γλιμεπιρίδη
- Τολμπουταμίδη
- Χλωροπροπαμίδη
- Κ.α.

Max Glu serum(Glucose serum test result) □ Το attribute αυτό δηλώνει αν ο ασθενής έχει τα επίπεδα γλυκόζης του σε φυσιολογικά επίπεδα η όχι , μέσω των τιμών normal και none.

Στους παρακάτω πίνακες βλέπουμε τα επίπεδα γλυκόζης βάση των οποίων πρέπει να δοθεί και η αντίστοιχη απάντηση εντός του συστήματος για να βρούμε σωστά αποτελέσματα.

ΠΙΝΑΚΑΣ

Test Γλυκόζης	Κανονικό (Φυσιολογικές τιμές)	Προ- διαβήτη(Επικίνδυνα επίπεδα)	Διαβήτης (Επίπεδα με απαραίτητη την ιατρική περίθαλψη)
Τυχαία στιγμή εντός της ημέρας	Κάτω του 11.1 mmol/l Κάτω του 200 mg/dl	-	Περισσότερο από 11.1 mmol/l Περισσότερο από 200 mg/dl
χωρίς φαγητό	Κάτω του 5.5 mmol/l Κάτω του 100 mg/dl	5.5 – 6.9 mmol/l 100 – 125 mg/dl	Περισσότερο από 7.0 mmol/l Περισσότερο από 126 mg/dl
2 ώρες μετά τον ύπνο	Κάτω του 7.8 mmol/l Κάτω του 140 mg/dl	7.8 – 11.0 mmol/l 140 – 199 mg/dl	Περισσότερο από 11.1 mmol/l Περισσότερο από 200 mg/dl

Number inpatient □ Το attribute αυτό δηλώνει τον αριθμό των εσωτερικών επισκέψεων του ασθενή πριν από την λήψη αγωγής του. Οι εσωτερικές επισκέψεις τις οποίες θα πρέπει να μετρήσει ο ασθενής και να εισάγει στο σύστημα είναι οι επισκέψεις που γίνανε σχετικά με το πρόβλημα του διαβήτη που αντιμετωπίζει ο ασθενής. Στο σύστημα αυτό το δεδομένο είναι αρκετά σημαντικό καθώς βλέποντας πόσες φορές έκανε εσωτερική επίσκεψη ο ασθενής ο αλγόριθμος μπορεί να υπολογίσει καλύτερα αν θα χρειαστεί να επισκεφθεί το ιατρείο ο ασθενής. Αυτό συμβαίνει διότι όσες περισσότερες επισκέψεις κάνει ο ασθενής τόσο καλύτερη θα είναι η άποψη του γιατρού για την κατάσταση του και άρα θα μειωθεί ο κίνδυνος, σε διαφορετική περίπτωση εάν ο γιατρός κρίνει ότι υπάρχει μεγάλος κίνδυνος τότε θα του επισημάνει ότι πρέπει να γίνει επίσκεψη και άρα δεν υπάρχει λόγος για χρήση του συστήματος ακόμη από τον ασθενή.

Number Diagnoses □ Αυτό το attribute δηλώνει τον αριθμό διαγνώσεων του ασθενή που περάστηκαν μέσα στο σύστημα. Οι διαγνώσεις οι οποίες θα πρέπει να λάβει υπόψη ο ασθενής είναι οι εξής:

- Οποιοσδήποτε έχει δείκτη μάζας σώματος υψηλότερο από 25 (23 για Ασιάτες-Αμερικανούς), ανεξάρτητα από την ηλικία, ο οποίος έχει επιπλέον παράγοντες κινδύνου, όπως υψηλή αρτηριακή πίεση, ανώμαλα επίπεδα χοληστερόλης, καθιστικό τρόπο ζωής, ιστορικό συνδρόμου πολυκυστικών ωοθηκών ή καρδιακών παθήσεων και έχοντας στενό συγγενή με διαβήτη.
- Κάθε γυναίκα που είχε διαβήτη κύησης.
- Όποιος έχει διαγνωστεί με πρώιμου επιπέδου διαβήτη.
- Όποιος έχει κακές εξετάσεις σε γλυκοποιημένη αιμοσφαιρίνη A1C
- Τα υψηλότερα του κανονικού επίπεδα σακχάρου στο αίμα.
- Τα ανησυχητικά επίπεδα σακχάρου στο αίμα τα οποία είναι κοντά στο όριο σε άτομα άνω των 45 επίσης θα πρέπει να εισαχθεί στο σύστημα σαν επιπλέον δεδομένο

Num Procedures □ Αυτό το στοιχείο στέλνει στον αλγόριθμο τον αριθμό των εξετάσεων που έκανε ο ασθενής εκτός από τα εργαστηριακά τεστ. Τα που θα πρέπει να εισάγει ο ασθενής στο σύστημα είναι εξετάσεις όπως :

- Εξετάσεις A1C
- Βαριατρική Επέμβαση
- Ενδοσκοπική γαστροπλαστική
- Γαστρικό Bypass
- Δοκιμή ανοχής γλυκόζης

