



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Web-based Εφαρμογή για τη Διαχείριση Ασθενών με Διαβήτη βάσει Αλγορίθμων Μηχανικής Μάθησης

Διπλωματική Εργασία: Γκαμπλιάς Αθανάσιος

Επιβλέπων: Συμεωνίδης Παναγιώτης (Αν. Καθηγητής)

Μέλη εξεταστικής επιτροπής: Κρητικός Κυριάκος (Αν. Καθηγητής), Κωστούλας
Θεόδωρος (Αν. Καθηγητής)

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Παναγιώτη Συμεωνίδη, για την πολύτιμη καθοδήγηση, την εμπιστοσύνη και την αμέριστη στήριξη που μου παρείχε καθ' όλη τη διάρκεια εκπόνησης αυτής της διπλωματικής εργασίας. Η επιστημονική του κατάρτιση και οι συμβουλές του αποτέλεσαν πολύτιμο εφόδιο για την ολοκλήρωση του έργου.

Θα ήθελα, επίσης, να ευχαριστήσω την οικογένειά μου, για την αγάπη, την κατανόηση και την αδιάκοπη στήριξή τους σε κάθε μου βήμα. Η παρουσία και η ενθάρρυνσή τους ήταν για μένα πηγή δύναμης και αισιοδοξίας σε όλη τη διάρκεια των σπουδών μου.

Τέλος, ένα μεγάλο ευχαριστώ στους φίλους και συναδέλφους μου, που με τη παρέα, τη συνεργασία και τη στήριξή τους συνέβαλαν στο να δημιουργήσω όμορφες αναμνήσεις και να ολοκληρώσω με επιτυχία αυτό το σημαντικό κεφάλαιο της ζωής μου.

Σας ευχαριστώ όλους από καρδιάς!

Περίληψη

Στην παρούσα διπλωματική εργασία παρουσιάζεται η ανάπτυξη ενός ολοκληρωμένου συστήματος για τη διαχείριση του διαβήτη, το οποίο συνδυάζει παραδοσιακά πρωτόκολλα υπολογισμού ινσουλίνης (Sliding Scale Protocol) και αλγορίθμους Μηχανικής Μάθησης. Αρχικά, γίνεται επισκόπηση βασικών εννοιών γύρω από τον σακχαρώδη διαβήτη, τους τύπους (Τύπου 1 και Τύπου 2), καθώς και τις κρίσιμες παραμέτρους ρύθμισης, όπως η γλυκοζυλιωμένη αιμοσφαιρίνη (HbA1c) και η ολική ημερήσια δόση ινσουλίνης (TDD). Ακολούθως, περιγράφονται μέθοδοι υπολογισμού ινσουλίνης, όπως το Sliding Scale Protocol και το Carbo Calculation, καθώς και ο τρόπος εκτίμησης της γλυκόζης μετά το γεύμα.

Η εργασία επικεντρώνεται σε δύο κύρια σενάρια εκπαίδευσης: την πρόβλεψη μελλοντικών επιπέδων γλυκόζης και την πρόβλεψη της κατάλληλης δόσης ινσουλίνης, αξιοποιώντας δεδομένα τόσο από ανθρώπους που ζουν με διαβήτη και παρακολουθούν τη θεραπεία τους μέσω της εφαρμογής chatGLUCO, όσο και από διασωληνωμένους ασθενείς σε Μονάδες Εντατικής Θεραπείας. Για την ανάλυση των δεδομένων και την εκτίμηση των τιμών-στόχων, χρησιμοποιήθηκαν οι αλγόριθμοι Linear Regression, Random Forest και XGBoost, ενώ αξιολογήθηκαν μέσω δεικτών σφάλματος (MAE, RMSE) και στατιστικών μετρικών. Στα πειράματα, ο Random Forest εμφάνισε υψηλή ακρίβεια σε δόσεις ινσουλίνης και επίπεδα γλυκόζης όσον αφορά το κομμάτι που αναφέρεται στους ασθενείς που βρίσκονται σε ΜΕΘ. Ενώ ο XGBoost απέδωσε ιδιαίτερα καλά στην πρόβλεψη επιπέδων γλυκόζης, με μικρό σφάλμα, σχετικά με τα δεδομένα της web εφαρμογής (Artificial Dataset).

Στο πρακτικό κομμάτι, αναπτύχθηκε η web εφαρμογή chatGLUCO με backend σε Laravel και ξεχωριστό Flask API για τις προβλέψεις Μηχανικής Μάθησης. Το σύστημα υποστηρίζει βασικές λειτουργίες, όπως διαχείριση χρηστών, καταγραφή και επεξεργασία μετρήσεων, αυτόματους υπολογισμούς υδατανθράκων και ινσουλίνης, οπτικοποίηση δεδομένων μέσω διαγραμμάτων, καθώς και εξατομικευμένη πρόβλεψη γλυκόζης μετά το γεύμα.

Η παρούσα εργασία συμβάλλει στην αυτοματοποίηση και την καλύτερη διαχείριση του διαβήτη, ενσωματώνοντας καινοτόμες μεθόδους Μηχανικής Μάθησης και πρακτικές λειτουργικότητες. Μελλοντικά, η ενσωμάτωση πρόσθετων αλγορίθμων (π.χ. νευρωνικών δικτύων), η διασύνδεση με φορητές συσκευές και η κλινική επικύρωση του συστήματος μπορούν να βελτιώσουν περαιτέρω την ακρίβεια και την αξιοπιστία του, παρέχοντας μια ολοκληρωμένη λύση για την υποστήριξη ασθενών και ιατρών.

Λέξεις κλειδιά: Διαβήτης, Τύπος 1, Τύπος 2, γλυκοζυλιωμένη αιμοσφαιρίνη (HbA1c), ολική ημερήσια δόση ινσουλίνης (TDD), Sliding Scale Protocol, Carbo Calculation, γλυκόζη μετά το γεύμα, πρόβλεψη γλυκόζης, πρόβλεψη δόσης ινσουλίνης, Linear Regression, Random Forest, XGBoost, MAE (Mean Absolute Error), RMSE (Root Mean Square Error), Time In Range (TIR), Laravel, Flask API, διαχείριση χρηστών, καταγραφή μετρήσεων, υπολογισμός υδατανθράκων, οπτικοποίηση δεδομένων, εξατομικευμένη πρόβλεψη, αλγόριθμοι μηχανικής μάθησης, νευρωνικά δίκτυα, φορητές συσκευές, κλινική επικύρωση.

Abstract

This thesis presents the development of an integrated system for diabetes management, combining traditional insulin calculation protocols (e.g., the Sliding Scale Protocol) with machine learning (ML) algorithms. The study begins with a review of fundamental concepts related to diabetes mellitus, including its types (Type 1 and Type 2) and key regulation parameters such as glycated hemoglobin (HbA1c) and the Total Daily Dose of insulin (TDD). Subsequently, insulin calculation methods such as the Sliding Scale Protocol and Carbohydrate (Carbo) Calculation and postprandial glucose estimation are described.

The research focuses on two primary training scenarios: predicting future glucose levels and determining appropriate insulin doses. Data sources include individuals with diabetes who monitor their treatment via the chatGLUCO application and intubated patients in Intensive Care Units (ICUs). For data analysis and target value estimation, the Linear Regression, Random Forest, and XGBoost algorithms were evaluated using error metrics (MAE, RMSE) and statistical measurements. In experiments, the Random Forest algorithm demonstrated high accuracy in predicting insulin doses and glucose levels for ICU patients, while XGBoost excelled in glucose level prediction, achieving low error rates for the chatGLUCO dataset (artificial dataset).

On the implementation front, the chatGLUCO web application was developed with a Laravel backend and a separate Flask API for ML predictions. The system supports core functionalities such as user management, measurement recording/editing, automatic carbohydrate and insulin calculations, data visualization via graphs, and personalized postprandial glucose predictions.

This thesis contributes to the automation and improved management of diabetes by integrating innovative ML methods with practical tools. Future work could enhance accuracy and reliability through the incorporation of additional algorithms (e.g., neural networks), integration with wearable devices, and clinical validation of the system, ultimately providing a comprehensive solution for patients and healthcare professionals.

Keywords: Diabetes mellitus, Type 1 diabetes, Type 2 diabetes, glycated hemoglobin (HbA1c), total daily dose of insulin (TDD), Sliding Scale Protocol, Carbohydrate Calculation, postprandial glucose, glucose prediction, insulin dose prediction, Linear Regression, Random Forest, XGBoost, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Time in Range (TIR), Laravel, Flask API, user management, measurement recording, carbohydrate calculation, data visualization, personalized prediction, machine learning algorithms, neural networks, wearable devices, clinical validation

Contents

Ευχαριστίες.....	2
Περίληψη.....	3
Abstract	5
Λίστα Πινάκων.....	8
Λίστα Σχημάτων.....	8
Λίστα Ακρωνυμίων	9
1 Εισαγωγή	11
1.1 Πλαίσιο Διπλωματικής Εργασίας.....	11
1.2 Σκοπός Διπλωματικής Εργασίας.....	11
1.3 Δομή της Διπλωματικής Εργασίας	12
2 Θεωρητικό Υπόβαθρο.....	15
2.1 Βασικές Έννοιες και Ορισμοί	15
2.2 Μέθοδοι Υπολογισμού Ινσουλίνης.....	19
2.2.1 Sliding Scale Protocol	19
2.2.2 Carbohydrate Calculation.....	20
2.2.3 Weight-based Insulin Prediction.....	20
2.3 Μέθοδος Υπολογισμού Γλυκόζης Μετά Το Γεύμα.....	21
2.4 Επισκόπηση Υφιστάμενων Εφαρμογών	24
2.4.1 GlucoLog RapidCalc	24
2.4.2 Diabetes:M.....	25
2.4.3 FreeStyle LibreLink	26
2.4.4 Hedia Diabetes Assistant.....	26
2.4.5 Sugarmate	27
2.4.6 ChatGLUCO	28
3 Μέθοδοι Μοντέλων Μηχανικής Μάθησης	31
3.1 Αλγόριθμοι Μηχανικής Μάθησης.....	31
3.1.1 Linear Regression	32
3.1.2 Random Forest.....	37
3.1.3 XGBoost.....	43
3.2 Σενάρια Εκπαίδευσης	48
3.2.1 Σενάριο 1: Πρόβλεψη Επιπέδων Γλυκόζης	49
3.2.2 Σενάριο 2: Πρόβλεψη Δόσης Ινσουλίνης.....	52

4 Αξιολόγηση των Μοντέλων Μηχανικής Μάθησης.....	55
4.1 Μετρικές Αξιολόγησης Ακρίβειας	55
4.2 Συγκριτικά Αποτελέσματα	60
4.3 Το Example για τον ασθενή 533.....	62
5 Ανάπτυξη Εφαρμογής	66
5.1 Ανάλυση Απαιτήσεων.....	66
5.1.1 Λειτουργικές Απαιτήσεις	66
5.1.2 Μη Λειτουργικές Απαιτήσεις.....	73
5.2 Σχεδίαση Συστήματος	74
5.2.1 Διάγραμμα Περιπτώσεων Χρήσης / Use Case Diagram	75
5.2.2 Αρχιτεκτονική Συστήματος	77
5.2.3 Διάγραμμα Βάσεων Δεδομένων / Database Diagram.....	81
5.3 Υλοποίηση του Συστήματος.....	85
5.3.1 Υλοποίηση του Frontend	85
5.3.2 Υλοποίηση του Backend.....	87
5.3.4 Διεπαφή Χρήστη (UI)	92
6 Συμπεράσματα και Μελλοντική Εργασία	107
6.1 Συμπεράσματα	107
6.2 Μελλοντική Εργασία και Βελτιώσεις	108
7 Παράρτημα.....	112
7.1 Βιβλιοθήκες που χρησιμοποιήθηκαν	112
7.2 Θεωρητικό Υπόβαθρο για το Web App	115
7.3 Περιγραφή του dataset	116
7.4 Ανάλυση του Grid Search	120
7.5 Μέθοδοι υπολογισμού δόσεων ινσουλίνης και επιπέδων γλυκόζης: Υλοποίηση στον κώδικα	125
7.6 Υλοποίηση Διαγραμμάτων στην εφαρμογή (Κώδικας).....	128
8 Βιβλιογραφία.....	134

Λίστα Πινάκων

Πίνακας 1: Συγκριτική Μελέτη των Συστημάτων

Πίνακας 2: Συγκριτική ανάλυση προβλεπτικών μοντέλων με βάση το MAE και το RMSE σε 5659 ασθενείς για την πρόβλεψη επιπέδων γλυκόζης.

Πίνακας 3: Συγκριτική ανάλυση προβλεπτικών μοντέλων με βάση το MAE και το RMSE σε 5659 ασθενείς για την πρόβλεψη δόσης ινσουλίνης.

Λίστα Σχημάτων

Σχήμα 1: Before and After Feature Scaling

Σχήμα 2: Γραφική αναπαράσταση μιας Γραμμικής Παλινδρόμησης

Σχήμα 3: Γραφική Απεικόνιση της Λειτουργίας του Random Forest

Σχήμα 4: Επίπεδα Γλυκόζης Κατά τη Διάρκεια της Ημέρας για τον ασθενή 533

Σχήμα 5: Σχέση Επιπέδων Ινσουλίνης και Γλυκόζης για Ασθενή 533

Σχήμα 6: Χορήγηση Ινσουλίνης με την Πάροδο του Χρόνου για τον Ασθενή 533

Σχήμα 7: Σύγκριση Προβλέψεων Δόσεων Ινσουλίνης με Πραγματικά Δεδομένα για τον Ασθενή 533

Σχήμα 8: Διάγραμμα Περιπτώσεων Χρήσης

Σχήμα 9: Διάγραμμα Αρχιτεκτονικής

Σχήμα 10: Διάγραμμα Βάσης Δεδομένων

Σχήμα 11: Οθόνη Εισόδου (Log In)

Σχήμα 12: Αποτυχημένη προσπάθεια εισόδου με μήνυμα ενημέρωσης

Σχήμα 13: Οθόνη Εγγραφής (Sign Up)

Σχήμα 14: Προειδοποιητικό μήνυμα ανεπάρκειας στοιχείων στην οθόνη Diary

Σχήμα 15: Οθόνη Προφίλ Χρήστη. Συμπλήρωση και Ενημέρωση Πληροφοριών Υγείας για την αντιμετώπιση του προειδοποιητικού μηνύματος από το Σχήμα 14.

Σχήμα 16: Προβολή και Επεξεργασία του ποσοστού HbA1c.

Σχήμα 17: Διαχείριση Δεδομένων Χρήστη με την χρήση Ημερολογίου (Diary)

Σχήμα 18: Καταχώρηση Νέας Εγγραφής στο ημερολόγιο (Diary)

Σχήμα 19: Εξαγωγή Αναφοράς Δεδομένων Ασθενούς

Σχήμα 20: Διάγραμμα *Time In Range*

Σχήμα 21: Διάγραμμα *Glucose – Insulin Correlation*

Σχήμα 22: Διάγραμμα *Future Glucose Trend Prediction*

Σχήμα 23: Διεπαφή Πρόβλεψης Γλυκόζης

Σχήμα 24: Επεξήγηση Υπολογισμού Πρόβλεψης Γλυκόζης

Σχήμα 25: Υπολογισμός ηλικίας από ημερομηνία γέννησης

Σχήμα 26: SQL Ερώτημα για Εξαγωγή Δεδομένων *Patients* και *Patients Statistics*

Σχήμα 27: Παρακολούθηση Τελευταίων Ενημερώσεων στη Βάση Δεδομένων με *Timestamps*

Σχήμα 28: Προεπεξεργασία Δεδομένων: Συμπλήρωση Ελλιπών Τιμών στο Πεδίο *food_carbo*

Σχήμα 29: Μετατροπή Κατηγορικών Δεδομένων σε Αριθμητικές Τιμές για τα Πεδία *Gender* και *Diagnosis*

Σχήμα 30: *Data Cleaning*: Αντικατάσταση Κενών Τιμών στο Πεδίο *Weight* με τον Μέσο Όρο

Σχήμα 31: Ταξινόμηση Δεδομένων κατά Ημερομηνία Δημιουργίας

Σχήμα 32: Υπολογισμός Δόσης Ινσουλίνης με *Sliding Scale* και *Carbo Calculation*

Σχήμα 33: *Weight-based Insulin Prediction* για Ασθενείς με Λιγότερες από 4 Εγγραφές

Σχήμα 34: Υπολογισμός Γλυκόζης Μετά το Γεύμα - Ροή Υλοποίησης

Σχήμα 35: Υλοποίηση Διαγράμματος *Time In Range* στο *Blade Template*

Σχήμα 36: Φόρτωση και Φιλτράρισμα Δεδομένων για το Διάγραμμα *Time In Range* στον *Controller*

Σχήμα 37: Υλοποίηση Διαγράμματος *Glucose-Insulin Correlation* στο *Blade Template*

Σχήμα 38: Υπολογισμός Γραμμικής Παλινδρόμησης για Συσχέτιση Γλυκόζης-Ινσουλίνης

Σχήμα 39: Κλήση *Flask API* για Πρόβλεψη Μελλοντικών Τιμών Γλυκόζης

Σχήμα 40: Υλοποίηση Πρόβλεψης Μελλοντικών Τιμών Γλυκόζης με *XGBoost*

Λίστα Ακρωνυμίων

AI	Artificial Intelligence
API	Application Programming Interface
CGMs	Continuous Glucose Monitors
CI	Confidence Intervals
CR	Carbohydrate Ratio
CRUD	Create, Read, Update, Delete
CSRF	Cross-Site Request Forgery
CSS	Cascading Style Sheets

DB	Database
GDPR	General Data Protection Regulation
HTTP	Hypertext Transfer Protocol
HIPAA	Health Insurance Portability and Accountability Act
ICR	Insulin to Carbohydrate
IOB	Insulin on Board
ISF	Insulin Sensitivity Factor
JSON	JavaScript Object Notation
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
NFC	Near Field Communications
OOB	Out-of-Bag
OLS	Ordinary Least Squares
REST	Representational State Transfer
R²	R-squared (Coefficient of Determination)
RMSE	Root Mean Squared Error
TDD	Total Daily Dose
T1DM	Diabetes Mellitus Type-1
T2DM	Diabetes Mellitus Type-2
TIR	Time In Range
UI	User Interface
URIs	Uniform Resource Identifier
URL	Uniform Resource Locator
USDA	United States Department of Agriculture
UX	User Experience
XGBoost	eXtreme Gradient Boosting

1 Εισαγωγή

1.1 Πλαίσιο Διπλωματικής Εργασίας

Η ρύθμιση των επιπέδων γλυκόζης στο αίμα αποτελεί θεμελιώδη προτεραιότητα για τη διατήρηση της υγείας, καθώς η γλυκόζη είναι ο βασικός ενεργειακός πόρος των κυττάρων και του εγκεφάλου. Η αυξημένη ή μειωμένη συγκέντρωση γλυκόζης στο αίμα, υπεργλυκαιμία ή υπογλυκαιμία, αντίστοιχα, μπορεί να προκαλέσει σημαντικές διαταραχές στην υγεία, ιδίως σε διαβητικούς ασθενείς. Αυτοί οι ασθενείς χρειάζονται εντατική διαχείριση της γλυκόζης, καθώς η ανεπάρκεια ή αναποτελεσματική χρήση της ινσουλίνης οδηγεί σε διακυμάνσεις των επιπέδων της γλυκόζης στο αίμα, αυξάνοντας τον κίνδυνο σοβαρών επιπλοκών, όπως καρδιαγγειακές παθήσεις, νευρολογικές βλάβες και νεφρική ανεπάρκεια.

Η διατήρηση της γλυκόζης σε φυσιολογικά επίπεδα είναι κρίσιμη για την υγεία των διαβητικών ασθενών. Παρά την ύπαρξη γενικών πρωτοκόλλων (όπως το Sliding Scale, Carbohydrate Calculation και Weight-based model) που διευκολύνουν την προσαρμογή της δόσης ινσουλίνης, η ετερογένεια των ασθενών, λόγω διαφορών σε ηλικία, βάρος, διάγνωση ή καθημερινές διακυμάνσεις, καθιστά αναγκαία μια πιο εξατομικευμένη προσέγγιση.

Στο πλαίσιο αυτό, η παρούσα εργασία εφαρμόζει αλγορίθμους Μηχανικής Μάθησης (Linear Regression, Random Forest, XGBoost) με στόχο τη βελτίωση της πρόβλεψης δόσης ινσουλίνης και των επιπέδων γλυκόζης. Αξιοποιώντας τόσο δημογραφικά όσο και κλινικά δεδομένα, επιδιώκεται η ανάπτυξη ενός αυτοματοποιημένου και επεκτάσιμου συστήματος που θα ενισχύει την αυτονομία των ασθενών, μειώνοντας τα επεισόδια υπεργλυκαιμίας και υπογλυκαιμίας και προσφέροντας πιο αποτελεσματική θεραπευτική παρακολούθηση.

1.2 Σκοπός Διπλωματικής Εργασίας

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός συστήματος πρόβλεψης δόσεων ινσουλίνης και μελλοντικών επιπέδων γλυκόζης, το οποίο συνδυάζει αλγορίθμους μηχανικής μάθησης με μια λειτουργική και εύχρηστη web εφαρμογή για τη διαχείριση δεδομένων διαβητικών ασθενών. Ο κύριος στόχος είναι η εξατομικευμένη ρύθμιση των δόσεων ινσουλίνης, με στόχο τη μείωση των περιστατικών υπεργλυκαιμίας και υπογλυκαιμίας, διατηρώντας τη γλυκόζη εντός των φυσιολογικών ορίων. Το σύστημα στοχεύει στη βελτίωση της ποιότητας ζωής των ασθενών, παρέχοντάς τους αυτονομία και υποστήριξη στη διαχείριση του διαβήτη.

Πιο αναλυτικά, οι επιμέρους στόχοι της διπλωματικής εργασίας είναι οι εξής:

1. **Χρήση Αλγορίθμων Μηχανικής Μάθησης:** Εφαρμογή αλγορίθμων, όπως οι Linear Regression, Random Forest και XGBoost, για την πρόβλεψη δόσεων ινσουλίνης και επιπέδων γλυκόζης με βάση δημογραφικά, κλινικά και καθημερινά δεδομένα των ασθενών.
2. **Ενσωμάτωση Μοντέλων Υπολογισμού Δόσεων:** Χρήση μοντέλων όπως το Sliding Scale Protocol, το Carbo Calculation και το Weight-based Insulin Prediction για την προσαρμογή της δόσης ανάλογα με τα επίπεδα γλυκόζης, την πρόσληψη υδατανθράκων και το βάρος του ασθενούς.
3. **Ανάπτυξη Διαδραστικής Web Εφαρμογής σε PHP Laravel:** Δημιουργία μιας εύχρηστης πλατφόρμας που επιτρέπει στους ασθενείς:
 - Να καταγράφουν και να παρακολουθούν μετρήσεις όπως τα επίπεδα γλυκόζης και τις δόσεις ινσουλίνης.
 - Να λαμβάνουν προτάσεις δόσεων βασισμένες σε εξατομικευμένες προβλέψεις.
 - Να αναλύουν την πορεία της γλυκόζης τους μέσω διαδραστικών γραφημάτων, ενισχύοντας την κατανόηση και τη λήψη αποφάσεων.
 - Να λαμβάνουν προτάσεις μελλοντικών επιπέδων γλυκόζης.

Η παρούσα εργασία φιλοδοξεί να συμβάλει στη διευκόλυνση της καθημερινής διαχείρισης του διαβήτη, προσφέροντας ένα εργαλείο για την παρακολούθηση και βελτιστοποίηση της ρύθμισης της ινσουλίνης. Παράλληλα, αποσκοπεί στη μείωση των επιπλοκών που σχετίζονται με τις διακυμάνσεις της γλυκόζης και την ενίσχυση της ασφάλειας και της αυτοπεποίθησης των ασθενών.

1.3 Δομή της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία δομείται σε οκτώ κεφάλαια, ακολουθώντας μια πορεία από τη βασική θεωρία και τις υπάρχουσες προσεγγίσεις, έως την πρακτική ανάπτυξη και αξιολόγηση του συστήματος:

- **Κεφάλαιο 1: Εισαγωγή**
Παρουσιάζεται το γενικό πλαίσιο της διπλωματικής εργασίας, ο σκοπός και οι στόχοι της. Δίνονται οι βάσεις για την αναγκαιότητα ανάπτυξης ενός ολοκληρωμένου συστήματος διαχείρισης του σακχαρώδους διαβήτη και υπογραμμίζονται οι προκλήσεις που επιδιώκει να αντιμετωπίσει η παρούσα προσέγγιση.
- **Κεφάλαιο 2: Θεωρητικό Υπόβαθρο**
Γίνεται αναλυτική παρουσίαση των βασικών εννοιών του διαβήτη, όπως οι τύποι T1DM και T2DM, η γλυκοζυλιωμένη αιμοσφαιρίνη (HbA1c) και η ολική ημερήσια

δόση ινσουλίνης (TDD). Παράλληλα, περιγράφονται οι παραδοσιακές μέθοδοι υπολογισμού δόσεων ινσουλίνης (Sliding Scale Protocol, Carbo Calculation, Weight-based Insulin Prediction) και η μέθοδος υπολογισμού της γλυκόζης μετά το γεύμα, ενώ παρουσιάζονται και αντίστοιχες υφιστάμενες εφαρμογές.

- **Κεφάλαιο 3: Μέθοδοι Μοντέλων Μηχανικής Μάθησης**

Επεξηγούνται οι αλγόριθμοι Linear Regression, Random Forest και XGBoost, καθώς και η διαδικασία εκπαίδευσής τους. Γίνεται διάκριση των σεναρίων χρήσης των μοντέλων σε πρόβλεψη επιπέδων γλυκόζης και πρόβλεψη δόσης ινσουλίνης, με αναφορά στα αντίστοιχα χαρακτηριστικά δεδομένων.

- **Κεφάλαιο 4: Αξιολόγηση των Μοντέλων Μηχανικής Μάθησης**

Παρουσιάζονται οι μετρικές αξιολόγησης (MAE, RMSE, Time In Range κ.ά.) και πραγματοποιείται συγκριτική ανάλυση των αποτελεσμάτων που προέκυψαν από τα μοντέλα. Στη συνέχεια, παρατίθεται ένα «Toy Example» για την ανάδειξη της συμπεριφοράς των αλγορίθμων σε πραγματικές συνθήκες, εστιάζοντας σε μεμονωμένο ασθενή.

- **Κεφάλαιο 5: Ανάπτυξη Εφαρμογής**

Γίνεται η ανάλυση απαιτήσεων (λειτουργικών και μη λειτουργικών) και η σχεδίαση του συστήματος, όπου παρουσιάζονται τα βασικά διαγράμματα (Use Case, Αρχιτεκτονική, Βάση Δεδομένων). Ακολουθώντας, περιγράφεται η υλοποίηση τόσο του frontend (Laravel, Blade Templates, JavaScript) όσο και του backend (Laravel Controllers, επικοινωνία με Flask API), καθώς και η διεπαφή χρήστη.

- **Κεφάλαιο 6: Συμπεράσματα και Μελλοντική Εργασία**

Συνοψίζονται τα κύρια ευρήματα και η συνεισφορά της παρούσας εργασίας στη βελτίωση της αυτοματοποίησης και της εξατομίκευσης της διαχείρισης του διαβήτη. Επιπλέον, προτείνονται πιθανά βήματα για περαιτέρω βελτίωση και επεκτάσεις, όπως η αξιοποίηση νευρωνικών δικτύων ή η κλινική αξιολόγηση σε μεγαλύτερο δείγμα ασθενών.

- **Κεφάλαιο 7: Παράρτημα**

Συμπληρωματικό υλικό που περιλαμβάνει τις βιβλιοθήκες που χρησιμοποιήθηκαν, αναλυτικό θεωρητικό υπόβαθρο για το web app, περιγραφή του dataset, παρουσίαση του Grid Search, ενδεικτικές υλοποιήσεις κώδικα (υπολογισμός δόσεων ινσουλίνης και γλυκόζης, δημιουργία διαγραμμάτων).

- **Κεφάλαιο 8: Βιβλιογραφία**

Παρατίθεται η σχετική βιβλιογραφία και οι πηγές στις οποίες στηρίχθηκε η εκπόνηση της διπλωματικής εργασίας.

Αυτή η δομή διασφαλίζει ότι ο αναγνώστης μπορεί να κατανοήσει πλήρως τη μεθοδολογία, τα αποτελέσματα, και τη λειτουργικότητα της αναπτυχθείσας εφαρμογής.

2 Θεωρητικό Υπόβαθρο

2.1 Βασικές Έννοιες και Ορισμοί

Η διαχείριση του διαβήτη απαιτεί σαφή κατανόηση βασικών εννοιών που αφορούν τη νόσο. Σε αυτήν την ενότητα, παρουσιάζονται οι βασικές έννοιες και ορισμοί που αποτελούν το θεμέλιο για την κατανόηση της παρούσας εργασίας.

Θεωρητικό Υπόβαθρο Διαβήτη

Ο διαβήτης είναι μια χρόνια μεταβολική διαταραχή που χαρακτηρίζεται από την αδυναμία του σώματος να ρυθμίζει αποτελεσματικά τα επίπεδα γλυκόζης στο αίμα. Η γλυκόζη είναι η κύρια πηγή ενέργειας για τον οργανισμό, ενώ η ινσουλίνη, μια ορμόνη που παράγεται από το πάγκρεας, παίζει κεντρικό ρόλο στη ρύθμισή της, επιτρέποντας τη μεταφορά της γλυκόζης από το αίμα στα κύτταρα. Η έλλειψη ή αναποτελεσματική χρήση της ινσουλίνης οδηγεί σε συσσώρευση γλυκόζης στο αίμα, προκαλώντας διαταραχές που, αν δεν ρυθμιστούν, μπορούν να οδηγήσουν σε σοβαρές επιπλοκές, όπως καρδιαγγειακά νοσήματα, νευρολογικές βλάβες και νεφρική ανεπάρκεια.

Τύποι Διαβήτη

Ο διαβήτης διακρίνεται κυρίως σε δύο κατηγορίες: τον διαβήτη τύπου 1 (T1DM) και τον διαβήτη τύπου 2 (T2DM).

- 1. Διαβήτης Τύπου 1 (T1DM):** Ο διαβήτης τύπου 1 είναι μια αυτοάνοση πάθηση κατά την οποία το ανοσοποιητικό σύστημα του ασθενούς επιτίθεται στα β-κύτταρα του παγκρέατος, που είναι υπεύθυνα για την παραγωγή ινσουλίνης. Αυτή η καταστροφή των β-κυττάρων σημαίνει ότι οι ασθενείς με T1DM αδυνατούν να παράγουν ινσουλίνη, καθιστώντας τη χορήγηση ινσουλίνης απαραίτητη για την επιβίωσή τους. Οι ασθενείς αυτοί πρέπει να παρακολουθούν συνεχώς τα επίπεδα γλυκόζης τους και να προσαρμόζουν τις δόσεις ινσουλίνης ανάλογα με τις ανάγκες τους.
- 2. Διαβήτης Τύπου 2 (T2DM):** Ο διαβήτης τύπου 2 αποτελεί τη συχνότερη μορφή διαβήτη και εμφανίζεται λόγω της αντίστασης των κυττάρων στην ινσουλίνη ή της ανεπαρκούς παραγωγής της από το πάγκρεας. Ο T2DM συνδέεται συχνά με παράγοντες όπως η παχυσαρκία, ο καθιστικός τρόπος ζωής και η ηλικία. Η διαχείριση του T2DM απαιτεί κυρίως αλλαγές στον τρόπο ζωής, όπως διατροφή και άσκηση, και, σε πολλές περιπτώσεις, χορήγηση φαρμακευτικής αγωγής ή εξωτερικής ινσουλίνης.

Γλυκοζυλιωμένη Αιμοσφαιρίνη (HbA1c)

Η γλυκοζυλιωμένη αιμοσφαιρίνη (HbA1c) αποτελεί έναν σημαντικό βιολογικό δείκτη που αντικατοπτρίζει τις μέσες τιμές γλυκόζης αίματος των τελευταίων 2-3 μηνών. Δημιουργείται όταν η γλυκόζη συνδέεται με την αιμοσφαιρίνη στα ερυθρά αιμοσφαίρια, και το ποσοστό της στο αίμα παρέχει μια συνολική εικόνα της γλυκαιμικής ρύθμισης του ασθενούς. Οι φυσιολογικές τιμές της HbA1c για υγιή άτομα είναι κάτω από 5,7%, ενώ τιμές μεταξύ 5,7% και 6,4% υποδεικνύουν προδιαβήτη. Τιμές ίσες ή μεγαλύτερες του 6,5% χρησιμοποιούνται ως διαγνωστικό κριτήριο για τον σακχαρώδη διαβήτη.

Η μέτρηση της HbA1c είναι κρίσιμη για τη διαχείριση του διαβήτη, καθώς βοηθά στην αξιολόγηση της αποτελεσματικότητας της θεραπείας και στην πρόληψη επιπλοκών. Επιπλέον, οι στόχοι της HbA1c εξατομικεύονται ανάλογα με την ηλικία, τη διάρκεια της νόσου, και τη γενικότερη κατάσταση της υγείας του ασθενούς, καθιστώντας τη βασικό εργαλείο για την παρακολούθηση και βελτίωση της μακροχρόνιας ρύθμισης της γλυκόζης.

Η χρήση και προσαρμογή της ινσουλίνης

Η ινσουλίνη είναι βασικό εργαλείο για τη ρύθμιση των επιπέδων γλυκόζης στους διαβητικούς ασθενείς, καθώς λειτουργεί ως “κλειδί” για την είσοδο της γλυκόζης στα κύτταρα του σώματος. Για τους διαβητικούς ασθενείς, ιδιαίτερα για αυτούς με διαβήτη τύπου 1, η ανάγκη προσαρμογής της δόσης ινσουλίνης είναι ζωτικής σημασίας. Ο καθορισμός της κατάλληλης δόσης ινσουλίνης είναι περίπλοκος και βασίζεται σε παράγοντες όπως:

- **Τρέχοντα επίπεδα γλυκόζης:** Οι ασθενείς πρέπει να ελέγχουν τα επίπεδα γλυκόζης τους πριν από τα γεύματα και να προσαρμόζουν την δόση ινσουλίνης αναλόγως.
- **Υδατάνθρακες στα γεύματα:** Η κατανάλωση υδατανθράκων απαιτεί αύξηση της δόσης ινσουλίνης, ανάλογα με την ποσότητα που πρόκειται να καταναλωθεί.
- **Σωματική άσκηση:** Η φυσική δραστηριότητα αυξάνει την ευαισθησία στην ινσουλίνη, μειώνοντας τις ανάγκες για χορήγηση. Ανάλογα με την ένταση και τη διάρκεια της άσκησης, οι ασθενείς μπορεί να χρειαστούν προσαρμογές στη δόση για να αποφύγουν την υπογλυκαιμία.

Η προσαρμογή της ινσουλίνης γίνεται κυρίως με τη χρήση πρωτοκόλλων, όπως το Sliding Scale Protocol και το Carbohydrate Calculation. Το Sliding Scale Protocol βασίζεται αποκλειστικά στα τρέχοντα επίπεδα γλυκόζης για τον καθορισμό της δόσης, ενώ το Carbohydrate Calculation λαμβάνει υπόψη την ποσότητα των υδατανθράκων που καταναλώνει ο ασθενής. Οι μέθοδοι αυτοί βοηθούν τους ασθενείς να προσαρμόζουν τη δόση τους, διατηρώντας τα επίπεδα γλυκόζης σε ασφαλή όρια.

Total Daily Dose (TDD)

Η Συνολική Ημερήσια Δόση (TDD) αναφέρεται στη συνολική ποσότητα ινσουλίνης που λαμβάνει ένας ασθενής με σακχαρώδη διαβήτη μέσα σε ένα 24ωρο. Η TDD αποτελείται από δύο βασικά μέρη:

1. **Basal (Βασική) ινσουλίνη:** Πρόκειται για την ινσουλίνη μακράς δράσης που χορηγείται ώστε να καλύπτει τις βασικές ανάγκες του οργανισμού σε όλη τη διάρκεια του 24ώρου. Σκοπός της basal ινσουλίνης είναι να διατηρεί τα επίπεδα γλυκόζης σταθερά όταν δεν λαμβάνονται γεύματα, όπως κατά τη διάρκεια της νύχτας ή μεταξύ των γευμάτων.
2. **Bolus (Γευματική) ινσουλίνη:** Πρόκειται για την ινσουλίνη ταχείας ή βραχείας δράσης, η οποία χορηγείται πριν ή κατά τη διάρκεια των γευμάτων. Χρησιμοποιείται για την κάλυψη της πρόσληψης υδατανθράκων (meal bolus) ή για τη διόρθωση υψηλών επιπέδων γλυκόζης (correction bolus).

Σημασία της TDD:

- Η γνώση της TDD βοηθά στον καθορισμό του καθημερινού θεραπευτικού σχήματος, τόσο για ασθενείς με διαβήτη τύπου 1 όσο και για κάποιες περιπτώσεις διαβήτη τύπου 2.
- Αποτελεί τη βάση για τον υπολογισμό σημαντικών παραμέτρων, όπως ο συντελεστής ευαισθησίας στην ινσουλίνη (ISF) και ο λόγος ινσουλίνης-υδατανθράκων (Insulin-to-Carbohydrate Ratio).
- Εφόσον ο οργανισμός κάθε ατόμου διαφέρει, η TDD είναι εξατομικευμένη και μπορεί να τροποποιείται ανάλογα με παράγοντες όπως η διατροφή, η φυσική δραστηριότητα και τα επίπεδα στρες.

Insulin Sensitivity Factor (ISF)

Ο Συντελεστής Ευαισθησίας στην Ινσουλίνη (ISF) προσδιορίζει πόσο (σε mg/dL) μειώνεται η γλυκόζη στο αίμα για κάθε επιπλέον μονάδα ινσουλίνης που χορηγείται. Για παράδειγμα, αν ο ISF είναι 50, σημαίνει ότι 1 μονάδα ινσουλίνης αναμένεται να μειώσει το σάκχαρο κατά περίπου 50 mg/dL.

Χρήση του ISF στη διαχείριση του διαβήτη:

- **Υπολογισμός δόσης διόρθωσης (correction dose):** Όταν τα επίπεδα γλυκόζης είναι υψηλότερα από το επιθυμητό όριο, ο ISF βοηθά να υπολογιστεί πόση επιπλέον ινσουλίνη χρειάζεται ο ασθενής για να επανέλθει σε ασφαλή επίπεδα.
- **Εξατομίκευση θεραπείας:** Ο ISF δεν είναι ίδιος για όλους τους ασθενείς και μπορεί να μεταβάλλεται με την πάροδο του χρόνου, ανάλογα με αλλαγές στο καθημερινό πρόγραμμα, τη διατροφή, τη φυσική δραστηριότητα και τη φαρμακευτική αγωγή.

Σύνδεση με την TDD:

Συχνά, ο ISF υπολογίζεται βάσει της Συνολικής Ημερήσιας Δόσης (TDD). Υπάρχουν εμπειρικοί τύποι (όπως ο “100/TDD” ή “1500/TDD”, ανάλογα με το χρησιμοποιούμενο πρωτόκολλο) που παρέχουν μια πρώτη εκτίμηση του ISF, η οποία ωστόσο πρέπει να επαληθευτεί και να προσαρμοστεί εξατομικευμένα.

Machine Learning in Diabetes Management

Η μηχανική μάθηση (Machine Learning - ML) είναι ένας κλάδος της τεχνητής νοημοσύνης που επικεντρώνεται στη δημιουργία αλγορίθμων που επιτρέπουν στους υπολογιστές να μαθαίνουν, να προβλέπουν και να λαμβάνουν αποφάσεις χωρίς άμεσο προγραμματισμό. Στη διαχείριση του διαβήτη, το ML χρησιμοποιείται για:

1. **Πρόβλεψη επιπέδων γλυκόζης (Glucose Prediction):** Βάσει ιστορικών δεδομένων και άλλων παραγόντων ασθενούς.
2. **Πρόβλεψη δόσεων ινσουλίνης (Insulin Dose Prediction):** Υπολογισμός της απαιτούμενης δόσης σε δεδομένη στιγμή.

Η παρούσα εργασία χρησιμοποιεί αλγορίθμους όπως Linear Regression, Random Forest, και XGBoost για την επίτευξη προβλέψεων που βοηθούν τους ασθενείς να διαχειρίζονται καλύτερα τη νόσο τους.

Data Preprocessing (Προεπεξεργασία Δεδομένων)

Η προεπεξεργασία δεδομένων είναι ένα κρίσιμο στάδιο στη μηχανική μάθηση, που διασφαλίζει ότι τα δεδομένα είναι καθαρά, συνεπή και κατάλληλα για χρήση στους αλγόριθμους. Αποτελείται από τα εξής βασικά βήματα:

- **Καθαρισμός Δεδομένων (Data Cleaning):** Αντιμετώπιση ελλিপών, λανθασμένων ή μη χρήσιμων δεδομένων. Αυτό περιλαμβάνει τη διαγραφή ή αντικατάσταση κενών τιμών, τη διόρθωση λανθασμένων τιμών και την απομάκρυνση θορύβου (noise).