Readmission □ Το readmission έχει τροποποιηθεί από το αρχικό dataset και πλέον υπάρχουν 2 τιμές που περνούν στο σύστημα οι οποίες δείχνουν εάν ο ασθενής έχει κάποιο ραντεβού για επανεξέταση στο μέλλον σε διάστημα >30 ημερών ή όχι. Το δεδομένο αυτό είναι σημαντικό για το σύστημα καθώς βοηθά στην καλύτερη πρόβλεψη μέσω των επιλογών που δόθηκαν στο σύστημα. Εάν κάποιος ασθενής έχει επανεξέταση με τον γιατρό του σε λιγότερες από 30 ημέρες τότε αυτό σημαίνει ότι θα διατρέχει μικρότερο κίνδυνο αφού ο γιατρός έχει κρίνει την κατάσταση του σχετικά σύντομα, ωστόσο ανάλογα με τα υπόλοιπα δεδομένα αυτό μπορεί να είναι και αρνητικό διότι αν τα υπόλοιπα δεδομένα δεν είναι καλά τότε μπορεί και να σημαίνει ότι ο ασθενής είναι σε επικίνδυνη ζώνη και για αυτό το λόγο πρέπει να γίνει η επανεισδοχή του στην κλινική τόσο σύντομα. Αντίστοιχα χρησιμοποιείται και το μεγαλύτερο των 30 ημερών όπου σημαίνει ότι θα αργήσει να εξετάσει τον ασθενή ο ιατρός ωστόσο μεγάλο ρόλο παίζουν και τα υπόλοιπα δεδομένα καθώς με προβληματικά τα υπόλοιπα δεδομένα σημαίνει ότι υπάρχει μεγάλο περιθώριο μεταξύ των επισκέψεων ενώ με θετικά σημαίνει ότι ο ασθενής δεν διατρέχει άμεσο κίνδυνο. Η επιλογή όχι συνήθως συνοδεύεται από καλά δεδομένα καθώς ο ασθενής έχει σταθερή κατάσταση ή δεν αντιμετωπίζει κάποιο άμεσο κίνδυνο.

A1Cresult □ Το attribute αυτό παίρνει 4 τιμές οι οποίες είναι >8 δηλαδή οι εξετάσεις δείχνουν >8 , $>7\%$ εάν οι εξετάσεις είναι <8 αλλά >7 , normal εάν οι εξετάσεις είναι <7 και άρα σε φυσιολογικά επίπεδα και τέλος none εάν δεν έχει μετρηθεί καθόλου το συγκεκριμένο attribute.

Το τεστ A1C - επίσης γνωστό ως τεστ αιμοσφαιρίνης A1C ή HbA1c - είναι ένα απλό τεστ αίματος που μετρά τα μέσα επίπεδα σακχάρου στο αίμα τους τελευταίους 3 μήνες. Είναι μια από τις συχνότερα χρησιμοποιούμενες εξετάσεις για τη διάγνωση του προδιαβήτη και του διαβήτη. Τα υψηλότερα επίπεδα A1C συνδέονται με επιπλοκές του διαβήτη, επομένως η επίτευξη και η διατήρηση του ατομικού στόχου A1C είναι πραγματικά σημαντική εάν ένας ασθενής πάσχει από διαβήτη. Όταν η ζάχαρη εισέρχεται στην κυκλοφορία του αίματός, προσκολλάται στην αιμοσφαιρίνη, μια πρωτεΐνη στα ερυθρά αιμοσφαίρια. Όλοι έχουν ζάχαρη

συνδεδεμένη με την αιμοσφαιρίνη τους, αλλά τα άτομα με υψηλότερα επίπεδα σακχάρου στο αίμα έχουν περισσότερα. Η δοκιμή A1C μετρά το ποσοστό των ερυθρών αιμοσφαιρίων που έχουν αιμοσφαιρίνη επικαλυμμένη με ζάχαρη.

Τα άτομα τα οποία θα πρέπει να κάνουν αυτό το τεστ και να το εισάγουν στο σύστημα τα αποτελέσματα είναι εάν :

- Εάν το αποτέλεσμα είναι φυσιολογικό, αλλά είναι άνω των 45 ετών, υπάρχουν παράγοντες κινδύνου ή ο ασθενής έπασχε ποτέ από διαβήτη κύησης.
- Εάν το αποτέλεσμα δείχνει ότι ο ασθενής είναι σε επίπεδα προ διαβήτη.
- Εάν δεν υπάρχουν συμπτώματα, αλλά το αποτέλεσμα δείχνει ότι υπάρχουν αποτελέσματα προ διαβήτη ή διαβήτη, πρέπει να γίνει δοκιμή σε άλλη μέρα για αν επιβεβαιωθούν τα αποτελέσματα.
- Εάν το τεστ δείχνει ότι υπάρχει διαβήτης.

ΠΙΝΑΚΑΣ

Φυσιολογικές τιμές	Κάτω του 5.7%
Τιμές Προδιαβήτη	Από 5.7% έως 6.4%
Τιμές Διαβήτη	Τιμές από 6.5% και άνω

Num Emergency □ Αριθμός των επειγουσών επισκέψεων του ασθενή στο νοσοκομείο / κλινική την χρονιά πριν την εξέταση του.

Οι ερευνητές διαπίστωσαν ότι το 2015 άτομα με διαβήτη ηλικίας 45 ετών και άνω πραγματοποίησαν περίπου 12 εκατομμύρια επισκέψεις στο τμήμα έκτακτης ανάγκης, ποσοστό 92 ανά 1.000 άτομα. Από το 2012 έως το 2015, το ποσοστό όλων των επισκέψεων έκτακτης ανάγκης από ασθενείς ηλικίας 45 ετών και άνω αυξήθηκε για άτομα με διαβήτη. Εκείνοι ηλικίας 75 ετών και άνω είχαν ποσοστό επίσκεψης στο τμήμα επειγόντων περιστατικών

διαβήτη που ήταν περίπου 2,5 φορές υψηλότερο από το ποσοστό για άτομα ηλικίας 45 έως 64 ετών.