- **Κανονικοποίηση (Normalization):** Εξισορρόπηση των τιμών των χαρακτηριστικών σε παρόμοια κλίμακα, ώστε να μην υπερισχύουν χαρακτηριστικά με υψηλές τιμές έναντι αυτών με χαμηλότερες. Για παράδειγμα, χαρακτηριστικά όπως η ηλικία ή τα επίπεδα γλυκόζης μπορεί να κανονικοποιηθούν σε ένα εύρος 0-1.
- **Κατηγοριοποίηση (Categorical Encoding):** Μετατροπή κατηγορικών δεδομένων, όπως το φύλο ή η διάγνωση (Τύπου 1 ή Τύπου 2), σε μορφή που μπορεί να επεξεργαστεί ένας αλγόριθμος (π.χ., one-hot encoding ή label encoding).
- **Επιλογή Χαρακτηριστικών (Feature Selection):** Αναγνώριση και διατήρηση μόνο των χαρακτηριστικών που συμβάλλουν σημαντικά στην πρόβλεψη. Για παράδειγμα, δεδομένα όπως η ηλικία, το βάρος, τα επίπεδα γλυκόζης και η ποσότητα υδατανθράκων θεωρούνται κρίσιμα για τη διαχείριση της ινσουλίνης.

Η προεπεξεργασία διασφαλίζει ότι τα δεδομένα είναι συνεπή και ποιοτικά, επιτρέποντας στους αλγόριθμους να λειτουργούν με μεγαλύτερη ακρίβεια και αποτελεσματικότητα.

2.2 Μέθοδοι Υπολογισμού Ινσουλίνης

Υπάρχουν διάφορες μέθοδοι υπολογισμού της δόσης ινσουλίνης, οι οποίες βοηθούν στον καθορισμό των αναγκαίων μονάδων βάσει των τρεχουσών συνθηκών και της διατροφής του ασθενούς. Στο πλαίσιο της παρούσας διπλωματικής εργασίας, εφαρμόζονται οι εξής μέθοδοι:

2.2.1 Sliding Scale Protocol

Το Sliding Scale Protocol είναι ένα σύστημα προσαρμογής της δόσης ινσουλίνης που χρησιμοποιείται συνήθως για την άμεση αντιμετώπιση αυξημένων επιπέδων γλυκόζης στο αίμα. Βασίζεται αποκλειστικά στα τρέχοντα επίπεδα γλυκόζης, χωρίς να λαμβάνει υπόψη το φαγητό ή άλλες μεταβλητές. Η δόση της ινσουλίνης αυξάνεται ανάλογα με το πόσο υψηλότερη είναι η τιμή της γλυκόζης από τη φυσιολογική τιμή (π.χ. 100 mg/dL).

Για παράδειγμα αν ο ασθενής έχει επίπεδο γλυκόζης 250 mg/dL, τότε ο ασθενής είναι 150 mg/dL πάνω από το φυσιολογικό. Σε αυτήν την περίπτωση, με βάση το πρωτόκολλο Sliding Scale, η επιπλέον δόση ινσουλίνης υπολογίζεται με τη μέθοδο των τριών. Συγκεκριμένα, για κάθε 50 mg/dL πάνω από τα 100 mg/dL, αντιστοιχούν 2 μονάδες ινσουλίνης. Αν η υπερβάλλουσα γλυκόζη είναι 150 mg/dL, τότε εφαρμόζουμε τη μέθοδο των τριών:

50mg/dL → 2 μονάδες ινσουλίνης

150mg/dL → x μονάδες ινσουλίνης

Υπολογίζουμε:

$$x = \frac{150 \times 2}{50} = 6$$

Έτσι, ο ασθενής χρειάζεται 6 μονάδες ινσουλίνης. Αυτή η προσέγγιση επιτρέπει στον ασθενή να προσαρμόζει τη δόση του σύμφωνα με τις ημερήσιες διακυμάνσεις της γλυκόζης χωρίς να λαμβάνει υπόψη άλλους παράγοντες, όπως τα γεύματα ή τη φυσική δραστηριότητα.

2.2.2 Carbohydrate Calculation

Το Carbohydrate Calculation είναι μια μέθοδος υπολογισμού της ινσουλίνης που εξαρτάται από την ποσότητα των υδατανθράκων που πρόκειται να καταναλώσει ο ασθενής. Η μέθοδος αυτή βασίζεται στη σχέση υδατανθράκων-ινσουλίνης (Insulin-to-Carb Ratio, ICR), που υπολογίζει τις μονάδες ινσουλίνης που απαιτούνται για να διατηρηθούν τα επίπεδα γλυκόζης σταθερά μετά το γεύμα.

Για παράδειγμα, ένας ασθενής με αναλογία 1:10 (δηλαδή 1 μονάδα ινσουλίνης ανά 10 γραμμάρια υδατανθράκων) θα χρειαστεί 6 μονάδες ινσουλίνης για ένα γεύμα 60 γραμμαρίων υδατανθράκων. Επιπλέον, αν ο ασθενής έχει αυξημένα επίπεδα γλυκόζης πριν από το γεύμα, η ποσότητα της ινσουλίνης μπορεί να προσαρμοστεί, προσθέτοντας τις μονάδες από το Sliding Scale Protocol στο αποτέλεσμα του Carbo Calculation για μια συνολική δόση.

2.2.3 Weight-based Insulin Prediction

Το Weight-based Insulin Prediction είναι μια μέθοδος βασικής χορήγησης ινσουλίνης που λαμβάνει υπόψη το βάρος του ασθενούς, χωρίς να υπολογίζει άλλες παραμέτρους όπως η κατανάλωση υδατανθράκων ή τα τρέχοντα επίπεδα γλυκόζης. Ο υπολογισμός αυτός βασίζεται στη σχέση: 0,5 μονάδες ινσουλίνης ανά κιλό σωματικού βάρους.

Για παράδειγμα, ένας ασθενής με βάρος 70 κιλά θα χρειαστεί περίπου 35 μονάδες ινσουλίνης ημερησίως. Αυτή η δόση μπορεί να χρησιμοποιηθεί για τη βασική συντήρηση των επιπέδων γλυκόζης, ιδίως σε ασθενείς που δεν έχουν επαρκή ιστορικά δεδομένα για ακριβέστερες μεθόδους πρόβλεψης.

2.3 Μέθοδος Υπολογισμού Γλυκόζης Μετά Το Γεύμα

Η μέθοδος “Postprandial Glucose Calculation” που ακολουθεί υπολογίζει την τιμή γλυκόζης αίματος μετά την κατανάλωση ενός γεύματος και τη χορήγηση ινσουλίνης, λαμβάνοντας υπόψη συγκεκριμένες παραμέτρους που σχετίζονται με το σωματικό βάρος του ασθενούς, τον τύπο του διαβήτη (Type I ή Type II) και την ποσότητα των υδατανθράκων του γεύματος.

1. Ολική Ημερήσια Δόση Ινσουλίνης (Total Daily Dose - TDD)

Το TDD αποτελεί μια εκτίμηση της συνολικής ποσότητας ινσουλίνης που απαιτείται μέσα σε ένα 24ωρο, ώστε να καλυφθούν οι καθημερινές ανάγκες του οργανισμού σε Basal και Bolus ινσουλίνη. Διαφέρει μεταξύ ασθενών με Type I και Type II διαβήτη, κυρίως λόγω διαφορετικής αντίστασης στην ινσουλίνη:

- Για **Type I** διαβήτη:

$$TDD = \text{Βάρος}(kg) \times 0.5$$

- Για **Type II** διαβήτη:

$$TDD = \text{Βάρος}(kg) \times 0.8$$

Μόλις υπολογίσουμε το TDD, μπορούμε να το αξιοποιήσουμε για να βρούμε τους σημαντικούς παράγοντες που περιγράφονται παρακάτω.

2. Συντελεστής Ευαισθησίας στην Ινσουλίνη (Insulin Sensitivity Factor - ISF)

Ο ISF εκφράζει πόση μείωση (σε mg/dL) αναμένεται στην τιμή γλυκόζης αίματος για κάθε 1 μονάδα ινσουλίνης που χορηγείται. Πρόκειται ουσιαστικά για δείκτη του πόσο “ευαίσθητος” είναι ο οργανισμός στην ινσουλίνη και καθορίζεται διαφορετικά για τον Type I και τον Type II διαβήτη:

- Για **Type I**:

$$ISF = \frac{1800}{TDD}$$

- Για **Type II**:

$$ISF = \frac{1500}{TDD}$$

Όσο υψηλότερος είναι ο ISF, τόσο μεγαλύτερη η ευαισθησία του ατόμου στην ινσουλίνη, άρα τόσο πιο έντονη η μείωση στη γλυκόζη για μια δεδομένη δόση.

3. Συντελεστής Υδατανθράκων (Carbohydrate Ratio - CR)

Ο CR δείχνει πόσα γραμμάρια υδατανθράκων καλύπτει 1 μονάδα ινσουλίνης. Παρέχει τη βάση για να προσαρμόζουμε τη δόση ινσουλίνης ανάλογα με την ποσότητα υδατανθράκων που καταναλώνουμε σε κάθε γεύμα. Υπολογίζεται ως:

$$CR = \frac{500}{TDD}$$

Για παράδειγμα, αν ο CR είναι 10, σημαίνει ότι 1 μονάδα ινσουλίνης καλύπτει 10 γραμμάρια υδατανθράκων.

4. Υπολογισμός της Νέας Τιμής Γλυκόζης

1. Ινσουλίνη για την κάλυψη των υδατανθράκων

$$insulin_for_carb = \frac{food_carbo}{CR}$$

Εδώ, *food_carbo* είναι τα γραμμάρια υδατανθράκων που καταναλώθηκαν.

2. Συνολική μείωση στη γλυκόζη λόγω ινσουλίνης

$$glucose_reduction = insulin_dose \times ISF$$

Όπου *insulin_dose* είναι η πραγματική δόση ινσουλίνης που χορηγήθηκε (σε μονάδες).

3. Τελική Προβλεπόμενη Τιμή Γλυκόζης

$$predicted_glucose_new = glucose_old + insulin_for_carb - glucose_reduction$$

- *glucose_old*: Η τιμή γλυκόζης πριν το γεύμα.

- *insulin_for_carb*: Το ποσό της ινσουλίνης που αντιστοιχεί στην ποσότητα υδατανθράκων, μεταφρασμένη στη μελλοντική επίδρασή της στη γλυκόζη.
- *glucose_reduction*: Η μείωση της γλυκόζης από τη συνολική ινσουλίνη που χορηγήθηκε.

Παράδειγμα Εφαρμογής

Έστω ασθενής με Type I διαβήτη και βάρος 70 κιλά, που καταναλώνει 26 γραμμάρια υδατανθράκων (*food_carbo* = 26), έχει τιμή γλυκόζης πριν το γεύμα *glucose_old* = 152 *mg/dL*, και χορηγεί 1 μονάδα ινσουλίνης (*insulin_dose* = 1).

1. Υπολογισμός TDD (Ολική Ημερήσια Δόση Ινσουλίνης)
Για Type I διαβήτη:

$$TDD = 70 \times 0.5 = 35 \text{ μονάδες/ημέρα}$$

2. Υπολογισμός ISF (Συντελεστής Ευαισθησίας στην Ινσουλίνη)

$$ISF = \frac{1800}{TDD} = \frac{1800}{35} \approx 51.43 \text{ mg/dL ανά μονάδα}$$

3. Υπολογισμός CR (Συντελεστής Υδατανθράκων)

$$CR = \frac{500}{TDD} = \frac{500}{35} \approx 14.29 \text{ g ανά μονάδα ινσουλίνης}$$

4. Υπολογισμός ινσουλίνης για τους υδατάνθρακες (Insulin for Carb)

$$insulin_for_carb = \frac{food_carbo}{CR} = \frac{26}{14.29} \approx 1.82 \text{ μονάδες}$$

5. Υπολογισμός της συνολικής μείωσης γλυκόζης λόγω ινσουλίνης (Glucose Reduction)

$$glucose_reduction = insulin_dose \times ISF = 1 \times 51.43 \approx 51.43 \text{ mg/dL}$$

6. Τελικός Υπολογισμός της Νέας Τιμής Γλυκόζης (Predicted Glucose New)

$$predicted_glucose_new = glucose_old + insulin_for_carb - glucose_reduction$$

$$predicted_glucose_new = 152 + 1.82 - 51.43 \approx 102.39 \text{ mg/dL}$$

η τελική προβλεπόμενη τιμή γλυκόζης είναι:

$$predicted_glucose_new = 102.39 \text{ mg/dL}$$

2.4 Επισκόπηση Υφιστάμενων Εφαρμογών

Η τεχνολογία έχει φέρει επανάσταση στη διαχείριση του διαβήτη, παρέχοντας καινοτόμες εφαρμογές που διευκολύνουν την παρακολούθηση των επιπέδων γλυκόζης και τη ρύθμιση της δόσης ινσουλίνης. Οι εφαρμογές αυτές αξιοποιούν σύγχρονα εργαλεία, όπως Τεχνητή Νοημοσύνη (AI), συνεχής παρακολούθηση γλυκόζης (CGM), και ανάλυση μοτίβων, για να προσφέρουν στους ασθενείς εξατομικευμένη υποστήριξη. Σε αυτή την ενότητα, παρουσιάζονται ορισμένες εφαρμογές, αναλύονται οι δυνατότητές τους και συγκρίνονται με την προτεινόμενη εφαρμογή chatGLUCO. Η σύγκριση πραγματοποιείται μέσω ενός συγκριτικού πίνακα, αναδεικνύοντας τα μοναδικά χαρακτηριστικά της chatGLUCO και τη συμβολή της στη βελτίωση της διαχείρισης του διαβήτη.

2.4.1 GlucoLog RapidCalc

Το GlucoLog RapidCalc είναι μια προηγμένη εφαρμογή υπολογισμού bolus ινσουλίνης, σχεδιασμένη για άτομα με διαβήτη που ακολουθούν το σχήμα bolus-basal. Η εφαρμογή προσφέρει έξυπνους και ακριβείς υπολογισμούς δόσεων ινσουλίνης, λαμβάνοντας υπόψη κρίσιμους παράγοντες όπως την ενεργή ινσουλίνη (Insulin on Board - IOB), την αναλογία ινσουλίνης-υδατανθράκων (ICR) και τους στόχους γλυκόζης. Παράλληλα, ενσωματώνει το Sliding Scale Protocol, επιτρέποντας τη διόρθωση των επιπέδων γλυκόζης όταν αυτά υπερβαίνουν τα φυσιολογικά όρια. Η λειτουργία Planned Carbohydrate Intake υπολογίζει τις απαραίτητες δόσεις ινσουλίνης με βάση την προγραμματισμένη κατανάλωση υδατανθράκων.

Επιπλέον, η εφαρμογή διαθέτει ένα ολοκληρωμένο ημερολόγιο καταγραφής (Diary), όπου οι χρήστες μπορούν να εισάγουν δεδομένα για τα επίπεδα γλυκόζης, τις δόσεις ινσουλίνης και τις δραστηριότητές τους. Παρέχει επίσης στατιστικά στοιχεία, αναφορές και γραφήματα, διευκολύνοντας την παρακολούθηση και την κατανόηση των δεδομένων. Μέσω της λειτουργίας Reminders, επιτρέπει την οργάνωση δόσεων και γευμάτων, συμβάλλοντας σε καλύτερη διαχείριση του διαβήτη στην καθημερινότητα.

Τέλος, η δυνατότητα εξαγωγής και διαμοιρασμού δεδομένων μέσω email διευκολύνει την επικοινωνία με επαγγελματίες υγείας, παρέχοντας τους πολύτιμες πληροφορίες για την προσαρμογή της θεραπείας. Συνολικά, το GlucoLog RapidCalc αποτελεί ένα ισχυρό εργαλείο για την απλοποίηση της διαχείρισης του διαβήτη, προσφέροντας ακριβείς υπολογισμούς και αναλυτική υποστήριξη.

2.4.2 Diabetes:M

Το Diabetes:M είναι μια ολοκληρωμένη πλατφόρμα για τη διαχείριση του διαβήτη, προσφέροντας ένα ευρύ φάσμα εργαλείων για υπολογισμό, παρακολούθηση και ανάλυση δεδομένων. Ο υπολογιστής δόσης ινσουλίνης (Bolus Calculator) υπολογίζει την απαιτούμενη ποσότητα ινσουλίνης με βάση την κατανάλωση υδατανθράκων, τα επίπεδα γλυκόζης και την ενεργή ινσουλίνη (IOB). Ενσωματώνει λειτουργίες όπως διορθωτικές δόσεις (Correction Bolus) και υπολογισμούς για γεύματα, προσφέροντας εξατομικευμένες συστάσεις.

Η εφαρμογή διαθέτει ένα πλήρως λειτουργικό ημερολόγιο καταγραφής (Logbook), όπου οι χρήστες μπορούν να εισάγουν δεδομένα όπως μετρήσεις γλυκόζης, πρόσληψη υδατανθράκων, δόσεις ινσουλίνης, φυσική δραστηριότητα και άλλες παραμέτρους υγείας. Μέσω της λειτουργίας ανάλυσης μοτίβων (Pattern Analysis), οι χρήστες μπορούν να αναγνωρίσουν επαναλαμβανόμενες τάσεις και να προσαρμόσουν τη διαχείριση του διαβήτη τους.

Το Diabetes:M προσφέρει επίσης διαγράμματα και στατιστικά στοιχεία, παρέχοντας γραφικές απεικονίσεις για δεδομένα όπως ημερήσια κατανομή μετρήσεων, μέσες τιμές γλυκόζης και αναλύσεις γευμάτων. Επιπλέον, η λειτουργία Remote Monitor επιτρέπει στους επαγγελματίες υγείας να παρακολουθούν πολλαπλούς χρήστες σε πραγματικό χρόνο, να δημιουργούν αναφορές και να αναλύουν δεδομένα για καλύτερη προσαρμογή της θεραπείας.

Η εφαρμογή υποστηρίζει εισαγωγή και εξαγωγή δεδομένων από εξωτερικές συσκευές και εφαρμογές (π.χ. Medtronic, Dexcom, MySugr), καθιστώντας την ιδιαίτερα ευέλικτη. Το ενσωματωμένο σύστημα υπενθυμίσεων (Reminders) διασφαλίζει την τακτική παρακολούθηση των μετρήσεων, δόσεων και άλλων σημαντικών δραστηριοτήτων.

Επιπλέον, η βάση δεδομένων τροφίμων επιτρέπει την παρακολούθηση της διατροφής μέσω ενσωμάτωσης εξωτερικών πηγών, όπως το USDA, διευκολύνοντας τον ακριβή υπολογισμό υδατανθράκων. Η εφαρμογή είναι επίσης ιδανική για χρήση από επαγγελματίες υγείας και κλινικές, με λειτουργίες ομαδικής διαχείρισης που διευκολύνουν την παρακολούθηση και την αναφορά δεδομένων.

Συνολικά, το Diabetes:M αποτελεί ένα ισχυρό εργαλείο διαχείρισης διαβήτη, προσφέροντας ακρίβεια, εξατομίκευση και ολοκληρωμένη υποστήριξη για την καθημερινή ζωή των χρηστών.

2.4.3 FreeStyle LibreLink

Το FreeStyle LibreLink της Abbott είναι μια πρωτοποριακή εφαρμογή παρακολούθησης γλυκόζης, σχεδιασμένη για ασθενείς με διαβήτη. Η εφαρμογή επιτρέπει στους χρήστες να σαρώνουν τον αισθητήρα FreeStyle Libre απευθείας με το smartphone τους, παρέχοντας άμεσα δεδομένα γλυκόζης σε πραγματικό χρόνο, χωρίς την ανάγκη για παραδοσιακές μετρήσεις με τσιμπήματα. Οι βασικές λειτουργίες περιλαμβάνουν:

- Τρέχουσα ένδειξη γλυκόζης
- Βέλος τάσης που δείχνει τη ροή και τον ρυθμό αλλαγής της γλυκόζης
- Γραφήματα με δεδομένα των τελευταίων 8 ωρών

Οι χρήστες έχουν επίσης τη δυνατότητα να προσθέτουν σημειώσεις σχετικά με τη διατροφή, την άσκηση και τις δόσεις ινσουλίνης, δημιουργώντας ένα πλήρες αρχείο παραμέτρων που επηρεάζουν τα επίπεδα γλυκόζης.

Η εφαρμογή προσφέρει προηγμένες αναφορές και γραφήματα, όπως:

- Ημερήσια μοτίβα γλυκόζης
- Χρόνος εντός στόχου (Time in Range)
- Συμβάντα χαμηλής γλυκόζης (υπογλυκαιμίας)

Όλες οι αναφορές αποθηκεύονται για έως και 90 ημέρες, επιτρέποντας στους χρήστες την παρακολούθηση μακροπρόθεσμων τάσεων και την αναγνώριση μοτίβων.

Με τη συνεχή παρακολούθηση, την ευκολία χρήσης και τη δυνατότητα καταγραφής σημαντικών παραμέτρων, το FreeStyle LibreLink βοηθά τους ασθενείς με διαβήτη να βελτιώσουν τη διαχείριση της γλυκόζης τους. Η εφαρμογή ενισχύει την ευκολία (Convenience), τη σύνδεση (Connection) και τον έλεγχο (Control) της υγείας τους.

2.4.4 Hedia Diabetes Assistant

Η Hedia Diabetes Assistant είναι μια καινοτόμα εφαρμογή διαχείρισης διαβήτη που αξιοποιεί Τεχνητή Νοημοσύνη (AI) για την εξατομίκευση της θεραπείας, βελτιώνοντας τον έλεγχο της γλυκόζης σε άτομα με διαβήτη τύπου 1 και τύπου 2. Η εφαρμογή παρέχει ένα προηγμένο Bolus Calculator, ο οποίος υπολογίζει δόσεις ινσουλίνης με βάση δεδομένα όπως:

- Επίπεδα γλυκόζης
- Κατανάλωση υδατανθράκων

- Σωματική δραστηριότητα
- Ενεργή ινσουλίνη (IOB)

Η ενσωματωμένη βάση δεδομένων τροφίμων διευκολύνει την εκτίμηση της πρόσληψης υδατανθράκων, ενώ η εφαρμογή συγχρονίζεται με μετρητές γλυκόζης μέσω Bluetooth ή NFC για αυτόματη μεταφορά μετρήσεων, εξοικονομώντας χρόνο και αποτρέποντας σφάλματα.

Η Τεχνητή Νοημοσύνη της Hedia μαθαίνει δυναμικά από τα δεδομένα των χρηστών, προσαρμόζοντας τις συστάσεις της ώστε να παρέχει εξατομικευμένες και ακριβείς συμβουλές. Αυτή η προσέγγιση βοηθά στη βελτίωση της διαχείρισης της γλυκόζης, μειώνοντας τον κίνδυνο υπογλυκαιμίας. Επιπλέον, λαμβάνει υπόψη τον αντίκτυπο της φυσικής δραστηριότητας στα επίπεδα γλυκόζης, ενσωματώνοντας αυτήν την παράμετρο στους υπολογισμούς της.

Η Hedia προσφέρει αναφορές και αναλύσεις, συμπεριλαμβανομένου του eA1c (εκτιμώμενη γλυκοζυλιωμένη αιμοσφαιρίνη - HbA1c), το οποίο παρέχει μια επικαιροποιημένη εικόνα του ελέγχου του διαβήτη, βασισμένη σε συνεχή δεδομένα γλυκόζης. Επιπλέον, μέσω της ενσωμάτωσης με την πλατφόρμα Glooko®, οι επαγγελματίες υγείας μπορούν να παρακολουθούν την πρόοδο των ασθενών σε πραγματικό χρόνο, καθιστώντας την Hedia ένα απαραίτητο εργαλείο για την εξατομικευμένη διαχείριση του διαβήτη.

2.4.5 Sugarmate

Η Sugarmate είναι μια προηγμένη εφαρμογή παρακολούθησης διαβήτη, σχεδιασμένη για να παρέχει πληροφορίες σε πραγματικό χρόνο σχετικά με τα επίπεδα γλυκόζης, μέσω σύνδεσης με συσκευές όπως το Dexcom G6. Ξεχωρίζει για την ακρίβεια των ειδοποιήσεων και τη χρήση αλγορίθμων πρόβλεψης, οι οποίοι εντοπίζουν επικίνδυνες τάσεις στα επίπεδα γλυκόζης. Οι χρήστες προειδοποιούνται εγκαίρως για πιθανή υπογλυκαιμία ή υπεργλυκαιμία, βελτιώνοντας την ασφάλεια και τη διαχείριση του διαβήτη.

Η εφαρμογή προσφέρει προσαρμοσμένα γραφήματα και στατιστικά, επιτρέποντας την ανάλυση δεδομένων και την παρακολούθηση σημαντικών παραμέτρων όπως:

- Φυσική δραστηριότητα
- Διατροφή
- Λήψη ινσουλίνης

Όλα αυτά οργανώνονται σε ένα Activity Feed με διαδραστικά γραφήματα, διευκολύνοντας την κατανόηση των δεδομένων.

Η Sugarmate υποστηρίζει συγχρονισμό με συσκευές όπως:

- Apple Watch
- Apple Health
- Amazon Alexa

Επιπλέον, διαθέτει βάση δεδομένων διατροφής για την εκτίμηση υδατανθράκων και θερμίδων, μέσω της ενσωμάτωσης με το FatSecret. Οι χρήστες μπορούν να εξάγουν αναφορές, να διαχειρίζονται πολλαπλούς λογαριασμούς και να μοιράζονται δεδομένα με φροντιστές ή επαγγελματίες υγείας.

Οι μοναδικές λειτουργίες περιλαμβάνουν:

- Προειδοποιήσεις μέσω τηλεφωνικών κλήσεων όταν τα επίπεδα γλυκόζης είναι κάτω από το φυσιολογικό
- Φωνητική αναγνώριση μέσω Siri για άμεση πρόσβαση στα δεδομένα

Η Sugarmate ενισχύει την καθημερινή διαχείριση του διαβήτη, προσφέροντας ασφάλεια, ευκολία και συνεχή πρόσβαση σε ζωτικής σημασίας πληροφορίες.

2.4.6 ChatGLUCO

Η ChatGLUCO είναι μία ολοκληρωμένη εφαρμογή διαχείρισης διαβήτη που προσφέρει λειτουργίες καταγραφής, ανάλυσης και πρόβλεψης των επιπέδων γλυκόζης, βοηθώντας τους ασθενείς να παρακολουθούν και να διαχειρίζονται τον διαβήτη τους με ακρίβεια και ευκολία. Οι χρήστες μπορούν να εγγραφούν και να συνδεθούν για να προσθέσουν προσωπικά στοιχεία, να επεξεργαστούν υγειονομικά δεδομένα όπως βάρος, HbA1c και διάγνωση (Type I ή Type II), και να διατηρούν ημερολόγιο με τιμές γλυκόζης, πρόσληψη υδατανθράκων και δόσεις ινσουλίνης. Το ημερολόγιο παρέχει τη δυνατότητα αυτόματου υπολογισμού της δόσης ινσουλίνης μέσω διαθέσιμων αλγορίθμων (Linear Regression, XGBoost, Random Forest, Sliding Scale + Carbo Calculation) και ενσωματώνει αναζήτηση τροφών μέσω API για υπολογισμό υδατανθράκων. Οι χρήστες μπορούν να προβάλλουν τα δεδομένα τους σε διαγράμματα όπως Time in Range και Glucose-Insulin Correlation, ενώ η λειτουργία Glucose Prediction προβλέπει τα επίπεδα γλυκόζης μετά τη λήψη τροφής και ινσουλίνης, με χρωματική κατηγοριοποίηση για άμεση κατανόηση των αποτελεσμάτων. Επίσης, η λειτουργία Future Glucose Trend Prediction παρέχει τη δυνατότητα απεικόνισης ενός

διαγράμματος που συνδυάζει ιστορικά δεδομένα γλυκόζης με έξι προβλεπόμενες τιμές γλυκόζης για τις επόμενες δύο ημέρες, τρεις χρονικές στιγμές ανά ημέρα. Αυτή η προβολή δίνει τη δυνατότητα στους χρήστες να παρακολουθήσουν τόσο την παρελθοντική πορεία των επιπέδων γλυκόζης τους όσο και την προβλεπόμενη τάση στο εγγύς μέλλον. Επιπλέον, προσφέρεται δυνατότητα εξαγωγής δεδομένων σε μορφή XLS και πλήρης διαχείριση του λογαριασμού, με αλλαγή κωδικού και διαγραφή προφίλ. Με τις λειτουργίες αυτές, η ChatGLUCO παρέχει ένα εύχρηστο και επιστημονικά τεκμηριωμένο εργαλείο για την καθημερινή παρακολούθηση και ανάλυση των δεδομένων διαβήτη.

Πίνακας Συγκριτικής Μελέτης των Συστημάτων

Στον Πίνακα 1 που ακολουθεί, παρουσιάζονται 15 βασικά χαρακτηριστικά που αναδείχθηκαν κατά την ανάλυση των εφαρμογών διαχείρισης διαβήτη. Τα χαρακτηριστικά αυτά αφορούν τόσο λειτουργικές πτυχές των εφαρμογών (π.χ. υπολογισμός bolus ινσουλίνης, παρακολούθηση γλυκόζης σε πραγματικό χρόνο) όσο και πιο προηγμένες δυνατότητες (π.χ. AI/ML-based προβλέψεις, ανάλυση μοτίβων, συγχρονισμός με εξωτερικές συσκευές). Με βάση τις λειτουργίες αυτές, είναι ευκολότερο να γίνει μια συνοπτική αξιολόγηση του τρόπου με τον οποίο καθεμία εφαρμογή εξυπηρετεί τις ανάγκες των ασθενών με διαβήτη, αλλά και των επαγγελματιών υγείας που τους υποστηρίζουν.

Η εφαρμογή μας, ChatGLUCO, διαφοροποιείται από τις υπόλοιπες, καθώς συνδυάζει ολοκληρωμένη καταγραφή δεδομένων (γλυκόζης, πρόσληψης υδατανθράκων, δόσεων ινσουλίνης) με προηγμένους αλγορίθμους που βασίζονται σε Τεχνητή Νοημοσύνη. Στο πλαίσιο αυτό, παρέχει λειτουργία Glucose Prediction για την πρόβλεψη των επιπέδων γλυκόζης, αξιοποιώντας τεχνικές Machine Learning (Linear Regression, XGBoost, Random Forest) και ενσωματώνοντας το Sliding Scale Protocol σε συνδυασμό με την ανάλυση υδατανθράκων (Carbo Calculation) για την πρόβλεψη των δόσεων ινσουλίνης. Επίσης, υποστηρίζει Food Database μέσω εξωτερικού API, εξασφαλίζοντας πιο ακριβή προσέγγιση στον υπολογισμό υδατανθράκων. Επιπλέον, προσφέρει γραφήματα και αναλύσεις για την οπτικοποίηση σημαντικών μεταβλητών (Time in Range, Glucose-Insulin Correlation, Future Glucose Trend Prediction), ενώ διευκολύνει την αποθήκευση, επεξεργασία και εξαγωγή των δεδομένων. Συνολικά, η ChatGLUCO επιδιώκει να καλύψει πολλαπλές ανάγκες των χρηστών, παρέχοντας μια εύχρηστη και επιστημονικά τεκμηριωμένη λύση, σε σύγκριση με τις ανταγωνιστικές εφαρμογές που παρουσιάστηκαν προηγουμένως.

Παρακάτω ακολουθεί ο συγκριτικός πίνακας με τα επιλεγμένα χαρακτηριστικά (Πίνακας 1).

Feature \ Systems	chatGLUCO	Diabetes:M	FreeStyle LibreLink	Hedia	Sugarmate	GlucoLog RapidCalc
Bolus Insulin Calculation	+	+	-	+	-	+
Insulin on board (IOB)	-	+	-	+	-	+
Sliding Scale Protocol	+	+	-	-	-	+
Carbohydrate Calculation	+	+	-	+	-	+
Diary Management	+	+	+	+	+	+
Pattern Analysis	-	+	+	-	-	-
Real-Time Glucose Monitoring (CGM)	-	+	+	-	+	-
AI/ML-based Predictions	+	-	-	+	-	-
Glucose Prediction	+	-	-	-	-	-
Food Database	+	+	-	+	+	-
Activity Tracking	-	+	-	+	+	+
Charts	+	+	+	+	+	+
Data Export	+	+	+	+	+	+
Notifications	-	+	+	+	+	+
Device Synchronization	-	+	+	+	+	+

Πίνακας 1: Συγκριτική Μελέτη των Συστημάτων

3 Μέθοδοι Μοντέλων Μηχανικής Μάθησης

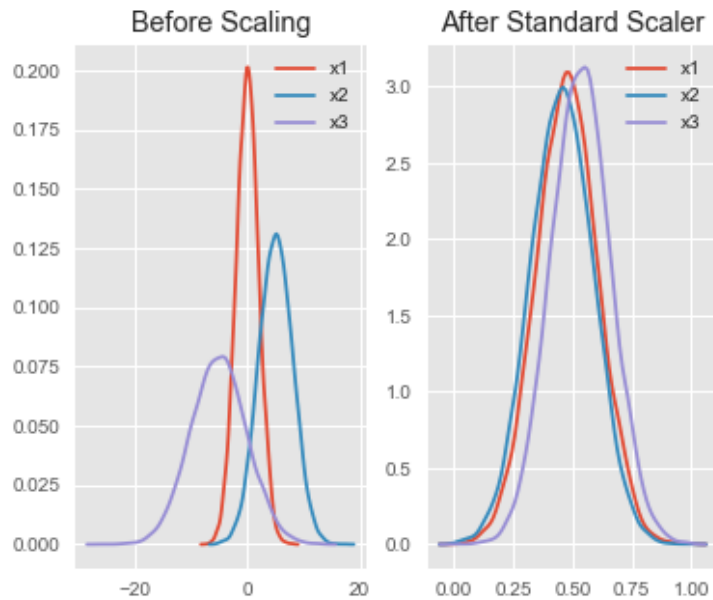
3.1 Αλγόριθμοι Μηχανικής Μάθησης

Προτού προχωρήσουμε στην παρουσίαση των αλγορίθμων επιβλεπόμενης μάθησης (supervised learning), είναι χρήσιμο να επιστημόνουμε συνοπτικά τη βασική μεθοδολογία που ακολουθείται στην ανάπτυξη και αξιολόγηση μοντέλων. Στο πλαίσιο της παρούσας εργασίας, επιλέγονται τεχνικές παλινδρόμησης (regression) με γνώμονα την ικανότητά τους να μοντελοποιούν με ακρίβεια σύνθετες σχέσεις μεταξύ δεδομένων εισόδου (features) και τιμών στόχου (target values). Οι αλγόριθμοι που θα παρουσιάσουμε (Linear Regression, Random Forest, XGBoost) καλύπτουν ένα ευρύ φάσμα εφαρμογών και επιτυγχάνουν ένα καλό επίπεδο γενίκευσης.

Ένα κρίσιμο βήμα στην ανάπτυξη μοντέλων μηχανικής μάθησης, το οποίο εφαρμόζεται σε όλα τα σενάρια που μελετώνται, είναι η κανονικοποίηση των χαρακτηριστικών (feature scaling). Στην παρούσα εργασία, χρησιμοποιήθηκε ο StandardScaler της βιβλιοθήκης *scikit-learn*, ώστε όλα τα χαρακτηριστικά να έχουν μηδενική μέση τιμή (mean) και τυπική απόκλιση ίση με 1 (unit variance). Με τον τρόπο αυτό:

1. Διευκολύνεται η σύγκλιση των αλγορίθμων που στηρίζονται σε βελτιστοποίηση (π.χ. Γραμμική Παλινδρόμηση).
2. Παρόλο που τα μοντέλα βασισμένα σε δέντρα (π.χ. Random Forest, XGBoost) δεν απαιτούν κανονικοποίηση, η χρήση του StandardScaler βοηθά στην ομοιογένεια και συγκρισιμότητα των χαρακτηριστικών όταν συνδυάζονται διαφορετικοί αλγόριθμοι.

Στην παρακάτω εικόνα [Σχήμα 1] απεικονίζονται οι κατανομές τριών χαρακτηριστικών (x_1 , x_2 , x_3) πριν και μετά την εφαρμογή του StandardScaler. Στο αριστερό διάγραμμα (*Before Scaling*), οι μεταβλητές έχουν διαφορετικά εύρη τιμών και μέσες τιμές. Στο δεξιό διάγραμμα (*After Standard Scaler*), διακρίνεται ότι πλέον έχουν ευθυγραμμιστεί σε κοινή κλίμακα, καθώς αποκτούν μηδενική μέση τιμή και τυπική απόκλιση ίση με 1, βελτιστοποιώντας έτσι τη σύγκλιση και την απόδοση των αλγορίθμων μηχανικής μάθησης.



Σχήμα 1: Before and After Feature Scaling [13]

Στις επόμενες υποενότητες, παρουσιάζονται λεπτομερώς οι βασικές αρχές λειτουργίας, η μαθηματική θεμελίωση, τα πλεονεκτήματα και οι περιορισμοί τριών δημοφιλών αλγορίθμων: Linear Regression, Random Forest και XGBoost.

3.1.1 Linear Regression

Η Γραμμική Παλινδρόμηση (Linear Regression) [6] [7] είναι ένας από τους πιο θεμελιώδεις αλγόριθμους επιβλεπόμενης μάθησης, ιδιαίτερα χρήσιμος σε προβλήματα παλινδρόμησης. Η βασική ιδέα είναι ότι η εξαρτημένη μεταβλητή y μπορεί να προσεγγιστεί ως γραμμικός συνδυασμός των ανεξάρτητων μεταβλητών $x = (x_1, x_2, \dots, x_n)$, όπου κάθε x_i αναπαριστά μία μεταβλητή εισόδου:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon,$$

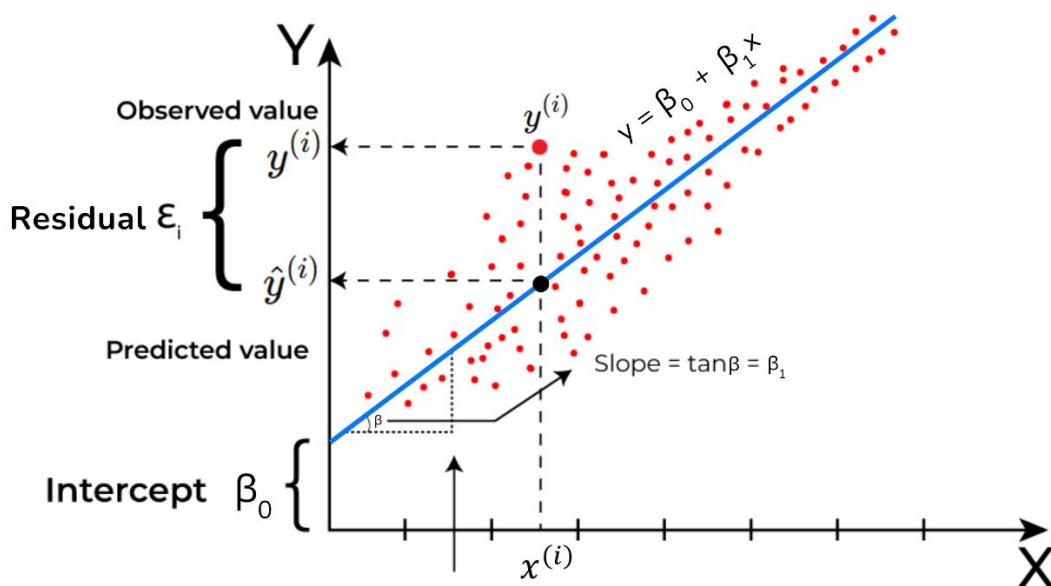
όπου:

- β_0 είναι ο σταθερός όρος (intercept),
- β_i είναι οι συντελεστές ή βάρη της γραμμικής παλινδρόμησης (regression coefficients), που αντιπροσωπεύουν τη συμβολή κάθε ανεξάρτητης μεταβλητής x_i στην πρόβλεψη της εξαρτημένης μεταβλητής y .

- ε είναι ο όρος σφάλματος (error term), που αντιπροσωπεύει τη διαφορά μεταξύ της πραγματικής τιμής y και της προβλεπόμενης τιμής.

Στην παρακάτω εικόνα [Σχήμα 2] παρουσιάζεται μια γραφική απεικόνιση της απλής Γραμμικής Παλινδρόμησης (με μία μόνο μεταβλητή x):

- Στον οριζόντιο άξονα παριστάνεται η ανεξάρτητη μεταβλητή x , ενώ στον κατακόρυφο άξονα η εξαρτημένη μεταβλητή y .
- Οι κόκκινες κουκκίδες αντιπροσωπεύουν τα πραγματικά, παρατηρούμενα δεδομένα $(x^{(i)}, y^{(i)})$.
- Η μπλε γραμμή είναι η «γραμμή παλινδρόμησης» $y = \beta_0 + \beta_1 x$, η οποία προκύπτει από την εκτίμηση των παραμέτρων β_0 (σταθερός όρος) και β_1 (κλίση).
- Το σημείο τομής της μπλε γραμμής με τον κατακόρυφο άξονα αντιστοιχεί στο β_0 , ενώ η κλίση της γραμμής είναι ίση με β_1 .
- Για κάθε σημείο δεδομένων, η κατακόρυφη απόσταση ανάμεσα στην παρατηρούμενη τιμή $y^{(i)}$ και στην προβλεπόμενη τιμή $\hat{y}^{(i)}$ στο σχήμα αναφέρεται ως residual και συμβολίζεται με ε_i ($\varepsilon_i = y_i - \hat{y}_i$). Τα residuals είναι εκτιμήσεις του θεωρητικού error term (ε_i), το οποίο αναφέρεται στη διαφορά από την πραγματική υποκείμενη σχέση.