Συνολικά, το 24% των επισκέψεων στο τμήμα έκτακτης ανάγκης για διαβήτη για άτομα ηλικίας 45 έως 64 ετών έχουν τα τμήματα έκτακτης ανάγκης ως την κύρια αναμενόμενη πηγή πληρωμής, σε σύγκριση με το 14% αυτών για ασθενείς χωρίς διαβήτη στην ίδια ηλικιακή ομάδα. Το είκοσι δύο τοις εκατό των επισκέψεων στο τμήμα επειγόντων περιστατικών για διαβήτη για άτομα ηλικίας 45 έως 64 ετών χρειαστήκανε ιατρική περίθαλψη στο νοσοκομείο σε σύγκριση με το 12% των επισκέψεων από ασθενείς χωρίς διαβήτη σε αυτήν την ηλικιακή ομάδα.

3.2 Χρήση Του WEKA Στο Σύστημα

Στο συγκεκριμένο σύστημα δουλεύει μέσα από Classification μιας κλάσης οπότε επιλέχθηκε η επιλογή εργαλείου WEKA.

Ανοίγοντας το WEKA θα διαλέξουμε το αρχείο με τα δεδομένα που θέλουμε να δουλέψουμε στο σύστημα, αφού πρώτα μετατρέψουμε το αρχείο από excel σε αρχείο WEKA.

Στη συνέχεια Επιλέγοντας την βάση δεδομένων που δημιουργήθηκε έγινε Classify με Class to Change το οποίο δείχνει αν θα χρειαστεί να αλλάξει η θεραπεία ή αν δεν του έχει χορηγηθεί, να μας δείξει ότι χρειάζεται να του χορηγηθεί.

Για να το κάνουμε αυτό αφού επιλέξουμε την καρτέλα Classify θα επιλέξουμε την επιλογή Choose και από την επιλογή Trees θα διαλέξουμε τον αλγόριθμο J48. Πριν γίνει αυτό ωστόσο θα πρέπει να επιλέξουμε τις επιλογές του j48 και να επιλέξουμε τις καταλλήλες επιλογές. Τρέχοντας τον αλγόριθμο θα πάρουμε τα παρακάτω αποτελέσματα.

3.3 κώδικας – Δέντρο WEKA

```
insulin = Down: Ch (12218.0)
```

```
insulin != Down
```

```
| insulin = Up: Ch (11316.0)
```

```

|  insulin != Up
|  |  metformin = No
|  |  |  insulin = No: No (37407.0/3136.0)
|  |  |  insulin != No
|  |  |  |  age = [10-20): No (308.0/4.0)
|  |  |  |  age != [10-20)
|  |  |  |  |  age = [20-30): No (430.0/25.0)
|  |  |  |  |  age != [20-30)
|  |  |  |  |  |  age = [30-40): No (874.0/145.0)
|  |  |  |  |  |  age != [30-40)
|  |  |  |  |  |  |  age = [40-50): No (2125.0/546.0)
|  |  |  |  |  |  |  age != [40-50)
|  |  |  |  |  |  |  |  num_medications <= 14.0: No (9542.0/2845.0)
|  |  |  |  |  |  |  |  num_medications > 14.0
|  |  |  |  |  |  |  |  |  max_glu_serum = Norm: No (226.0/75.0)
|  |  |  |  |  |  |  |  |  max_glu_serum != Norm
|  |  |  |  |  |  |  |  |  |  number_inpatient <= 1.0
|  |  |  |  |  |  |  |  |  |  |  age = [50-60): No (1756.0/677.0)
|  |  |  |  |  |  |  |  |  |  |  age != [50-60)
|  |  |  |  |  |  |  |  |  |  |  |  num_medications <= 27.0
|  |  |  |  |  |  |  |  |  |  |  |  |  age = [90-100): Ch (263.0/128.0)
|  |  |  |  |  |  |  |  |  |  |  |  |  age != [90-100)
|  |  |  |  |  |  |  |  |  |  |  |  |  |  number_diagnoses <= 8.0
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  readmitted = <30: Ch (260.0/116.0)

```

```

| | | | | | | | | | | | | | | readmitted != <30
| | | | | | | | | | | | | | | time_in_hospital <= 3.0
| | | | | | | | | | | | | | | num_procedures <= 0.0: Ch (238.0/100.0)
| | | | | | | | | | | | | | | num_procedures > 0.0
| | | | | | | | | | | | | | | time_in_hospital <= 2.0: No (346.0/145.0)
| | | | | | | | | | | | | | | time_in_hospital > 2.0: Ch (211.0/89.0)
| | | | | | | | | | | | | | | time_in_hospital > 3.0: No (1185.0/526.0)
| | | | | | | | | | | | | | | number_diagnoses > 8.0
| | | | | | | | | | | | | | | number_inpatient <= 0.0
| | | | | | | | | | | | | | | age = [60-70): No (910.0/355.0)
| | | | | | | | | | | | | | | age != [60-70)
| | | | | | | | | | | | | | | gender = Female: No (1137.0/477.0)
| | | | | | | | | | | | | | | gender != Female
| | | | | | | | | | | | | | | time_in_hospital <= 5.0: Ch (483.0/225.0)
| | | | | | | | | | | | | | | time_in_hospital > 5.0: No (315.0/139.0)
| | | | | | | | | | | | | | | number_inpatient > 0.0: No (1097.0/413.0)
| | | | | | | | | | | | | | | num_medications > 27.0
| | | | | | | | | | | | | | | readmitted = >30: Ch (314.0/134.0)
| | | | | | | | | | | | | | | readmitted != >30
| | | | | | | | | | | | | | | number_diagnoses <= 8.0: Ch (264.0/114.0)
| | | | | | | | | | | | | | | number_diagnoses > 8.0: No (438.0/199.0)
| | | | | | | | | | | | | | | number_inpatient > 1.0: No (1944.0/707.0)
| | metformin != No
| | | insulin = No

```

```

| | | | metformin = Steady
| | | | | A1Cresult = >8: Ch (618.0/188.0)
| | | | | A1Cresult != >8
| | | | | | num_medications <= 11.0
| | | | | | | num_medications <= 6.0: No (576.0/210.0)
| | | | | | | num_medications > 6.0
| | | | | | | | number_diagnoses <= 8.0
| | | | | | | | | num_procedures <= 0.0
| | | | | | | | | | gender = Female
| | | | | | | | | | | num_medications <= 8.0: No (217.0/99.0)
| | | | | | | | | | | num_medications > 8.0: Ch (412.0/193.0)
| | | | | | | | | | | gender != Female: Ch (539.0/235.0)
| | | | | | | | | | | num_procedures > 0.0: No (565.0/256.0)
| | | | | | | | | | | number_diagnoses > 8.0: No (628.0/256.0)
| | | | | | | num_medications > 11.0
| | | | | | | | number_emergency <= 2.0
| | | | | | | | | A1Cresult = Norm: No (261.0/121.0)
| | | | | | | | | A1Cresult != Norm
| | | | | | | | | | number_inpatient <= 1.0: Ch (4873.0/1975.0)
| | | | | | | | | | | number_inpatient > 1.0
| | | | | | | | | | | readmitted = >30: No (279.0/129.0)
| | | | | | | | | | | readmitted != >30: Ch (262.0/128.0)
| | | | | | | | | | | number_emergency > 2.0: No (80.0/31.0)
| | | | | | | | metformin != Steady: Ch (666.0)

```

| | | insulin != No: Ch (6183.0)

3.4 Πειράματα Και Αποτελέσματα Έρευνας

Η εφαρμογή όπως προαναφέρθηκε και παραπάνω έχει ως κύριο στόχο να εισάγει κάποια πολύ σημαντικά δεδομένα για τον ασθενή και να προβλέψει αν στο μέλλον θα χρειαστεί να αλλάξει την αγωγή του και να επισκεφθεί τον γιατρό που εξετάζει την κατάστασή του ή εάν δεν θα έχει επιπλοκές στο μέλλον σύμφωνα με τα τωρινά δεδομένα. Το σύστημα κατάφερε να πετύχει ένα ποσοστό επιτυχίας 85% μέσα σε ένα Dataset περίπου 100000 ασθενών από 13 διαφορετικά νοσοκομεία της Αμερικής. Για την λειτουργία του έγινε 66% training και το υπόλοιπο testing χρησιμοποιώντας το k fold validation με τα folds, δηλαδή το k να ισούται με 10, δηλαδή 10 φορές. Ο αλγόριθμος ο οποίος χρησιμοποιήθηκε είναι ο j48 και το εργαλείο που χρησιμοποιήθηκε για την υλοποίηση του συστήματος είναι το WEKA.

Το τελικό σύστημα χρησιμοποίησε τον αλγόριθμο j48 ο οποίος δημιουργεί δέντρα αποφάσεων. Ο j48 είναι ο γνωστός αλγόριθμος c4.5 γραμμένος ξανά στην γλώσσα προγραμματισμού Java για την ενσωμάτωση του στο WEKA.

Στο σύστημα που πραγματοποιήθηκε πριν τρέξουμε τα δεδομένα έγινε αρχικά αφαίρεση κενών τιμών. Λόγω του Dataset το οποίο χρησιμοποιήθηκε το οποίο είναι ένα Dataset που δημιουργήθηκε για διαφορετική χρήση έπρεπε να γίνει αφαίρεση αρκετών δεδομένων. Συνεπώς αφού χρησιμοποιήθηκαν κάποιοι αλγόριθμοι και βγήκαν τα σχετικά αποτελέσματα παρατηρήθηκε ποια από τα δεδομένα δεν χρειάζονται στο σύστημα που αναπτύχθηκε και έτσι έγινε ένα είδος φιλτραρίσματος των δεδομένων και μείνανε μόνο τα απαραίτητα δεδομένα έτσι ώστε να δημιουργηθεί το κατάλληλο δέντρο που χρησιμοποιήθηκε για την δημιουργία της μετέπειτα εφαρμογής.

Το J48 δημιουργεί δέντρα αποφάσεων από ένα σύνολο δεδομένων με τον ίδιο τρόπο όπως το c4.5. Τα δεδομένα στα οποία γίνεται το training είναι ένα σύνολο $S = s_1, s_2, \dots$ ήδη ταξινομημένων δειγμάτων. Κάθε δείγμα S_i αποτελείται από ένα p-διαστατικό φορέα $(X_{1,j}, X_{2,j}, \dots, X_{p,i})$ όπου το χαρακτηριστικό X_j αντιπροσωπεύει τιμές ή χαρακτηριστικά του δείγματος, καθώς και την κλάση στην οποία πέφτει το S_i .

Σε κάθε κόμβο του δέντρου, το J48 επιλέγει το χαρακτηριστικό των δεδομένων που χωρίζει αποτελεσματικότερα το σύνολο των δειγμάτων του σε υποσύνολα εμπλουτισμένα σε

μια κατηγορία ή στην άλλη. Το κριτήριο διαχωρισμού είναι το ομαλοποιημένο κέρδος πληροφοριών. Το χαρακτηριστικό με το υψηλότερο κέρδος πληροφοριών επιλέγεται για τη λήψη της απόφασης. Ο αλγόριθμος J48 επαναλαμβάνεται στη συνέχεια στις κατατομημένες δευτερεύουσες λίστες.

Αυτός ο αλγόριθμος έχει μερικές βασικές περιπτώσεις.