Σχήμα 2: Γραφική αναπαράσταση μιας Γραμμικής Παλινδρόμησης [14]

Μέσω αυτού του διαγράμματος επιδιώκουμε να τονίσουμε τον τρόπο με τον οποίο η γραμμική παλινδρόμηση προσεγγίζει τα δεδομένα με μια ευθεία γραμμή. Ουσιαστικά, απεικονίζουμε τη διαδικασία υπολογισμού του σταθερού όρου (β_0) και της κλίσης (β_1), καθώς και το residual (ε_i), μεταξύ προβλεπόμενων και πραγματικών τιμών. Με τον τρόπο αυτόν, μπορούμε να κατανοήσουμε καλύτερα πώς τα σημεία (δηλαδή οι πραγματικές τιμές) κατανέμονται γύρω από την εκτιμώμενη γραμμή παλινδρόμησης, ποσοτικοποιώντας τη συνολική απόκλιση. Τελικά, το διάγραμμα μας βοηθά να ελέγξουμε εάν η γραμμική υπόθεση φαίνεται εύλογη για τα συγκεκριμένα δεδομένα και να αξιολογήσουμε οπτικά την καταλληλότητα του μοντέλου.

Θεωρητικό Υπόβαθρο

- **Κριτήριο βέλτιστης εκτίμησης – Μέθοδος Ελαχίστων Τετραγώνων (Ordinary Least Squares, OLS)**

Η πιο διαδεδομένη μέθοδος εκτίμησης των β_i στη Γραμμική Παλινδρόμηση είναι η μέθοδος των Ελαχίστων Τετραγώνων, όπου επιδιώκεται η ελαχιστοποίηση του αθροίσματος των τετραγώνων των residuals. Δηλαδή, αν έχουμε m δείγματα $\{(x^{(i)}, y^{(i)})\}_{i=1}^m$, ο στόχος είναι να ελαχιστοποιηθεί η συνάρτηση κόστους:

$$S(\beta) = \sum_{i=1}^m (y^{(i)} - \hat{y}^{(i)})^2 = \sum_{i=1}^m (y^{(i)} - (\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_n x_n^{(i)}))^2$$

Με γραμμική άλγεβρα, αν θέσουμε X τον πίνακα των ανεξάρτητων μεταβλητών (διάστασης $m \times (n + 1)$ αν συμπεριλάβουμε τη στήλη του σταθερού όρου) και y το διάνυσμα-στήλη των τιμών του στόχου, τότε η λύση β που ελαχιστοποιεί το $S(\beta)$ δίνεται από τις κανονικές εξισώσεις (Normal Equations):

$$\beta = (X^T X)^{-1} X^T y$$

υπό την προϋπόθεση ότι $(X^T X)$ είναι αντιστρέψιμος πίνακας.

Υπολογιστική πολυπλοκότητα OLS:

Η αντιστροφή του πίνακα $(X^T X)$ έχει υπολογιστική πολυπλοκότητα $O(n^3)$, όπου n είναι ο αριθμός των χαρακτηριστικών (features) στο dataset. Το κόστος αυτό γίνεται ιδιαίτερα υψηλό σε datasets με μεγάλο αριθμό χαρακτηριστικών (n), καθώς αυξάνεται δραματικά με τον αριθμό τους. Σε datasets με πολύ μεγάλο αριθμό

χαρακτηριστικών ή παρατηρήσεων (m), η αντιστροφή μπορεί να είναι ακριβή ή και ασταθής, καθιστώντας την εφαρμογή της OLS λιγότερο αποδοτική.

Για την αντιμετώπιση αυτών των περιορισμών, επαναληπτικές μέθοδοι, όπως Gradient Descent ή Stochastic Gradient Descent (SGD), συχνά προτιμώνται για την εκτίμηση των συντελεστών.

- **Εναλλακτικές Μέθοδοι Εκτίμησης**

Εκτός από τη βασική OLS, σε πιο σύνθετα ή μεγάλα σύνολα δεδομένων χρησιμοποιούνται μέθοδοι κανονικοποίησης (regularization), όπως:

Ridge Regression (L2): Προσθήκη όρου ποινής (penalties) $\lambda \sum_{j=1}^n \beta_j^2$ στη συνάρτηση κόστους,

Lasso Regression (L1): Προσθήκη όρου ποινής (penalties) $\lambda \sum_{j=1}^n |\beta_j|$

Οι μέθοδοι αυτές μειώνουν την υπερπροσαρμογή (overfitting) και την πολυσυγγραμμικότητα (multicollinearity), θυσιάζοντας ελαφρώς την ακρίβεια των εκτιμήσεων (μεροληψία) για πιο σταθερά και αξιόπιστα αποτελέσματα (μείωση διακύμανσης). Στην παρούσα εργασία, όμως, εφαρμόζεται η βασική μέθοδος Γραμμικής Παλινδρόμησης (OLS).

- **Ερμηνευσιμότητα Συντελεστών**

Κάθε συντελεστής β_i δείχνει τη «σχετική βαρύτητα» ή την επίδραση της αντίστοιχης ανεξάρτητης μεταβλητής x_i στην έξοδο y . Για παράδειγμα, αν $\beta_i > 0$, η μεταβλητή x_i συσχετίζεται θετικά με την έξοδο. Αν $\beta_i < 0$, αυτό σημαίνει ότι η μεταβλητή x_i έχει αρνητική συσχέτιση. Σε περίπτωση που οι μεταβλητές έχουν πολύ διαφορετικές κλίμακες, η κανονικοποίηση (standardization) μπορεί να βοηθήσει στην άμεση συγκρισιμότητα των συντελεστών. Χωρίς κάποια μορφή κλιμάκωσης, οι συντελεστές ενδέχεται να έχουν «φαινομενικά» διαφορετική βαρύτητα, απλώς επειδή η κλίμακα του x_i είναι διαφορετική.

Ρόλος στο Παρόν Έργο

Στη συγκεκριμένη διπλωματική εργασία, η Γραμμική Παλινδρόμηση χρησιμοποιείται για την πρόβλεψη δύο διαφορετικών μεταβλητών (δύο σενάρια):

- Σενάριο Πρόβλεψης Δόσης Ινσουλίνης: Η εξαρτημένη μεταβλητή y είναι η `insulin_dose`, ενώ τα χαρακτηριστικά (ανεξάρτητες μεταβλητές) περιλαμβάνουν τα `glucose_old`, `food_carbo`, `glucose_new`, `weight`, `age`, `diagnosis`, `gender`.

- Σενάριο Πρόβλεψης Επιπέδων Γλυκόζης: Η εξαρτημένη μεταβλητή y είναι η `glucose_new`, ενώ τα χαρακτηριστικά (ανεξάρτητες μεταβλητές) περιλαμβάνουν τα `glucose_old`, `food_carbo`, `insulin_dose`, `weight`, `age`, `diagnosis`, `gender`

Εφαρμογή και Υλοποίηση

Η υλοποίηση της Γραμμικής Παλινδρόμησης στο παρόν σύστημα γίνεται μέσω της κλάσης `LinearRegression` της βιβλιοθήκης *scikit-learn* (Python). Σε αντίθεση με πιο σύνθετους αλγόριθμους (π.χ. `XGBoost`, `Random Forest`) που διαθέτουν πολλές υπερπαραμέτρους και ενδείκνυνται για *Grid Search* ή άλλες τεχνικές βελτιστοποίησης, η βασική (OLS) γραμμική παλινδρόμηση απαιτεί ελάχιστη ρύθμιση.

Πλεονεκτήματα και Περιορισμοί

Πλεονεκτήματα:

- **Απλότητα και ταχύτητα**: Η εκτίμηση είναι συχνά άμεση με την επίλυση των κανονικών εξισώσεων.
- **Υψηλή ερμηνευσιμότητα**: Οι συντελεστές β_i παρέχουν ευανάγνωστη εικόνα της συμβολής κάθε χαρακτηριστικού.
- **Καλή βάση σύγκρισης (baseline)**: Πολλοί πιο πολύπλοκοι αλγόριθμοι συγκρίνονται συνήθως ως προς την επίδοση που θα έδινε μια γραμμική παλινδρόμηση.

Περιορισμοί:

- **Απαίτηση γραμμικής σχέσης**: Εάν η πραγματική σχέση $x - y$ είναι μη γραμμική, η γραμμική παλινδρόμηση δύναται να εμφανίσει υψηλό σφάλμα.
- **Ευαισθησία σε εξωτερικές τιμές (outliers)**: Μερικές ακραίες τιμές μπορούν να επηρεάσουν σημαντικά τη λύση.
- **Πολυσυγγραμμικότητα**: Σε περιπτώσεις όπου οι ανεξάρτητες μεταβλητές (x) παρουσιάζουν πολυσυγγραμμικότητα (multicollinearity), δηλαδή είναι σχεδόν γραμμικά εξαρτημένες, η διαδικασία εκτίμησης των συντελεστών (β_i) γίνεται ασταθής. Αυτό μπορεί να οδηγήσει σε μη αξιόπιστες προβλέψεις και ευαισθησία στις μικρές αλλαγές των δεδομένων.

Συνοψίζοντας, η Γραμμική Παλινδρόμηση αποτελεί μια ιδιαίτερα χρήσιμη μέθοδο για την επίλυση προβλημάτων παλινδρόμησης και συνιστά ιδανική αφετηρία (baseline) πριν περάσουμε σε πιο πολύπλοκες προσεγγίσεις. Παρά την απλότητά της, παρέχει ισχυρή ερμηνευσιμότητα και γρήγορη εκπαίδευση. Στο πλαίσιο της διπλωματικής αυτής,

χρησιμοποιείται σε δύο ξεχωριστά σενάρια πρόβλεψης (δόση ινσουλίνης και επίπεδα γλυκόζης), επιτρέποντας να αναλυθεί η συσχέτιση των βασικών μεταβλητών με τις τιμές-στόχο. Αν και έχει ορισμένους περιορισμούς (π.χ. ευαισθησία σε outliers, απαιτούμενη γραμμικότητα), λειτουργεί αποτελεσματικά ως αρχικό εργαλείο εξερεύνησης και ποσοτικής εκτίμησης των σχέσεων μεταξύ των δεδομένων μας.

3.1.2 Random Forest

Ο Random Forest είναι ένας από τους πιο διαδεδομένους και αποτελεσματικούς αλγορίθμους Συνόλων (Ensemble Methods) στη Μηχανική Μάθηση. Βασίζεται σε πολλαπλά Δέντρα Απόφασης (Decision Trees) και αξιοποιεί την τεχνική που ονομάζεται bagging (bootstrap aggregating). Ουσιαστικά, αντί να εκπαιδεύουμε ένα μόνο Δέντρο Απόφασης σε ολόκληρο το σύνολο δεδομένων, εκπαιδεύουμε πολλαπλά Δέντρα, το καθένα σε ένα διαφορετικό υποσύνολο του αρχικού συνόλου εκπαίδευσης, και στη συνέχεια συνδυάζουμε τις προβλέψεις τους. Στην περίπτωση της ταξινόμησης (classification), ο Random Forest επιλέγει την κλάση που εμφανίζεται πιο συχνά (πλειοψηφική ψήφος), ενώ στην περίπτωση της παλινδρόμησης (regression) λαμβάνει τον μέσο όρο των επιμέρους προβλέψεων. Παρακάτω παρουσιάζονται το θεωρητικό και μαθηματικό υπόβαθρο του αλγορίθμου.

Η αποτελεσματικότητα του Random Forest έγκειται στο γεγονός ότι κάθε δέντρο:

1. Εκπαιδεύεται σε ένα διαφορετικό τυχαίο δείγμα (με επανάθεση) των δεδομένων εκπαίδευσης (Bagging).
2. Χρησιμοποιεί τυχαίο υποσύνολο χαρακτηριστικών για κάθε διακλάδωση (split), μειώνοντας τη συσχέτιση μεταξύ των δέντρων.

Με αυτόν τον τρόπο, τα δέντρα είναι λιγότερο συσχετισμένα μεταξύ τους (μικρότερη συσχέτιση σφαλμάτων) και, ως εκ τούτου, το τελικό σύνολο εμφανίζει χαμηλότερη διακύμανση (variance) και χαμηλότερο κίνδυνο overfitting.

Εφαρμογή του Random Forest

Στην παρούσα εργασία, ο Random Forest χρησιμοποιείται ως μοντέλο παλινδρόμησης (regression) σε δύο σενάρια:

1. Την πρόβλεψη των επιπέδων γλυκόζης,
2. Την πρόβλεψη των δόσεων ινσουλίνης.

Η έξοδος του αλγορίθμου είναι μία συνεχής τιμή, γεγονός που τον καθιστά κατάλληλο για προβλήματα παλινδρόμησης.

Θεωρητικό Υπόβαθρο

Βασική Ιδέα των Ensemble Methods

Ένας τρόπος βελτίωσης της απόδοσης ενός αλγορίθμου είναι η χρήση πολλαπλών «απλών» μοντέλων και ο συνδυασμός των αποτελεσμάτων τους. Αυτό είναι το θεμέλιο της οικογένειας των ensemble methods. Η φιλοσοφία είναι ότι πολλά αδύναμα μοντέλα (weak learners) μπορούν, όταν συνδυαστούν κατάλληλα, να παράγουν ένα ισχυρό μοντέλο με βελτιωμένη γενίκευση.

Ο Random Forest ανήκει στην κατηγορία των ensemble methods που χρησιμοποιούν την τεχνική bagging (Bootstrap Aggregating), η οποία λειτουργεί ως εξής:

1. Γίνεται δειγματοληψία με επανάθεση (bootstrap sampling) από το αρχικό σύνολο δεδομένων, ώστε να δημιουργηθούν υποσύνολα δεδομένων διαφορετικά μεταξύ τους.
2. Σε κάθε υποσύνολο εκπαιδεύεται ένας ανεξάρτητος αλγόριθμος (στην περίπτωση μας, ένα Δέντρο Απόφασης).
3. Τα αποτελέσματα συνδυάζονται: στην ταξινόμηση με πλειοψηφική ψήφο (majority voting) και στην παλινδρόμηση με μέσο όρο (averaging).

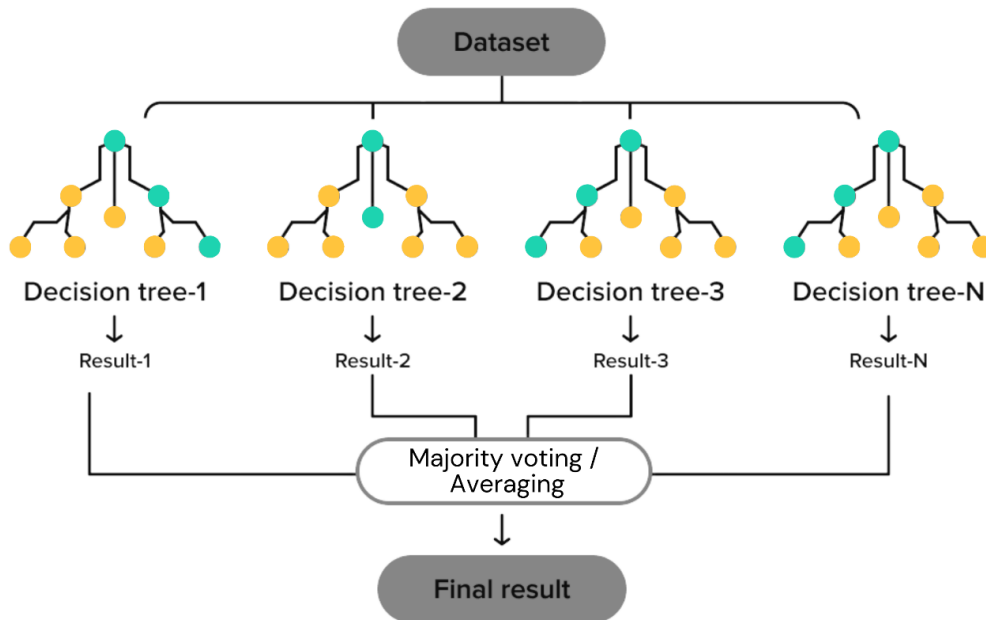
Δέντρα Απόφασης και Τυχαία Επιλογή Χαρακτηριστικών

Ο Random Forest εμπλουτίζει τη διαδικασία του bagging με την επιπλέον ιδέα της τυχαίας επιλογής χαρακτηριστικών. Κατά την κατασκευή ενός Δέντρου Αποφάσεων (Decision Tree), σε κάθε διακλάδωση (split) αντί να εξετάζονται όλα τα διαθέσιμα χαρακτηριστικά (features) για να βρεθεί ο βέλτιστος χωρισμός, εξετάζεται μόνο ένα τυχαίο υποσύνολο των χαρακτηριστικών. Έτσι:

- Μειώνεται η συσχέτιση μεταξύ των διαφόρων δέντρων του δάσους (forest).
- Αυξάνεται η ποικιλία (diversity) των δέντρων, στοιχείο που τελικά βελτιώνει την απόδοση του συνολικού μοντέλου.

Στην παρακάτω εικόνα απεικονίζεται γραφικά η διαδικασία λειτουργίας του Random Forest. Το dataset διαιρείται σε υποσύνολα, και σε κάθε ένα εκπαιδεύεται ένα Δέντρο Απόφασης.

Στο τέλος, τα αποτελέσματα συνδυάζονται μέσω πλειοψηφικής ψήφου (majority voting) ή μέσου όρου (averaging) για την τελική πρόβλεψη.



Σχήμα 3: Γραφική Απεικόνιση της Λειτουργίας του Random Forest [11]

Μαθηματική Περιγραφή της Διαδικασίας Bagging

Ας υποθέσουμε ότι έχουμε ένα σύνολο δεδομένων εκπαίδευσης:

$$D = \{(x_i, y_i)\}_{i=1}^N,$$

όπου $x_i \in \mathbb{R}^d$ είναι το διάνυσμα χαρακτηριστικών και y_i η ετικέτα ή η τιμή-στόχος (ανάλογα με το αν λύνουμε πρόβλημα ταξινόμησης ή παλινδρόμησης).

1. **Bootstrap Samples:** Δημιουργούμε B διαφορετικά υποσύνολα D_b (με $b = 1, 2, \dots, B$) από το D μέσω δειγματοληψίας με επανάθεση (bootstrap). Κάθε D_b έχει το ίδιο μέγεθος με το D , δηλαδή N , αλλά ορισμένα δείγματα μπορεί να επαναλαμβάνονται, ενώ κάποια άλλα ενδέχεται να μην επιλεγούν καθόλου.
2. **Κατασκευή Δέντρων:** Για κάθε δείγμα D_b εκπαιδεύεται ένα Δέντρο Αποφάσεων $h_b(x)$. Όταν εξετάζουμε ένα split σε κάποιο κόμβο του δέντρου, διαλέγουμε τυχαία $m < d$ χαρακτηριστικά (όπου m είναι υπερπαράμετρος) και βρίσκουμε το βέλτιστο split μόνο ανάμεσα στα m χαρακτηριστικά.
3. **Συνδυασμός Αποτελεσμάτων:**

- ο Για **ταξινόμηση**, το τελικό αποτέλεσμα $H(x)$ είναι η κατηγορία που ψηφίζεται από τα περισσότερα δέντρα:

$$H(\mathbf{x}) = \arg \max_c \sum_{b=1}^B \mathbb{I}(h_b(\mathbf{x}) = c),$$

όπου $\mathbb{I}(\cdot)$ είναι η ενδεικτική συνάρτηση που ισούται με 1 όταν η συνθήκη ικανοποιείται και 0 διαφορετικά.

- ο Για **παλινδρόμηση**, το τελικό αποτέλεσμα $H(x)$ είναι ο μέσος όρος των εξόδων όλων των δέντρων:

$$H(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^B h_b(\mathbf{x}).$$

Μείωση Διασποράς και Κίνδυνος Overfitting

Η κατασκευή πολλών Δέντρων Αποφάσεων σε διαφορετικά bootstrap δείγματα οδηγεί σε μείωση της διασποράς (variance) του τελικού μοντέλου, διατηρώντας ταυτόχρονα χαμηλή μεροληψία (bias). Τα Δέντρα Αποφάσεων, ως αλγόριθμοι, τείνουν να έχουν χαμηλό bias αλλά υψηλό variance. Με τον Random Forest αντιμετωπίζουμε ακριβώς αυτό το πρόβλημα του υψηλού variance:

- Η **διασπορά** μειώνεται επειδή η εξαγωγή τελικής απόφασης βασίζεται στον συνδυασμό πολλών ανεξάρτητων (ή «σχεδόν» ανεξάρτητων) δέντρων.
- Ο **κίνδυνος overfitting** μειώνεται διότι κανένα δέντρο δεν «μαθαίνει» ολόκληρο το dataset: το καθένα βλέπει ένα επιλεγμένο με επανάθεση υποσύνολο δεδομένων και κάθε split σε κόμβο βασίζεται σε τυχαία υποσύνολα χαρακτηριστικών.

Out-of-Bag (OOB) Error

Κατά τη δημιουργία κάθε bootstrap δείγματος D_b , κάποια δείγματα από το αρχικό σύνολο D δεν συμπεριλαμβάνονται (κατά μέσο όρο περίπου το 36,8% των δειγμάτων «μένουν έξω» λόγω επαναλαμβανόμενης επιλογής άλλων δειγμάτων). Τα δείγματα αυτά λέγονται Out-of-Bag (OOB) samples.

Ο Random Forest μας παρέχει μια εκτίμηση του σφάλματος γενίκευσης, υπολογίζοντας το σφάλμα μόνο σε αυτά τα OOB δείγματα για κάθε δέντρο. Συγκεκριμένα, για ένα OOB δείγμα

x_i , προβλέπουμε την ετικέτα του χρησιμοποιώντας μόνο τα δέντρα που δε το είχαν στο bootstrap δείγμα τους. Η μέση απόδοση στα OOB δείγματα αποτελεί και μία εκτίμηση της OOB error.

Μαθηματικές Πτυχές της Κατασκευής των Δέντρων

Κριτήριο Διαχωρισμού (Split Criterion)

Τα Δέντρα Αποφάσεων χρειάζονται ένα κριτήριο για να επιλέξουν σε κάθε κόμβο ποιο χαρακτηριστικό και ποια τιμή-κατώφλι (threshold) θα διαχωρίσει τα δεδομένα. Συνήθη κριτήρια είναι:

1. **Gini Impurity (Για ταξινόμηση):** Χρησιμοποιείται για την αξιολόγηση του κατά πόσο ένας διαχωρισμός δεδομένων σε έναν κόμβο βελτιώνει την ομοιογένεια των υποκατηγοριών.

$$G_{split} = \sum_{j=1}^J p_j (1 - p_j)$$

όπου p_j είναι η σχετική συχνότητα της κατηγορίας j στον κόμβο. Όσο μικρότερη είναι η τιμή του G_{split} , τόσο μεγαλύτερη είναι η ομοιογένεια του κόμβου.

2. **Εντροπία (Για ταξινόμηση):** Η εντροπία (entropy) βασίζεται στη θεωρία πληροφορίας και μετρά την αβεβαιότητα στον κόμβο. Ορίζεται ως:

$$H_{split} = - \sum_{j=1}^J p_j \log_2(p_j)$$

Όσο μεγαλύτερη είναι η μείωση της εντροπίας μετά από μια διάσπαση, τόσο καλύτερη θεωρείται η διάσπαση.

3. **MSE (Για παλινδρόμηση):** Στην παλινδρόμηση, το κριτήριο διαχωρισμού εστιάζει στη μείωση του Μέσου Τετραγωνικού Σφάλματος (MSE), το οποίο ορίζεται ως:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

όπου y_i είναι οι πραγματικές τιμές στον κόμβο και $\hat{y}_i$ η εκτίμηση της παλινδρόμησης (συνήθως ο μέσος όρος των τιμών εκπαίδευσης του κόμβου). Όσο μικρότερη είναι η τιμή του MSE, τόσο καλύτερο θεωρείται το split.

Πλεονεκτήματα και Μειονεκτήματα

Πλεονεκτήματα

- **Υψηλή Ακρίβεια:** Ο συνδυασμός πολλών δέντρων προσφέρει καλύτερη απόδοση από ένα μεμονωμένο δέντρο.
- **Μειωμένος Κίνδυνος Overfitting:** Η τυχαία δειγματοληψία (στα δεδομένα) μειώνει τη συσχέτιση μεταξύ των δέντρων.
- **Εκτίμηση Σφάλματος OOB:** Παρέχει μια εγγενή εκτίμηση του σφάλματος γενίκευσης χωρίς τη χρήση εξωτερικού validation set.
- **Δυνατότητα Χειρισμού Μεγάλου Αριθμού Χαρακτηριστικών:** Μέσω της τυχαίας επιλογής χαρακτηριστικών, ο Random Forest παραμένει αποτελεσματικός ακόμα και με μεγάλο αριθμό χαρακτηριστικών.
- **Ανθεκτικότητα σε Missing Values:** Σε πολλές υλοποιήσεις, μπορεί να χειριστεί ελλιπή δεδομένα.

Μειονεκτήματα

- **Μεγάλο Υπολογιστικό Κόστος:** Η εκπαίδευση πολλών δέντρων μπορεί να είναι χρονοβόρα.
- **Δυσκολία Ερμηνείας:** Παρά τη δυνατότητα εξαγωγής μέτρων «Feature Importance», η ερμηνεία ενός δάσους από πολλά δέντρα είναι πιο περίπλοκη σε σχέση με ένα απλό Δέντρο Αποφάσεων.
- **Ευαισθησία σε Θόρυβο:** Αν τα δεδομένα έχουν πολύ θόρυβο, η απόδοση μπορεί να επηρεαστεί ελαφρώς.

Η παραπάνω ανάλυση για τα πλεονεκτήματα και τα μειονεκτήματα του Random Forest επιβεβαιώνεται και στην πρακτική εφαρμογή του στον τομέα της πρόβλεψης δόσεων ινσουλίνης και επιπέδων γλυκόζης, όπως υλοποιήθηκε στο Jupyter Notebook. Ο Random Forest αποδείχθηκε ότι είναι ο πιο κατάλληλος αλγόριθμος, παρέχοντας τις καλύτερες μετρικές απόδοσης. Συγκεκριμένα, πέτυχε το μικρότερο Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE) σε σύγκριση με τους άλλους αλγορίθμους που δοκιμάστηκαν, ενώ και οι υπόλοιπες μετρικές αξιολόγησης υποδεικνύουν την εξαιρετική του απόδοση. Αυτή η επίδοση αποδίδεται στην ικανότητα του Random Forest να διαχειρίζεται πολύπλοκα δεδομένα με πολλαπλά χαρακτηριστικά, να μειώνει τη διασπορά μέσω του bagging και να αντιμετωπίζει τον θόρυβο του dataset, που είναι συχνός στις ιατρικές εφαρμογές.

3.1.3 XGBoost

Ο XGBoost (eXtreme Gradient Boosting) [15] [16] [17] αποτελεί έναν από τους πιο δημοφιλείς και αποδοτικούς αλγορίθμους μηχανικής μάθησης για προβλήματα ταξινόμησης και παλινδρόμησης. Η φιλοσοφία του βασίζεται στην τεχνική Gradient Boosting, η οποία συνδυάζει πολλά απλά, ασθενή μοντέλα (συνήθως δέντρα αποφάσεων), με κάθε νέο μοντέλο να μαθαίνει από τα λάθη των προηγούμενων, ώστε να δημιουργηθεί ένα ενιαίο, ισχυρότερο μοντέλο. Η ονομασία “eXtreme” οφείλεται στην έμφαση που δίνεται στην υψηλή υπολογιστική απόδοση και αποδοτικότητα μνήμης μέσω βελτιστοποιημένων υλοποιήσεων και παραλληλοποίησης.

Στην παρούσα διπλωματική εργασία, ο XGBoost χρησιμοποιείται για την πρόβλεψη δόσεων ινσουλίνης και επιπέδων γλυκόζης, αξιοποιώντας τη δυνατότητά του να χειρίζεται μεγάλα και πολύπλοκα σύνολα δεδομένων με υψηλή ακρίβεια. Η παραλληλοποίηση επιτρέπει την ταυτόχρονη επεξεργασία πολλαπλών δεδομένων, μειώνοντας τον χρόνο εκπαίδευσης, ενώ η αποδοτικότητα μνήμης εξασφαλίζει τη δυνατότητα επεξεργασίας μεγάλων datasets ακόμα και σε περιορισμένους υπολογιστικούς πόρους.

Θεωρητικό και Μαθηματικό Υπόβαθρο XGBoost

1. Βασικές Αρχές Gradient Boosting

Ο XGBoost ανήκει στην οικογένεια των αλγορίθμων Gradient Boosting. Στο Gradient Boosting, ο στόχος είναι να δημιουργήσουμε ένα «ισχυρό» τελικό μοντέλο μέσα από τον συνδυασμό πολλών απλών (ασθενών) μοντέλων (weak learners), συνήθως Δέντρων Αποφάσεων (Decision Trees). Η γενική ιδέα συνοψίζεται ως εξής:

1. Ξεκινάμε με ένα αρχικό μοντέλο $\hat{y}^{(0)}$, το οποίο συχνά είναι ένας απλός μέσος όρος των εξαρτημένων μεταβλητών (στην περίπτωση παλινδρόμησης) ή μια αρχική πρόβλεψη για την ταξινόμηση.
2. Σε κάθε βήμα t , προσαρμόζουμε (fit) ένα νέο Δέντρο Αποφάσεων f_t σε κάποια μορφή υπολοίπων (residuals). Τα «υπόλοιπα» αυτά ορίζονται από τη ροή του Gradient Boosting ως οι τιμές της πρώτης παραγώγου της συνάρτησης κόστους (loss function) ως προς τις τρέχουσες προβλέψεις του μοντέλου.
3. Το νέο δέντρο προστίθεται στο σύνολο των επιμέρους «ασθενών μοντέλων», βελτιώνοντας σταδιακά την απόδοση του τελικού συνδυαστικού μοντέλου.

Σε μαθηματική μορφή, αν $\phi(x)$ είναι το συνολικό μοντέλο που αποτελείται από K ασθενή μοντέλα (δέντρα), τότε:

$$\phi(x) = \sum_{k=1}^K f_k(x),$$

όπου κάθε f_k ανήκει σε ένα σύνολο επιτρεπτών συναρτήσεων (π.χ. σύνολο Δέντρων Αποφάσεων). Σκοπός είναι η ελαχιστοποίηση κάποιας συνάρτησης κόστους L σε σχέση με τις παραμέτρους των δέντρων.

2. Στόχος (Objective Function) στον XGBoost

Στον XGBoost, ορίζεται μια αντικειμενική συνάρτηση (objective function) που συνδυάζει:

1. Τη συνάρτηση κόστους $\ell(y_i, \hat{y}_i)$, η οποία εξαρτάται από το σφάλμα πρόβλεψης (π.χ. Mean Squared Error για παλινδρόμηση ή log-loss για ταξινόμηση).
2. Έναν όρο κανονικοποίησης (regularization) $\Omega(f_k)$, ο οποίος ελέγχει την πολυπλοκότητα των δέντρων και περιλαμβάνει:
 - ο Έναν όρο ανάλογο με τον αριθμό των φύλλων T (π.χ. γT), για να αποφεύγονται υπερβολικά μεγάλα δέντρα.
 - ο Όρους L1/L2 κανονικοποίησης για τα βάρη των φύλλων (leaf weights).

Η συνολική αντικειμενική συνάρτηση στο βήμα t γράφεται ως:

$$L(\phi) = \sum_{i=1}^n \ell(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t),$$

όπου $\hat{y}_i^{(t-1)}$ είναι η πρόβλεψη του μοντέλου μέχρι το βήμα $t - 1$, ενώ $f_t(x)$ είναι το νέο δέντρο που προστίθεται στο βήμα t . Ο όρος κανονικοποίησης $\Omega(f_t)$ διατυπώνεται συνήθως ως:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2,$$

όπου T είναι ο αριθμός των φύλλων, γ είναι υπερπαραμέτρος που επιβαρύνει το μοντέλο για κάθε επιπλέον φύλλο και λ είναι υπερπαραμέτρος που ελέγχει την L2 κανονικοποίηση στα βάρη $\{w_j\}$.

3. Δεύτερης Τάξης Προσέγγιση (Second-Order Approximation)

Ένα από τα σημεία-κλειδιά του XGBoost είναι ότι, προκειμένου να βρει το καλύτερο επόμενο δέντρο f_t , βασίζεται όχι μόνο στην πρώτη παράγωγο της συνάρτησης κόστους, αλλά και στη

δεύτερη παράγωγο. Με τον τρόπο αυτό, κατασκευάζεται μια τοπική (second-order) προσεγγιστική μορφή του σφάλματος, η οποία επιτρέπει ταχύτερους και ακριβέστερους υπολογισμούς.

1. Ορίζουμε τις παραγώγους της συνάρτησης κόστους για κάθε δείγμα i :

$$g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}$$

- ο g_i : βαθμίδα (gradient) της συνάρτησης κόστους για το δείγμα i .
 - ο h_i : καμπυλότητα (hessian) της συνάρτησης κόστους για το δείγμα i .
2. Χρησιμοποιώντας τον δεύτερο όρο Taylor επέκταση της L , το πρόβλημα βελτιστοποίησης ανάγεται στη μεγιστοποίηση μιας συνάρτησης κέρδους (gain) που περιλαμβάνει τους όρους g_i και h_i , συν τον όρο κανονικοποίησης $\Omega(f_t)$. Αν θεωρήσουμε ότι το δέντρο f_t έχει T φύλλα και κάθε φύλλο j αντιστοιχεί στο σύνολο δειγμάτων I_j , τότε η προσεγγιστική μορφή για το βήμα t είναι:

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i (f_t(x_i))^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

3. Το βάρος w_j του φύλλου j (δηλαδή η τιμή που αντιστοιχεί στο συγκεκριμένο φύλλο) βρίσκεται με το κλείσιμο τύπου (closed-form solution):

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

4. Η βελτίωση (gain) που προκύπτει από τη διάσπαση ενός κόμβου υπολογίζεται επίσης με αναλυτικούς τύπους βάσει του αθροίσματος των g_i και h_i των δειγμάτων που εμπίπτουν σε κάθε κλάδο, μειώνοντας σημαντικά τον χρόνο που απαιτείται για να βρεθεί η βέλτιστη διάσπαση.

Με τον τρόπο αυτό, ο XGBoost επιτυγχάνει:

- Εκμετάλλευση πληροφοριών δεύτερης τάξης (δεύτερη παράγωγος), με αποτέλεσμα καλύτερη σύγκλιση και πιο αξιόπιστους διαχωρισμούς στα δέντρα.
- Ταχύτητα και παραλληλοποίηση, χάρη στις αθροιστικές δομές δεδομένων που χρησιμοποιούνται για την υπολογιστικά αποδοτική αναζήτηση του βέλτιστου split.

4. Κανονικοποίηση (Regularization) και Έλεγχος της Πολυπλοκότητας

Η εισαγωγή των υπερπαραμέτρων γ (complexity cost) και λ (L2 regularization) στη συνάρτηση $\Omega(f_t)$ επιβάλλει ποινές (penalties) στα πολύπλοκα μοντέλα (δέντρα με πολλά φύλλα και μεγάλα βάρη). Επιπλέον, ο XGBoost υποστηρίζει και L1 κανονικοποίηση (παράμετρος α), η οποία μηδενίζει ορισμένα βάρη φύλλων, ενισχύοντας τη σποραδικότητα (sparsity) και οδηγώντας σε πιο αραιά μοντέλα που συχνά έχουν καλύτερη γενίκευση.

Αυτές οι τεχνικές κανονικοποίησης μειώνουν δραστικά το φαινόμενο του overfitting, καθιστώντας το XGBoost κατάλληλο ακόμη και για περιπτώσεις όπου ο αριθμός των γνωρισμάτων (features) και των δειγμάτων είναι μεγάλος.

5. Διαδικασία Αναζήτησης (Greedy Splitting) και Σημασία Σπανιότητας (Handling of Sparse Data)

1. Greedy Splitting: Σε κάθε επίπεδο δέντρου, ο αλγόριθμος αναζητά τη διάσπαση (split) που μεγιστοποιεί το κέρδος (gain). Η λογική είναι παρόμοια με αυτή των Δέντρων Αποφάσεων, με τη διαφορά ότι η μέτρηση του gain περιλαμβάνει και τους δευτέρους όρους παραγώγου (h_i).
2. Sparse Features: Ο XGBoost διαθέτει έναν καινοτόμο τρόπο χειρισμού των σπάνιων (sparse) χαρακτηριστικών, όπως απουσία τιμών (missing values) ή πολλά μηδενικά. Προτείνεται ένας μηχανισμός που επιτρέπει στο δέντρο να αποφασίσει «αυτόματα» ποια θα είναι η καλύτερη διακλάδωση για τα μηδενικά ή κενά δεδομένα. Αυτό βελτιώνει την ακρίβεια και την ταχύτητα σε περιβάλλοντα με αραιή πληροφορία.

6. Παραλληλοποίηση και Αποδοτικότητα Μνήμης

Ο XGBoost έχει υλοποιηθεί με τρόπο που αξιοποιεί στο μέγιστο τη διαθέσιμη υπολογιστική ισχύ:

- Χρήση πολλαπλών πυρήνων (multi-threading): Κάθε υπολογισμός gain για πιθανή διάσπαση μπορεί να εκτελεστεί παράλληλα, επιταχύνοντας σημαντικά την εκπαίδευση.
- Block structure & cache-awareness: Η αποθήκευση δεδομένων σε μορφή “blocks” βελτιστοποιεί την πρόσβαση στη μνήμη, μειώνοντας τις καθυστερήσεις από προσπέλαση σε πίνακες/χαρακτηριστικά.

Η σχεδίαση αυτή καθιστά τον XGBoost αποτελεσματικό τόσο σε μεγάλες βάσεις δεδομένων όσο και σε συστήματα περιορισμένων πόρων (π.χ. περιβάλλοντα cloud με επιλεγμένους υπολογιστικούς περιορισμούς).

Πλεονεκτήματα και Μειονεκτήματα

1. Πλεονεκτήματα:

- **Ταχύτητα:** Το XGBoost έχει σχεδιαστεί για να είναι γρήγορο και αποδοτικό, ακόμη και για μεγάλα σύνολα δεδομένων. Είναι βελτιστοποιημένο για επεξεργασία τόσο σε μονοπύρηνια όσο και σε πολυπύρηνια συστήματα, γεγονός που το καθιστά εξαιρετική επιλογή για εργασίες που απαιτούν γρήγορες προβλέψεις.
- **Υψηλή ακρίβεια:** Το XGBoost είναι γνωστό για την υψηλή του ακρίβεια. Λειτουργεί συνδυάζοντας πολλαπλά δέντρα αποφάσεων για να δημιουργήσει πιο ακριβείς προβλέψεις, καθιστώντας το αποτελεσματικό για εργασίες όπως η αναγνώριση εικόνων και ομιλίας, η επεξεργασία φυσικής γλώσσας και τα συστήματα συστάσεων.
- **Ευελιξία:** Το XGBoost είναι ένας ευέλικτος αλγόριθμος που μπορεί να χρησιμοποιηθεί για ποικίλες εργασίες μηχανικής μάθησης, όπως ταξινόμηση, παλινδρόμηση και κατάταξη. Είναι επίσης συμβατό με μια ευρεία γκάμα γλωσσών προγραμματισμού, όπως Python, R και Java.
- **Κανονικοποίηση:** Το XGBoost περιλαμβάνει τεχνικές κανονικοποίησης που βοηθούν στην αποφυγή του overfitting, το οποίο αποτελεί συχνό πρόβλημα στη μηχανική μάθηση. Χρησιμοποιεί έναν συνδυασμό κανονικοποίησης L1 και L2 για να μειώσει την πολυπλοκότητα του μοντέλου, οδηγώντας σε πιο αξιόπιστες και ακριβείς προβλέψεις.

2. Μειονεκτήματα:

- **Πολυπλοκότητα:** Το XGBoost είναι ένας πολύπλοκος αλγόριθμος που απαιτεί κάποιο επίπεδο τεχνικής εξειδίκευσης για να εφαρμοστεί και να βελτιστοποιηθεί αποτελεσματικά. Μπορεί να είναι δύσκολο να διαμορφωθούν και να ρυθμιστούν οι πολλές υπερπαραμέτροι που εμπλέκονται, γεγονός που μπορεί να το καθιστά χρονοβόρο στη χρήση.
- **Overfitting:** Παρόλο που το XGBoost περιλαμβάνει τεχνικές κανονικοποίησης για την αποφυγή του overfitting, είναι ακόμη δυνατό ο αλγόριθμος να προσαρμοστεί υπερβολικά στα δεδομένα εκπαίδευσης. Αυτό μπορεί να οδηγήσει σε ανακριβείς προβλέψεις για νέα δεδομένα.
- **Χρήση μνήμης:** Το XGBoost μπορεί να καταναλώσει πολλή μνήμη, ειδικά για μεγάλα σύνολα δεδομένων. Αυτό μπορεί να δυσκολέψει τη λειτουργία του σε υπολογιστές με περιορισμένη μνήμη, οδηγώντας σε χαμηλότερη απόδοση.

- ο **Έλλειψη διαφάνειας:** Το XGBoost συχνά θεωρείται ένας αλγόριθμος «μαύρου κουτιού», που σημαίνει ότι μπορεί να είναι δύσκολο να ερμηνευθεί και να κατανοηθεί πώς καταλήγει στις προβλέψεις του. Αυτό μπορεί να κάνει τη διαδικασία εντοπισμού σφαλμάτων και τη βελτίωση πιο δύσκολη.