Όλα τα δείγματα στη λίστα ανήκουν στην ίδια κλάση. Όταν συμβαίνει αυτό, δημιουργεί απλά έναν κόμβο φύλλων για το δέντρο αποφάσεων που λέει ότι επιλέγει αυτήν την κλάση.

Καμία από τις δυνατότητες δεν παρέχει κέρδος πληροφοριών. Σε αυτήν την περίπτωση, το j48 δημιουργεί έναν κόμβο αποφάσεων ψηλότερα στο δέντρο χρησιμοποιώντας την αναμενόμενη τιμή της κλάσης.

Εάν παρουσιαστεί κλάση που δεν βρέθηκε προηγουμένως τότε και πάλι, το J48 δημιουργεί έναν κόμβο απόφασης ψηλότερα στο δέντρο χρησιμοποιώντας την αναμενόμενη τιμή.

Στο σύστημα το οποίο αναπτύχθηκε γίνανε διάφορες μετρήσεις χρησιμοποιώντας αυτόν τον αλγόριθμο με τις πιο πετυχημένες να είναι κάνοντας binary split τα φύλλα των δέντρων έτσι ώστε να βγει ένα σωστότερο δέντρο αποφάσεων και να γίνει καλύτερη επεξεργασία των δεδομένων. Σχετικά με την επεξεργασία των δεδομένων, προκειμένου να βελτιωθεί το δέντρο δοκιμαστηκαν διαφορετικές τιμές στην παράμετρο minNumObj του WEKA η οποία είναι υπεύθυνη για τον αριθμό των δεδομένων που επεξεργάζονται σε κάθε φύλλο του δέντρου. Λόγω του μεγάλου όγκου δεδομένων δοκιμαστηκαν 3 διαφορετικές τιμές οι οποίες είναι 100, 200 και 500. Δοκιμάζοντας 100 ο αριθμός των φύλλων είναι 68 που μας δείχνει ότι είναι πολύ μικρή η τιμή που δόθηκε καθώς δεν μπορεί να γίνει η σωστή επεξεργασία του δέντρου λόγω του όγκου των φύλλων πάνω στο δέντρο. Συνεπώς δοκιμάστηκε η τιμή 500 η οποία δημιούργησε ένα δέντρο 12 φύλλων η οποία είναι πάρα πολύ μικρή με πάρα πολλά προβλήματα διότι μας δείχνει ότι λόγω των πολλών δεδομένων που επεξεργάζεται σε κάθε φύλλο και κάθε στιγμή δεν έγιναν σωστές μετρήσεις καθώς πολλά σημαντικά δεδομένα δεν χρησιμοποιήθηκαν για την δημιουργία του δέντρου. Η τιμή για την οποία υπάρχει ένα πολύ καλό δέντρο και αποτελέσματα είναι χρησιμοποιώντας την τιμή 200 η οποία μας έδωσε ένα πολύ ευανάγνωστο δέντρο με 39 φύλλα και ένα πολύ καλό ποσοστό 85% και είναι το πείραμα το οποίο χρησιμοποιήθηκε για την τελική κατασκευή του συστήματος και της εφαρμογής.

Εκτός από τον j48 χρησιμοποιήθηκαν και άλλοι γνωστή αλγόριθμοι όπως ο IBK η αλλιώς k nearest neighbor με Linear Search, με τιμές 1, 2, 3 σαν κοντινότερους γείτονες, τα αποτελέσματα ήταν αρκετά καλό ωστόσο δεν ξεπερνούσε τα αποτελέσματα που μας έδωσε ο j48. Επίσης ο j48 μας δίνει το δέντρο το οποίο κάνει πιο κατανοητή την δομή που θα πρέπει να ακολουθήσουμε για την κατασκευή της εφαρμογής βάση του συστήματος και των αλγορίθμων που χρησιμοποιήθηκαν.

Για $k = 1$ είχαμε τα παρακάτω αποτελέσματα:

Correctly Classified Instances	81640	80.2233 %
Incorrectly Classified Instances	20126	19.7767 %
Kappa statistic	0.6016	
Mean absolute error	0.2011	
Root mean squared error	0.443	
Relative absolute error	40.4591 %	
Root relative squared error	88.8532 %	
Total Number of Instances	101766	

Για $k = 2$ έχουμε ακόμη καλύτερα αποτελέσματα όπως φαίνεται παρακάτω :

Correctly Classified Instances	84395	82.9304 %
Incorrectly Classified Instances	17371	17.0696 %
Kappa statistic	0.6514	
Mean absolute error	0.203	
Root mean squared error	0.387	
Relative absolute error	40.8295 %	
Root relative squared error	77.6206 %	

Total Number of Instances	101766
---------------------------	--------

Για $k = 3$ έχω σχεδόν τα ίδια αποτελέσματα με $k = 2$:

Correctly Classified Instances	83706	82.2534 %
--------------------------------	-------	-----------

Incorrectly Classified Instances	18060	17.7466 %
----------------------------------	-------	-----------

Kappa statistic	0.6409
-----------------	--------

Mean absolute error	0.2038
---------------------	--------

Root mean squared error	0.3664
-------------------------	--------

Relative absolute error	40.9891 %
-------------------------	-----------

Root relative squared error	73.494 %
-----------------------------	----------

Total Number of Instances	101766
---------------------------	--------

Επίσης για τα πειράματα του συστήματος χρησιμοποιήθηκε ο αλγόριθμος Naïve Bayes ο οποίος έδειξε επίσης πολύ καλά αποτελέσματα όπως φαίνεται από το matrix παρακάτω:

Correctly Classified Instances	83284	81.8387 %
--------------------------------	-------	-----------

Incorrectly Classified Instances	18482	18.1613 %
----------------------------------	-------	-----------

Kappa statistic	0.63
-----------------	------

Mean absolute error	0.2368
---------------------	--------

Root mean squared error	0.3572
-------------------------	--------

Relative absolute error	47.6267 %
-------------------------	-----------

Root relative squared error	71.6379 %
-----------------------------	-----------

Total Number of Instances	101766
---------------------------	--------

ΠΙΝΑΚΑΣ

	Min num Obj test 100, binary split true	Min num Obj test 200, binary split true	Min num Obj test 500, binary split true	Binary Split False, Min num Obj test 200, K = 10	Simple Estimator, K2, K = 10	Simple Estimator, Repeated , K = 10	KNN = 1, K = 10	KNN = 2, K = 10	KNN = 5, K = 10	K = 2,	K = 3	K = 5
IBK	-	-	-	-	-	-	81%	83%	83%	KNN = 5 / 83%	KNN = 5 / 83%	KNN = 5 / 83%
Naive Bayes	-	-	-	-	83%	83%	-	-	-	Simple Estimator, K2 / 83%	Simple Estimator, K2 / 83%	Simple Estimator, K2 / 83%
J48	85%	85%	85%	85%						Min num Obj test 200, binary split true / 85%	Min num Obj test 200, binary split true / 85%	Min num Obj test 200, binary split true / 85%

3.5 Περιγραφή Εφαρμογής

Η εφαρμογή σχεδιάστηκε ώστε να υποστηρίζει τους γιατρούς σχετικά με τις αποφάσεις που θα πρέπει να λάβουν πάνω στο θέμα του διαβήτη, αλλά και για προσωπική χρήση των ασθενών ώστε να παρακολουθούν την κατάστασή τους και να προσέχουν, έτσι ώστε να επισκεφθούν τον γιατρό που τους παρακολουθεί αν χρειαστεί.

Η εφαρμογή είναι αρκετά εύκολη στην χρήση της και τα βήματα για την λειτουργία της είναι τα εξής

- 1) Αρχικά ο χρήστης οποίος μπορεί να είναι γιατρός ή ασθενής θα εισέλθει στην εφαρμογή και θα δει την αρχική οθόνη που φαίνεται παρακάτω

3.1 Αρχική οθόνη συστήματος

11:57

Diabetes Medication Support App

Option 1 - Yes Option 2 - No

Normal Insulin Blood Levels ☒ ☐

Insulin Drug Medication ☐ ☒

Normal Glycose Blood Levels ☒ ☐

Metformin Drug Medication ☐ ☒

Hospital Readmission Issued ☐ ☒

Are you a Male ☐ ☒

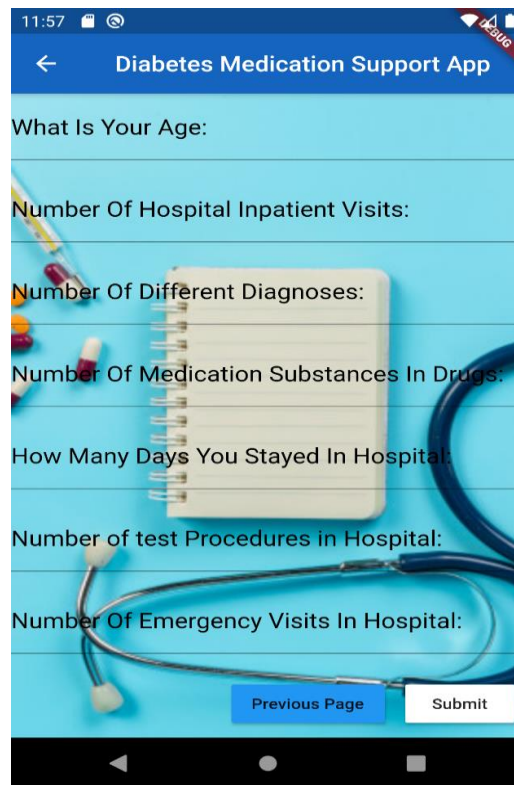
Normal A1C test Results ☒ ☐

Steady Dosage Of Metformin ☒ ☐

Next Page

Στην παραπάνω πρώτη οθόνη ο χρήστης θα επιλέξει ανάμεσα στις 2 επιλογές που υπάρχουν (δηλαδή yes ή no) ανάλογα με τις ερωτήσεις που υπάρχουν και την κατάσταση του, αφού επιλέξει τις επιλογές του τότε θα επιλέξει το κουμπί next page για να προχωρήσει στην επόμενη οθόνη.

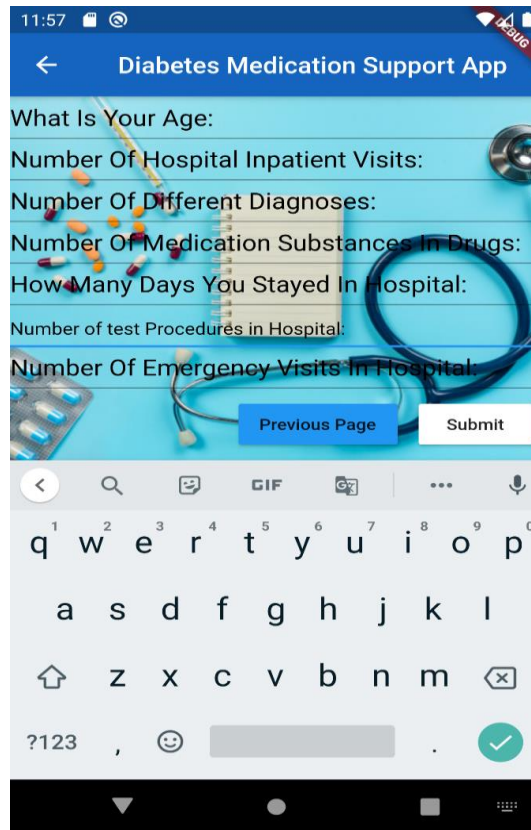
- 2) Έπειτα ο χρήστης θα περάσει στην δεύτερη οθόνη της εφαρμογής όπως φαίνεται παρακάτω