Στο πλαίσιο της παρούσας εργασίας, ο XGBoost αποδείχθηκε ο αποτελεσματικότερος αλγόριθμος για το σενάριο πρόβλεψης επιπέδων γλυκόζης μετά το γεύμα στο κομμάτι της web εφαρμογής, παρουσιάζοντας το μικρότερο σφάλμα MAE (Mean Absolute Error) συγκριτικά με τους λοιπούς αλγορίθμους και γενικά τις καλύτερες μετρικές απόδοσης. Αυτή η υπεροχή μπορεί να αποδοθεί στην ικανότητα του XGBoost να εκμεταλλεύεται τους όρους της πρώτης και δεύτερης παραγωγού, σε συνδυασμό με το ρυθμιζόμενο επίπεδο κανονικοποίησης (regularization), επιτυγχάνοντας πιο ακριβή εκπαίδευση χωρίς σημαντική υπερ-προσαρμογή στα δεδομένα.

Με τα κατάλληλα ρυθμισμένα hyperparameters (π.χ. βάθος δέντρων, learning rate, regularization), το XGBoost μπορεί να προσαρμοστεί σε πληθώρα εφαρμογών, από ανταγωνιστικές προκλήσεις (Kaggle competitions) μέχρι παραγωγικά περιβάλλοντα σε επιχειρηματικές λύσεις, δίνοντας αξιόπιστα και ισχυρά μοντέλα πρόβλεψης.

3.2 Σενάρια Εκπαίδευσης

Στην παρούσα εργασία υλοποιήθηκαν δύο βασικά σενάρια εκπαίδευσης, τα οποία εφαρμόστηκαν σε δύο διαφορετικά σύνολα δεδομένων:

1. Dataset A (chatGLUCO Web App):

- ο Περιλαμβάνει **38 ασθενείς που ζουν κανονική καθημερινή ζωή (εκτός ΜΕΘ).**
- ο Οι τιμές καταγράφονται σε πραγματικό χρόνο μέσω της εφαρμογής chatGLUCO, η οποία συνδέεται με μια βάση δεδομένων και κάνει live ανάκτηση των μετρήσεων (μέσω Flask API).
- ο Ακολουθεί τα πρότυπα του *Sliding Scale Protocol* και της μεθόδου *Carbo Calculation*.
- ο Κύρια χαρακτηριστικά:
 - glucose_old (επίπεδα γλυκόζης πριν το γεύμα)
 - glucose_new (επίπεδα γλυκόζης μετά το γεύμα)

- food_carbo (ποσότητα υδατανθράκων στο γεύμα)
- insulin_dose (χορηγούμενη δόση ινσουλίνης)
- Δημογραφικά στοιχεία: age, gender, weight, diagnosis (Type I ή Type II)

2. Dataset B (Jupyter Notebook – Ασθενείς ΜΕΘ):

- Περιλαμβάνει **5659 ασθενείς**, οι οποίοι βρίσκονται **διασωληνωμένοι σε Μονάδα Εντατικής Θεραπείας (ΜΕΘ)**.
- Εφαρμόστηκε για την αρχική διερεύνηση των αλγορίθμων και αποτελεί σημαντικό τμήμα της διπλωματικής, καθώς καλύπτει μεγαλύτερο εύρος περιπτώσεων και συνθηκών.
- Δεν περιλαμβάνει στήλη food_carbo, καθώς τα γεύματα (αν υπάρχουν) σε ΜΕΘ είναι διαφορετικής μορφής και δεν ακολουθείται τυπικά το «συμβατικό» πλαίσιο διατροφής.
- Περιλαμβάνεται η στήλη TIME, η οποία αναπαριστά τον χρόνο (σε ώρες) που έχει περάσει από την εισαγωγή του ασθενούς στη ΜΕΘ.
- Κύρια χαρακτηριστικά (ενδεικτικά):
 - GLUCOSE_OLD (επίπεδα γλυκόζης σε προηγούμενη μέτρηση)
 - GLUCOSE_NEW (επίπεδα γλυκόζης στην επόμενη μέτρηση)
 - INSULIN_DOSE (χορηγούμενη δόση ινσουλίνης)
 - Δημογραφικά στοιχεία: AGE, GENDER, WEIGHT, DIAGNOSIS
 - TIME (χρόνος από την εισαγωγή του ασθενούς στη ΜΕΘ, σε sec)

Παρόλο που και στα δύο έργα ακολουθούνται δύο βασικά σενάρια εκπαίδευσης (Πρόβλεψη Επιπέδων Γλυκόζης και Πρόβλεψη Δόσης Ινσουλίνης), υπάρχουν διαφορές στα χαρακτηριστικά (features) που χρησιμοποιούνται, λόγω της διαφορετικής φύσης των δεδομένων. Στις επόμενες ενότητες περιγράφονται αναλυτικά τα σενάρια αυτά

3.2.1 Σενάριο 1: Πρόβλεψη Επιπέδων Γλυκόζης

Η δυνατότητα ακριβούς πρόβλεψης της μελλοντικής γλυκόζης σημαντική, καθώς δίνει τη δυνατότητα στον ιατρό ή στον ίδιο τον ασθενή να αξιολογεί την αποτελεσματικότητα της θεραπείας και να προσαρμόζει έγκαιρα τη φαρμακευτική αγωγή ή/και τις διατροφικές συνήθειες.

Στόχος της Πρόβλεψης

- Εκτίμηση της νέας τιμής γλυκόζης (glucose_new) βάσει της τιμής γλυκόζης πριν το γεύμα, της ποσότητας υδατανθράκων που καταναλώθηκε (food_carbo) και της χορηγούμενης δόσης ινσουλίνης (insulin_dose).
- Λαμβάνονται υπόψη επίσης χαρακτηριστικά όπως η ηλικία, το βάρος, το φύλο και το είδος διάγνωσης (Type I ή Type II).

Συλλογή και Προεπεξεργασία Δεδομένων

- Τα δεδομένα προέρχονται από καταγραφές επιπέδων γλυκόζης και περιλαμβάνουν παραμέτρους σχετικές με την καθημερινή ρουτίνα του ασθενούς.
- Η διαδικασία προεπεξεργασίας περιλαμβάνει τη διαχείριση ελλιπών τιμών, τη μετατροπή των κατηγορικών πεδίων (π.χ. φύλο, διάγνωση) σε αριθμητική μορφή, καθώς και μεθόδους κλιμάκωσης (π.χ. κανονικοποίηση ή τυποποίηση), με στόχο τη βέλτιστη απόδοση των μοντέλων.

Διαδικασία Εκπαίδευσης και Αξιολόγησης

- Το διαθέσιμο σύνολο δεδομένων διαχωρίζεται σε σύνολο εκπαίδευσης (80%) και σύνολο δοκιμής (20%), τηρώντας τη χρονολογική σειρά για τη διατήρηση της χρονικής συνέπειας.
- Χρησιμοποιείται ένα σύνολο αλγορίθμων που εκπαιδεύονται με τεχνικές βελτιστοποίησης υπερπαραμέτρων (Grid Search), ώστε να προσαρμόζονται βέλτιστα στα δεδομένα.
- Η αξιολόγηση των αποτελεσμάτων πραγματοποιείται μέσω δεικτών σφάλματος, όπως το Μέσο Απόλυτο Σφάλμα (MAE) και το Μέσο Τετραγωνικό Σφάλμα (MSE). Αυτές οι μετρικές επιτρέπουν τον προσδιορισμό του πόσο κοντά βρίσκονται οι προβλέψεις στα πραγματικά επίπεδα γλυκόζης.

Προσαρμογή Προβλέψεων για Εξατομίκευση

- Επιπρόσθετα, εφαρμόζεται ένας μηχανισμός που χρησιμοποιεί τις προηγούμενες αποκλίσεις κάθε ασθενούς, ώστε να γίνεται διόρθωση των μελλοντικών προβλέψεων.
- Πρακτικά, το μοντέλο «μαθαίνει» από τα λάθη του σε κάθε ασθενή και προσαρμόζει ανάλογα την τελική εκτίμηση για τη γλυκόζη, βελτιώνοντας έτσι περαιτέρω την ακρίβεια.

Dataset B (Jupyter Notebook - ΜΕΘ)

Στόχος της Πρόβλεψης

- Εκτίμηση της νέας τιμής γλυκόζης (GLUCOSE_NEW) βάσει της προηγούμενης τιμής γλυκόζης (GLUCOSE_OLD) και της χορηγούμενης δόσης ινσουλίνης (INSULIN_DOSE).
- Λαμβάνονται υπόψη επίσης χαρακτηριστικά όπως η ηλικία, το βάρος, το φύλο, το είδος διάγνωσης (Type I ή Type II) και το TIME (χρόνος σε sec από την εισαγωγή στη ΜΕΘ).

Συλλογή και Προεπεξεργασία Δεδομένων

- Τα δεδομένα προέρχονται από νοσοκομειακά αρχεία ασθενών σε ΜΕΘ, όπου είναι διαθέσιμες πολλαπλές καταγραφές γλυκόζης και ινσουλίνης σε διάφορες χρονικές στιγμές.
- Ακολουθείται αντίστοιχη διαδικασία προεπεξεργασίας (διαχείριση ελλειπών τιμών, κανονικοποίηση). Επίσης, δίνεται έμφαση στον έλεγχο ακραίων τιμών, καθώς σε ΜΕΘ μπορεί να υπάρξουν ιδιαίτερα υψηλές ή χαμηλές μετρήσεις.

Διαδικασία Εκπαίδευσης και Αξιολόγησης

- Το διαθέσιμο σύνολο δεδομένων διαχωρίζεται σε σύνολο εκπαίδευσης (80%) και σύνολο δοκιμής (20%), τηρώντας τη χρονολογική σειρά για τη διατήρηση της χρονικής συνέπειας.

- Χρησιμοποιείται ένα σύνολο αλγορίθμων που εκπαιδεύονται με τεχνικές βελτιστοποίησης υπερπαραμέτρων (Grid Search), ώστε να προσαρμόζονται βέλτιστα στα δεδομένα.
- Η αξιολόγηση των αποτελεσμάτων πραγματοποιείται μέσω δεικτών σφάλματος, όπως το Μέσο Απόλυτο Σφάλμα (MAE) και το Μέσο Τετραγωνικό Σφάλμα (MSE). Αυτές οι μετρικές επιτρέπουν τον προσδιορισμό του πόσο κοντά βρίσκονται οι προβλέψεις στα πραγματικά επίπεδα γλυκόζης.

3.2.2 Σενάριο 2: Πρόβλεψη Δόσης Ινσουλίνης

Η εκτίμηση της δόσης ινσουλίνης πριν το γεύμα ή πριν από μία περίοδο χορήγησης (π.χ. σε ΜΕΘ) είναι κρίσιμη για την αποφυγή υπογλυκαιμίας ή υπεργλυκαιμίας. Σκοπός είναι ένα εξατομικευμένο πλαίσιο θεραπείας που προσαρμόζεται στις ανάγκες κάθε ασθενούς.

Dataset A (chatGLUCO)

Στόχος της Πρόβλεψης

- Εκτίμηση της κατάλληλης δόσης ινσουλίνης (insulin_dose), με γνώμονα τη γλυκόζη πριν το γεύμα (glucose_old), την ποσότητα των υδατανθράκων (food_carbo), καθώς και άλλες πληροφορίες, όπως η ηλικία, το βάρος του ασθενούς, το φύλο και το είδος διάγνωσης (Type I ή Type II).
- Σκοπός είναι η αποφυγή τόσο της υπογλυκαιμίας όσο και της υπεργλυκαιμίας, προσφέροντας ένα πλαίσιο εξατομικευμένης θεραπείας.

Συλλογή και Προεπεξεργασία Δεδομένων

- Χρησιμοποιούνται καταγραφές δόσης ινσουλίνης που έχουν διασυνδεθεί με ιστορικά δεδομένα γλυκόζης, αλλά και δημογραφικές πληροφορίες κάθε ασθενούς.
- Η προεπεξεργασία περιλαμβάνει τον έλεγχο ορίων (π.χ. αποκλεισμός τιμών ινσουλίνης πολύ πέρα από ρεαλιστικές ποσότητες), τη διαχείριση ακραίων τιμών και την κλιμάκωση των χαρακτηριστικών.

Διαδικασία Εκπαίδευσης και Αξιολόγησης

- Όπως και στο πρώτο σενάριο, τα δεδομένα διαχωρίζονται σε εκπαίδευση (80%) και δοκιμή (20%), τηρώντας τη χρονολογική σειρά.
- Εκπαιδεύονται μοντέλα με διάφορες παραμετροποιήσεις, ώστε να διερευνηθεί η απόδοσή τους και να επιλεγεί η καταλληλότερη αρχιτεκτονική.
- Τα αποτελέσματα αξιολογούνται μετρώντας, μεταξύ άλλων, το Μέσο Απόλυτο Σφάλμα (MAE) και τη Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE). Τα συγκεκριμένα κριτήρια επιτρέπουν την ποσοτική αποτίμηση του σφάλματος στις προβλέψεις δόσης ινσουλίνης.

Προσαρμογή Προβλέψεων για Εξατομίκευση

- Εφαρμόζεται επίσης μηχανισμός που αξιοποιεί τις διαχρονικές αποκλίσεις ανά ασθενή, ώστε η τελική πρόβλεψη να προσαρμόζεται στο προσωπικό προφίλ και στις ανάγκες του.
- Με αυτόν τον τρόπο, οι προβλέψεις λαμβάνουν υπόψη τυχόν συστηματικές διαφορές που παρατηρούνται στον εκάστοτε ασθενή, οδηγώντας σε πιο αποτελεσματική αντιμετώπιση του διαβήτη.

Dataset B (Jupyter Notebook - ΜΕΘ)

Στόχος της Πρόβλεψης

- Εκτίμηση της κατάλληλης δόσης ινσουλίνης (INSULIN_DOSE), με γνώμονα την προηγούμενη γλυκόζη (GLUCOSE_OLD), καθώς και άλλες πληροφορίες, όπως η ηλικία, το βάρος του ασθενούς, το φύλο, το είδος διάγνωσης (Type I ή Type II) και το TIME.
- Σκοπός είναι η αποφυγή τόσο της υπογλυκαιμίας όσο και της υπεργλυκαιμίας, προσφέροντας ένα πλαίσιο εξατομικευμένης θεραπείας.

Συλλογή και Προεπεξεργασία Δεδομένων

- Σε ΜΕΘ, η χορήγηση ινσουλίνης μπορεί να διαφέρει ριζικά λόγω της κατάστασης του ασθενούς (διασωλήνωση, πολλαπλές μεταβολικές παράμετροι).

- Γίνεται αυστηρή διερεύνηση ακραίων τιμών, καθώς σε ορισμένες περιπτώσεις οι απαιτούμενες δόσεις μπορεί να είναι πολύ υψηλές ή πολύ χαμηλές σε σχέση με έναν μέσο διαβητικό ασθενή.

Διαδικασία Εκπαίδευσης και Αξιολόγησης

- Όπως και στο πρώτο σενάριο, τα δεδομένα διαχωρίζονται σε εκπαίδευση (80%) και δοκιμή (20%), τηρώντας τη χρονολογική σειρά.
- Εκπαιδεύονται μοντέλα με διάφορες παραμετροποιήσεις (Grid Search), ώστε να διερευνηθεί η απόδοσή τους και να επιλεγεί η καταλληλότερη αρχιτεκτονική.
- Τα αποτελέσματα αξιολογούνται μετρώντας, μεταξύ άλλων, το Μέσο Απόλυτο Σφάλμα (MAE) και τη Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE). Τα συγκεκριμένα κριτήρια επιτρέπουν την ποσοτική αποτίμηση του σφάλματος στις προβλέψεις δόσης ινσουλίνης.

Συνοψίζοντας, τα δύο προτεινόμενα σενάρια εκπαίδευσης συνιστούν έναν ενιαίο μηχανισμό πρόβλεψης για τους ασθενείς με διαβήτη, εστιάζοντας:

- **Στην πρόβλεψη επιπέδων γλυκόζης (Σενάριο 1).**
- **Στην πρόβλεψη της απαιτούμενης δόσης ινσουλίνης (Σενάριο 2).**

Η αξιοποίηση προηγμένων δεδομένων (δημογραφικών, διατροφικών, κλινικών και χρονικών) και η εφαρμογή σύγχρονων μεθόδων Μηχανικής Μάθησης προσφέρουν ένα προσωποποιημένο σύστημα προσαρμοσμένο στις πραγματικές ανάγκες των ασθενών. Οι μετρικές αξιολόγησης, σε συνδυασμό με τους μηχανισμούς προσαρμογής, ενισχύουν την ακρίβεια και τη σταθερότητα των προβλέψεων, και στις δύο υλοποιήσεις, καλύπτοντας ευρύ φάσμα εφαρμογών, από την καθημερινή αυτοδιαχείριση έως τις εξειδικευμένες μονάδες εντατικής θεραπείας.

4 Αξιολόγηση των Μοντέλων Μηχανικής Μάθησης

4.1 Μετρικές Αξιολόγησης Ακρίβειας

Οι μετρικές αυτές βοηθούν να αξιολογήσουμε πόσο καλά ή πόσο άστοχα προβλέπει ένα μοντέλο τις τιμές μιας μεταβλητής-στόχου. Στην περίπτωση μας της πρόβλεψης δόσεων ινσουλίνης και των επιπέδων γλυκόζης. Μικρότερες τιμές σφάλματος σε μετρικές όπως MAE, MSE και RMSE δείχνουν καλύτερη προσαρμογή του μοντέλου στα δεδομένα.

Παράλληλα, η τυπική απόκλιση (std) των αντίστοιχων τιμών (MAE std, MSE std, RMSE std) αποτυπώνει τη διασπορά των σφαλμάτων γύρω από τον μέσο όρο τους. Ο υψηλότερος συντελεστής R^2 (κοντά στο 1) υποδεικνύει ότι το μοντέλο εξηγεί καλύτερα τη διακύμανση της μεταβλητής-στόχου. Επιπλέον, οι στατιστικοί δείκτες p-value και T-statistic χρησιμεύουν για να διαπιστώσουμε την στατιστική σημαντικότητα των αποτελεσμάτων ή/και παραμέτρων που λαμβάνουμε. Τέλος, η μετρική TIR% (Time In Range) δείχνει το ποσοστό του χρόνου που τα επίπεδα γλυκόζης βρίσκονται εντός του επιθυμητού εύρους (80-180 mg/dL).

Παρακάτω ακολουθεί αναλυτική παρουσίαση και ερμηνεία κάθε μετρικής [21]:

- **MAE (Mean Absolute Error - Μέσο Απόλυτο Σφάλμα):**

Υπολογίζει τη μέση απόσταση μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Εάν R_{ui} είναι η πραγματική τιμή ινσουλίνης για έναν ασθενή u σε μία τιμή γλυκόζης i στο R_{test} και $\hat{R}_{ui}$ η πρόβλεψή του χρησιμοποιώντας το μοντέλο που προκύπτει από το R_{train} , τότε το MAE ορίζεται ως εξής:

$$MAE = \frac{1}{|R_{test}|} \sum_{R_{ui} \in R_{test}} |R_{ui} - \hat{R}_{ui}|$$

Η MAE δείχνει πόσο μακριά κατά μέσο όρο βρισκόμαστε από την πραγματική τιμή, χωρίς να μεγεθύνει τα μεγαλύτερα σφάλματα (καθώς δεν υψώνεται στο τετράγωνο). Έχει το ίδιο πεδίο μονάδων με την εκτιμώμενη μεταβλητή (π.χ. αν οι τιμές είναι σε μονάδες δόσεων ινσουλίνης, τότε και η MAE εκφράζεται σε μονάδες δόσεων ινσουλίνης). Επίσης, όσο μικρότερη είναι η MAE, τόσο πιο κοντά είναι το μοντέλο στις πραγματικές τιμές.

- **MSE (Mean Squared Error - Μέσο Τετραγωνικό Σφάλμα):**

Το Mean Squared Error (Μέσο Τετραγωνικό Σφάλμα) υπολογίζεται ως ο μέσος όρος των τετραγώνων των διαφορών μεταξύ των πραγματικών τιμών και των προβλεπόμενων. Μαθηματικά εκφράζεται ως:

$$MSE = \frac{1}{|R_{test}|} \sum_{R_{ui} \in R_{test}} (R_{ui} - \hat{R}_{ui})^2$$

Το MSE μας δίνει μια τιμή που αυξάνεται σημαντικά όταν εμφανίζονται μεγάλες αποκλίσεις (σφάλματα) ανάμεσα σε πραγματικές και προβλεπόμενες τιμές. Με άλλα λόγια, μεγεθύνει τα μεγάλα σφάλματα (δίνει μεγαλύτερη βαρύτητα), λόγω της ύπαρξης του τετραγώνου, κάτι που μπορεί να είναι επιθυμητό σε εφαρμογές όπου μεγάλες αποκλίσεις θεωρούνται πιο επικίνδυνες ή ανεπιθύμητες. Ωστόσο, ένα μειονέκτημα του MSE είναι ότι η μονάδα μέτρησης του είναι το τετράγωνο της μονάδας του εκτιμώμενου μεγέθους (π.χ. αν οι τιμές είναι σε μονάδες δόσεων ινσουλίνης, τότε το MSE εκφράζεται σε $(\text{μονάδες δόσεων ινσουλίνης})^2$), γεγονός που μπορεί να δυσχεραίνει την ερμηνεία. Ακόμα ένα σημαντικό μειονέκτημα είναι ότι είναι ευαίσθητο σε ακραίες τιμές (outliers).

- **RMSE (Root Mean Squared Error - Τετραγωνική Ρίζα του Μέσου Τετραγωνικού Σφάλματος):** Αποτελεί την τετραγωνική ρίζα του MSE, επαναφέροντας έτσι το μέγεθος του σφάλματος στην ίδια τάξη μεγέθους με την αρχική μεταβλητή:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{|R_{test}|} \sum_{R_{ui} \in R_{test}} (R_{ui} - \hat{R}_{ui})^2}$$

Με τον τρόπο αυτό, είναι ευκολότερο να ερμηνευτεί η απόκλιση των προβλέψεων από τις πραγματικές τιμές, καθώς η RMSE εκφράζεται στην ίδια μονάδα μέτρησης της μεταβλητής-στόχου.

Τόσο το MAE όσο και το RMSE εκφράζουν έναν δείκτη σφάλματος πρόβλεψης. Όσο χαμηλότερη είναι η τιμή τους, τόσο καλύτερη είναι η πρόβλεψη. Τα κύρια αποτελέσματά μας είναι η διαφορά στις μετρικές MAE και RMSE για τα σενάρια πρόβλεψης: (i) των επιπέδων γλυκόζης μετά το γεύμα και (ii) της δόσης ινσουλίνης, συγκρίνοντας τους αλγορίθμους Linear Regression, Random Forest και XGBoost.

- **Τυπική Απόκλιση (std):**

Η Τυπική Απόκλιση (std) αποτελεί έναν δείκτη διασποράς των δεδομένων γύρω από τον μέσο όρο τους. Παρέχει μια μέτρηση του πόσο αραιωμένα ή συμπυκνωμένα είναι τα δεδομένα σε σχέση με την κεντρική τους τάση.

Ο μαθηματικός τύπος για την τυπική απόκλιση είναι ο εξής:

$$std = \sqrt{\frac{1}{|R_{test}|} \sum_{R_{ui} \in R_{test}} (R_{ui} - \bar{R})^2}$$

όπου:

- R_{ui} : η πραγματική τιμή (π.χ. δόση ινσουλίνης ή επίπεδο γλυκόζης) για τον ασθενή u σε μία μέτρηση i ,
- $\bar{R}$: ο μέσος όρος των πραγματικών τιμών στο R_{test} ,
- $|R_{test}|$: το πλήθος των τιμών στο σύνολο ελέγχου (Test Set).

Σε εφαρμογές πρόβλεψης ινσουλίνης ή επιπέδων γλυκόζης, η std μπορεί να χρησιμοποιηθεί συμπληρωματικά προς τις μετρικές σφάλματος (π.χ. MAE, RMSE) ώστε να εξετάσουμε τη συνολική διασπορά των τιμών σφάλματος (ή των ίδιων των προβλέψεων). Μία χαμηλή τιμή std υποδηλώνει ότι οι εκτιμήσεις του μοντέλου συγκεντρώνονται κοντά σε έναν μέσο όρο (συνήθως τον πραγματικό), ενώ μία υψηλή τιμή υποδεικνύει μεγαλύτερη μεταβλητότητα και, δυνητικά, αυξημένη αβεβαιότητα ή ασταθή απόδοση. Με αυτό τον τρόπο, η γνώση της τυπικής απόκλισης συμβάλλει στην πληρέστερη αξιολόγηση της συνολικής συμπεριφοράς ενός μοντέλου, συμπληρώνοντας τη «στατιστική εικόνα» που προκύπτει από μετρικές όπως το MAE, το RMSE, τα Διαστήματα Εμπιστοσύνης και τη στατιστική ανάλυση (p-value, t-statistic).

- **R^2 (Συντελεστής Προσδιορισμού):**

Ο Συντελεστής Προσδιορισμού (R^2) δείχνει πόσο καλά το μοντέλο μπορεί να εξηγήσει τη διακύμανση των πραγματικών τιμών, δηλαδή πόσο κοντά βρίσκονται οι προβλέψεις στις πραγματικές τιμές. Στο πλαίσιο πρόβλεψης δόσεων ινσουλίνης (ή επιπέδων γλυκόζης), αν R_{ui} είναι η πραγματική τιμή (π.χ. δόση ινσουλίνης) για τον ασθενή u σε μία μέτρηση i στο σύνολο ελέγχου R_{test} και $\hat{R}_{ui}$ η αντίστοιχη πρόβλεψη του μοντέλου (που προκύπτει από το R_{train}), τότε ορίζουμε τον μέσο όρο των πραγματικών τιμών στο R_{test} ως:

$$\bar{R} = \frac{1}{|R_{test}|} \sum_{R_{ui} \in R_{test}} R_{ui},$$

και ο Συντελεστής Προσδιορισμού υπολογίζεται ως:

$$R^2 = 1 - \frac{\sum_{R_{ui} \in R_{test}} (R_{ui} - \hat{R}_{ui})^2}{\sum_{R_{ui} \in R_{test}} (R_{ui} - \bar{R})^2}$$

Μια τιμή του R^2 κοντά στο 1 υποδηλώνει ότι το μοντέλο εξηγεί μεγάλο ποσοστό της διακύμανσης των πραγματικών τιμών (δεδομένων), αποτελώντας ένδειξη καλής προσαρμογής. Αντίθετα, τιμές του R^2 κοντά στο 0 υποδηλώνουν ότι το μοντέλο δεν μπορεί σχεδόν καθόλου να εξηγήσει τη διακύμανση των δεδομένων. Επίσης, αρνητικές τιμές του R^2 υποδεικνύουν ότι το μοντέλο αποδίδει χειρότερα ακόμη και από μια απλή πρόβλεψη βάσει του μέσου όρου $\bar{R}$.

- **Στατιστική ανάλυση (T-statistic & p-value):**

Για τη στατιστική ανάλυση των πειραμάτων μας, χρησιμοποιήσαμε τη βιβλιοθήκη SciPy. Πραγματοποιήσαμε ένα διπλής όψεως t-test (two-tailed t-test) προκειμένου να διαπιστωθεί αν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών ινσουλίνης/γλυκόζης, και να ελεγχθεί αν ανήκουν στην ίδια κατανομή. Οι υποθέσεις πάνω στις οποίες βασίζεται η ανάλυση ήταν οι εξής:

- **H0:** Η διαφορά μεταξύ των μέσων τιμών των προβλεπόμενων και των πραγματικών τιμών ινσουλίνης/γλυκόζης δεν είναι στατιστικά σημαντική, συνεπώς οι μέσες τιμές θεωρούνται «ίσες».
- **H1:** Η διαφορά μεταξύ των μέσων τιμών των προβλεπόμενων και των πραγματικών τιμών ινσουλίνης/γλυκόζης είναι στατιστικά σημαντική, συνεπώς οι μέσες τιμές δεν θεωρούνται «ίσες» μεταξύ τους.

Αφού ορίσαμε τη μηδενική και την εναλλακτική υπόθεση, εξετάζουμε ποια από τις δύο ισχύει χρησιμοποιώντας το διπλής όψεως t-test, το οποίο ελέγχει τον βαθμό ομοιογένειας της διακύμανσης μεταξύ των δύο δειγμάτων. Εάν η τιμή p-value είναι μεγαλύτερη ή ίση με 0.05, αποδεχόμαστε τη μηδενική υπόθεση. Εάν η p-value είναι μικρότερη ή ίση με 0.05, απορρίπτουμε τη μηδενική υπόθεση και αποδεχόμαστε την εναλλακτική (H1).

- **Confidence Intervals:**

Τα Διαστήματα Εμπιστοσύνης (Confidence Intervals) παρέχουν ένα εύρος τιμών εντός του οποίου εκτιμάται ότι βρίσκεται η πραγματική τιμή μιας υπό εξέταση μετρικής (π.χ. MAE, MSE, RMSE), με ένα συγκεκριμένο επίπεδο εμπιστοσύνης, κατά κανόνα 95%. Στα αρχεία Excel που χρησιμοποιήθηκαν για την αξιολόγηση του μοντέλου, οι στήλες “CI FROM” και “CI TO” αντιστοιχούν στα κάτω και άνω όρια του εκάστοτε διαστήματος εμπιστοσύνης.

Ένα διάστημα εμπιστοσύνης 95% υποδηλώνει ότι, εάν επαναλαμβάναμε το πείραμα πολλές φορές σε δείγματα με παρόμοια στατιστικά χαρακτηριστικά, στο 95% αυτών των επαναλήψεων η πραγματική τιμή της μετρικής θα βρισκόταν εντός του συγκεκριμένου εύρους τιμών.

Ένα μικρό διάστημα εμπιστοσύνης σηματοδοτεί μικρότερη αβεβαιότητα στην εκτίμηση της μετρικής, υποδηλώνοντας ότι η τιμή της (π.χ. του MAE) είναι σχετικά σταθερή. Ένα μεγάλο διάστημα εμπιστοσύνης υποδηλώνει πιθανή αστάθεια της εκτίμησης ή/και ανεπαρκή αριθμό δεδομένων.

Η ενσωμάτωση διαστημάτων εμπιστοσύνης στις μετρικές αξιολόγησης παρέχει μια πληρέστερη εικόνα για την ακρίβεια των αποτελεσμάτων και τη στατιστική σημασία των διαφορών μεταξύ μοντέλων (π.χ. Linear Regression, Random Forest, XGBoost). Για παράδειγμα, αν τα 95% διαστήματα εμπιστοσύνης για τις τιμές RMSE δύο μοντέλων δεν επικαλύπτονται, μπορούμε να συμπεράνουμε με μεγάλη βεβαιότητα ότι τα μοντέλα παρουσιάζουν πραγματική διαφορά στην απόδοσή τους. Αντίθετα, όταν επικαλύπτονται σημαντικά, η διαφορά μπορεί να μην είναι στατιστικά σημαντική.

- **Time in Range (TIR):**

Ο Χρόνος Εντός Φυσιολογικού Εύρους (Time in Range, TIR) ορίζεται ως το ποσοστό των τιμών γλυκόζης που διατηρούνται εντός ενός αποδεκτού εύρους (π.χ., 80–180 mg/dL) κατά τη διάρκεια της πρόβλεψης. Η μετρική TIR εφαρμόζεται αποκλειστικά στο σενάριο πρόβλεψης επιπέδων γλυκόζης μετά το γεύμα καθώς μετρά το ποσοστό των προβλεπόμενων τιμών που βρίσκονται εντός ενός καθορισμένου φυσιολογικού εύρους (80–180 mg/dL).

Αντίθετα, στο σενάριο πρόβλεψης της δόσης ινσουλίνης, η TIR δεν υπολογίζεται, επειδή το ζητούμενο είναι η ίδια η τιμή της δόσης και όχι τα επίπεδα γλυκόζης που προκύπτουν έπειτα από αυτήν. Εν κατακλείδι, ο υπολογισμός του TIR σε επίπεδο συνόλου δεδομένων χρησιμεύει στη σύγκριση των αλγορίθμων πρόβλεψης γλυκόζης, αξιολογώντας την ικανότητά τους να διατηρούν το επίπεδο γλυκόζης του ασθενούς εντός των φυσιολογικών ορίων.

Συνολικά, ο συνδυασμός όλων αυτών των μετρικών παρέχει μια ολοκληρωμένη εικόνα της απόδοσης των μοντέλων, ενισχύοντας την αξιοπιστία των συμπερασμάτων. Επιπλέον, διευκολύνει τον σχεδιασμό και τη βελτιστοποίηση μοντέλων που ανταποκρίνονται με συνέπεια στις απαιτήσεις της πρόβλεψης του φαινομένου που μελετάται.

4.2 Συγκριτικά Αποτελέσματα

Στην υποενότητα αυτή παρατίθενται και αναλύονται τα αποτελέσματα που προέκυψαν από την εφαρμογή των τριών αλγορίθμων μηχανικής μάθησης (Linear Regression, Random Forest, XGBoost), καθώς και της κλασικής μεθόδου SSI (Sliding Scale Insulin), στα δύο σενάρια πρόβλεψης:

1. Πρόβλεψη των επιπέδων γλυκόζης.
2. Πρόβλεψη της δόσης ινσουλίνης.

Ακολουθεί η συγκριτική παρουσίαση των μετρικών απόδοσης που υπολογίστηκαν και η ερμηνεία των αποτελεσμάτων.

1. Πρόβλεψη των επιπέδων γλυκόζης

Model	MAE	Std	CI FROM	CI TO	RMSE	Std	CI FROM	CI TO	p-value	T-statistic	TIR%
LR	37.36	23.36	13.99	60.72	44.77	23.36	21.40	68.13	0.23	0.56	73.39
XGBoost	31.26	21.51	9.75	52.77	38.38	21.51	16.86	59.89	0.31	0.08	85.15
RF	31.08	21.35	9.72	52.43	38.18	21.35	16.82	59.53	0.33	0.07	83.09
SSI	32.25	24.09	8.16	56.35	40.74	24.09	16.63	64.83	0.22	1.77	91.10

Πίνακας 2: Συγκριτική ανάλυση προβλεπτικών μοντέλων με βάση το MAE και το RMSE σε 5659 ασθενείς για την πρόβλεψη επιπέδων γλυκόζης.

Σχολιασμός και Συμπεράσματα

Σφάλμα Πρόβλεψης (MAE, RMSE)

Τα μοντέλα Random Forest (RF) και XGBoost παρουσιάζουν παρόμοια και σχετικά χαμηλά σφάλματα (MAE γύρω στο 31 mg/dL και RMSE ~38 mg/dL), βελτιώνοντας την απόδοση έναντι της Γραμμικής Παλινδρόμησης (LR), η οποία εμφανίζει υψηλότερες τιμές (MAE=37.36, RMSE=44.77). Το μοντέλο SSI (Sliding Scale Insulin) εμφανίζει ενδιάμεσα σφάλματα (MAE=32.25, RMSE=40.74), ελαφρώς υψηλότερα από RF και XGBoost αλλά χαμηλότερα από τη Γραμμική Παλινδρόμηση.

TIR% (Time In Range)

Παρατηρείται ότι το SSI έχει το υψηλότερο ποσοστό TIR (91.10%), γεγονός που υποδηλώνει ότι διατηρεί τις προβλεπόμενες τιμές γλυκόζης περισσότερο χρόνο εντός του εύρους 80–180 mg/dL. Τα μοντέλα RF και XGBoost ακολουθούν με TIR=83.09% και 85.15% αντίστοιχα, ενώ η Γραμμική Παλινδρόμηση εμφανίζει σημαντικά χαμηλότερο TIR (73.39%).

Διαστήματα Εμπιστοσύνης (CI)

Τα διαστήματα εμπιστοσύνης (CI FROM–CI TO) σε MAE και RMSE δείχνουν ότι όλα τα μοντέλα έχουν συγκρίσιμη κλίμακα αβεβαιότητας. Ωστόσο, οι τιμές για το RF και το XGBoost είναι ελαφρώς πιο «συμπιεσμένες» και συνεπείς (μικρότερη διασπορά), κάτι που συνάδει με τη βελτιωμένη απόδοσή τους σε σχέση με τη Γραμμική Παλινδρόμηση.

Στατιστική Σημασία (p-value, T-statistic)

Οι τιμές p-value σε όλα τα μοντέλα είναι μεγαλύτερες από 0.05, γεγονός που δεν επιτρέπει την απόρριψη της μηδενικής υπόθεσης (H_0). Αυτό σημαίνει ότι, βάσει του t-test που πραγματοποιήθηκε, δεν προκύπτει στατιστικά σημαντική διαφορά ανάμεσα στις μέσες πραγματικές τιμές γλυκόζης και τις αντίστοιχες προβλέψεις (με το συγκεκριμένο δείγμα δεδομένων και τρόπο δοκιμής).

Συνολικά, για την πρόβλεψη των επιπέδων γλυκόζης, τα μοντέλα RF και XGBoost εμφανίζουν αρκετά ικανοποιητικές και παρόμοιες επιδόσεις όσον αφορά το σφάλμα (MAE, RMSE) και τη διασπορά τους. Ωστόσο, παρά το ελαφρώς μεγαλύτερο MAE, το SSI καταγράφει υψηλότερο Time In Range (TIR), υποδεικνύοντας ότι για αρκετές μετρήσεις διατηρεί το επίπεδο γλυκόζης σε πιο ασφαλή περιοχή.

1. Πρόβλεψη της δόσης ινσουλίνης

Model	MAE	Std	CI FROM	CI TO	RMSE	Std	CI FROM	CI TO
LR	1.89	1.15	0.73	3.04	2.25	1.15	1.09	3.40
XGBoost	1.86	1.25	0.61	3.11	2.28	1.25	1.03	3.53
RF	1.85	1.18	0.67	3.04	2.23	1.18	1.05	3.42
SSI	0.90	0.57	0.32	1.47	1.08	0.57	0.51	1.66

Πίνακας 3: Συγκριτική ανάλυση προβλεπτικών μοντέλων με βάση το MAE και το RMSE σε 5659 ασθενείς για την πρόβλεψη δόσης ινσουλίνης.

Σφάλμα Πρόβλεψης (MAE, RMSE)

Η μέθοδος SSI παρουσιάζει σαφώς καλύτερες επιδόσεις στην πρόβλεψη της δόσης ινσουλίνης, καταγράφοντας $MAE=0.90$ και $RMSE=1.08$ — τιμές πολύ χαμηλότερες συγκριτικά με τα μοντέλα ML ($MAE>1.80$, $RMSE>2.20$). Τα τρία μοντέλα μηχανικής μάθησης (LR, XGBoost, RF) εμφανίζουν παρόμοια αποτελέσματα και υψηλότερο σφάλμα, γεγονός που υποδηλώνει ότι, στο συγκεκριμένο dataset, η κλασική μέθοδος SSI ανταποκρίνεται με μεγαλύτερη ακρίβεια για την εκτίμηση της δόσης ινσουλίνης.

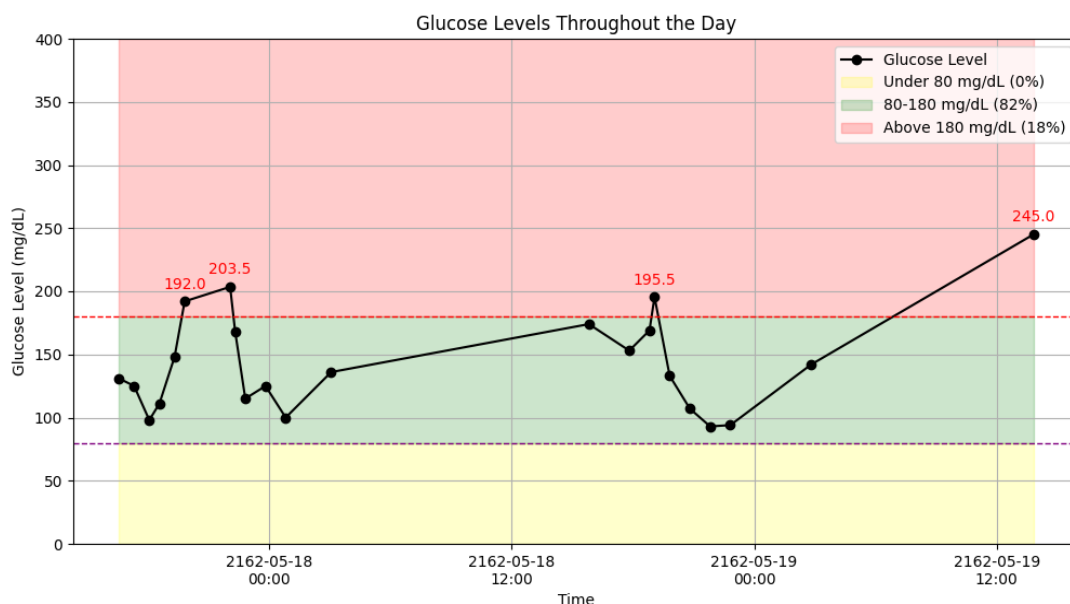
Διαστήματα Εμπιστοσύνης (CI)

Τα διαστήματα εμπιστοσύνης για τη μέθοδο SSI είναι αρκετά στενά (0.32–1.47 για το MAE), επιβεβαιώνοντας τη χαμηλή αβεβαιότητα και την υψηλή σταθερότητα των προβλέψεων. Αντίθετα, τα διαστήματα εμπιστοσύνης στις μετρικές των μοντέλων ML είναι εμφανώς μεγαλύτερα (π.χ. CI for MAE έως 3.04), καταδεικνύοντας μεγαλύτερη διασπορά στα σφάλματα πρόβλεψης.