3.2 Δεύτερη οθόνη συστήματος

The screenshot displays the 'Diabetes Medication Support App' interface on a mobile device. The app has a blue header with a back arrow and the title 'Diabetes Medication Support App'. The main content area is light blue and contains a form with the following fields: 'What Is Your Age:', 'Number Of Hospital Inpatient Visits:', 'Number Of Different Diagnoses:', 'Number Of Medication Substances In Drugs:', 'How Many Days You Stayed In Hospital:', 'Number of test Procedures in Hospital:', and 'Number Of Emergency Visits In Hospital:'. The form is overlaid on a background image of medical supplies, including a stethoscope, a syringe, and a spiral notebook. At the bottom of the form, there are two buttons: 'Previous Page' (blue) and 'Submit' (white). The mobile status bar at the top shows the time as 11:57 and various icons.

Η δεύτερη οθόνη για τον χρήστη είναι μια φόρμα η οποία όταν την επιλέγει ο χρήστης μπορεί να συμπληρώσει τα στοιχεία που πρέπει σύμφωνα με τις οδηγίες ώστε να περαστούν στον αλγόριθμο,

3.3 Εισαγωγή στοιχείων στην εφαρμογή

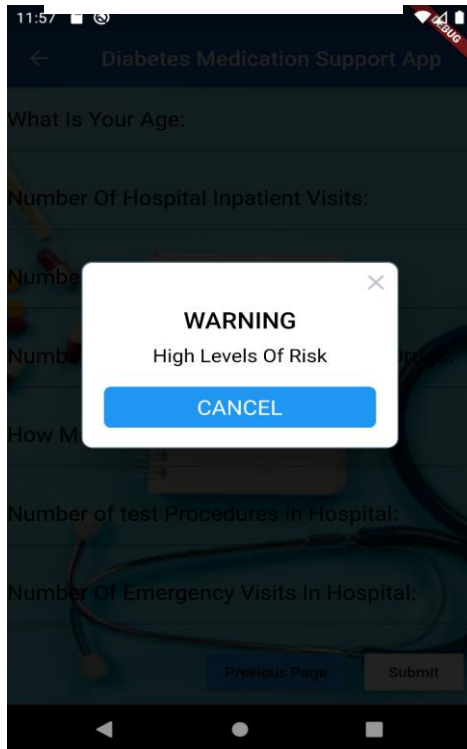


The screenshot displays the 'Diabetes Medication Support App' interface on a mobile device. The app has a blue header with a back arrow and the title. Below the header, there is a list of input fields for user data: 'What Is Your Age:', 'Number Of Hospital Inpatient Visits:', 'Number Of Different Diagnoses:', 'Number Of Medication Substances In Drugs:', 'How Many Days You Stayed In Hospital:', 'Number of test Procedures in Hospital:', and 'Number Of Emergency Visits In Hospital:'. At the bottom of the form, there are two buttons: 'Previous Page' (blue) and 'Submit' (white). The background of the app interface features a medical-themed image with pills and a stethoscope. Below the app interface, a standard Android keyboard is visible, with the letter 'p' highlighted on the top row.

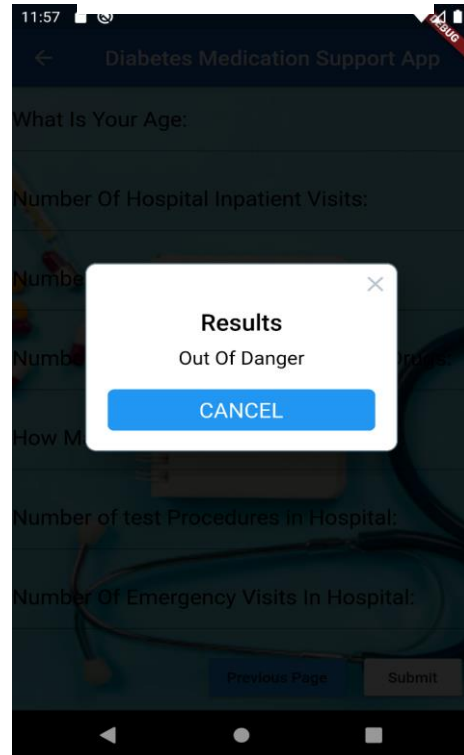
Στην παραπάνω εικόνα ο χρήστης έχει διαλέξει την ερώτηση Number of test Procedures in Hospital και ετοιμάζεται να εισάγει τα στοιχεία του.

Αφού ο χρήστης εισάγει τα στοιχεία του τότε θα πατήσει το κουμπί submit ώστε να τρέξει ο αλγόριθμος που είδαμε προηγουμένως στο Weka και να μας εμφανίσει το κατάλληλο αποτέλεσμα όπως φαίνεται στις παρακάτω εικόνες

3.4 Θετικό αποτέλεσμα εφαρμογής



3.5 Θετικό αποτέλεσμα εφαρμογής



Τρέχοντας τον αλγόριθμο ο χρήστης θα καταλήξει σε 1 από τις 2 παραπάνω εικόνες – αποτελέσματα ανάλογα με τα στοιχεία που εισήγαγε. Τα αποτελέσματα είναι θετικά(εικ. 1) όταν ο αλγόριθμος καταλήγει στο συμπέρασμα ότι ο ασθενής πρέπει να συνεχίσει την θεραπεία του ως έχειν ή θα καταλήξει στο συμπέρασμα ότι πρέπει να προσέχει η ακόμα και ότι θα πρέπει να επισκεφθεί τον γιατρό για να ξαναδούνε την θεραπεία του.