Λαμβάνοντας υπόψη το χαμηλό σφάλμα πρόβλεψης και τη στενή διακύμανσή του, η μέθοδος SSI αποδεικνύεται πιο κατάλληλη στο δεύτερο σενάριο (πρόβλεψη δόσης ινσουλίνης). Τα μοντέλα ML δεν εμφανίζουν σημαντικές διαφορές μεταξύ τους, αλλά συνολικά υπολείπονται της SSI ως προς το σφάλμα πρόβλεψης.

4.3 Toy Example για τον ασθενή 533

Στην ενότητα αυτή παρουσιάζουμε ένα απλό παράδειγμα (Toy Example) εφαρμογής της προτεινόμενης μεθοδολογίας στον ασθενή με SUBJECT_ID = 533, ο οποίος είναι 61 ετών και έχει βάρος 115 κιλά, νοσηλευόμενος στη Μονάδα Εντατικής Θεραπείας (ΜΕΘ). Επιλέχθηκε ο συγκεκριμένος ασθενής, καθώς διαθέτουμε αρκετές καταγραφές επιπέδων γλυκόζης και δόσεων ινσουλίνης, επιτρέποντας να επιδείξουμε βήμα-βήμα τη μεθοδολογία σε ένα μικρό, ελεγχόμενο δείγμα.

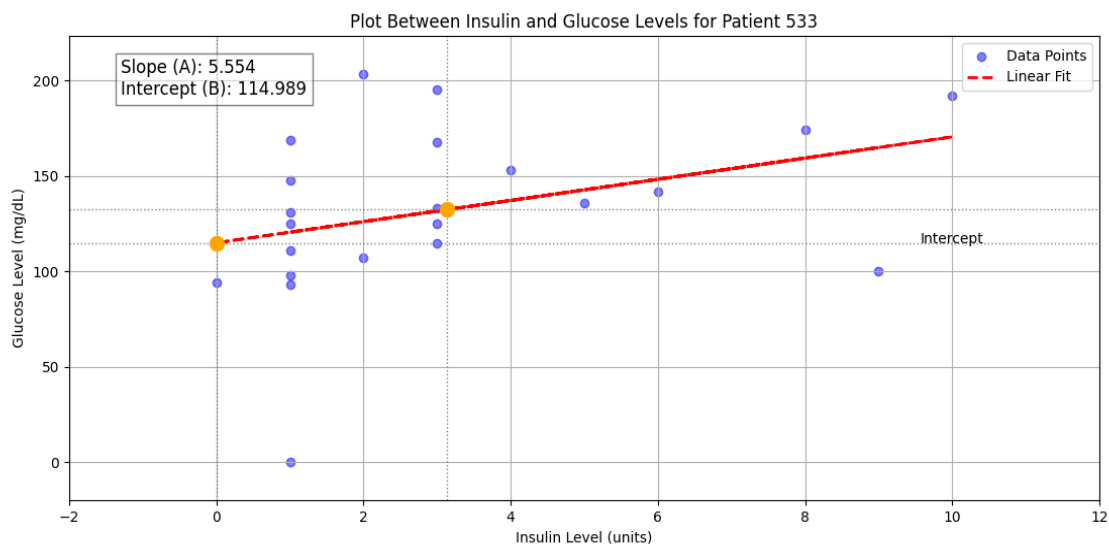


Σχήμα 4: Επίπεδα Γλυκόζης Κατά τη Διάρκεια της Ημέρας για τον ασθενή 533

Στο πρώτο διάγραμμα (Σχήμα 4), απεικονίζονται οι πραγματικές μετρήσεις γλυκόζης του ασθενούς καθ' όλη τη διάρκεια της ημέρας. Οι τιμές χωρίζονται ενδεικτικά σε τρεις ζώνες:

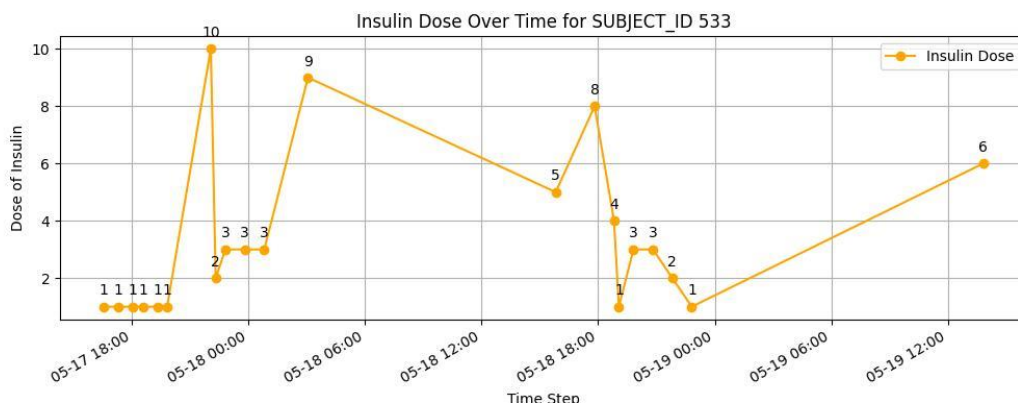
- **Κάτω από 80 mg/dL (κίτρινη):** Υπογλυκαιμικές τιμές, όπου απαιτείται προσοχή.
- **80–180 mg/dL (πράσινη):** Φυσιολογικό ή στόχος ρύθμισης.
- **Πάνω από 180 mg/dL (κόκκινη):** Υπεργλυκαιμικές τιμές που ενδέχεται να απαιτούν παρέμβαση.

Παρατηρείται ότι ο ασθενής περνά αρκετό διάστημα εντός του επιθυμητού εύρους (πράσινη ζώνη), ωστόσο σημειώνονται και περίοδοι υπεργλυκαιμίας (κόκκινη ζώνη), με αιχμές που αναδεικνύουν τη δυναμική μεταβολή της γλυκόζης σε συνθήκες ΜΕΘ.



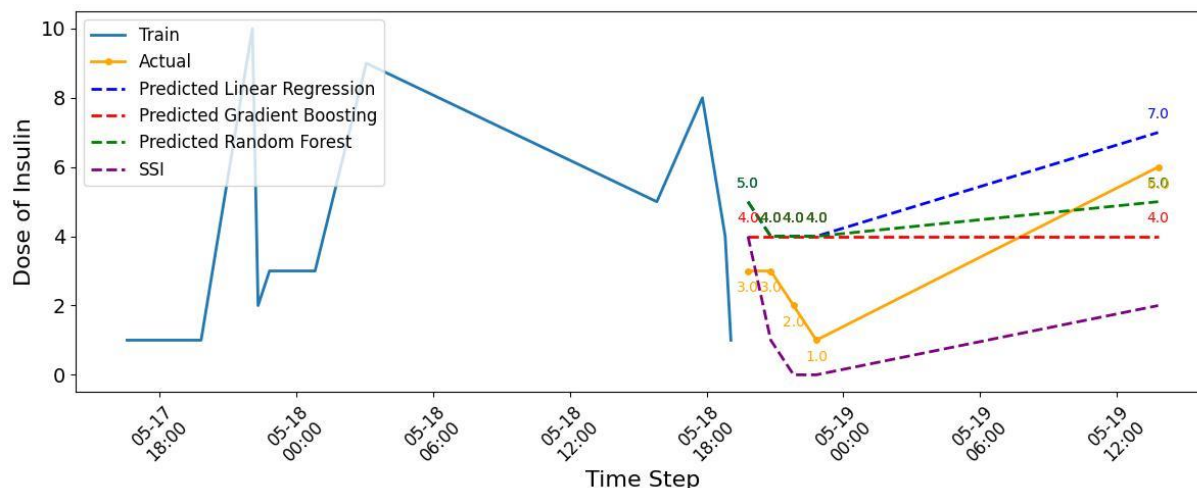
Σχήμα 5: Σχέση Επιπέδων Ινσουλίνης και Γλυκόζης για Ασθενή 533

Στο επόμενο διάγραμμα (Σχήμα 5) εξετάζεται η σχέση μεταξύ των επιπέδων ινσουλίνης που χορηγούνται και των αντίστοιχων επιπέδων γλυκόζης. Παρουσιάζεται και η γραμμική παλινδρόμηση, η οποία έχει κλίση (slope) = 5.554 και τομή στον άξονα (intercept) = 114.989. Παρότι θεωρητικά αναμένεται αρνητική συσχέτιση (όσο περισσότερη ινσουλίνη χορηγείται, τόσο πιο χαμηλή θα είναι η γλυκόζη), το παρατηρούμενο θετικό slope μπορεί να οφείλεται στο ότι οι δόσεις ινσουλίνης αυξάνονται όταν ήδη ανιχνευτούν υψηλές τιμές γλυκόζης, οπότε η απλή γραμμική παλινδρόμηση αποτυπώνει συσχέτιση αλλά όχι αιτιατή επίδραση. Επιπλέον, παράγοντες όπως τα γεύματα ή άλλες φαρμακευτικές αγωγές πιθανόν επηρεάζουν σημαντικά τη γλυκόζη, κάνοντας τη σχέση μη μονοδιάστατη.



Σχήμα 6: Χορήγηση Ινσουλίνης με την Πάροδο του Χρόνου για τον Ασθενή 533

Ιδιαίτερα ενδιαφέρον είναι το χρονικό διάγραμμα της ινσουλίνης (Σχήμα 6), όπου βλέπουμε στα πρώτα στάδια μικρές δόσεις (π.χ. 1 μονάδα) οι οποίες αυξάνονται απότομα σε υψηλότερα επίπεδα (π.χ. 10 μονάδες) καθώς εξελίσσεται η θεραπεία. Στη συνέχεια ακολουθούν αυξομειώσεις, δείχνοντας τη δυναμική ρύθμιση της ινσουλίνης σε πραγματικό χρόνο, καθώς το ιατρικό προσωπικό προσαρμόζει διαρκώς τη θεραπεία με βάση τις τρέχουσες ενδείξεις και τη συνολική κλινική κατάσταση του ασθενούς.



Σχήμα 7: Σύγκριση Προβλέψεων Δόσεων Ινσουλίνης με Πραγματικά Δεδομένα για τον Ασθενή 533

Στο τελευταίο διάγραμμα, απεικονίζονται οι πραγματικές δόσεις ινσουλίνης του ασθενούς (πορτοκαλί γραμμή) σε σχέση με τις προβλεπόμενες δόσεις από τους τρεις διαφορετικούς αλγορίθμους μηχανικής μάθησης (Linear Regression, XGBoost, Random Forest) και του Sliding Scale Insulin - SSI. Παρατηρούμε ότι οι προβλέψεις των αλγορίθμων αποκλίνουν σε κάποιο βαθμό από τις πραγματικές τιμές, ιδιαίτερα στις περιπτώσεις αιχμών ή απότομων μεταβολών. Το Sliding Scale Insulin (SSI) φαίνεται να ακολουθεί καλύτερα τη γενική τάση των πραγματικών δεδομένων, αν και σε μερικές περιπτώσεις οι πιο σύνθετοι αλγόριθμοι, όπως το XGBoost ή το Random Forest, παρουσιάζουν μικρότερες αποκλίσεις σε συγκεκριμένα χρονικά σημεία. Η σύγκριση αναδεικνύει τις προκλήσεις στην ακριβή πρόβλεψη των δόσεων ινσουλίνης σε δυναμικά και πολυπαραγοντικά περιβάλλοντα, όπως μια ΜΕΘ. Αξιολογώντας την επίδοση μέσω μετρικών σφάλματος (π.χ. MAE, RMSE) μπορεί να δοθεί μια πιο ποσοτική εικόνα της ακρίβειας κάθε μεθόδου.

5 Ανάπτυξη Εφαρμογής

5.1 Ανάλυση Απαιτήσεων

Η εφαρμογή chatGLUCO έχει σχεδιαστεί για να προσφέρει ένα ολοκληρωμένο σύστημα διαχείρισης και πρόβλεψης δόσεων ινσουλίνης και επιπέδων γλυκόζης, παρέχοντας ένα φιλικό προς τον χρήστη περιβάλλον με λειτουργίες που ενισχύουν τη διαχείριση του διαβήτη. Οι βασικές λειτουργικές απαιτήσεις περιγράφονται παρακάτω.

5.1.1 Λειτουργικές Απαιτήσεις

Διαχείριση Χρηστών

Απαίτηση 1: Εγγραφή νέου χρήστη στο σύστημα (Sign Up)

Περιγραφή: Ο χρήστης έχει τη δυνατότητα να εγγραφεί στην εφαρμογή συμπληρώνοντας τα παρακάτω υποχρεωτικά πεδία:

- First Name
- Last Name
- Email
- Date of Birth
- Gender
- Password
- Confirm Password

Εάν κάποια από τις παραπάνω πληροφορίες είναι ελλιπής ή μη έγκυρη, το σύστημα εμφανίζει μηνύματα λάθους, υποδεικνύοντας τι πρέπει να διορθωθεί. Με την επιτυχή εγγραφή, ο χρήστης ανακατευθύνεται στο Dashboard, όπου μπορεί να περιηγηθεί στις λειτουργίες της εφαρμογής ή να αποσυνδεθεί μέσω του Logout.

Απαίτηση 2: Σύνδεση χρήστη στο σύστημα (Log In)

Περιγραφή: Ο χρήστης έχει τη δυνατότητα να συνδεθεί στην εφαρμογή εισάγοντας τα παρακάτω στοιχεία:

- Email

- Password

Σε περίπτωση λανθασμένων διαπιστευτηρίων, εμφανίζεται μήνυμα λάθους. Με την επιτυχή σύνδεση, ο χρήστης μεταφέρεται στην αρχική σελίδα (Dashboard) με δυνατότητα αποσύνδεσης (Logout).

Απαίτηση 3: Είδη χρηστών της εφαρμογής

Περιγραφή: Η εφαρμογή υποστηρίζει μόνο έναν τύπο χρήστη, ο οποίος είναι ο ασθενής (Patient). Κάθε χρήστης έχει πρόσβαση στις ίδιες λειτουργίες της εφαρμογής και μπορεί να διαχειρίζεται τα προσωπικά του δεδομένα, το ημερολόγιο και τις προβλέψεις.

Διαχείριση Προσωπικών Στοιχείων

Απαίτηση 4: Τροποποίηση Υγειονομικών Στοιχείων (Health Information)

Περιγραφή: Ο χρήστης μπορεί να επεξεργαστεί τα παρακάτω πεδία στην ενότητα Health Information:

- Diagnosis: Επιλογή μεταξύ Type I και Type II.
- Weight: Εισαγωγή του σωματικού βάρους (αριθμητική τιμή).
- Latest HbA1c, Second Latest HbA1c, Third Latest HbA1c: Εισαγωγή των τριών τελευταίων μετρήσεων HbA1c (%), όπου:
 - Latest HbA1c: Η πιο πρόσφατη μέτρηση.
 - Second Latest HbA1c: Η προτελευταία μέτρηση.
 - Third Latest HbA1c: Η προ-προτελευταία μέτρηση.

Κατά την εισαγωγή τιμών HbA1c, εμφανίζεται κάτω από το πεδίο το αντίστοιχο ισοδύναμο της μέσης γλυκόζης (mg/dL) για διευκόλυνση του χρήστη. Η τιμή χρωματίζεται ανάλογα με την κατηγορία (π.χ., πράσινο για φυσιολογικά επίπεδα, πορτοκαλί για προδιαβήτη, κόκκινο για διαβήτη).

Αρχικά, εμφανίζεται προειδοποιητικό μήνυμα στην κορυφή της σελίδας που ενημερώνει τον χρήστη να συμπληρώσει το Weight και το Diagnosis, καθώς είναι απαραίτητα για τη χρήση των προβλέψεων. Οι αλλαγές αποθηκεύονται μέσω του κουμπιού “Save Changes”.

Απαίτηση 5: Αλλαγή Εικόνας Προφίλ

Περιγραφή: Ο χρήστης μπορεί να ανεβάσει ή να αντικαταστήσει την εικόνα του προφίλ μέσω του κουμπιού “Add Photo”. Με την αποθήκευση της νέας εικόνας μέσω του “Save”, εμφανίζεται μήνυμα επιβεβαίωσης επιτυχίας.

Απαίτηση 6: Αλλαγή Κωδικού Πρόσβασης

Περιγραφή: Ο χρήστης έχει τη δυνατότητα να αλλάξει τον κωδικό πρόσβασης συμπληρώνοντας τα εξής:

- Current Password
- New Password
- Confirm New Password

Σε περίπτωση σφάλματος (π.χ., λάθος κωδικός ή μη ταυτιζόμενοι νέοι κωδικοί), εμφανίζεται κατάλληλο μήνυμα λάθους και ο χρήστης πρέπει να διορθώσει τα πεδία. Αν η διαδικασία είναι επιτυχής, ο χρήστης αποσυνδέεται αυτόματα και μεταφέρεται στη σελίδα σύνδεσης, όπου εμφανίζεται μήνυμα επιτυχίας ότι ο κωδικός πρόσβασης άλλαξε επιτυχώς. Ο χρήστης πρέπει να συνδεθεί ξανά με τον νέο κωδικό.

Απαίτηση 7: Διαγραφή Λογαριασμού

Περιγραφή: Ο χρήστης μπορεί να διαγράψει οριστικά τον λογαριασμό του εισάγοντας τον τρέχοντα κωδικό πρόσβασης για επιβεβαίωση. Πριν τη διαγραφή, εμφανίζεται προειδοποιητικό μήνυμα ότι η ενέργεια είναι μη αναστρέψιμη. Αν εισαχθεί λανθασμένος κωδικός, εμφανίζεται κατάλληλο μήνυμα λάθους και η διαδικασία διαγραφής πρέπει να επαναληφθεί. Με την επιτυχή ολοκλήρωση της διαγραφής, όλα τα προσωπικά δεδομένα και οι εγγραφές ημερολογίου διαγράφονται από το σύστημα, και εμφανίζεται μήνυμα επιβεβαίωσης που ενημερώνει τον χρήστη ότι ο λογαριασμός διαγράφηκε με επιτυχία.

Διαχείριση Ημερολογίου (Diary)

Απαίτηση 8: Δημιουργία νέας εγγραφής

Περιγραφή: Ο χρήστης έχει τη δυνατότητα να προσθέσει μια νέα εγγραφή ημερολογίου μέσω της φόρμας "Add New Record" συμπληρώνοντας τα παρακάτω πεδία:

- Date & Time: Αυτόματη εμφάνιση της τρέχουσας ημερομηνίας και ώρας με δυνατότητα τροποποίησης της.

- **Glucose Before (glucose_old):** Υποχρεωτικό πεδίο για την εισαγωγή τιμής γλυκόζης πριν το γεύμα. Το πεδίο διαθέτει έγχρωμη κωδικοποίηση για την άμεση κατανόηση των επιπέδων:
 - Κίτρινο χρώμα: Για τιμές <80 mg/dL (χαμηλά επίπεδα - υπογλυκαιμία).
 - Πράσινο χρώμα: Για τιμές μεταξύ 80-180 mg/dL (φυσιολογικά επίπεδα).
 - Κόκκινο χρώμα: Για τιμές >180 mg/dL (υψηλά επίπεδα - υπεργλυκαιμία).
- **Food Carbohydrates (food_carbo):** Προαιρετικό πεδίο για την εισαγωγή υδατανθράκων σε γραμμάρια. Η τιμή μπορεί να υπολογιστεί αυτόματα μέσω του εργαλείου αναζήτησης φαγητών.
- **Insulin Dose in Units (insulin_dose):** Προαιρετικό πεδίο για τη συμπλήρωση δόσης ινσουλίνης. Το πεδίο μπορεί να λάβει αυτόματα τιμή από τον υπολογισμό των αλγορίθμων πρόβλεψης ή του sliding scale protocol + carbo calculation.
- **Glucose After (glucose_new):** Προαιρετική εισαγωγή τιμής γλυκόζης μετά το γεύμα. Το πεδίο διαθέτει την ίδια έγχρωμη κωδικοποίηση με το πεδίο Glucose Before για εύκολη κατανόηση.

Υπολογισμός Δόσης Ινσουλίνης:

Ο χρήστης μπορεί να επιλέξει αλγόριθμο για την πρόβλεψη της δόσης ινσουλίνης από το dropdown menu με τις παρακάτω επιλογές:

1. Linear Regression
2. XGBoost
3. Random Forest
4. Sliding Scale + Carbo Calculation

Το σύστημα υπολογίζει τη δόση ινσουλίνης και εμφανίζει την τιμή στο πεδίο Insulin Dose. Ο χρήστης έχει τη δυνατότητα να απορρίψει την προτεινόμενη τιμή μέσω του X εικονιδίου που εμφανίζεται δίπλα στο πεδίο και να εισάγει δική του τιμή.

Υπολογισμός Υδατανθράκων:

Ο χρήστης μπορεί να αναζητήσει φαγητά μέσω του εργαλείου "Search Food". Αυτή η λειτουργία βασίζεται στο Nutritionix API για την αναζήτηση τροφών και τον αυτόματο υπολογισμό των υδατανθράκων. Πληκτρολογώντας το όνομα του φαγητού (π.χ. apple) και επιλέγοντας το είδος και την ποσότητα, το σύστημα υπολογίζει αυτόματα τους υδατάνθρακες και συμπληρώνει το πεδίο Food Carbo in Grams. Ο χρήστης μπορεί να

απορρίψει την προτεινόμενη τιμή υδατανθράκων μέσω του X εικονιδίου και να συμπληρώσει δική του τιμή.

Επεξεργασία Πεδίων:

Όλα τα παραπάνω πεδία (Date & Time, Food Carbo, Insulin Dose, Glucose After) είναι προαιρετικά και επεξεργάσιμα. Ο χρήστης έχει τη δυνατότητα να αποδεχθεί ή να απορρίψει αυτόματα υπολογισμένες τιμές και να εισάγει τις δικές του μετρήσεις. Το μόνο υποχρεωτικό πεδίο είναι το Glucose Before.

Αποθήκευση Εγγραφής:

Μετά τη συμπλήρωση του Glucose Before και την προαιρετική εισαγωγή τιμών στα υπόλοιπα πεδία, ο χρήστης μπορεί να αποθηκεύσει τη νέα εγγραφή μέσω του κουμπιού "Save".

Απαίτηση 9: Ενημέρωση υφιστάμενων εγγραφών

Περιγραφή: Ο χρήστης μπορεί να ενημερώσει μία υπάρχουσα εγγραφή ημερολογίου, τροποποιώντας οποιοδήποτε από τα διαθέσιμα πεδία (Date & Time, Glucose Before, Food Carbohydrates, Insulin Dose, Glucose After). Η διαδικασία ενημέρωσης περιλαμβάνει τις ίδιες δυνατότητες με τη δημιουργία νέας εγγραφής, όπως πρόβλεψη δόσης ινσουλίνης, αναζήτηση φαγητού, αυτόματους υπολογισμούς, απόρριψη τιμών και έγχρωμη κωδικοποίηση των επιπέδων γλυκόζης.

Απαίτηση 10: Διαγραφή εγγραφών

Περιγραφή: Ο χρήστης μπορεί να διαγράψει μία εγγραφή ημερολογίου πατώντας το εικονίδιο του κάδου απορριμμάτων. Πριν τη διαγραφή, εμφανίζεται pop-up παράθυρο με μήνυμα επιβεβαίωσης.

Απαίτηση 11: Εξαγωγή δεδομένων

Περιγραφή: Ο χρήστης μπορεί να εξάγει όλες τις εγγραφές του ημερολογίου σε αρχείο XLS, επιτρέποντας τη διατήρηση αρχείου των δεδομένων.

Διαχείριση Διαγραμμάτων (Diagrams)

Απαίτηση 12: Οπτικοποίηση Δεδομένων (Diagrams)

Περιγραφή: Ο χρήστης μπορεί να επιλέξει μεταξύ δύο διαγραμμάτων:

- Time in Range
- Glucose-Insulin Correlation

Μπορεί να ορίσει συγκεκριμένες ημερομηνίες για την εμφάνιση των δεδομένων και να φορτώσει το διάγραμμα μέσω του κουμπιού “Load Diagram”.

Διαχείριση Πρόβλεψης Γλυκόζης

Απαίτηση 13: Δυνατότητα Πρόβλεψης Γλυκόζης

Περιγραφή: Η εφαρμογή επιτρέπει στον χρήστη να προβλέψει την τιμή των επιπέδων γλυκόζης μετά την κατανάλωση υδατανθράκων και τη χορήγηση ινσουλίνης. Η διαδικασία της πρόβλεψης περιλαμβάνει τα παρακάτω βήματα:

1. Συμπλήρωση δεδομένων εισόδου:

- Ο χρήστης εισάγει την τιμή του Current Glucose Level (Glucose Before) και την δόση ινσουλίνης που χορηγήθηκε (Insulin Dose (units)), τα οποία αποτελούν υποχρεωτικά πεδία.
- Προαιρετικά, ο χρήστης μπορεί να συμπληρώσει τα πεδία:
 - Food Carbohydrates (grams): Υδατάνθρακες που καταναλώθηκαν.

2. Επιλογή τρόπου υπολογισμού (Calculation Mode):

- Ο χρήστης επιλέγει μεταξύ δύο διαθέσιμων μεθόδων:
 - Postprandial Glucose Calculation: Αυτή η μέθοδος προβλέπει τα επίπεδα γλυκόζης μετά από γεύμα, λαμβάνοντας υπόψη την αρχική γλυκόζη, την αύξηση λόγω υδατανθράκων (CR) και τη μείωση από τη χορηγούμενη ινσουλίνη (ISF).
 - Machine Learning-Based Glucose Prediction: Χρησιμοποιεί έναν από τους διαθέσιμους αλγόριθμους μηχανικής μάθησης για την πρόβλεψη. Οι διαθέσιμοι αλγόριθμοι είναι:
 - Linear Regression
 - Random Forest
 - XGBoost

3. Προβολή αποτελέσματος:

- Ο χρήστης πατάει το κουμπί "Predict Glucose" για να ξεκινήσει ο υπολογισμός.
- Η προβλεπόμενη τιμή Glucose After εμφανίζεται σε ένα πλαίσιο στην κορυφή της σελίδας με την εξής έγχρωμη κωδικοποίηση:
 - Κίτρινο χρώμα: Glucose Levels κάτω από 80 mg/dL (Low).
 - Πράσινο χρώμα: Glucose Levels μεταξύ 80 και 180 mg/dL (Normal).
 - Κόκκινο χρώμα: Glucose Levels πάνω από 180 mg/dL (High).
- Αυτή η χρωματική κατηγοριοποίηση βοηθά τον χρήστη να κατανοήσει άμεσα τα αποτελέσματα και να λάβει τα κατάλληλα μέτρα αν χρειαστεί.

4. Εκκαθάριση πεδίων:

- Ο χρήστης έχει τη δυνατότητα να διαγράψει όλα τα δεδομένα που έχουν εισαχθεί πατώντας το κουμπί "Clear Fields".

Απαίτηση 14: Προβολή Μελλοντικής Τάσης Γλυκόζης (Future Glucose Trend Prediction)

Περιγραφή: Ο χρήστης μπορεί να μεταβεί στη σελίδα "Future Glucose Trend Prediction" μέσω του μενού της εφαρμογής, όπου εμφανίζεται ένα διάγραμμα με τις εξής πληροφορίες:

1. Ιστορικές τιμές γλυκόζης (Historical Glucose) σε μορφή συνεχούς (μπλε) γραμμής.
2. Προβλεπόμενες μελλοντικές τιμές γλυκόζης (Predicted Glucose) για τις επόμενες δύο ημέρες, σε τρεις χρονικές στιγμές ανά ημέρα (πρωί, μεσημέρι, βράδυ). Οι τιμές αυτές αποδίδονται με διακεκομμένη πορτοκαλί γραμμή που ξεκινά ακριβώς μετά το τελευταίο ιστορικό δεδομένο.

Οι προβλεπόμενες τιμές παράγονται μέσω ενός Flask API, το οποίο χρησιμοποιεί ιστορικά δεδομένα του χρήστη και αλγόριθμο μηχανικής μάθησης. Το διάγραμμα δίνει στον χρήστη μια συνολική εικόνα της τρέχουσας κατάστασης και της μελλοντικής τάσης των επιπέδων γλυκόζης του.

5.1.2 Μη Λειτουργικές Απαιτήσεις

Οι μη λειτουργικές απαιτήσεις της εφαρμογής chatGLUCO καθορίζουν τα βασικά χαρακτηριστικά και τους περιορισμούς της, εστιάζοντας στην απόδοση, την ασφάλεια, τη χρηστικότητα και τη συντηρησιμότητά της. Μέσω αυτών διασφαλίζεται η παροχή υψηλής ποιότητας υπηρεσιών, μια ομαλή εμπειρία χρήστη και η δυνατότητα μελλοντικής επέκτασης και βελτίωσης.

Η **απόδοση και η αξιοπιστία** αποτελούν βασικές προτεραιότητες. Κάθε λειτουργία έχει σχεδιαστεί ώστε να ολοκληρώνεται εντός 2-5 δευτερολέπτων, εξασφαλίζοντας άμεση απόκριση στις αιτήσεις των χρηστών. Παράλληλα, η εφαρμογή πρέπει να είναι αξιόπιστη, αποφεύγοντας σφάλματα κατά την εκτέλεση λειτουργιών. Σε περιπτώσεις προβλημάτων, παρέχονται κατανοητά και ακριβή μηνύματα λάθους, βοηθώντας τους χρήστες να αναγνωρίσουν και να διορθώσουν πιθανά προβλήματα.

Η **χρηστικότητα και η διαδραστικότητα** είναι κρίσιμες για την εμπειρία του χρήστη. Η διεπαφή χρήστη έχει σχεδιαστεί ώστε να είναι προσαρμόσιμη σε διάφορες συσκευές και μεγέθη οθόνης, διασφαλίζοντας μια ενιαία και φιλική εμπειρία χρήσης. Ο σχεδιασμός περιλαμβάνει δυναμικά στοιχεία, όπως pop-ups για επαλήθευση ενεργειών (π.χ., διαγραφή εγγραφών) και χρωματική κωδικοποίηση για την κατηγοριοποίηση των αποτελεσμάτων πρόβλεψης. Η λειτουργία αυτή βοηθά τους χρήστες να λαμβάνουν άμεσα οπτική ανατροφοδότηση, ενισχύοντας τη διαδραστικότητα και την ευχρηστία.

Η **ασφάλεια** της εφαρμογής είναι θεμελιώδης, προστατεύοντας τα προσωπικά δεδομένα των χρηστών. Το σύστημα εφαρμόζει μέτρα ασφαλείας, όπως προστασία από επιθέσεις CSRF. Οι χρήστες ταυτοποιούνται μέσω μοναδικών διαπιστευτηρίων και η πρόσβαση στις λειτουργίες της εφαρμογής πραγματοποιείται βάσει καθορισμένων κανόνων. Επιπλέον, η αυτόματη αποσύνδεση χρηστών μετά από περιόδους αδράνειας προστατεύει από μη εξουσιοδοτημένη πρόσβαση.

Η **ποιότητα του κώδικα και η συντηρησιμότητα** της εφαρμογής διασφαλίζονται μέσω της χρήσης σύγχρονων τεχνολογιών και βέλτιστων πρακτικών ανάπτυξης. Το backend της εφαρμογής έχει υλοποιηθεί με αρχιτεκτονική REST, παρέχοντας ευελιξία στη διαχείριση δεδομένων και επεκτασιμότητα για μελλοντικές λειτουργίες. Οι φόρμες και τα δεδομένα επικυρώνονται τόσο στο frontend όσο και στο backend, εξασφαλίζοντας την αποτροπή διείσδυσης κακόβουλων δεδομένων και την αξιοπιστία της εφαρμογής. Η διεπαφή χρήστη έχει σχεδιαστεί με δυναμικά στοιχεία, όπως η χρήση Ajax και δυναμικής HTML, τα οποία επιτρέπουν την άμεση απόκριση σε ενέργειες, όπως η διαγραφή εγγραφών ή η επικύρωση δεδομένων, βελτιώνοντας έτσι την εμπειρία του χρήστη.

Ο σχεδιασμός του συστήματος περιλαμβάνει τον πλήρη διαχωρισμό του frontend από το backend, κάτι που εξασφαλίζει καθαρή και οργανωμένη δομή κώδικα, διευκολύνοντας τη συντήρηση και την ανάπτυξη του συστήματος. Επιπλέον, ο κώδικας έχει διαμορφωθεί με στόχο τη μακροπρόθεσμη συντηρησιμότητα, επιτρέποντας εύκολες επεκτάσεις ή τροποποιήσεις στο μέλλον χωρίς την ανάγκη ριζικών αλλαγών, γεγονός που διασφαλίζει τη βιωσιμότητα και την προσαρμοστικότητα της εφαρμογής.

Η **συναλλακτικότητα** διασφαλίζεται μέσω της χρήσης κατάλληλων μηχανισμών στην υποκείμενη βάση δεδομένων. Οι λειτουργίες που επηρεάζουν τη βάση δεδομένων είναι σχεδιασμένες ώστε να αποτρέπουν διπλές ή εσφαλμένες εγγραφές, διατηρώντας την ακεραιότητα των δεδομένων. Για παράδειγμα, αν ο χρήστης προσπαθήσει να καταχωρίσει την ίδια εγγραφή στο Diary δύο φορές, το σύστημα διασφαλίζει ότι μόνο μία θα αποθηκευτεί.

Η **επεκτασιμότητα** της εφαρμογής επιτρέπει την ενσωμάτωση νέων λειτουργιών, όπως επιπλέον τύπους γραφημάτων ή νέους αλγορίθμους πρόβλεψης. Το modular design που υιοθετήθηκε διασφαλίζει ότι νέες δυνατότητες μπορούν να προστεθούν χωρίς να επηρεάζονται οι υφιστάμενες λειτουργίες.

Συνολικά, οι μη λειτουργικές απαιτήσεις της chatGLUCO έχουν σχεδιαστεί για να προσφέρουν ασφάλεια, απόδοση, χρηστικότητα και επεκτασιμότητα, διασφαλίζοντας μια φιλική και αξιόπιστη εμπειρία χρήστη, ενώ παράλληλα υποστηρίζουν τη συντήρηση και την περαιτέρω ανάπτυξη του συστήματος.

5.2 Σχεδίαση Συστήματος

Η ανάλυση των λειτουργικών και μη λειτουργικών απαιτήσεων που προηγήθηκε ανέδειξε την ανάγκη για έναν ολοκληρωμένο και καλά δομημένο σχεδιασμό του συστήματος chatGLUCO. Ο προσεκτικός σχεδιασμός αποτελεί το θεμέλιο για την επιτυχή υλοποίηση και ομαλή λειτουργία της εφαρμογής, καθώς εξασφαλίζει την κατανόηση των απαιτήσεων και την αποφυγή πιθανών προβλημάτων κατά τη φάση της ανάπτυξης.

Σε αυτή την ενότητα παρουσιάζουμε τα βασικά σχεδιαστικά μοντέλα που αποτυπώνουν τις κύριες σχεδιαστικές αποφάσεις και τη δομή της εφαρμογής.

5.2.1 Διάγραμμα Περιπτώσεων Χρήσης / Use Case Diagram

Το Use Case Diagram που έχει δημιουργηθεί για την εφαρμογή chatGLUCO παρουσιάζει με σαφήνεια τις λειτουργίες που μπορεί να εκτελέσει ο Patient (Χρήστης). Το διάγραμμα απεικονίζει τις λειτουργικές απαιτήσεις που καλύπτουν τη διαχείριση ημερολογίου, διαγραμμάτων, προβλέψεων γλυκόζης, καθώς και τη διαχείριση προσωπικών στοιχείων. Παρακάτω περιγράφονται συνοπτικά οι βασικές ενότητες και οι επιμέρους λειτουργίες:

1. Διαχείριση Ημερολογίου

- Δημιουργία νέας εγγραφής
Ο χρήστης συμπληρώνει υποχρεωτικά το πεδίο "glucose_before" και προαιρετικά μπορεί να προβεί σε:
 - Πρόβλεψη δόσης ινσουλίνης
 - Αναζήτηση φαγητού για αυτόματη συμπλήρωση υδατανθράκων.
- Ενημέρωση υφιστάμενων εγγραφών
Δίνει τη δυνατότητα επεξεργασίας των ίδιων πεδίων με τη δημιουργία εγγραφής.
- Διαγραφή εγγραφών
Με include σχέση, η διαγραφή απαιτεί επιβεβαίωση της ενέργειας από τον χρήστη.
- Εξαγωγή δεδομένων XLS
Ο χρήστης μπορεί να εξαγάγει τα δεδομένα του ημερολογίου σε μορφή XLS για περαιτέρω ανάλυση.

2. Προβολή Διαγραμμάτων

- Προβολή διαγραμμάτων
Ο χρήστης μπορεί να προβάλλει τα δεδομένα του ορίζοντας:
 - Επιλογή τύπου διαγράμματος (Time in Range, Glucose-Insulin Correlation).
 - Ορισμός εύρους ημερομηνίας για τη στοχευμένη εμφάνιση των δεδομένων.

3. Διαχείριση Πρόβλεψης Γλυκόζης

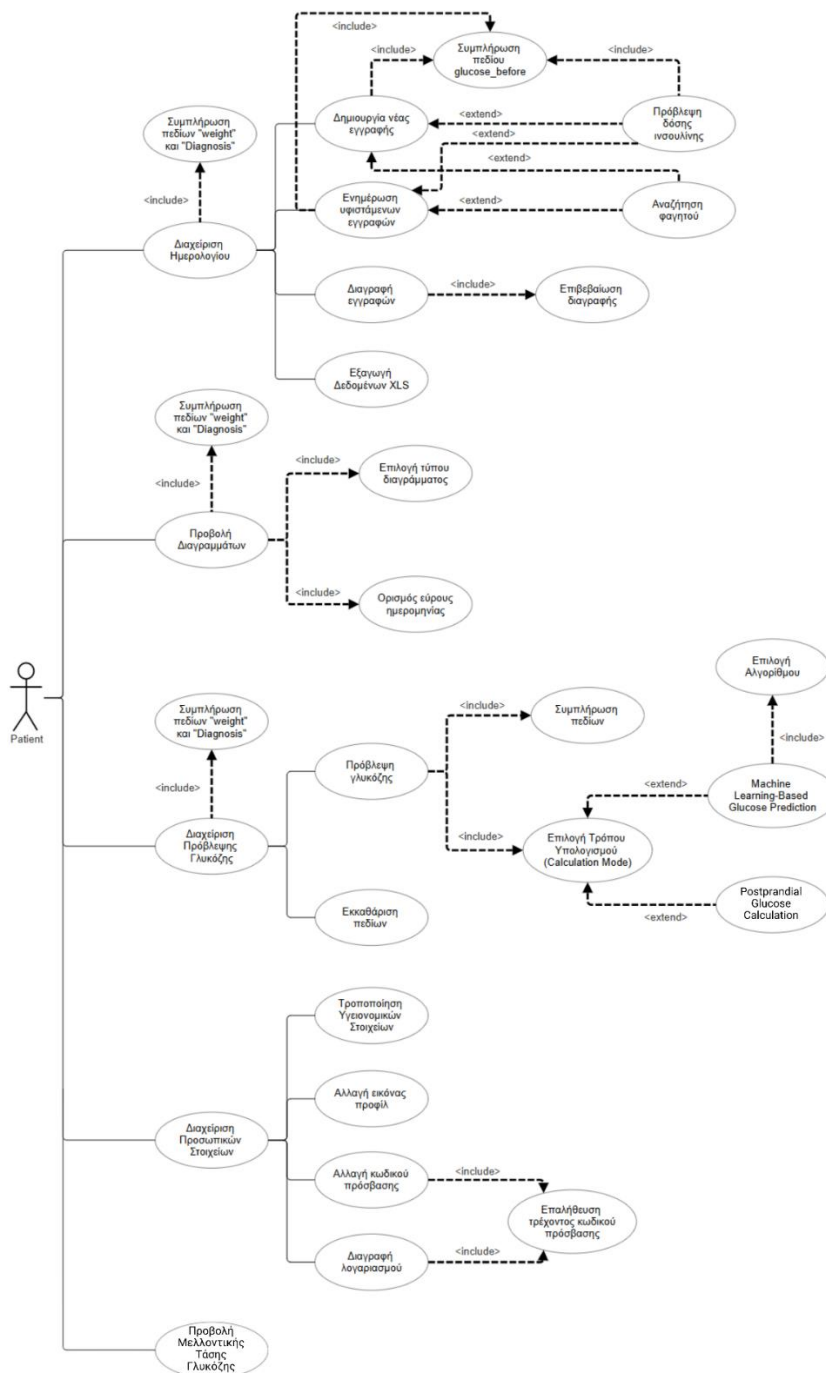
- Πρόβλεψη γλυκόζης
Ο χρήστης επιλέγει Calculation Mode, συμπληρώνει τα απαραίτητα πεδία και εκτελεί την πρόβλεψη.
 - Επεκτάσεις:
 - Postprandial Glucose Calculation
 - Machine Learning-Based Glucose Prediction με επιλογή αλγορίθμου (Linear Regression, Random Forest, XGBoost).
- Εκκαθάριση πεδίων
Η λειτουργία αυτή διαγράφει όλα τα δεδομένα που έχουν καταχωρηθεί στη φόρμα.

4. Διαχείριση Προσωπικών Στοιχείων

- Τροποποίηση υγειονομικών στοιχείων
Ο χρήστης μπορεί να επεξεργαστεί τα πεδία "Weight" και "Diagnosis", καθώς και τα σχετικά πεδία της γλυκοζυλιωμένης αιμοσφαιρίνης (HbA1c).
- Αλλαγή εικόνας προφίλ
Παρέχεται δυνατότητα αντικατάστασης ή προσθήκης εικόνας προφίλ.
- Αλλαγή κωδικού πρόσβασης
Απαιτείται επιβεβαίωση του τρέχοντος κωδικού, όπως φαίνεται από τη <include> σχέση.
- Διαγραφή λογαριασμού
Η διαγραφή απαιτεί επιβεβαίωση μέσω εισαγωγής του τρέχοντος κωδικού πρόσβασης.

5. Προβολή Μελλοντικής Τάσης Γλυκόζης

Επιτρέπει στον χρήστη να δει ένα διάγραμμα που παρουσιάζει τόσο τις ιστορικές τιμές γλυκόζης του όσο και τις προβλεπόμενες τιμές για τις επόμενες δύο ημέρες. Οι προβλέψεις αφορούν τρεις χρονικές στιγμές ανά ημέρα (πρωί, μεσημέρι, βράδυ) και ξεκινούν αμέσως μετά το τελευταίο διαθέσιμο ιστορικό δεδομένο.



Σχήμα 8: Διάγραμμα Περιπτώσεων Χρήσης

5.2.2 Αρχιτεκτονική Συστήματος

Η αρχιτεκτονική του chatGLUCO σχεδιάστηκε με βάση την τριών επιπέδων αρχιτεκτονική (three-tier architecture), προσφέροντας καθαρό διαχωρισμό ευθυνών, επεκτασιμότητα και ευκολία συντήρησης. Επιπλέον, ενσωματώνει ένα Machine Learning Tier μέσω ενός

RESTful API υλοποιημένου με Flask για την εκτέλεση προβλέψεων. Παρακάτω περιγράφονται αναλυτικά τα τέσσερα βασικά επίπεδα της αρχιτεκτονικής:

1. Επίπεδο Παρουσίασης (Presentation Tier)

Το Presentation Tier είναι το επίπεδο που βρίσκεται πιο κοντά στον τελικό χρήστη και αφορά την παρουσίαση των δεδομένων και των διεπαφών της εφαρμογής. Στο πλαίσιο της αρχιτεκτονικής του chatGLUCO, έχουμε τα εξής:

- **Blade Templating Engine:** Χρησιμοποιείται για τη δυναμική δημιουργία σελίδων μέσω του Laravel.
- **Static Assets:** Περιλαμβάνει τα στατικά αρχεία όπως CSS, JavaScript και εικόνες, που χρησιμοποιούνται για τη διαμόρφωση του UI/UX.
- **Views:** Είναι τα τελικά HTML αρχεία που αποστέλλονται στον χρήστη και διαχειρίζονται τη ροή πληροφοριών και δεδομένων προς και από το backend.

Το επίπεδο αυτό δέχεται HTTP Requests από τον χρήστη και αποστέλλει HTTP Responses με τα επεξεργασμένα δεδομένα που παρουσιάζονται στον τελικό χρήστη.

2. Επίπεδο Επιχειρησιακής Λογικής (Business Logic Tier)

Το Business Logic Tier αναλαμβάνει την επεξεργασία των αιτημάτων του χρήστη και την εκτέλεση των κανόνων της εφαρμογής. Αποτελεί τον σύνδεσμο μεταξύ του Presentation Tier και του Data Tier, ενώ λειτουργεί και ως διαμεσολαβητής για την επικοινωνία με το Machine Learning Tier (Flask API).

Κύρια Στοιχεία:

- **Web Routes (routes/web.php):** Καθορίζουν τις διαδρομές που χρησιμοποιούνται από τον χρήστη για την αποστολή αιτημάτων προς τους Controllers.
- **Controllers:** Διαχειρίζονται τα αιτήματα του χρήστη, εφαρμόζουν την επιχειρησιακή λογική και επικοινωνούν με το Flask API για τη λήψη προβλέψεων. Επίσης, καλούν τα Models για την εκτέλεση CRUD λειτουργιών στη βάση δεδομένων.
- **Models και Query Builder:** Είναι υπεύθυνα για την επικοινωνία με το Data Tier. Χρησιμοποιούν το Laravel's Query Builder για την ανάκτηση, ενημέρωση και αποθήκευση δεδομένων.

Ροή Επικοινωνίας:

1. Τα Controllers λαμβάνουν ένα αίτημα από τον χρήστη μέσω των Routes.
2. Επεξεργάζονται τα δεδομένα και, αν χρειάζεται πρόβλεψη, αποστέλλουν HTTP Requests προς το FLASK API στο Machine Learning Tier.
3. Το Flask API επιστρέφει JSON Responses με τα αποτελέσματα της πρόβλεψης.
4. Τα Controllers ανανεώνουν τα Views και παρουσιάζουν τα δεδομένα στον χρήστη.

3. Επίπεδο Μηχανικής Μάθησης (Machine Learning Tier)

Το Machine Learning Tier αποτελεί ξεχωριστό κομμάτι της αρχιτεκτονικής και έχει υλοποιηθεί μέσω ενός RESTful API χρησιμοποιώντας το Flask Framework.

Λειτουργίες:

- Το Flask API εκτελεί προβλέψεις χρησιμοποιώντας τρία μοντέλα Μηχανικής Μάθησης:
 - Linear Regression
 - Random Forest
 - XGBoost
- Συνδέεται με το Data Tier για να αντλήσει τα απαραίτητα δεδομένα εκπαίδευσης και δοκιμής των μοντέλων.
- Επιστρέφει τα αποτελέσματα των προβλέψεων στο Laravel backend σε JSON format.

Ροή Δεδομένων:

1. Το Laravel Controller στέλνει HTTP Requests προς το Flask API (π.χ., /predict_insulin ή /predict_glucose).
2. Το Flask API λαμβάνει τα δεδομένα, εκτελεί την πρόβλεψη μέσω των μοντέλων και επιστρέφει HTTP Responses σε JSON μορφή.
3. Τα δεδομένα παρουσιάζονται στον χρήστη μέσω των Views στο Presentation Tier.

4. Επίπεδο Δεδομένων (Data Tier)

Το Data Tier είναι υπεύθυνο για τη διαχείριση και αποθήκευση των δεδομένων της εφαρμογής. Περιλαμβάνει μία βάση δεδομένων (MySQL), η οποία χρησιμοποιείται τόσο από το Laravel Backend όσο και από το Flask API.

Ρόλοι της Βάσης Δεδομένων:

1. Laravel Backend:

- Διαχειρίζεται τα δεδομένα χρηστών και τις εγγραφές του ημερολογίου.
- Εκτελεί CRUD λειτουργίες μέσω των Models και του Query Builder.

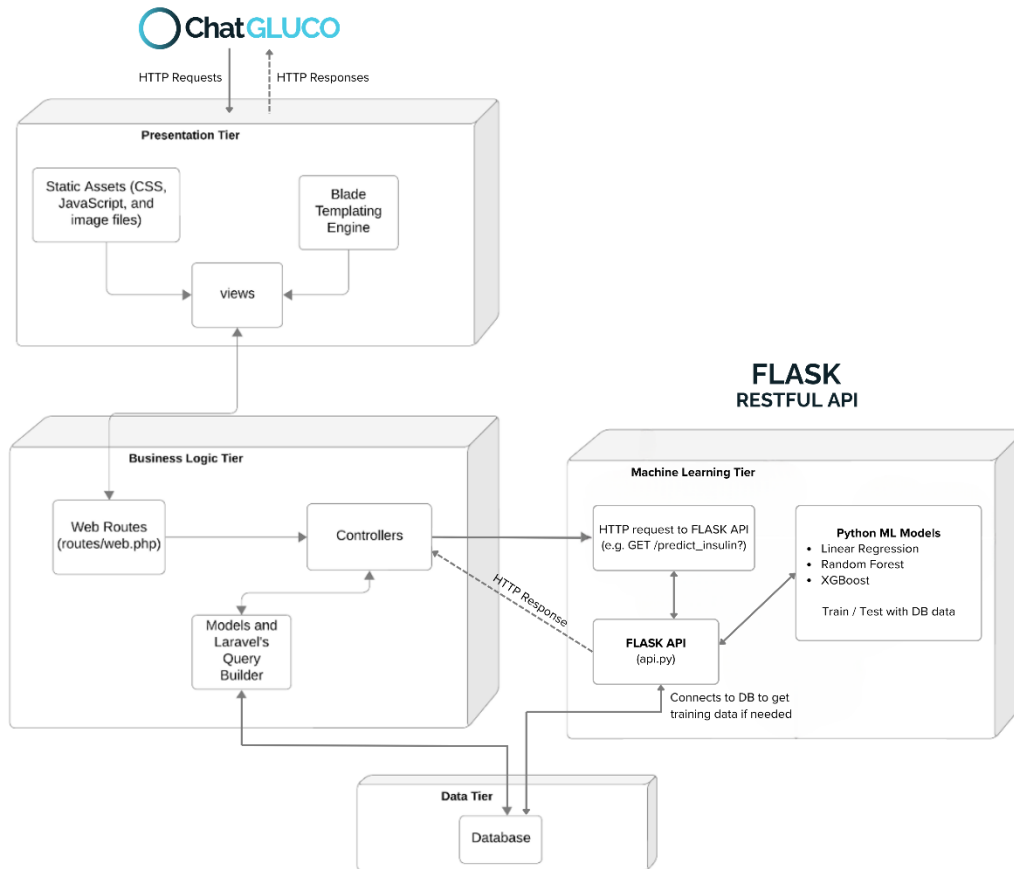
2. Flask API:

- Συνδέεται με τη βάση δεδομένων για να αντλήσει δεδομένα εκπαίδευσης για τα ML μοντέλα.
- Εξασφαλίζει ότι οι προβλέψεις βασίζονται σε ενημερωμένα δεδομένα.

Πλεονεκτήματα της Αρχιτεκτονικής

Η αρχιτεκτονική του chatGLUCO, βασισμένη στα τρία επίπεδα με την προσθήκη ενός Machine Learning Tier, προσφέρει τα εξής πλεονεκτήματα:

1. **Διαχωρισμός Ευθυνών:** Κάθε επίπεδο έχει ξεχωριστό ρόλο, διευκολύνοντας τη συντήρηση και την αναβάθμιση της εφαρμογής.
2. **Επεκτασιμότητα:** Η ενσωμάτωση του Flask API επιτρέπει την ανεξάρτητη ανάπτυξη και βελτίωση του Machine Learning Tier.
3. **Αποδοτική Επικοινωνία:** Η χρήση HTTP Requests/Responses και JSON επιτρέπει την ομαλή συνεργασία μεταξύ Laravel και Flask.
4. **Ευκολία Συντήρησης και Επεκτασιμότητας:** Η αρχιτεκτονική επιτρέπει τη μελλοντική προσθήκη νέων λειτουργιών ή τη βελτίωση των υπαρχόντων.



Σχήμα 9: Διάγραμμα Αρχιτεκτονικής

5.2.3 Διάγραμμα Βάσεων Δεδομένων / Database Diagram

Το παρακάτω διάγραμμα απεικονίζει τη δομή της βάσης δεδομένων της εφαρμογής chatGLUCO, η οποία έχει σχεδιαστεί για την αποθήκευση και διαχείριση των δεδομένων των χρηστών και των υγειονομικών εγγραφών.

Για την υλοποίηση του συστήματος ταυτοποίησης και αυθεντικοποίησης των χρηστών, χρησιμοποιούμε το Laravel UI Auth. Με την εντολή “`php artisan ui:auth`”, δημιουργούνται οι πίνακες που σχετίζονται με τις λειτουργίες επαναφοράς κωδικών πρόσβασης, την υλοποίηση OAuth και τη διαχείριση προσωπικών access tokens. Οι πίνακες που παράγονται από αυτή τη διαδικασία περιλαμβάνονται στην περιγραφή που ακολουθεί.

1. Table: patients

Ο πίνακας patients περιλαμβάνει πληροφορίες που σχετίζονται με το προφίλ των ασθενών. Είναι συνδεδεμένος με τον πίνακα users, (σχέση ένα προς ένα), μέσω του πεδίου user_id, το οποίο χρησιμεύει ως ξένο κλειδί (foreign key).

Πεδία:

- id (*bigint, unsigned*): Το μοναδικό αναγνωριστικό του ασθενούς (primary key).
- user_id (*bigint, unsigned*): Ξένο κλειδί που συνδέει τον πίνακα patients με τον πίνακα users.
- gender (*varchar(255)*): Το φύλο του ασθενούς (π.χ., male ή female).
- birth_at (*date*): Η ημερομηνία γέννησης του ασθενούς.
- weight (*decimal(5,2)*): Το σωματικό βάρος του ασθενούς, αποθηκευμένο με δύο δεκαδικά ψηφία.
- diagnosis (*varchar(255)*): Ο τύπος διάγνωσης του ασθενούς (π.χ., Type I, Type II).
- created_at (*timestamp*): Η χρονική στιγμή δημιουργίας της εγγραφής.
- updated_at (*timestamp*): Η τελευταία χρονική στιγμή τροποποίησης της εγγραφής.
- hba1c1 (*decimal (5,2)*): Η πιο πρόσφατη μέτρηση της γλυκοζυλιωμένης αιμοσφαιρίνης (HbA1c), που αντικατοπτρίζει τη ρύθμιση της γλυκόζης αίματος των τελευταίων 2-3 μηνών.
- hba1c2 (*decimal (5,2)*): Η προτελευταία μέτρηση της HbA1c, που βοηθά στην παρακολούθηση της ιστορικής εξέλιξης της γλυκαιμικής ρύθμισης.
- hba1c3 (*decimal (5,2)*): Η μέτρηση της HbA1c πριν από την προτελευταία, παρέχοντας πιο μακροπρόθεσμη εικόνα της πορείας των επιπέδων γλυκόζης αίματος.

2. Table: patients_statistics

Ο πίνακας patients_statistics αποθηκεύει χρονικά ευαίσθητες πληροφορίες που σχετίζονται με τις καθημερινές μετρήσεις των ασθενών, όπως τα επίπεδα γλυκόζης, τη δόση ινσουλίνης και τους υδατάνθρακες που καταναλώθηκαν. Είναι συνδεδεμένος με τον πίνακα patients, (σχέση ένα προς πολλά), μέσω του πεδίου patient_id.

Πεδία:

- *id (bigint, unsigned)*: Το μοναδικό αναγνωριστικό της εγγραφής (primary key).
- *patient_id (bigint, unsigned)*: Ξένο κλειδί που συνδέει τον πίνακα *patients_statistics* με τον πίνακα *patients*.
- *glucose_old (double(5,2))*: Η τιμή γλυκόζης πριν το γεύμα (Glucose Before).
- *insulin_dose (double(5,2))*: Η ποσότητα δόσης ινσουλίνης που χορηγήθηκε (σε μονάδες).
- *food_carbo (double(5,2))*: Η ποσότητα υδατανθράκων που καταναλώθηκε (σε γραμμάρια).
- *glucose_new (double(5,2))*: Η τιμή γλυκόζης μετά το γεύμα (Glucose After).
- *weight (decimal(5,2))*: Το σωματικό βάρος του ασθενούς κατά την καταγραφή της εγγραφής.
- *created_at (datetime)*: Η χρονική στιγμή δημιουργίας της εγγραφής.
- *updated_at (datetime)*: Η τελευταία χρονική στιγμή τροποποίησης της εγγραφής.

3. Table: users

Ο πίνακας *users* αποθηκεύει τα βασικά στοιχεία σύνδεσης και ταυτοποίησης των χρηστών της εφαρμογής.

Πεδία:

- *id (bigint, unsigned)*: Το μοναδικό αναγνωριστικό του χρήστη (primary key).
- *firstname (varchar(255))*: Το όνομα του χρήστη.
- *lastname (varchar(255))*: Το επώνυμο του χρήστη.
- *email (varchar(255))*: Η διεύθυνση email του χρήστη (μοναδικό πεδίο).
- *email_verified_at (timestamp)*: Η χρονική στιγμή επιβεβαίωσης του email.
- *password (varchar(255))*: Ο κρυπτογραφημένος κωδικός πρόσβασης του χρήστη.
- *role (varchar(50))*: Ο ρόλος του χρήστη (π.χ., Patient).
- *profile_pic (text)*: Η εικόνα προφίλ του χρήστη (αποθηκευμένη ως URL ή binary).
- *remember_token (varchar(100))*: Το διακριτικό που χρησιμοποιείται για να διατηρεί τον χρήστη συνδεδεμένο.

- `created_at (timestamp)`: Η χρονική στιγμή δημιουργίας της εγγραφής.
- `updated_at (timestamp)`: Η τελευταία χρονική στιγμή τροποποίησης της εγγραφής.

Οι πίνακες `patients`, `patients_statistics` και `users` αποτελούν τον πυρήνα της βάσης δεδομένων της εφαρμογής `chatGLUCO`. Η διασύνδεση αυτών των πινάκων εξασφαλίζει τη συνοχή και την ορθή διαχείριση των προσωπικών και υγειονομικών δεδομένων των χρηστών, επιτρέποντας στην εφαρμογή να παρέχει λειτουργίες όπως πρόβλεψη γλυκόζης, καταγραφή ημερολογίου και ανάλυση στατιστικών στοιχείων.

Αφού είδαμε τους τρεις βασικότερους πίνακες της βάσης δεδομένων, θα προχωρήσουμε και σε μία πιο συνοπτική ανάλυση των υπόλοιπων πινάκων της εφαρμογής.

4. Table: password_reset_tokens

Αποθηκεύει πληροφορίες για αιτήματα επαναφοράς κωδικού πρόσβασης. Περιέχει τις στήλες `email` (διεύθυνση email του χρήστη), `token` (το μοναδικό token για την επαναφορά του κωδικού), και `created_at` (ημερομηνία και ώρα δημιουργίας του token).

5. Table: oauth_personal_access_clients

Χρησιμοποιείται για τον ορισμό προσωπικών εφαρμογών OAuth που έχουν πρόσβαση στο σύστημα. Στήλες όπως `client_id` και `created_at` καταγράφουν τον αναγνωριστικό αριθμό του πελάτη και την ημερομηνία δημιουργίας.

6. Table: personal_access_tokens

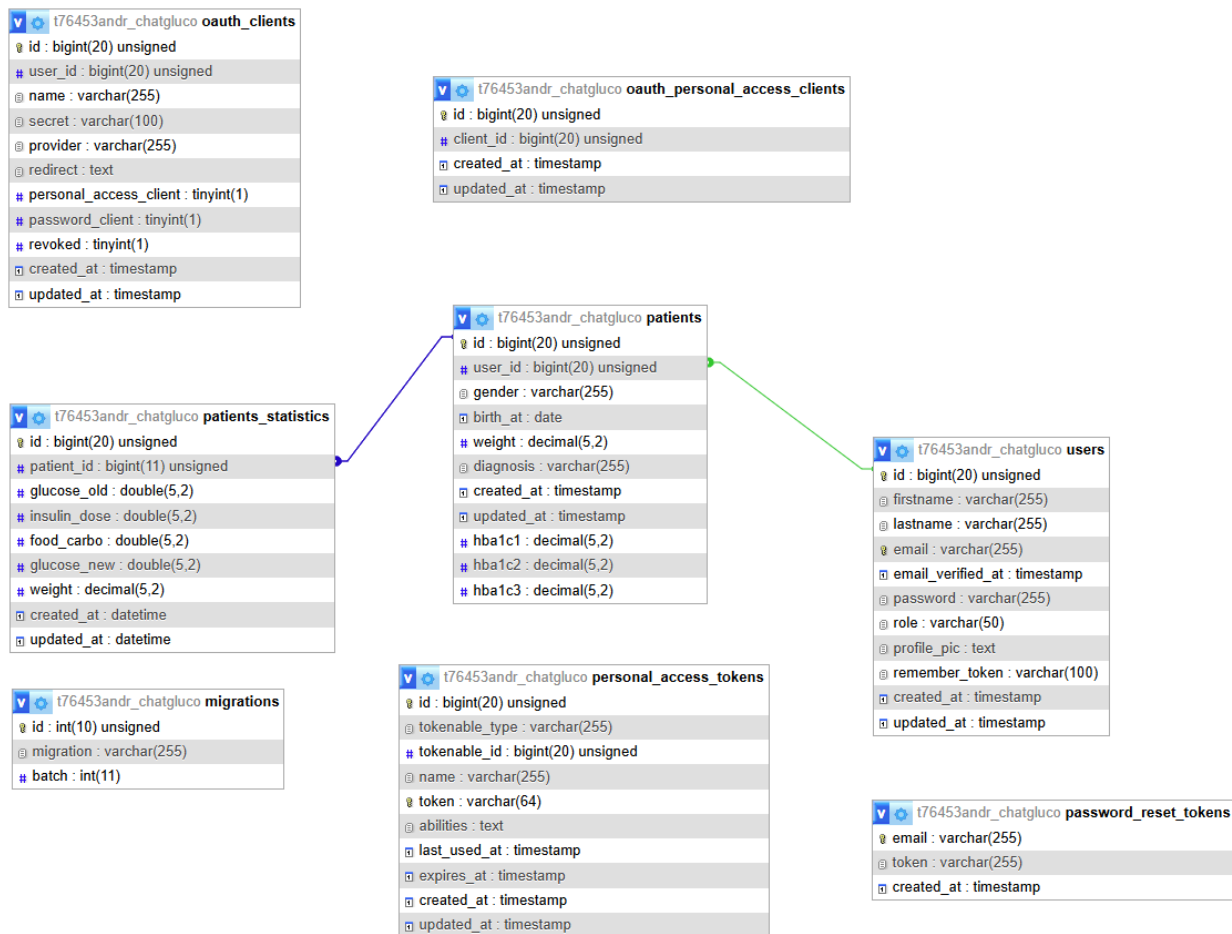
Αποθηκεύει τα προσωπικά tokens που επιτρέπουν στους χρήστες να έχουν πρόσβαση μέσω API. Περιλαμβάνει στήλες όπως `tokenable_type` (ο τύπος του στοιχείου), `tokenable_id` (ο αναγνωριστικός αριθμός του στοιχείου), `token` (το token), `expires_at` (ημερομηνία λήξης), και άλλες σχετικές με την πρόσβαση.

7. Table: oauth_clients

Διαχειρίζεται πληροφορίες για πελάτες OAuth που είναι συνδεδεμένοι με την εφαρμογή. Στήλες όπως `user_id` (το ID του χρήστη), `name` (όνομα πελάτη), `secret` (το μυστικό του πελάτη), `redirect` (διεύθυνση ανακατεύθυνσης), καθώς και πληροφορίες για το αν είναι `personal_access_client` ή `password_client`.

8. Table: migrations

Διαχειρίζεται τις εγγραφές των μεταναστεύσεων (migrations) για την βάση δεδομένων. Περιλαμβάνει στήλες όπως id, migration (όνομα migration), και batch (ο αριθμός της παρτίδας που εκτελέστηκε).



Σχήμα 10: Διάγραμμα Βάσης Δεδομένων

5.3 Υλοποίηση του Συστήματος

5.3.1 Υλοποίηση του Frontend

Το Frontend της εφαρμογής chatGLUCO υλοποιήθηκε με τη χρήση του PHP Laravel [18] framework σε συνδυασμό με σύγχρονες τεχνολογίες και βιβλιοθήκες που εξασφάλισαν ένα

λειτουργικό, διαδραστικό και φιλικό προς τον χρήστη περιβάλλον. Η χρήση του Laravel Blade επέτρεψε τη δυναμική διαχείριση των δεδομένων μέσω επαναχρησιμοποιούμενων templates, μειώνοντας τον πλεονάζοντα κώδικα και διευκολύνοντας τη συντηρησιμότητα της εφαρμογής.

Για τη διαμόρφωση της διεπαφής χρήστη χρησιμοποιήθηκε το Bootstrap, ένα ισχυρό CSS framework, το οποίο επέτρεψε τη δημιουργία ενός responsive design. Χάρη στο grid system και τα components που παρέχει το Bootstrap, η εμφάνιση της εφαρμογής προσαρμόζεται αυτόματα σε διαφορετικές συσκευές και μεγέθη οθονών, προσφέροντας ομοιόμορφη εμπειρία χρήσης. Στην υλοποίηση των λειτουργιών του Diary, οι εγγραφές εμφανίζονται σε έναν πίνακα με δυναμικά κουμπιά διαγραφής, ενώ η ενότητα Diagrams περιλαμβάνει φόρμες επιλογής ημερομηνιών και τύπων διαγραμμάτων, σχεδιασμένες με Bootstrap elements.

Η λειτουργικότητα του frontend ενισχύθηκε με τη χρήση JavaScript και jQuery για τη διαχείριση των συμβάντων και τη διαδραστικότητα. Συγκεκριμένα, οι φόρμες επαληθεύονται μέσω client-side validation, ενώ δυναμικά μηνύματα επιτυχίας ή σφάλματος εμφανίζονται για την καθοδήγηση του χρήστη. Σε συνδυασμό με την τεχνολογία AJAX, η εφαρμογή επιτρέπει την ανταλλαγή δεδομένων τύπου JSON με τον server χωρίς την ανάγκη επαναφόρτωσης της σελίδας. Αυτό υλοποιήθηκε σε λειτουργίες όπως ο υπολογισμός της δόσης ινσουλίνης και ο αυτόματος υπολογισμός των υδατανθράκων μέσω της αναζήτησης φαγητών.

Για την οπτικοποίηση των δεδομένων, χρησιμοποιήθηκε η βιβλιοθήκη Chart.js, η οποία προσέφερε τη δυνατότητα δημιουργίας διαδραστικών διαγραμμάτων. Στην ενότητα Diagrams, ο χρήστης μπορεί να επιλέξει ανάμεσα σε δύο τύπους διαγραμμάτων: Time in Range και Glucose-Insulin Correlation, όπου τα δεδομένα απεικονίζονται με σαφήνεια και ευκρίνεια. Η έγχρωμη κωδικοποίηση για τις τιμές γλυκόζης (πράσινο για φυσιολογικά επίπεδα, κίτρινο για χαμηλά και κόκκινο για υψηλά) εξασφαλίζει την άμεση κατανόηση των αποτελεσμάτων.

Η εφαρμογή περιλαμβάνει επίσης Popper.js, το οποίο χρησιμοποιήθηκε για τη δημιουργία tooltips και dropdowns, προσθέτοντας μικρά δυναμικά στοιχεία που ενισχύουν την εμπειρία χρήσης. Παράλληλα, η έγχρωμη κωδικοποίηση σε πεδία όπως το Glucose Before και Glucose After προσφέρει στον χρήστη σαφή οπτική ανατροφοδότηση των δεδομένων που καταχωρεί.

Συνολικά, το Frontend της εφαρμογής chatGLUCO είναι σχεδιασμένο για να είναι καθαρό, σύγχρονο και λειτουργικό. Η χρήση του Laravel σε συνδυασμό με HTML5, CSS, Bootstrap, JavaScript και AJAX συνέβαλε στη δημιουργία ενός περιβάλλοντος που παρέχει υψηλή

διαδραστικότητα και φιλικότητα στον χρήστη. Η εφαρμογή είναι πλήρως responsive, προσφέροντας απρόσκοπτη εμπειρία χρήσης σε κάθε συσκευή, ενώ παράλληλα διασφαλίζεται η ομαλή ροή δεδομένων και η αισθητική συνέπεια.

5.3.2 Υλοποίηση του Backend

Στην παρούσα ενότητα παρουσιάζεται η υλοποίηση του Backend για την εφαρμογή chatGLUCO, η οποία επιλέχθηκε να αναπτυχθεί με το δημοφιλές framework Laravel (έκδοση 10.48.22). Παράλληλα, περιγράφεται και το δεύτερο project του συστήματος, το οποίο αφορά την υλοποίηση και ενσωμάτωση των αλγορίθμων μηχανικής μάθησης μέσω ενός Flask API. Παρακάτω ακολουθεί η αναλυτική περιγραφή της αρχιτεκτονικής, των τεχνολογιών που χρησιμοποιήθηκαν και του τρόπου επικοινωνίας μεταξύ των δύο projects.

1. Χωρισμός σε Δύο Έργα (Projects)

Στο πλαίσιο ανάπτυξης της εφαρμογής πραγματοποιήθηκε διαχωρισμός σε δύο ξεχωριστά projects:

1. chatGLUCO (Laravel Project)

- Αναλαμβάνει όλες τις λειτουργίες της εφαρμογής, όπως η αυθεντικοποίηση χρηστών, η εμφάνιση, διαγραφή και διαχείριση δεδομένων, καθώς και οι βασικές ροές χρήσης (use cases).
- Στο backend χρησιμοποιεί τους μηχανισμούς και τα εργαλεία του Laravel, όπως Controllers, Models και Routes, ενώ για την επικοινωνία με το frontend αξιοποιεί views που βασίζονται στο Blade Templating Engine.

2. Flask API (api.py)

- Επικεντρώνεται στην εκπαίδευση αλγορίθμων μηχανικής μάθησης (Linear Regression, Random Forest, XGBoost) και στην παροχή προβλέψεων.
- Παρέχει RESTful endpoints (π.χ. GET /predict_glucose, GET /predict_insulin) τα οποία δέχονται δεδομένα από το Laravel project και επιστρέφουν αποτελέσματα σε μορφή JSON.

Η επιλογή αυτή επιτρέπει την ανεξάρτητη συντήρηση και ανάπτυξη των δύο τμημάτων, διατηρώντας τον κώδικα καθαρό και κλιμακώσιμο.

2. Λειτουργία του Laravel Backend (chatGLUCO)

Το Laravel project συνδέεται με τη βάση δεδομένων MySQL που έχει στηθεί στο περιβάλλον XAMPP, τροποποιώντας τις ρυθμίσεις στο αρχείο περιβάλλοντος .env. Εκεί καθορίζονται στοιχεία όπως ο τύπος της βάσης δεδομένων, στην περίπτωση μας MySQL, η διεύθυνση IP του διακομιστή (server) που φιλοξενεί τη βάση δεδομένων, την πόρτα (port), το όνομα της βάσης δεδομένων, ο χρήστης και ο κωδικός πρόσβασης. Με αυτόν τον τρόπο, η εφαρμογή αποκτά πρόσβαση για την εκτέλεση ερωτημάτων και την αποθήκευση δεδομένων.

Η δημιουργία και διαχείριση των πινάκων της βάσης δεδομένων πραγματοποιήθηκε μέσω του Migration μηχανισμού του Laravel. Τα αρχεία migrations, γραμμένα σε PHP, παρέχουν ευκολία στη δημιουργία και τροποποίηση της δομής της βάσης, διασφαλίζοντας τις αλλαγές που σχετίζονται με την ανάπτυξη και εξέλιξη του project.

Η επικοινωνία της εφαρμογής (chatGLUCO) με τη βάση δεδομένων πραγματοποιείται μέσα από τους controllers του Laravel. Εκεί υλοποιούνται οι βασικές λειτουργίες που αφορούν τα δεδομένα, όπως η διαγραφή, η εμφάνιση και η τροποποίησή τους, μέσα σε μεθόδους (συναρτήσεις) των αντίστοιχων κλάσεων. Αφού εκτελεστεί η επιλεγμένη λειτουργία σε έναν controller, ο χρήστης ανακατευθύνεται αυτόματα στην κατάλληλη σελίδα (view) για να ενημερωθεί ή να αλληλεπιδράσει περαιτέρω.

Το Laravel παρέχει τη δυνατότητα δημιουργίας RESTful API μέσω του αρχείου routes/web.php, όπου ορίζονται διαδρομές (URIs). Κάθε διαδρομή αντιστοιχεί σε μία συνάρτηση ενός controller, επιτρέποντας στο front-end να ζητά την προβολή ή την αλλαγή δεδομένων. Όταν ο χρήστης αλληλεπιδρά από το περιβάλλον του front-end, αποστέλλεται ένα αίτημα HTTP στο back-end. Το back-end επεξεργάζεται το αίτημα, αξιοποιώντας για παράδειγμα τη βάση δεδομένων για ανάκτηση ή ενημέρωση στοιχείων, και κατόπιν συντάσσει μια απόκριση (συνήθως σε μορφή JSON) που επιστρέφεται στο front-end. Έτσι, το front-end λαμβάνει τις αναγκαίες πληροφορίες ή μια επιβεβαίωση και ενημερώνει τη διεπαφή χρήστη, αναλόγως των δεδομένων που έλαβε.

Τέλος, αξίζει να αναφερθούν δύο δημοφιλή εργαλεία που συνδέονται συχνά με το Laravel:

1. Maatwebsite\Excel [19]: Ένα πακέτο PHP που επιτρέπει την εισαγωγή και εξαγωγή δεδομένων σε/από αρχεία Excel, διευκολύνοντας τη διαχείριση μεγάλου όγκου πληροφοριών.
2. Axios: Μια JavaScript βιβλιοθήκη για την πρακτική δημιουργία HTTP αιτημάτων από το front-end σε έναν server, η οποία χρησιμοποιείται ευρέως σε σύγχρονες διαδικτυακές εφαρμογές χάρη στη φιλική προς τον προγραμματιστή σύνταξή της και την αποδοτική διαχείριση των αιτημάτων.

3. Flask API – Μηχανική Μάθηση και Προβλέψεις

Στο δεύτερο project που βασίζεται στο Flask, το αρχείο `api.py` συνδέεται με τη βάση δεδομένων MySQL χρησιμοποιώντας τη βιβλιοθήκη `pymysql`. Αρχικά, δημιουργείται μια σύνδεση (connection) προς τη βάση, ορίζοντας παραμέτρους όπως τη διεύθυνση (host), τον χρήστη, τον κωδικό πρόσβασης και το όνομα της βάσης. Στη συνέχεια, μέσω ενός cursor εκτελούνται ερωτήματα και ανακτώνται ή τροποποιούνται δεδομένα. Επίσης, χρησιμοποιείται η βιβλιοθήκη `pandas` για τη μετατροπή των δεδομένων σε `DataFrame`, διευκολύνοντας έτσι τις διαδικασίες ανάλυσης.

Το Flask API περιέχει συναρτήσεις που είναι υπεύθυνες για:

- Την εκπαίδευση των μοντέλων (Linear Regression, Random Forest, XGBoost) σε δεδομένα που ανακτώνται από τη βάση.
- Την παροχή προβλέψεων μέσω συγκεκριμένων endpoints (π.χ. `/predict_glucose`, `/predict_insulin`).

Κατά την επίκληση αυτών των endpoints, το Flask API λαμβάνει τις κατάλληλες παραμέτρους από το αίτημα, εκτελεί τον αλγόριθμο πρόβλεψης και επιστρέφει το αποτέλεσμα σε μορφή JSON.

Χρήσιμες Βιβλιοθήκες στο Flask API (`api.py`):

- **Flask & Flask-CORS:** Το Flask αποτελεί τη βάση του API, ορίζοντας τα endpoints για την επικοινωνία με το Laravel. Η Flask-CORS επιτρέπει αιτήματα από διαφορετικές πηγές (π.χ. την εφαρμογή Laravel) διασφαλίζοντας τη σωστή ρύθμιση των Cross-Origin Resource Sharing πολιτικών.
- **pymysql:** Παρέχει τη λειτουργικότητα σύνδεσης και επικοινωνίας με MySQL.
- **NumPy & Pandas:** Χρησιμοποιούνται για την αποτελεσματική προετοιμασία και επεξεργασία δεδομένων (NumPy για αριθμητικούς υπολογισμούς, Pandas για αναπαράσταση και διαχείριση δεδομένων ως `DataFrame`).
- **scikit-learn (sklearn) & XGBoost:** Περιλαμβάνουν αλγόριθμους μηχανικής μάθησης (Linear Regression, Random Forest, XGBRegressor) και βοηθητικές λειτουργίες όπως η τυποποίηση (StandardScaler) ή η υπερπαραμετροποίηση (GridSearchCV).
- **statsmodels και scipy.stats:** Προσφέρουν προχωρημένες στατιστικές μεθόδους και αναλύσεις (π.χ. υπολογισμός p-values, t-statistics), επιτρέποντας ακριβέστερη αξιολόγηση των μοντέλων.

- **logging:** Χρησιμοποιείται για την καταγραφή σημαντικών γεγονότων (logs), όπως η επιτυχής εκπαίδευση αλγορίθμων ή ενδεχόμενα σφάλματα, βελτιώνοντας τη συντηρησιμότητα του API.

Με την αξιοποίηση των παραπάνω βιβλιοθηκών, το Flask API μπορεί να εκπαιδεύει μοντέλα μηχανικής μάθησης, να τα προσαρμόζει στα δεδομένα που συλλέγονται από τη βάση και να παρέχει προβλέψεις ή άλλα αποτελέσματα στο Laravel project, όλα σε ένα δομημένο και επεκτάσιμο περιβάλλον.

4. Ασφάλεια και Προστασία Δεδομένων στο Laravel

Για την ασφάλεια στην ανταλλαγή δεδομένων, το Laravel ενσωματώνει λειτουργίες όπως η αυθεντικοποίηση, η προστασία από CSRF, η αποτροπή επιθέσεων XSS, καθώς και η προστασία από SQL injection, συμβάλλοντας στη δημιουργία ασφαλέστερων εφαρμογών.

- Προστασία από CSRF: Μέσω της οδηγίας @csrf στο Blade (HTML φόρμες), προστίθεται αυτόματα ένα CSRF token, το οποίο είναι μοναδικό για κάθε συνεδρία χρήστη και διασφαλίζει ότι τα αιτήματα υποβάλλονται όντως από την εφαρμογή. Αν το token λείπει ή είναι εσφαλμένο, το Laravel απορρίπτει το αίτημα προφυλάσσοντας την εφαρμογή από επιθέσεις Cross-Site Request Forgery.
- Αντιμετώπιση Cross-Site Scripting (XSS): Με τη χρήση του Blade templating engine, οι μεταβλητές που εισάγονται με τη σύνταξη {{ }} υφίστανται αυτόματο escaping, αποτρέποντας την ενσωμάτωση κακόβουλου JavaScript ή HTML. Για περιπτώσεις εκτός Blade, μπορεί επιπλέον να εφαρμοστεί η συνάρτηση htmlspecialchars, εξασφαλίζοντας τη σωστή απόδοση των δεδομένων ως απλού κειμένου.
- SQL Injection: Το Laravel αξιοποιεί το Eloquent ORM και τον query builder που βασίζονται σε prepared statements και bound parameters. Έτσι, αποτρέπεται η δυνατότητα εκτέλεσης αυθαίρετων εντολών SQL από τρίτους, προσφέροντας μεγαλύτερη ασφάλεια σε σχέση με παραδοσιακές προσεγγίσεις.

5. Συνοπτική Ροή Δεδομένων

1. Ο χρήστης αλληλεπιδρά με την εφαρμογή μέσα από τις σελίδες (views) του Laravel.
2. Το frontend δημιουργεί αιτήματα (Routes) προς το Laravel, ζητώντας ενέργειες όπως εμφάνιση ή μεταβολή δεδομένων.

3. Ο Controller του Laravel αναλαμβάνει τη λογική επεξεργασία, συνδέεται με τα Models (μέσω Eloquent ή Query Builder) ή καλεί το Flask API για διαδικασίες που αφορούν προβλέψεις.
4. Το Flask API αποκρίνεται με τα αποτελέσματα των αλγορίθμων μηχανικής μάθησης σε μορφή JSON.
5. Το Laravel λαμβάνει την απάντηση και την επιστρέφει στον χρήστη με την κατάλληλη μορφή (σελίδα, μηνύματα, ενημερώσεις).

Αυτή η ροή εξασφαλίζει την ευελιξία και τον διαχωρισμό των ευθυνών, διατηρώντας τη λογική της επιχειρησιακής εφαρμογής (business logic) στο Laravel και τη λογική μηχανικής μάθησης (machine learning) στο Flask.