3.6 Απαιτήσεις Προγράμματος

- Να εμφανίζει μήνυμα το αποτέλεσμα το οποίο τρέχει ο αλγόριθμος σύμφωνα με τα εισαχθέντα στοιχεία.
- Ο χρήστης πρέπει να διαθέτει λειτουργικό σύστημα Android ή iOS
- Ο χρόνος απόκρισης να είναι μικρός
- Το σύστημα πρέπει να λειτουργεί αδιάλειπτα
- Τα προσωπικά στοιχεία των χρηστών πρέπει να προστατεύονται
- Το πρόγραμμα θα πρέπει να είναι <<ελαφρύ>>
- Ευκολία Χρήσης καθώς απευθύνεται κυρίως σε ανθρώπους μεγαλύτερης ηλικίας
- Ευελιξία με κάθε αναβάθμιση / Δυνατότητα συντήρησης
- Το σύστημα πρέπει να κάνει έλεγχο των εισαχθέντων στοιχείων(π.χ. αν ο χρήστης βάλει ως ηλικία 200 ετών).
- Ο χρόνος επανεκκίνησης του προγράμματος μετά από αναγκαστική διακοπή να μην είναι μεγαλύτερη των 3 δευτερολέπτων.

3.7 Επεκτασιμότητα

Το συγκεκριμένο πρόγραμμα θα μπορούσε να επεκταθεί μελλοντικά με τους εξής τρόπους. Αρχικά θα μπορούσε να γίνει μια βάση δεδομένων σε firebase όπου θα κρατάει τα στοιχεία των ασθενών που τρέχουν την εφαρμογή. Προϋπόθεση για να γίνει αυτό θα είναι να δημιουργεί λογαριασμό ο χρήστης ή να συνδέεται με τον λογαριασμό του στο google. Στη συνέχεια θα μπορούσε να δημιουργηθεί μενού επιλογών με μια από τις επιλογές να είναι η σύγκριση αποτελεσμάτων στην καρτέλα του χρήστη, όπου ο χρήστης θα μπορεί να συγκρίνει τα στοιχεία που έβαλε πριν να τρέξει τον αλγόριθμο και τα αποτελέσματα με μελλοντικά αποτελέσματα. Τέλος θα μπορούσε να συνδεθεί με την βάση δεδομένων του γιατρού, να τρέχει περισσότερους υπολογισμούς με βάση τα δεδομένα που έδωσε ο χρήστης στην εφαρμογή και με την συγκατάθεση του χρήστη να στέλνει αυτά τα δεδομένα στον γιατρό όταν βρίσκονται σε

ανησυχητικά επίπεδα. Έτσι θα μπορούσα οι ασθενής του γιατρού να τρέχουν την εφαρμογή συχνά και όταν εμφανιζόταν ανησυχητικά αποτελέσματα ο γιατρός να ενημερώνεται.

5. Βιβλιογραφία

- 1. Artificial Intelligence in Medicine (2019), Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes, ScienceDirect
- 2. Applied Computing and Informatics (2018), Predictive modelling and analytics for diabetes using a machine learning approach, ScienceDirect
- 3. Journal of King Saud University - Computer and Information Sciences (2020), Smart home health monitoring system for predicting type 2 diabetes and hypertension, ScienceDirect
- 4. Journal of King Saud University - Computer and Information Sciences (2017), Application of data mining: Diabetes health care in young and old patients, ScienceDirect
- 5. Journal of Biomedical Informatics (2019), Mixed effect machine learning: A framework for predicting longitudinal change in hemoglobin A1c, ScienceDirect
- 6. IEEE Journal of Biomedical and Health Informatics (2015), Identification of Type 2 Diabetes Risk Factors Using Phenotypes Consisting of Anthropometry and Triglycerides based on Machine Learning, IEEE Xplore
- 7. IEEE Transactions on Knowledge and Data Engineering (2016), Prediction of Fasting Plasma Glucose Status Using Anthropometric Measures for Diagnosing Type 2 Diabetes, IEEE Xplore

- 8. Healthcare Technology Letters (2019), Data mining process for predicting diabetes mellitus based model about other chronic diseases: a case study of the northwestern part of Nigeria, IEEE Xplore
- 9. IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS) (2016), An Ontology-Based Interpretable Fuzzy Decision Support System for Diabetes Diagnosis, IEEE Xplore
- 10. IEEE Access (2020), Development of Disease Prediction Model Based on Ensemble Learning Approach for Diabetes and Hypertension, IEEE Xplore
- 11. IEEE 2nd Colombian Conference on Automatic Control (CCAC) (2015), Advanced Techniques for Predicting the Future Progression of Type 2 Diabetes, IEEE Xplore
- 12. International Joint Conference on Neural Networks (IJCNN) (2015), Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers, IEEE Xplore
- 13. (2019), Data-Driven Blood Glucose Pattern Classification and Anomalies Detection: Machine-Learning Applications in Type 1 Diabetes, PubMed
- 14. (2018), Decision Support for Diabetes in Scotland: Implementation and Evaluation of a Clinical Decision Support System, PubMed
- 15. (2017), A Mobile Computerized Decision Support System to Prevent Hypoglycemia in Hospitalized Patients With Type 2 Diabetes Mellitus, PubMed