6. Συμπεράσματα

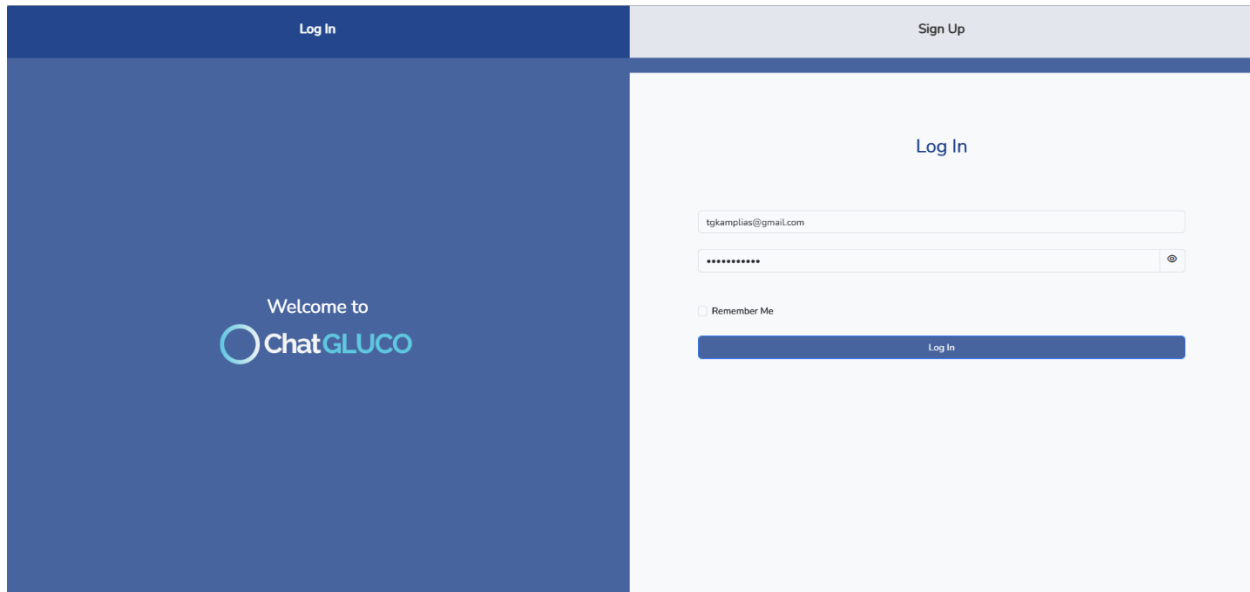
Η υλοποίηση του Backend στην εφαρμογή chatGLUCO βασίστηκε στη συνεργασία μεταξύ Laravel και Flask, με ξεκάθαρο διαχωρισμό των λειτουργιών:

- Το Laravel αναλαμβάνει την επιχειρησιακή λογική, τις βασικές CRUD διεργασίες, την επικοινωνία με τη βάση δεδομένων μέσω migrations και την ασφάλεια (π.χ. CSRF, XSS).
- Το Flask εστιάζει στην εκπαίδευση των αλγορίθμων μηχανικής μάθησης και στην παροχή προβλέψεων.

Η αρχιτεκτονική αυτή επιτρέπει ευκολότερη συντήρηση, κλιμάκωση και αναβάθμιση της εφαρμογής, ενώ παράλληλα κάνει χρήση καθιερωμένων πρακτικών ασφαλείας, εξασφαλίζοντας την ομαλή λειτουργία και προστασία των δεδομένων.

5.3.4 Διεπαφή Χρήστη (UI)

Για την καλύτερη κατανόηση της διεπαφής χρήστη στην εφαρμογή *chatGLUCO*, θα παρουσιαστεί ένα υποθετικό σενάριο χρήσης.



Σχήμα 11: Οθόνη Εισόδου (Log In)

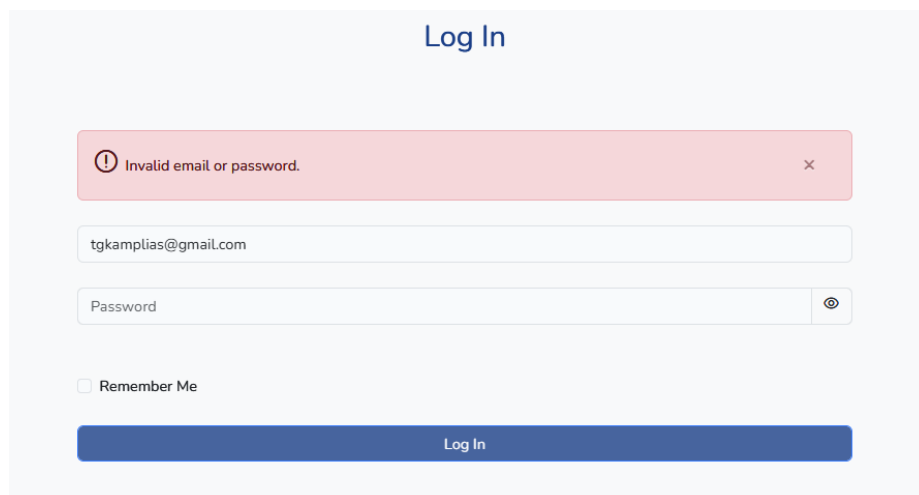
Οθόνη Εισόδου (Log In)

Ο χρήστης μεταβαίνει αρχικά στη σελίδα εισόδου της εφαρμογής, όπου εμφανίζεται μια φόρμα με τα πεδία:

- Email
- Password

και ένα προαιρετικό Remember Me checkbox.

Στο παράδειγμά μας, ο χρήστης συμπληρώνει το πεδίο Email με τη διεύθυνσή του (π.χ. *user@example.com*) και το πεδίο Password με τον κωδικό του. Στη συνέχεια, πατάει το κουμπί Log In για να αιτηθεί πρόσβαση στο σύστημα.



Log In

Invalid email or password.

tgkamplias@gmail.com

Password

☐ Remember Me

Log In

Σχήμα 12: Αποτυχημένη προσπάθεια εισόδου με μήνυμα ενημέρωσης

Αποτυχημένη προσπάθεια εισόδου με μήνυμα ενημέρωσης

Σε περίπτωση που τα διαπιστευτήρια του χρήστη (Email ή Password) δεν είναι σωστά, εμφανίζεται ένα κόκκινο πλαίσιο στην σελίδα με το μήνυμα “Invalid email or password.”. Ο χρήστης βλέπει ξεκάθαρα ότι πρέπει να ελέγξει τα στοιχεία του και να τα επανεισάγει.

- Αν πατηθεί ξανά το κουμπί Log In με τα ορθά στοιχεία, η εφαρμογή επιτρέπει την πρόσβαση στο *Diary* (την αρχική σελίδα του συστήματος).
- Αν τα στοιχεία παραμένουν λανθασμένα, η ίδια ειδοποίηση εξακολουθεί να εμφανίζεται, προτρέποντας τον χρήστη να ελέγξει ξανά την ορθότητα του email ή του κωδικού πρόσβασης.

Η συγκεκριμένη λειτουργικότητα ακολουθεί την απαίτηση 2 (Σύνδεση χρήστη στο σύστημα), όπως περιγράφεται στην *Ενότητα 5.1.1*, καθώς παρέχει έναν απλό και κατανοητό τρόπο επαλήθευσης των στοιχείων του χρήστη πριν του επιτραπεί η είσοδος στην εφαρμογή. Επιπλέον, δίνει άμεσα στον χρήστη σαφή ενημέρωση σε περίπτωση σφάλματος, ώστε να προχωρήσει σε διορθωτικές ενέργειες.

Στη συνέχεια, στο Σχήμα 13 απεικονίζεται η οθόνη εγγραφής ενός χρήστη:

The screenshot shows a web application interface for user registration. The top navigation bar has 'Log In' on the left and 'Sign Up' on the right. The main content area is divided into two panels. The left panel, titled 'Sign Up', contains a form with the following fields: 'First Name' (filled with 'Thanos'), 'Last Name' (filled with 'Gkamplias'), 'Email' (filled with 'tgkamplias@gmail.com'), 'Date of Birth' (filled with '10/05/1978'), 'Gender' (a dropdown menu showing 'Male'), 'Password' (masked with dots), and 'Confirm Password' (masked with dots). A blue 'Register' button is positioned below the form. The right panel has a dark blue background with the ChatGLUCO logo and the text 'Welcome to ChatGLUCO'.

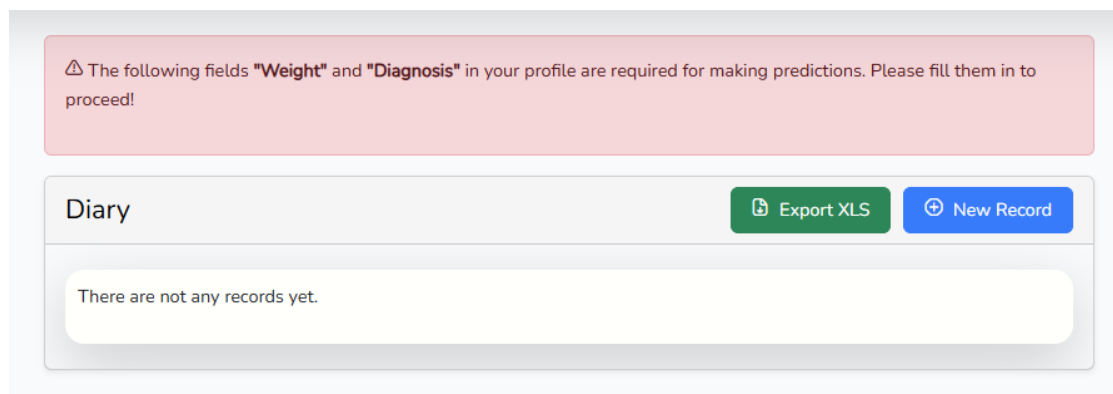
Σχήμα 13: Οθόνη Εγγραφής (Sign Up)

Οθόνη Εγγραφής (Sign Up)

Εφόσον ο χρήστης δεν διαθέτει λογαριασμό στην εφαρμογή, μεταφέρεται στη σελίδα εγγραφής. Εκεί καλείται να συμπληρώσει τα υποχρεωτικά πεδία που προβλέπει η *Απαίτηση 1 (Εγγραφή νέου χρήστη)*, δηλαδή:

- First Name
- Last Name
- Email
- Date of Birth
- Gender
- Password
- Confirm Password

Στο παράδειγμα της εικόνας, ο χρήστης συμπληρώνει τα αντίστοιχα πεδία (π.χ., Thanos, Gkamplias, ...). Πατώντας το κουμπί Register, το σύστημα επαληθεύει αν τα στοιχεία είναι ορθά και, εφόσον όλα είναι έγκυρα, δημιουργεί τον νέο λογαριασμό.



Σχήμα 14: Προειδοποιητικό μήνυμα ανεπάρκειας στοιχείων στην οθόνη Diary

Προειδοποιητικό μήνυμα ανεπάρκειας στοιχείων στην οθόνη Diary

Αμέσως μετά την επιτυχή εγγραφή, ο χρήστης ανακατευθύνεται στο *Dashboard* (στην προεπιλεγμένη καρτέλα “Diary”), όπως ορίζει η *Απαίτηση 1* περί ανακατεύθυνσης μετά το Sign Up. Στο συγκεκριμένο στιγμιότυπο, επάνω από την περιοχή του ημερολογίου εμφανίζεται ένα κόκκινο προειδοποιητικό πλαίσιο με το μήνυμα:

*“The following fields **Weight** and **Diagnosis** in your profile are required for making predictions. Please fill them in to proceed!”*

Αυτό συμβαδίζει με την *Απαίτηση 4 (Τροποποίηση Υγειονομικών Στοιχείων)*, σύμφωνα με την οποία τα πεδία “Weight” και “Diagnosis” είναι κρίσιμα για τη λειτουργία πρόβλεψης δόσεων ινσουλίνης και επιπέδων γλυκόζης. Επομένως, η εφαρμογή ενημερώνει τον χρήστη ότι πρέπει να μεταβεί στην ενότητα “My Profile” για να συμπληρώσει αυτά τα δεδομένα πριν χρησιμοποιήσει τις λειτουργίες πρόβλεψης.

Επίσης, σε αυτό το πρώτο άνοιγμα της σελίδας “Diary” εμφανίζεται η ένδειξη

“There are not any records yet.”

αποδεικνύοντας ότι ο ημερολογιακός πίνακας του χρήστη είναι ακόμα άδειος, καθώς ο λογαριασμός δημιουργήθηκε μόλις τώρα.

Με αυτόν τον τρόπο, η εφαρμογή καθοδηγεί ομαλά τον νέο χρήστη:

- Πρώτα, τον βοηθά να δημιουργήσει λογαριασμό (Sign Up).
- Στη συνέχεια, τον μεταφέρει αυτόματα στο ημερολόγιο (Diary), όπου του υπενθυμίζει ότι πρέπει να συμπληρώσει βασικά υγειονομικά στοιχεία στο προφίλ του πριν ξεκινήσει να καταχωρεί μετρήσεις γλυκόζης ή να λαμβάνει προτάσεις δόσης ινσουλίνης.

Ακολουθώντας, στο Σχήμα 15 απεικονίζεται η σελίδα “My Profile”, η οποία αποτελεί την κύρια ενότητα για την επεξεργασία των ιατρικών στοιχείων και την αντιμετώπιση του προειδοποιητικού μηνύματος που εμφανίστηκε στη σελίδα *Diary*.

The screenshot displays a user profile interface. At the top, there is a header section with a circular profile picture placeholder (a blue circle) and the name 'Thanos Gkamplias' with the role 'Patient' below it. To the right of the name are two buttons: 'Add Photo' and 'Save'. Below this header is a section titled 'Personal Information' containing a table with the following data: First Name: Thanos, Last Name: Gkamplias, Email Address: tgkamplias@gmail.com, Date of Birth: 1978-05-10, and Gender: Male. Below the personal information is a section titled 'Health Information' which contains two input fields: 'Diagnosis:' with a dropdown menu showing '--Select Diagnosis--' and 'Weight:' with a text input field labeled 'Enter Weight'.

Σχήμα 15: Οθόνη Προφίλ Χρήστη. Συμπλήρωση και Ενημέρωση Πληροφοριών Υγείας για την αντιμετώπιση του προειδοποιητικού μηνύματος από το Σχήμα 14.

Συμπλήρωση και Ενημέρωση Πληροφοριών Υγείας

- Στην ενότητα Health Information, ο χρήστης καλείται να συμπληρώσει υποχρεωτικά το πεδίο Diagnosis (Type I ή Type II) και το πεδίο Weight (σε κιλά). Αυτά τα δύο στοιχεία είναι απαραίτητα για να λειτουργήσουν οι μηχανισμοί πρόβλεψης της εφαρμογής (Απαίτηση 4).
- Με την επιλογή της κατάλληλης τιμής στη λίστα Diagnosis και την εισαγωγή του βάρους στο πεδίο Weight, ο χρήστης μπορεί να προχωρήσει στην ενημέρωση του προφίλ πατώντας το κουμπί Save Changes.

Personal Information

First Name: Thanos

Date of Birth: 1978-05-10

Last Name: Gkamplias

Gender: Male

Email Address: tgkamplias@gmail.com

Health Information

Diagnosis:

Type I

Weight:

72.00

HbA1c Measurements*

Latest HbA1c (%):

7.1

Equivalent Avg Glucose: 157.1 mg/dL

Second Latest HbA1c (%):

6.2

Equivalent Avg Glucose: 131.2 mg/dL

Third Latest HbA1c (%):

4.7

Equivalent Avg Glucose: 88.2 mg/dL

*Significance of HbA1c Levels: Normal Levels: Below 5.7% (Normal),
Prediabetes: 5.7% - 6.4% (Prediabetes), Diabetes: 6.5% and above (Diabetes).

Save Changes

Σχήμα 16: Προβολή και Επεξεργασία του ποσοστού HbA1c.

Προβολή και Επεξεργασία του ποσοστού HbA1c

Επιπρόσθετα, ο χρήστης έχει τη δυνατότητα να καταχωρίσει τις τρεις τελευταίες τιμές της εξέτασης HbA1c:

- Latest HbA1c (%)
- Second Latest HbA1c (%)
- Third Latest HbA1c (%)

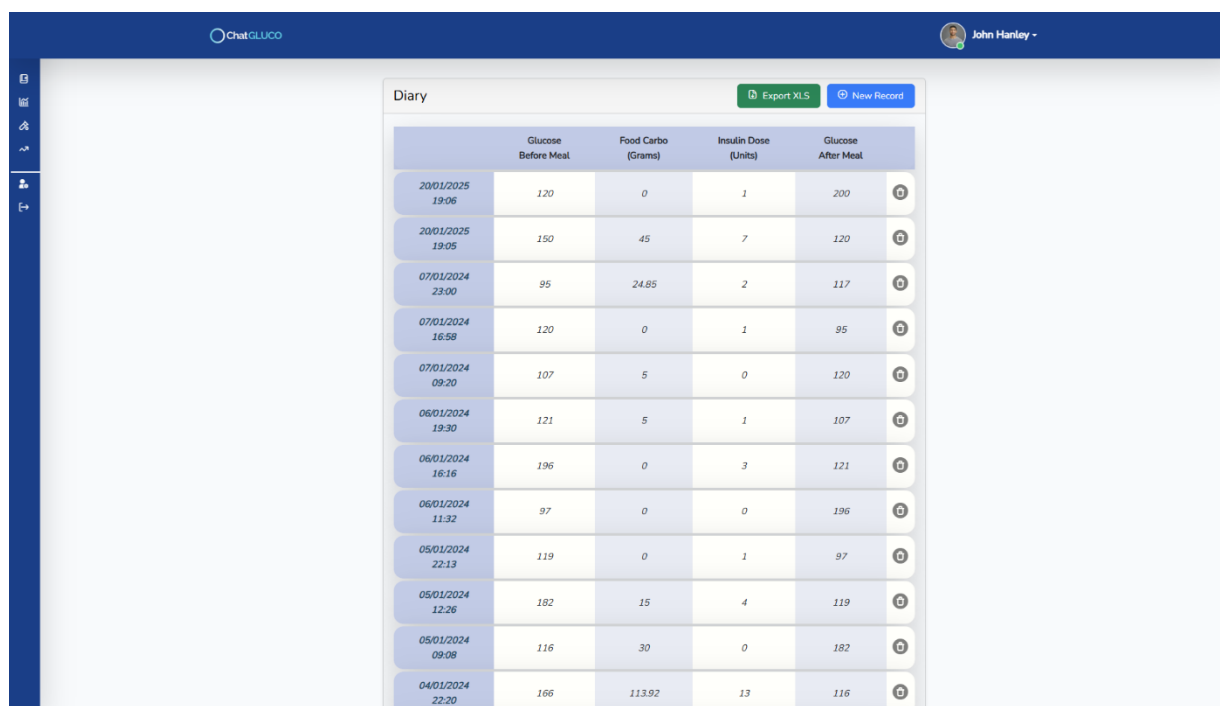
Για κάθε τιμή HbA1c, το σύστημα εμφανίζει ακριβώς από κάτω την αντίστοιχη ισοδύναμη μέση τιμή γλυκόζης (mg/dL), κάνοντας χρήση έγχρωμης επισήμανσης:

- Πράσινο για φυσιολογικά επίπεδα (Normal)
- Πορτοκαλί για προδιαβήτη (Prediabetes)
- Κόκκινο για διαβήτη (Diabetes)

Η χρήσιμη αυτή λειτουργία διευκολύνει τον χρήστη να ερμηνεύει τα αποτελέσματά του, βλέποντας με μια ματιά σε ποια κατηγορία εμπίπτουν οι μετρήσεις του. Παράλληλα, κάτω από τις τιμές HbA1c εμφανίζεται και μια σύντομη αναφορά στις τυπικές κατηγορίες φυσιολογικών και παθολογικών τιμών.

Με την ολοκλήρωση της επεξεργασίας στο προφίλ και την αποθήκευσή της, ο χρήστης καλύπτει τις ανάγκες της *Απαίτησης 4 (Τροποποίηση Υγειονομικών Στοιχείων)*. Πλέον, η εφαρμογή διαθέτει τα απαραίτητα στοιχεία για να πραγματοποιήσει ακριβέστερες προβλέψεις δόσης ινσουλίνης και επιπέδων γλυκόζης, ενώ το σύστημα ενημερώνει τον χρήστη με μήνυμα επιτυχίας για την ανανέωση του προφίλ του.

Παρακάτω, στα Σχήματα 17, 18, 19 αποτυπώνονται τρία βασικά βήματα στη ροή διαχείρισης δεδομένων της εφαρμογής, όπως αυτά καλύπτουν τις απαιτήσεις για το Diary και την Εξαγωγή Δεδομένων:



The screenshot displays the 'Diary' section of the ChatGLUCC application. At the top, there are buttons for 'Export XLS' and 'New Record'. The main area contains a table with the following columns: 'Date & Time', 'Glucose Before Meal', 'Food Carbo (Grams)', 'Insulin Dose (Units)', and 'Glucose After Meal'. Each row represents a log entry with a delete icon on the right.

Date & Time	Glucose Before Meal	Food Carbo (Grams)	Insulin Dose (Units)	Glucose After Meal
20/01/2025 19:06	120	0	1	200
20/01/2025 19:05	150	45	7	120
07/01/2024 23:00	95	24.85	2	117
07/01/2024 16:58	120	0	1	95
07/01/2024 09:20	107	5	0	120
06/01/2024 19:30	121	5	1	107
06/01/2024 16:16	196	0	3	121
06/01/2024 11:32	97	0	0	196
05/01/2024 22:13	119	0	1	97
05/01/2024 12:26	182	15	4	119
05/01/2024 09:08	116	30	0	182
04/01/2024 22:20	166	113.92	13	116

Σχήμα 17: Διαχείριση Δεδομένων Χρήστη με την χρήση Ημερολογίου (Diary)

Ημερολόγιο (Diary)

Εφόσον έχουν συμπληρωθεί τα βασικά ιατρικά στοιχεία, ο χρήστης βλέπει πλέον έναν πίνακα με όλες τις υπάρχουσες εγγραφές στο ημερολόγιό του. Σε κάθε γραμμή αναγράφονται η ημερομηνία και η ώρα καταγραφής (Date & Time), τα επίπεδα γλυκόζης πριν το γεύμα (Glucose Before), οι υδατάνθρακες του γεύματος (Food Carbo), η δόση ινσουλίνης (Insulin Dose), και τυχόν τιμή γλυκόζης μετά το γεύμα (Glucose After). Στη δεξιά στήλη διακρίνεται ένα εικονίδιο κάδου (Delete), με το οποίο ο χρήστης μπορεί να διαγράψει μία εγγραφή, ικανοποιώντας την *Απαίτηση 10 (Διαγραφή εγγραφών)*.

Στο επάνω μέρος της καρτέλας υπάρχουν δύο κουμπιά:

- **Export XLS:** Δίνει τη δυνατότητα εξαγωγής όλων των εγγραφών του ημερολογίου σε αρχείο Excel (XLS), *Απαίτηση 11 (Εξαγωγή δεδομένων)*.
- **New Record:** Μεταφέρει τον χρήστη στη φόρμα δημιουργίας νέας εγγραφής (περιγράφεται στο επόμενο βήμα).

Add New Record

Date & Time
26/01/2025 22:18

Glucose Before Meal	Food Carbo in Grams	Insulin Dose in Units	Glucose After Meal
182	60.67	3	121

Save

Calculate Insulin Dose
Sliding Scale + Carbo Calculation

Search Food
For the calculation of the carbohydrates you have consumed, please search for the foods you ate in the list below.

salad

sandwich

slices of bread 2 56.46 Calculate Carbohydrates

salad

cups 1 4.21 Calculate Carbohydrates

Σχήμα 18: Καταχώρηση Νέας Εγγραφής στο ημερολόγιο (Diary)

Καταχώρηση Νέας Εγγραφής

Πατώντας το κουμπί New Record, εμφανίζεται ένα pop-up για την εισαγωγή μιας καινούργιας εγγραφής ημερολογίου, βάσει της *Απαίτησης 8 (Δημιουργία νέας εγγραφής)*.

Ο χρήστης ορίζει την ημερομηνία και ώρα (Date & Time) και υποχρεωτικά εισάγει τιμή γλυκόζης πριν το γεύμα (Glucose Before Meal). Η εφαρμογή χρησιμοποιεί χρώματα (π.χ.

κόκκινο) για να επισημάνει υψηλές, πράσινο για φυσιολογικές, και κίτρινο για πολύ χαμηλές τιμές.

Επιπλέον, ο χρήστης έχει τη δυνατότητα να καταχωρίσει προαιρετικά:

Food Carbo in Grams: Είτε με μη αυτόματο υπολογισμό είτε μέσω του Search Food (όπως φαίνεται στο στιγμιότυπο, όπου ο χρήστης πληκτρολογεί “salad”, “sandwich” κ.λπ. και η εφαρμογή κάνει αυτόματο υπολογισμό υδατανθράκων).

Insulin Dose in Units: Είτε χειροκίνητα είτε επιλέγοντας αλγόριθμο από το dropdown (Calculate Insulin Dose), όπου διατίθενται οι διάφορες μέθοδοι πρόβλεψης (Linear Regression, XGBoost, Random Forest, Sliding Scale + Carbo Calculation).

Glucose After Meal: Προαιρετική τιμή της γλυκόζης μετά το γεύμα για σύγκριση και παρακολούθηση.

Μετά την εισαγωγή των επιθυμητών τιμών, ο χρήστης πατάει Save για να ολοκληρώσει τη δημιουργία της εγγραφής.

	A	B	C	D	E	F
1	Patient's Info					
2	Name	Hanley John				
3	Gender	male				
4	Birth date	1963-08-15				
5	Weight	80				
6	Diagnosis	Type II				
7						
8	Patient's Records					
9	Date	Glucose Before	Food Carbo	Insulin Dose	Glucose After	Weight
10	2025-01-20 19:06:00	120		1	200	80
11	2025-01-20 19:05:00	150	45	7	120	80
12	2024-01-07 23:00:00	95	24,85	2	117	80
13	2024-01-07 16:58:00	120		1	95	80
14	2024-01-07 09:20:00	107	5		120	80
15	2024-01-06 19:30:00	121	5	1	107	80
16	2024-01-06 16:16:00	196		3	121	80
17	2024-01-06 11:32:00	97			196	80
18	2024-01-05 22:13:00	119		1	97	80
19	2024-01-05 12:26:00	182	15	4	119	80
20	2024-01-05 09:08:00	116	30		182	80
21	2024-01-04 22:20:00	166	113,92	13	116	80
22	2024-01-04 15:46:00	132	8		166	80
23	2024-01-04 09:26:00	98	8		132	80
24	2024-01-03 22:42:00	108	26	3	98	80
25	2024-01-03 16:26:00	129	30	3	108	80
26	2024-01-03 08:18:00	107	38		129	80
27	2024-01-02 22:17:00	77	20		107	80
28	2024-01-02 16:06:00	92	12	2	77	80
29	2024-01-02 07:46:00	127	26	3	92	80
30	2024-01-01 19:37:00	89	15	2	127	80
31	2024-01-01 13:11:00	131	33	5	89	80
32	2024-01-01 07:27:00	117	42	1	131	80

Σχήμα 19: Εξαγωγή Αναφοράς Δεδομένων Ασθενούς

Εξαγωγή Αναφοράς Δεδομένων Ασθενούς

Η εφαρμογή επιτρέπει την εξαγωγή όλων των εγγραφών του ημερολογίου σε μορφή Excel. Το στιγμιότυπο δείχνει ένα παραγόμενο XLS αρχείο που περιέχει:

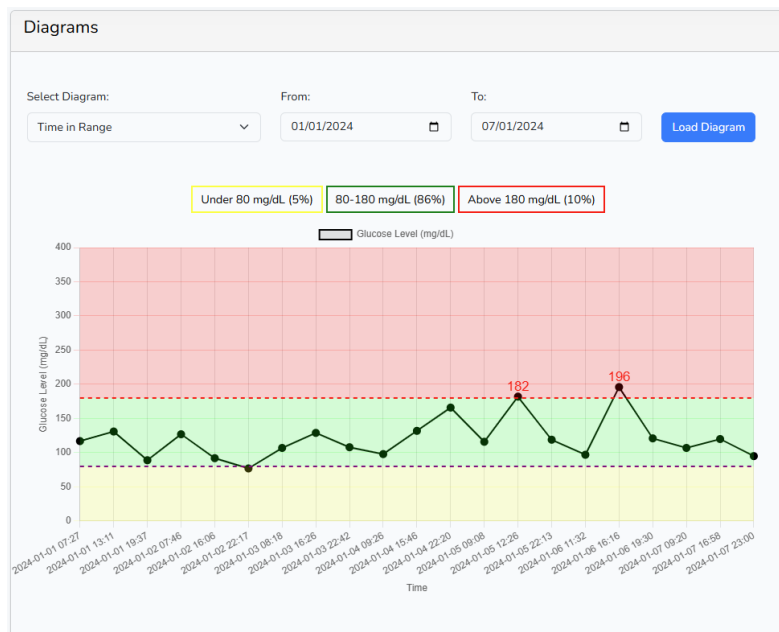
Τα βασικά στοιχεία του ασθενούς (Name, Gender, Birth Date, Weight, Diagnosis).

Τον πίνακα εγγραφών Date, Glucose Before, Food Carbo, Insulin Dose, Glucose After, κ.λπ., όπως έχουν καταγραφεί στην εφαρμογή.

Αυτή η δυνατότητα καλύπτει πλήρως την *Απαιτήση 11 (Εξαγωγή δεδομένων)*, εξασφαλίζοντας ότι ο χρήστης μπορεί να τηρήσει και να μοιραστεί το ιστορικό του με επαγγελματίες υγείας ή να το αποθηκεύσει για προσωπικό αρχείο.

Με αυτές τις λειτουργίες, το *Diary* γίνεται ο βασικός χώρος παρακολούθησης της διατροφής, της χορηγούμενης ινσουλίνης, αλλά και των μετρήσεων γλυκόζης, παρέχοντας μια ολοκληρωμένη εικόνα της διαχείρισης του διαβήτη στον χρήστη.

Στη συνέχεια, στα σχήματα 20, 21, 22 αναδεικνύονται οι δυνατότητες οπτικοποίησης των δεδομένων γλυκόζης και η πρόβλεψη μελλοντικών τιμών, όπως περιγράφονται στις *Απαιτήσεις 12 (Οπτικοποίηση Δεδομένων)* και *14 (Προβολή Μελλοντικής Τάσης Γλυκόζης)*.

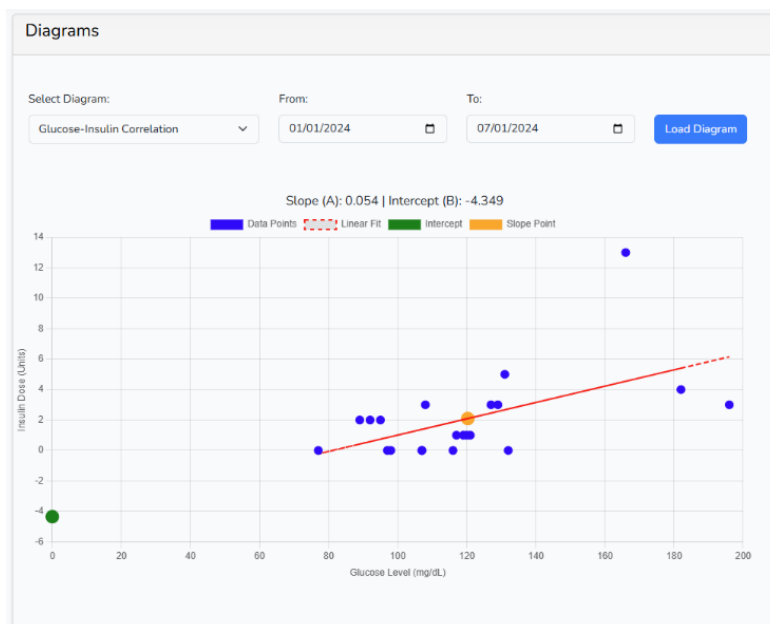


Σχήμα 20: Διάγραμμα Time In Range

Χρονικές Τάσεις Γλυκόζης (Time in Range)

Στην ενότητα “Diagrams” της εφαρμογής, ο χρήστης μπορεί να επιλέξει το διάγραμμα *Time in Range*, το οποίο απεικονίζει τα επίπεδα γλυκόζης του για ένα επιλεγμένο χρονικό διάστημα. Από την ειδική φόρμα επιλογής ημερομηνιών, ο χρήστης ορίζει το χρονικό εύρος (From, To) που επιθυμεί να εξετάσει.

Πατώντας το κουμπί Load Diagram, αποδίδεται γραφικά η πορεία της γλυκόζης, αναδεικνύοντας έτσι περιόδους με υπογλυκαιμία (κάτω από 80 mg/dL), φυσιολογικά επίπεδα (80-180 mg/dL) ή υπεργλυκαιμία (πάνω από 180 mg/dL).



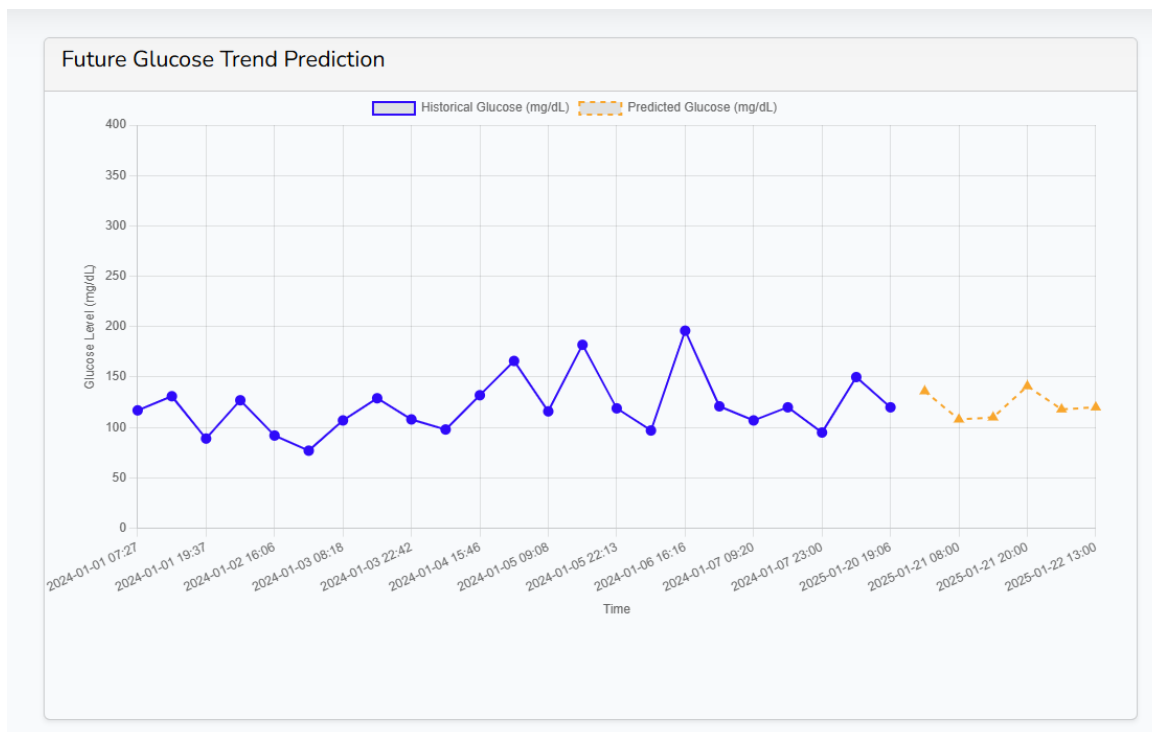
Σχήμα 21: Διάγραμμα Glucose – Insulin Correlation

Συσχέτιση Γλυκόζης-Ινσουλίνης (Glucose-Insulin Correlation)

Στην ίδια ενότητα υπάρχει η επιλογή *Glucose-Insulin Correlation*, που επικεντρώνεται στη σχέση των τιμών γλυκόζης με τη χορηγούμενη ποσότητα ινσουλίνης.

Το διάγραμμα αυτό βοηθά τους χρήστες να εντοπίσουν μοτίβα και να κατανοήσουν πώς οι αλλαγές στις δόσεις ινσουλίνης επηρεάζουν τα επίπεδα γλυκόζης με την πάροδο του χρόνου.

Ομοίως, το κουμπί Load Diagram εμφανίζει τα δεδομένα εντός του επιλεγμένου χρονικού διαστήματος, καθιστώντας εμφανείς τις τυχόν αυξομειώσεις και τις επιπτώσεις τους στον γλυκαιμικό έλεγχο.



Σχήμα 22: Διάγραμμα Future Glucose Trend Prediction

Προβολή Μελλοντικής Τάσης Γλυκόζης (Future Glucose Trend Prediction)

Σε ξεχωριστή σελίδα ο χρήστης έχει πρόσβαση στο διάγραμμα πρόβλεψης γλυκόζης για τις επόμενες δύο ημέρες. Αυτή η λειτουργία αναφέρεται στην *Απαίτηση 14*.

Η μπλε γραμμή απεικονίζει τα ιστορικά δεδομένα του χρήστη μέχρι την πιο πρόσφατη καταχώριση, ενώ η διακεκομμένη πορτοκαλί γραμμή ξεκινά αμέσως μετά το τελευταίο σημείο μέτρησης, παρουσιάζοντας τις προβλεπόμενες τιμές γλυκόζης (πρωί, μεσημέρι, βράδυ) για τις επόμενες δύο ημέρες.

Η πρόβλεψη παράγεται από αλγόριθμο XGBoost μέσω ενός *Flask API*, αξιοποιώντας τα ιστορικά δεδομένα του χρήστη. Με αυτόν τον τρόπο, ο χρήστης αποκτά μια σφαιρική εικόνα τόσο της παρούσας κατάστασης όσο και της πιθανής εξέλιξης των επιπέδων γλυκόζης, βοηθώντας τον στον προγραμματισμό της θεραπευτικής του αγωγής.

Συνολικά, αυτές οι οπτικοποιήσεις διευκολύνουν την παρακολούθηση της γλυκόζης, την κατανόηση της δράσης της ινσουλίνης και την έγκαιρη προσαρμογή της θεραπείας, αναδεικνύοντας τη σημασία του *chatGLUCO* ως ολοκληρωμένο εργαλείο διαχείρισης διαβήτη.

Τέλος, στα *σχήματα 23, 24* παρουσιάζεται η σελίδα όπου ο χρήστης μπορεί να προβλέψει τη μελλοντική τιμή της γλυκόζης του μετά από ένα γεύμα, αξιοποιώντας τόσο υπολογιστικές όσο και μηχανικές μεθόδους μάθησης, σύμφωνα με την *Απαίτηση 13 (Δυνατότητα Πρόβλεψης Γλυκόζης)*.

The screenshot shows a web application titled "Glucose Prediction". At the top, a green box displays "Predicted Glucose Level: 119 mg/dL" with a checkmark icon. Below this, there are several input fields: "Current Glucose Level (mg/dL):" with the value "152", "Food Carbohydrates (grams):" with the value "26", and "Insulin Dose (units):" with the value "1". There are also two dropdown menus: "Calculation Mode:" set to "Machine Learning-Based Glucose Prediction" and "Select Algorithm:" set to "Gradient Boosting". At the bottom left is a blue button labeled "Predict Glucose" with a plus icon, and at the bottom right is a grey button labeled "Clear Fields".

Σχήμα 23: Διεπαφή Πρόβλεψης Γλυκόζης

Διεπαφή Πρόβλεψης Γλυκόζης

Στην κορυφή εμφανίζεται ένα πλαίσιο που παρουσιάζει το Predicted Glucose Level (mg/dL), χρωματικά κωδικοποιημένο ανάλογα με την τιμή.

Ο χρήστης εισάγει τις παρακάτω τιμές:

- Current Glucose Level (mg/dL): Τιμή γλυκόζης πριν το γεύμα.
- Food Carbohydrates (grams): Οι υδατάνθρακες που πρόκειται να καταναλωθούν (προαιρετικό).
- Insulin Dose (units): Η δόση ινσουλίνης που χορηγείται (υποχρεωτικό)
- Calculation Mode:
 - *Postprandial Glucose Calculation*
 - *Machine Learning-Based Glucose Prediction*
- Select Algorithm: Εμφανίζεται μόνο στη δεύτερη περίπτωση (π.χ. Linear Regression, Random Forest, XGBoost), και αφορά τον αλγόριθμο που θα χρησιμοποιηθεί για την πρόβλεψη.

Με το κουμπί Predict Glucose, η εφαρμογή υπολογίζει τα επίπεδα γλυκόζης μετά το γεύμα.

Ανάλογα με τη διαμόρφωση της σελίδας, υπάρχει το κουμπί Clear Fields για γρήγορη διαγραφή όλων των τιμών από τα πεδία.

Glucose Prediction Categories:

- **Yellow Indicator:** Glucose Levels below 80 mg/dL (Low)
- **Green Indicator:** Glucose Levels between 80 and 180 mg/dL (Normal)
- **Red Indicator:** Glucose Levels above 180 mg/dL (High)

How is the new glucose level calculated?

This process, referred to as **Postprandial Glucose Calculation**, is based on the patient's weight, diagnosis, carbohydrate intake, and the insulin dose administered. These factors are combined to provide an accurate estimation of blood glucose levels after a meal.

- **Total Daily Dose (TDD):** Calculated based on the patient's weight:
 - For Type I: $TDD = Weight \times 0.5$
 - For Type II: $TDD = Weight \times 0.8$
- **Carbohydrate Ratio (CR):** Determines how many grams of carbohydrates 1 unit of insulin can cover. Calculated as: $CR = 500 / TDD$.
- **Insulin Sensitivity Factor (ISF):** Indicates how much glucose is reduced by 1 unit of insulin. Calculated as:
 - For Type I: $ISF = 1800 / TDD$
 - For Type II: $ISF = 1500 / TDD$

The predicted glucose level is calculated by considering:

- The glucose level before the meal.
- The increase caused by the carbohydrates consumed.
- The reduction due to the insulin dose.

Using these parameters, the system provides an accurate estimation of post-meal glucose levels, helping patients better manage their blood sugar.

Σχήμα 24: Επεξήγηση Υπολογισμού Πρόβλεψης Γλυκόζης

Επεξήγηση Υπολογισμού (Postprandial Glucose Calculation)

Κάτω από τη φόρμα εισαγωγής, υπάρχει ένα ενημερωτικό πλαίσιο που εξηγεί λεπτομερώς πώς υπολογίζεται η γλυκόζη μετά το γεύμα, αν επιλεγεί η μέθοδος *Postprandial Glucose Calculation*. Περιγράφεται, επίσης, η χρωματική κωδικοποίηση του αποτελέσματος: (*Yellow*: < 80 mg/dL (Low), *Green*: 80-180 mg/dL (Normal), *Red*: > 180 mg/dL (High)).

Χάρη σε αυτές τις λειτουργίες, το *chatGLUCO* δίνει στον χρήστη τη δυνατότητα να προβλέψει με ακρίβεια τις μελλοντικές τιμές της γλυκόζης. Είτε χρησιμοποιείται ο αλγόριθμος μηχανικής μάθησης είτε ο μεταγευματικός υπολογισμός (Postprandial), ο χρήστης λαμβάνει άμεση οπτική ανατροφοδότηση (χρωματικές ενδείξεις) και κατανοεί τις

βασικές αρχές που διέπουν τον υπολογισμό, κλείνοντας έτσι τον κύκλο των λειτουργιών που καλύπτουν τη διαχείριση και την πρόβλεψη των επιπέδων γλυκόζης.

6 Συμπεράσματα και Μελλοντική Εργασία

6.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία ανέπτυξε ένα ολοκληρωμένο σύστημα διαχείρισης του διαβήτη, συνδυάζοντας κλινικά πρωτόκολλα (Sliding Scale, Carbohydrate Calculation, Weight-based Insulin Prediction) με προηγμένους αλγορίθμους Μηχανικής Μάθησης. Βασικός στόχος ήταν η δημιουργία μιας πλατφόρμας που βελτιστοποιεί την πρόβλεψη δόσεων ινσουλίνης και επιπέδων γλυκόζης, προσφέροντας ταυτόχρονα ένα εύχρηστο εργαλείο για την καθημερινή αυτοδιαχείριση των ασθενών. Η εφαρμογή chatGLUCO, που αναπτύχθηκε σε Laravel και Flask, ενσωμάτωσε λειτουργίες όπως η καταγραφή δεδομένων, η αυτόματη πρόβλεψη μέσω μοντέλων ML και η οπτικοποίηση δεδομένων, αποτελώντας μια πρακτική λύση για ασθενείς και ιατρούς.

Κύρια Ευρήματα & Συνεισφορά

Τα πειράματα ανάλυσης δεδομένων ανέδειξαν σημαντικές διαφορές στην απόδοση των μοντέλων ανάλογα με το σενάριο και το dataset. Για τους ασθενείς της ΜΕΘ, ο Random Forest αποδείχθηκε ως ο καταλληλότερος αλγόριθμος μηχανικής μάθησης, με μέσο απόλυτο σφάλμα (MAE) 1.85 στις προβλέψεις δόσεων ινσουλίνης και MAE 31.08 mg/dL για τη γλυκόζη, δείχνοντας ισορροπία μεταξύ ακρίβειας και σταθερότητας. Αντίθετα, στο dataset της εφαρμογής chatGLUCO, το XGBoost διακρίθηκε στην πρόβλεψη γλυκόζης (MAE 31.26 mg/dL). Η παραδοσιακή μέθοδος Sliding Scale Insulin (SSI) παρέμεινε ανταγωνιστική, ιδίως στο Time In Range (TIR 91.1%) και στην πρόβλεψη των δόσεων ινσουλίνης με μέσο απόλυτο σφάλμα (MAE) 0.90, υπογραμμίζοντας τη σημασία της συνύπαρξης κλασικών και σύγχρονων μεθόδων. Η υλοποίηση του συστήματος επιβεβαίωσε ότι η ενσωμάτωση ML βελτιώνει την εξατομίκευση των προβλέψεων, ιδίως όταν λαμβάνονται υπόψη δημογραφικά και χρονικά χαρακτηριστικά.

Το σύστημα προσφέρει μια διπλή συνεισφορά:

1. **Κλινική:** Απλοποιεί τη ρύθμιση της ινσουλίνης μειώνοντας τον κίνδυνο υπογλυκαιμίας ή υπεργλυκαιμίας. Χρήστες μπορούν να αλληλεπιδρούν με γραφήματα και προβλέψεις σε πραγματικό χρόνο, ενισχύοντας τον έλεγχο της θεραπείας.
2. **Τεχνολογική:** Η τεσσάρων επιπέδων αρχιτεκτονική (Presentation Tier, Business Logic Tier, Machine Learning Tier και Data Tier) εξασφαλίζει επεκτασιμότητα, ενώ η

χρήση RESTful endpoints διευκολύνει την ενσωμάτωση νέων αλγορίθμων ή εξωτερικών συσκευών (π.χ., CGM).

Πρακτική Υλοποίηση και Συνεισφορά

Η ανάπτυξη της web εφαρμογής chatGLUCO (Laravel, Flask API) αποτελεί βασικό πρακτικό αποτέλεσμα της εργασίας. Το σύστημα ενσωματώνει λειτουργίες αυτόματων υπολογισμών (π.χ. υδατάνθρακες, δόσεις ινσουλίνης, επίπεδα γλυκόζης), οπτικοποίησης δεδομένων (Chart.js), και εξατομικευμένων προβλέψεων, προσφέροντας μια εύχρηστη πλατφόρμα για καθημερινή διαχείριση. Η συγκριτική ανάλυση με άλλες εφαρμογές (Πίνακας 1) υπενθυμίζει τη μοναδικότητα του chatGLUCO, ιδίως στην ενσωμάτωση ML με δυνατότητες όπως η πρόβλεψη γλυκόζης postprandial και η χρήση πληθώρας διαγραμμάτων.

Περιορισμοί και Προοπτικές

Παρά τα θετικά αποτελέσματα, η έρευνα αντιμετώπισε περιορισμούς, όπως ο μικρός όγκος δεδομένων (38 ασθενείς στο Dataset A) και η έλλειψη πραγματικών μετρήσεων από φορητές συσκευές.

Η εργασία επιτέλεσε με επιτυχία τους αρχικούς στόχους της, αναδεικνύοντας τη δυνατότητα συνύπαρξης μηχανικής μάθησης και κλινικών πρωτοκόλλων σε πλαίσια αυτοδιαχείρισης. Ωστόσο, η πλήρης αξιοποίηση του δυναμικού του συστήματος απαιτεί συνεχή βελτίωση, διεύρυνση των δεδομένων και διασύνδεση με το ευρύτερο οικοσύστημα υγείας. Με τις κατάλληλες επεκτάσεις, η chatGLUCO μπορεί να αποτελέσει ένα σημαντικό εργαλείο στην ιατρική, βοηθώντας ασθενείς και ιατρούς να λαμβάνουν εμπειριστατωμένες αποφάσεις σε πραγματικό χρόνο.

6.2 Μελλοντική Εργασία και Βελτιώσεις

Η παρούσα διπλωματική εργασία αποτελεί ένα πρώτο, ολοκληρωμένο βήμα για τη βέλτιστη και εξατομικευμένη διαχείριση του διαβήτη, συνδυάζοντας αλγορίθμους μηχανικής μάθησης και μια λειτουργική web εφαρμογή. Αν και έχει επιτευχθεί σημαντική πρόοδος, υπάρχει ευρύ πεδίο για περαιτέρω τεχνολογικές και κλινικές βελτιώσεις. Ακολουθούν ορισμένες προτάσεις για μελλοντικές επεκτάσεις:

Στοχευμένη Εκπαίδευση Μοντέλων μέσω Ομαδοποίησης Ασθενών

Ένας σημαντικός άξονας βελτίωσης αφορά τη στοχευμένη εκπαίδευση των μοντέλων, ανάλογα με συγκεκριμένα δημογραφικά και κλινικά χαρακτηριστικά. Στο πλαίσιο αυτό, προτείνεται η ομαδοποίηση των ασθενών με βάση τη διάγνωση (Type I ή Type II), την ηλικιακή ομάδα και το σωματικό βάρος (κάτω ή πάνω από ένα όριο — π.χ. 70 κιλά). Η διαφοροποίηση αυτή θα επέτρεπε την προσαρμοσμένη εκπαίδευση των μοντέλων στις ανάγκες κάθε ομάδας, οδηγώντας σε ακριβέστερη πρόβλεψη και βελτιστοποίηση των συστάσεων για τη δόση ινσουλίνης.

Ενσωμάτωση Επιπρόσθετων Αλγορίθμων Μηχανικής Μάθησης

Η χρήση μοντέλων όπως Linear Regression, Random Forest και XGBoost για την πρόβλεψη δόσεων ινσουλίνης και επιπέδων γλυκόζης προσφέρει ήδη υψηλό βαθμό ακρίβειας. Ωστόσο, θα μπορούσαν να μελετηθούν και να υλοποιηθούν πιο προχωρημένες τεχνικές, όπως:

- **Νευρωνικά Δίκτυα** (π.χ. Multi-Layer Perceptrons)
- **Deep Learning** (π.χ. LSTM για χρονικές ακολουθίες)

Η υλοποίηση τους προϋποθέτει αυξημένη υπολογιστική ισχύ και ενδεχομένως μεγαλύτερη ποικιλία δεδομένων (π.χ. σωματική δραστηριότητα, γενετικοί δείκτες). Παρά την πολυπλοκότητα που συνεπάγεται, η ενσωμάτωσή τους θα μπορούσε να βελτιώσει περαιτέρω την ακρίβεια και την ευρωστία του συστήματος.

Βελτίωση της οπτικοποίησης των δεδομένων με νέα διαγράμματα

Η βελτίωση της οπτικοποίησης των δεδομένων θα προσφέρει μια πιο ολοκληρωμένη και ευκρινή κατανόηση της δυναμικής μεταξύ γλυκόζης και ινσουλίνης. Συγκεκριμένα, προτείνεται η επέκταση των διαγραμμάτων με:

- **Διαγράμματα μπαρών** που θα απεικονίζουν τη μέση τιμή γλυκόζης και ινσουλίνης ανά μήνα, επιτρέποντας στον χρήστη ή τον ιατρό να εντοπίσει μακροπρόθεσμες τάσεις.
- **Βελτίωση του υπάρχοντος διαγράμματος Time In Range**, ώστε να απεικονίζει τόσο τα επίπεδα γλυκόζης όσο και τις δόσεις ινσουλίνης. Με αυτόν τον τρόπο, θα διευκολυνθεί η κατανόηση της συσχέτισης μεταξύ των επιπέδων γλυκόζης και των αντίστοιχων δόσεων ινσουλίνης στο χρόνο.

Σύνδεση με Εξωτερικές Συσκευές και Wearables

Η εφαρμογή θα ήταν ιδιαίτερα ωφέλιμη εάν συγχρονιζόταν με έξυπνες συσκευές μέτρησης γλυκόζης ή wearables που συλλέγουν δεδομένα σε πραγματικό χρόνο (π.χ. smartwatch, βηματομετρητές). Με τον τρόπο αυτό θα επιτυγχάνεται αδιάκοπη ροή δεδομένων,

μειώνοντας τα σφάλματα που προκύπτουν από χειροκίνητη καταχώριση τιμών και εμπλουτίζοντας παράλληλα τα δεδομένα που τροφοδοτούνται στους αλγορίθμους πρόβλεψης. Για τη διασύνδεση απαιτείται ανάπτυξη καταλλήλων API και πιθανώς τροποποίηση του backend, ωστόσο η βελτίωση στην αποτελεσματικότητα των αλγορίθμων μπορεί να είναι ουσιαστική.

Αυτοματοποιημένη Ανάλυση και Ειδοποιήσεις

Η ανάπτυξη ενός υποσυστήματος που εντοπίζει αυτόματα απότομες μεταβολές γλυκόζης και προβλέπει επικίνδυνες καταστάσεις θα μπορούσε να βελτιώσει την αυτονομία των ασθενών.

- **Ευφυείς ειδοποιήσεις:** Προειδοποίηση σε περιπτώσεις υπο- ή υπεργλυκαιμίας
- **Προληπτικά μέτρα:** Έγκαιρη ενημέρωση του ασθενούς (ή/και του ιατρού) ώστε να προβεί σε αντίστοιχες ρυθμίσεις δόσης ή διατροφής

Η υλοποίηση αυτής της λειτουργικότητας προϋποθέτει ρυθμίσεις “alert thresholds” και σύστημα επικύρωσης ειδοποιήσεων (π.χ. αποφυγή υπέρμετρων “false positives”). Συνολικά, ένα σύστημα προβλεπτικής ανάλυσης θα μπορούσε να προλάβει σοβαρά επεισόδια και να ενισχύσει τον ενεργό ρόλο των ασθενών στη θεραπεία τους.

Εμπλουτισμός της Βάσης Δεδομένων και Προσωποποιημένα Προφίλ

Η συλλογή περισσότερων και πιο σύνθετων δεδομένων με την πάροδο του χρόνου θα ενισχύσει την ακρίβεια των μοντέλων και θα επιτρέψει πιο ολοκληρωμένες αναλύσεις.

- **Λεπτομερέστερα Προφίλ:** Καταγραφή επιπέδων δραστηριότητας, ωραρίου ύπνου, αναλυτικής διατροφής
- **Ευρύτερη Συμμετοχή Χρηστών:** Δεδομένα από μεγαλύτερη γεωγραφική περιοχή ή διαφορετικές κλινικές συνθήκες

Η εφαρμογή αυτών των προτάσεων απαιτεί υποδομές αποθήκευσης και μεθόδους επεξεργασίας μεγάλου όγκου δεδομένων (π.χ. big data pipelines), ωστόσο ανοίγει τον δρόμο για πιο εξατομικευμένες προβλέψεις. Η ποικιλία και η ποιότητα των δεδομένων είναι καθοριστική για την επιτυχία οποιουδήποτε αλγορίθμου, δικαιολογώντας επενδύσεις σε ευέλικτες υποδομές βάσεων δεδομένων.

Συμμόρφωση με Νέους Κανονισμούς και Επικύρωση σε Κλινικά Περιβάλλοντα

Η εφαρμογή, εφόσον μεγαλώσει η κλίμακα χρήσης της, θα πρέπει:

- **Να συμμορφώνεται πλήρως** με τυχόν νέους νομικούς κανονισμούς (π.χ. GDPR, HIPAA), διασφαλίζοντας την προστασία των προσωπικών δεδομένων.

- **Να υποβληθεί σε αυστηρές κλινικές δοκιμές,** σε συνεργασία με επαγγελματίες υγείας και ερευνητικά κέντρα, για την αξιολόγηση της αποτελεσματικότητας και της ασφάλειας. Οι διαδικασίες κλινικής επικύρωσης (π.χ. randomized clinical trials (RCTs)) αυξάνουν την αξιοπιστία της εφαρμογής και αποκαλύπτουν τυχόν σημεία που χρήζουν περαιτέρω προσαρμογών ή βελτιώσεων. Με την ολοκλήρωση τέτοιων ελέγχων, η εφαρμογή θα μπορέσει να διεκδικήσει επίσημη έγκριση για χρήση σε κλινικά περιβάλλοντα, ενισχύοντας την εμπιστοσύνη των ασθενών και των ιατρών.

7 Παράρτημα

7.1 Βιβλιοθήκες που χρησιμοποιήθηκαν

Βιβλιοθήκες που χρησιμοποιήθηκαν για την κατασκευή του web app (Laravel)

1. **laravel/framework**

Το βασικό πακέτο του Laravel framework, το οποίο παρέχει όλα τα απαραίτητα εργαλεία για την ανάπτυξη μιας σύγχρονης web εφαρμογής, όπως routing, middleware, και ενσωμάτωση με βάσεις δεδομένων.

2. **maatwebsite/excel**

Μια βιβλιοθήκη που χρησιμοποιείται για την εξαγωγή και την εισαγωγή δεδομένων σε/από αρχεία Excel. Στο πλαίσιο της εφαρμογής, χρησιμοποιήθηκε για την εξαγωγή δεδομένων.

3. **laravel/ui**

Προσφέρει προκαθορισμένα authentication scaffolds και front-end utilities, όπως φόρμες εγγραφής και σύνδεσης, βασισμένα σε Bootstrap.

4. **laravel/sanctum**

Μια ελαφριά λύση authentication για SPA (Single Page Applications) και APIs. Χρησιμοποιήθηκε για την προστασία endpoints και την ασφαλή διαχείριση χρηστών.

5. **guzzlehttp/guzzle**

Ένας HTTP client για την αποστολή αιτημάτων HTTP.

6. **fruitcake/php-cors**

Χρησιμοποιείται για την υποστήριξη του CORS (Cross-Origin Resource Sharing), επιτρέποντας αιτήματα από διαφορετικά domains, κάτι απαραίτητο για την αλληλεπίδραση front-end και back-end.

7. **firebase/php-jwt**

Μια βιβλιοθήκη για την κωδικοποίηση και αποκωδικοποίηση JSON Web Tokens (JWTs), απαραίτητη για ασφαλή authentication.

8. **spatie/laravel-ignition**

Προσφέρει ένα όμορφο interface για τη διαχείριση σφαλμάτων και την παρακολούθηση των logs, διευκολύνοντας τη διαδικασία debugging.

9. **php-ai/php-ml**

Μια βιβλιοθήκη machine learning για PHP, η οποία χρησιμοποιήθηκε για την υλοποίηση της γραμμικής παλινδρόμησης (Linear Regression). Χρησιμοποιήθηκε στο

διάγραμμα "**Glucose-Insulin Correlation**" για τον υπολογισμό της σχέσης μεταξύ επιπέδων γλυκόζης και δόσεων ινσουλίνης, επιτρέποντας την ανάλυση και οπτικοποίηση αυτής της συσχέτισης μέσα από το Laravel framework.

Πρόσθετες βιβλιοθήκες/πακέτα (Frontend & Build Tools)

- **@popperjs/core@2.11.8**
Χρησιμοποιείται σε συνδυασμό με το Bootstrap για tooltip και popover components.
- **axios@1.7.7**
Βιβλιοθήκη που επιτρέπει την αποστολή αιτημάτων HTTP από τον browser και τον χειρισμό των αποκρίσεων, χρησιμοποιούμενη για AJAX/async επικοινωνίες με APIs.
- **bootstrap@5.3.3**
Δημοφιλές CSS framework για την ανάπτυξη responsive διεπαφών χρήστη.
- **laravel-vite-plugin@1.0.5**
Το επίσημο plugin του Laravel για την ενσωμάτωση του Vite, ενός σύγχρονου εργαλείου για bundling και building assets (CSS/JS).
- **sass@1.79.3**
Μεταγλωττιστής για τη γλώσσα CSS preprocessor (Sass), επιτρέποντας πιο οργανωμένη και επαναχρησιμοποιήσιμη διαχείριση style sheets.
- **vite@5.4.7**
Σύγχρονο εργαλείο ανάπτυξης (build tool) για Frontend εφαρμογές. Παρέχει ταχύτατο hot reloading και βελτιστοποιεί τελικά assets.

Βιβλιοθήκες που χρησιμοποιήθηκαν στο Flask API

1. **Flask**
Το βασικό framework για την ανάπτυξη του API. Παρέχει routing, διαχείριση αιτημάτων/απαντήσεων και δυνατότητες δημιουργίας RESTful APIs.
2. **Flask-CORS**
Επιτρέπει στο API να δέχεται αιτήματα από άλλες εφαρμογές (cross-origin), κάνοντας το backend προσιτό σε web εφαρμογές.
3. **pandas**
Χρησιμοποιήθηκε για την επεξεργασία δεδομένων, όπως τον καθαρισμό, την ανάλυση, και τη μετατροπή δεδομένων από τη βάση.

4. **numpy**

Παρέχει υποστήριξη για αριθμητικούς υπολογισμούς, διανύσματα και πίνακες.

5. **scikit-learn**

Περιλαμβάνει εργαλεία για την εκπαίδευση και αξιολόγηση των μοντέλων, όπως Linear Regression, Random Forest, και GridSearchCV.

6. **xgboost**

Χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση του XGBoost μοντέλου, ένα από τα βασικά μοντέλα για την πρόβλεψη δόσεων ινσουλίνης.

7. **pymysql**

Επιτρέπει τη σύνδεση και την εκτέλεση ερωτημάτων στη βάση δεδομένων MySQL.

8. **statsmodels**

Χρησιμοποιήθηκε για στατιστικές αναλύσεις και για την εξαγωγή μετρικών.

Βιβλιοθήκες που χρησιμοποιήθηκαν στο Jupyter Notebook

1. **matplotlib**

Παρέχει δυνατότητες δημιουργίας γραφημάτων για την οπτικοποίηση δεδομένων.

2. **openpyxl**

Χρησιμοποιήθηκε για τη δημιουργία, επεξεργασία και εξαγωγή αρχείων Excel με μορφοποίηση.

3. **scipy**

Παρέχει στατιστικές λειτουργίες όπως το T-test, απαραίτητες για την ανάλυση δεδομένων.

4. **LabelEncoder και StandardScaler (από scikit-learn)**

Χρησιμοποιήθηκαν για την προεπεξεργασία δεδομένων (κωδικοποίηση κατηγοριών και κανονικοποίηση τιμών).

5. **GridSearchCV**

Χρησιμοποιήθηκε για την εύρεση των καλύτερων υπερπαραμέτρων για τα μοντέλα machine learning.

6. **XGBoost και RandomForestRegressor**

Τα κύρια μοντέλα που χρησιμοποιήθηκαν για την πρόβλεψη της δόσης ινσουλίνης, αξιολογώντας την ακρίβεια τους.

7.2 Θεωρητικό Υπόβαθρο για το Web App

Όσον αφορά την ανάπτυξη ενός web app σε PHP Laravel για τη διαχείριση της γλυκόζης απαιτείται η κατανόηση τεχνικών εννοιών και τεχνολογιών που αφορούν την ασφάλεια, τη δομή και την επικοινωνία της εφαρμογής με τον χρήστη.

CSRF (Cross-Site Request Forgery)

Το CSRF είναι μια επίθεση ασφαλείας που εκμεταλλεύεται την εμπιστοσύνη που έχει ένας ιστότοπος στον browser ενός επισκέπτη. Σε μια επίθεση CSRF, ένας επιτιθέμενος προσπαθεί να αναγκάσει τον χρήστη να εκτελέσει μια ανεπιθύμητη ενέργεια σε μια εφαρμογή Web στην οποία έχει ήδη πιστοποιηθεί. Η επίθεση CSRF ακολουθεί κατά κανόνα τα εξής βήματα:

1. Ο χρήστης συνδέεται σε μια εφαρμογή και αποκτά πιστοποίηση μέσω cookies.
2. Ο χρήστης επισκέπτεται ένα κακόβουλο site που περιέχει HTML ή JavaScript για την εκτέλεση εντολών χωρίς τη γνώση του.
3. Ο κακόβουλος ιστότοπος εκμεταλλεύεται τα cookies της συνεδρίας του χρήστη για να εκτελέσει ενέργειες στην αρχική εφαρμογή, χωρίς τη συγκατάθεσή του.

Διεπαφή Χρήστη (User Interface - UI)

Το μέρος της εφαρμογής που αλληλεπιδρά άμεσα με τον χρήστη, παρέχοντας του έναν οπτικό και διαδραστικό τρόπο να πλοηγηθεί και να εκτελέσει τις λειτουργίες της εφαρμογής.

Ανταποκρινόμενος Σχεδιασμός (Responsive Design)

Ο ανταποκρινόμενος σχεδιασμός αφορά τη διαμόρφωση ιστοσελίδων για να προσαρμόζονται δυναμικά σε διαφορετικά μεγέθη και τύπους συσκευών (όπως desktop, tablet, smartphone). Οι βασικές αρχές περιλαμβάνουν:

- **Ευέλικτες Διατάξεις:** Τα στοιχεία της σελίδας αναδιατάσσονται και προσαρμόζονται σύμφωνα με το μέγεθος της οθόνης.
- **Ερωτήματα Πολυμέσων (Media Queries):** Εφαρμογή συγκεκριμένων στυλ ανάλογα με τις δυνατότητες της συσκευής και τον προσανατολισμό.
- **Προσέγγιση Mobile-First:** Ο σχεδιασμός ξεκινά από τις μικρότερες διαστάσεις και προσαρμόζεται για μεγαλύτερες οθόνες, εξασφαλίζοντας βελτιστοποιημένη εμπειρία χρήσης.

Προσαρμοστικός Σχεδιασμός (Adaptive Design)

Δημιουργία σταθερών διατάξεων που ενεργοποιούνται ανάλογα με τον τύπο της συσκευής και τις δυνατότητές της, βελτιώνοντας την εμπειρία χρήστη.

RESTful API

Το REST αποτελεί ένα αρχιτεκτονικό πρότυπο για τη δημιουργία ελαφριών και αποδοτικών εφαρμογών ιστού, βασισμένων στο πρωτόκολλο HTTP. Τα RESTful APIs επικοινωνούν χρησιμοποιώντας τυποποιημένες HTTP μεθόδους, όπως GET, POST, PUT και DELETE, διευκολύνοντας την ανταλλαγή δεδομένων σε μορφές όπως JSON και XML.

Η συγκεκριμένη υλοποίηση του Flask API στο πλαίσιο της εργασίας παρέχει υπηρεσίες πρόβλεψης για τα επίπεδα γλυκόζης (glucose_new) και τις δόσεις ινσουλίνης (insulin_dose) μέσω μοντέλων μηχανικής μάθησης. Βασίζεται στις αρχές του REST:

- **Ενιαίο Περιβάλλον Διεπαφής (Uniform Interface):** Απλοποιεί την επικοινωνία μέσω μοναδικών URIs για κάθε πόρο.
- **Κατάσταση Χωρίς Σύνδεση (Stateless):** Κάθε αίτημα περιλαμβάνει όλες τις απαραίτητες πληροφορίες χωρίς εξάρτηση από προηγούμενες αλληλεπιδράσεις.
- **Αποσύνδεση Πελάτη-Διακομιστή (Client-Server Separation):** Το API επικοινωνεί απρόσκοπτα με το frontend (Laravel) και το backend (Flask), επιτρέποντας ανεξαρτησία μεταξύ των συστημάτων.
- **Δυνατότητα Αποθήκευσης σε Cache (Cachable):** Ενισχύει την απόδοση με προσωρινή αποθήκευση απαντήσεων.

Η αρχιτεκτονική εξασφαλίζει την ακριβή πρόβλεψη μέσω κλιμακούμενων αλγορίθμων μηχανικής μάθησης, ενώ υποστηρίζει εύκολη συντήρηση και εξέλιξη της εφαρμογής.

Bootstraps

Εργαλείο front-end που διευκολύνει τη δημιουργία responsive εφαρμογών με χρήση προκαθορισμένων CSS και JavaScript στοιχείων.

7.3 Περιγραφή του dataset

Το dataset που χρησιμοποιείται για την εκπαίδευση των μοντέλων πρόβλεψης αποτελεί ένα Artificial Dataset, το οποίο δημιουργήθηκε ειδικά για την παρούσα εφαρμογή, με γνώμονα

την δομή του dataset που χρησιμοποιείται στο αρχείο του Jupyter Notebook. Η δημιουργία του σχεδιάστηκε με βάση τις αρχές του Sliding Scale Protocol και του Carbo Calculation, προκειμένου να προσομοιώνει ρεαλιστικά τις διαδικασίες διαχείρισης δεδομένων διαβήτη.

Περιλαμβάνει καθημερινές καταγραφές δεδομένων από ασθενείς με διαβήτη, με πληροφορίες σχετικά με τα επίπεδα γλυκόζης, την κατανάλωση υδατανθράκων, τις δόσεις ινσουλίνης και το σωματικό βάρος. Επιπλέον, δημογραφικές πληροφορίες, όπως η ηλικία, το φύλο και ο τύπος διάγνωσης, ενσωματώνονται για τη βελτίωση των μοντέλων πρόβλεψης. Η δομή του dataset διαμορφώνεται με τρόπο που να περιγράφει τη διαχείριση του διαβήτη για κάθε ασθενή με ακρίβεια.

Δομή του Dataset

Το dataset αποτελείται από δύο κύριους πίνακες στη βάση δεδομένων:

1. **patients_statistics**: Περιλαμβάνει δεδομένα καθημερινών μετρήσεων, όπως:

- **glucose_old**: Επίπεδα γλυκόζης πριν το γεύμα (mg/dL).
- **glucose_new**: Επίπεδα γλυκόζης μετά το γεύμα (mg/dL).
- **food_carbo**: Κατανάλωση υδατανθράκων (σε γραμμάρια).
- **insulin_dose**: Δόση ινσουλίνης μετά τη μέτρηση (μονάδες).
- **weight**: Σωματικό βάρος ασθενούς (σε κιλά).

2. **patients**: Περιλαμβάνει δημογραφικά στοιχεία ασθενών, όπως:

- **age**: Υπολογίζεται από την ημερομηνία γέννησης (birth_at).

```
# Calculate 'age' as the difference between today and 'birth_at' in years
df['age'] = ((pd.to_datetime('today') - df['birth_at']).dt.days / 365.25).astype(int)
```

Σχήμα 25: Υπολογισμός ηλικίας από ημερομηνία γέννησης

- **gender**: Το φύλο (αρσενικό ή θηλυκό).
- **diagnosis**: Ο τύπος διάγνωσης (π.χ., Type I ή Type II).

Συλλογή Δεδομένων

Η συλλογή δεδομένων πραγματοποιείται από τους ασθενείς τρεις φορές την ημέρα, με καταγραφή βασικών μετρήσεων διαχείρισης του διαβήτη. Συγκεκριμένα:

- Πριν από το πρωινό γεύμα.
- Πριν από το μεσημεριανό γεύμα.
- Πριν από το βραδινό γεύμα.

Μετά την ολοκλήρωση κάθε γεύματος και τη χορήγηση της δόσης ινσουλίνης, καταγράφεται επίσης το επίπεδο γλυκόζης μετά το γεύμα (Glucose After). Αυτή η πληροφορία αποτελεί κρίσιμο στοιχείο για την παρακολούθηση της αποτελεσματικότητας της δόσης και της κατανάλωσης υδατανθράκων.

Ενοποίηση Δεδομένων

Κατά τη διαδικασία εκπαίδευσης, τα δεδομένα από τους δύο παραπάνω πίνακες συνδυάζονται μέσω SQL ερωτημάτων. Το παρακάτω ερώτημα SQL αποτελεί παράδειγμα για τη σύνδεση των δεδομένων:

```
SELECT ps.patient_id, ps.glucose_old, ps.insulin_dose, ps.food_carbo, ps.glucose_new, ps.created_at,
       p.birth_at, p.weight, p.diagnosis, p.gender
FROM patients_statistics ps
JOIN patients p ON ps.patient_id = p.id
```

Σχήμα 26: SQL Ερώτημα για Εξαγωγή Δεδομένων Patients και Patients Statistics

Αυτό επιτρέπει την ενοποίηση δεδομένων στατιστικών (patients_statistics) και δημογραφικών (patients) για τη δημιουργία του τελικού συνόλου δεδομένων που χρησιμοποιείται για την εκπαίδευση των μοντέλων.

Κλίμακα του Dataset

Το dataset χαρακτηρίζεται από τα παρακάτω βασικά στοιχεία:

- Αριθμός Ασθενών: Περιλαμβάνονται δεδομένα για 38 ασθενείς, προσφέροντας ένα αντιπροσωπευτικό δείγμα.
- Διάρκεια Καταγραφής: Για κάθε ασθενή, τα δεδομένα εκτείνονται σε πολλές ημέρες, προσφέροντας πλούσιο ιστορικό διαχείρισης.
- Αριθμός Παρατηρήσεων: Κατά μέσο όρο, καταγράφονται 3 παρατηρήσεις την ημέρα για κάθε ασθενή.

Σημαντικές Παρατηρήσεις

1. **Συμπληρωματικά Δεδομένα:** Δημογραφικά δεδομένα (π.χ., age, gender, diagnosis) αντλούνται από τον πίνακα patients, καθώς δεν περιλαμβάνονται στον πίνακα patients_statistics.
2. **Συλλογή σε πραγματικό χρόνο:** Τα δεδομένα συλλέγονται σε πραγματικό χρόνο και αποθηκεύονται στη βάση. Τα πεδία created_at και updated_at διασφαλίζουν την παρακολούθηση ενημερώσεων.

Παράδειγμα κώδικα για τον έλεγχο της τελευταίας ενημέρωσης:

```
# Get the maximum of created_at and updated_at timestamps from patients_statistics table
cursor.execute("SELECT GREATEST(MAX(created_at), MAX(updated_at)) FROM patients_statistics")
stats_timestamp = cursor.fetchone()[0] or datetime.min

# Get the maximum of created_at and updated_at timestamps from patients table
cursor.execute("SELECT GREATEST(MAX(created_at), MAX(updated_at)) FROM patients")
patients_timestamp = cursor.fetchone()[0] or datetime.min # Use datetime.min if no timestamp is found
```

Σχήμα 27: Παρακολούθηση Τελευταίων Ενημερώσεων στη Βάση Δεδομένων με Timestamps

3. **Προεπεξεργασία Δεδομένων (Data Preprocessing):** Η προεπεξεργασία είναι κρίσιμη για τη διασφάλιση της ποιότητας των δεδομένων. Στο παρόν dataset εφαρμόζονται οι εξής τεχνικές:

1. **Συμπλήρωση ελλিপών δεδομένων (Data Cleaning):** Αν το πεδίο food_carbo είναι κενό, συμπληρώνεται με 0, όπως φαίνεται παρακάτω:

```
df['food_carbo'] = df['food_carbo'].fillna(0).astype(float)
```

Σχήμα 28: Προεπεξεργασία Δεδομένων: Συμπλήρωση Ελλিপών Τιμών στο Πεδίο food_carbo

2. **Μετατροπή κατηγορικών δεδομένων σε αριθμητικές τιμές (Categorical Encoding):** Για να είναι κατάλληλα τα δεδομένα για τα μοντέλα, τα κατηγορικά πεδία gender και diagnosis μετατρέπονται σε αριθμητικές τιμές:

```
# Map 'gender' from categorical to numerical values (male: 0, female: 1)
df['gender'] = df['gender'].map({'male': 0, 'female': 1}).fillna(0).astype(int)
# Map 'diagnosis' from categorical to numerical values (Type I: 0, Type II: 1)
df['diagnosis'] = df['diagnosis'].map({'Type I': 0, 'Type II': 1}).fillna(0).astype(int)
```

Σχήμα 29: Μετατροπή Κατηγορικών Δεδομένων σε Αριθμητικές Τιμές για τα Πεδία Gender και Diagnosis

3. **Συμπλήρωση ελλিপών τιμών βάρους (Data Cleaning):** Αν το πεδίο weight είναι κενό, συμπληρώνεται με τον μέσο όρο των τιμών:

```
# Handle missing 'weight' values by filling with the mean weight
df['weight'] = df['weight'].fillna(df['weight'].mean()).astype(float)
```

4. **Εξασφάλιση σειράς χρονολογικών δεδομένων:** Τα δεδομένα ταξινομούνται με βάση την ημερομηνία δημιουργίας τους:

```
# Sort the DataFrame by 'created_at' timestamp  
df = df.sort_values('created_at')
```

Σχήμα 31: Ταξινόμηση Δεδομένων κατά Ημερομηνία Δημιουργίας

Αυτός ο συνδυασμός πληροφοριών παρέχει ένα πλούσιο και ενοποιημένο dataset, το οποίο υποστηρίζει την εκπαίδευση και την αξιολόγηση ακριβών μοντέλων πρόβλεψης για τη διαχείριση διαβήτη.

7.4 Ανάλυση του Grid Search

Στο σημείο αυτό, θα αναλύσουμε τη θεωρητική διάσταση της μεθόδου Grid Search, η οποία εφαρμόζεται συχνά για τη βέλτιστη ρύθμιση υπερπαραμέτρων στους αλγορίθμους Random Forest και XGBoost. Η μέθοδος Grid Search αποτελεί ένα θεμελιώδες εργαλείο σε προβλήματα Μηχανικής Μάθησης και έχει ως στόχο να εντοπίσει τις παραμέτρους (ή συνδυασμό παραμέτρων) που βελτιώνουν την απόδοση ενός μοντέλου ως προς κάποια μετρική.

Θεωρητική Προσέγγιση του Grid Search

Το Grid Search είναι μια εξαντλητική διαδικασία αναζήτησης σε έναν ορισμένο «χώρο» υπερπαραμέτρων. Ο χώρος αυτός συνίσταται σε ένα πλέγμα (grid) από τιμές για κάθε υπερπαραμέτρο που μας ενδιαφέρει. Για παράδειγμα, αν θέλουμε να ρυθμίσουμε δύο υπερπαραμέτρους (π.χ. α και β) σε ένα μοντέλο, μπορούμε να ορίσουμε ένα σύνολο πιθανών τιμών για την καθεμιά ($\alpha = \{1, 2, 3\}$, $\beta = \{0.1, 0.01\}$) και κατόπιν να εκπαιδεύσουμε το μοντέλο μας για κάθε πιθανό συνδυασμό:

- ($\alpha = 1, \beta = 0.1$)
- ($\alpha = 1, \beta = 0.01$)
- ($\alpha = 2, \beta = 0.1$)
- ($\alpha = 2, \beta = 0.01$)
- ($\alpha = 3, \beta = 0.1$)

- ($\alpha = 3, \beta = 0.01$)

Στο τέλος, επιλέγουμε τον συνδυασμό που δίνει την καλύτερη απόδοση στο σύνολο επικύρωσης (validation set) ή βάσει κάποιας διαδικασίας cross-validation (GridSearchCV της βιβλιοθήκης scikit-learn). Παρ' ότι η μέθοδος μπορεί να γίνει υπολογιστικά ακριβή, ειδικά όταν ο αριθμός των υπερπαραμέτρων και των τιμών τους είναι μεγάλος, αποτελεί ένα ξεκάθαρο και συστηματικό εργαλείο εύρεσης βέλτιστων ρυθμίσεων.

▪ Grid Search για τον Random Forest [12]

Ο Random Forest είναι ένας αλγόριθμος που βασίζεται στη δημιουργία πολλών ανεξάρτητων Δέντρων Απόφασης (Decision Trees) και στη συνέχεια συνένωση των προβλέψεών τους (bagging). Οι βασικές υπερπαραμέτροι που μπορούν να ρυθμιστούν μέσω Grid Search για τον Random Forest είναι:

1. **n_estimators**

- Περιγραφή: Ο αριθμός των δέντρων που θα δημιουργηθούν.
- Τυπικές τιμές: 50, 100, 200, 500, 1000 (ανάλογα με τις ανάγκες και τους πόρους).
- Επίδραση: Ένας μεγαλύτερος αριθμός από δέντρα συνήθως βελτιώνει την απόδοση μέχρι ενός σημείου, αλλά αυξάνει τον χρόνο εκπαίδευσης.

2. **max_depth**

- Περιγραφή: Το μέγιστο επιτρεπτό βάθος για κάθε δέντρο στο δάσος.
- Τυπικές τιμές: 5, 10, 15, 20, ... ή None (χωρίς περιορισμό).
- Επίδραση: Μεγαλύτερο βάθος αυξάνει την ικανότητα του μοντέλου να μαθαίνει πολύπλοκα μοτίβα, αλλά και τον κίνδυνο υπερπροσαρμογής (overfitting).

3. **min_samples_split**

- Περιγραφή: Ο ελάχιστος αριθμός δειγμάτων για διαχωρισμό ενός κόμβου σε δύο νέους κόμβους.
- Τυπικές τιμές: 2, 5, 10, 20, ...
- Επίδραση: Μεγαλύτερη τιμή μειώνει την πολυπλοκότητα των δέντρων και μπορεί να βελτιώσει τη γενίκευση, μειώνοντας το overfitting.

4. **min_samples_leaf**

- Περιγραφή: Ο ελάχιστος αριθμός δειγμάτων σε ένα φύλλο (leaf) του δέντρου.
- Τυπικές τιμές: 1, 2, 5, 10, ...
- Επίδραση: Παρόμοια με το `min_samples_split`, όσο αυξάνει, τόσο μειώνεται η λεπτομέρεια του δέντρου.

5. **max_features**

- ο Περιγραφή: Το μέγιστο ποσοστό/αριθμός χαρακτηριστικών (features) που θα εξετάζονται σε κάθε split.
- ο Τυπικές τιμές: “auto” (ή “sqrt”), “log2”, ένας ακέραιος αριθμός ή ένα δεκαδικό < 1 που αντιστοιχεί σε ποσοστό του συνόλου των features.
- ο Επίδραση: Ελέγχει την ποικιλία των δέντρων. Η ρύθμιση αυτή μπορεί να βελτιώσει τη γενίκευση.

6. **bootstrap**

- ο Περιγραφή: Boolean παράμετρος που καθορίζει αν θα γίνει δειγματοληψία με επανάθεση (bootstrap samples) για την εκπαίδευση κάθε δέντρου.
- ο Τυπικές τιμές: True ή False.
- ο Επίδραση: Όταν είναι True (προεπιλογή), αυξάνει την τυχειότητα και σε πολλές περιπτώσεις βελτιώνει την απόδοση γενικά.

7. **criterion** (προαιρετικό)

- ο Περιγραφή: Το κριτήριο που χρησιμοποιείται για τη μέτρηση της ποιότητας των split (π.χ. “gini”, “entropy” για ταξινόμηση ή “mse”, “mae” για παλινδρόμηση).
- ο Επίδραση: Μικρότερη συνήθως η επίδραση στη συνολική απόδοση σε σχέση με άλλες παραμέτρους, αλλά χρήσιμο να δοκιμαστεί.

Χρησιμοποιώντας Grid Search, ορίζουμε ένα πλέγμα πιθανών τιμών για τις παραπάνω παραμέτρους και εκπαιδεύουμε τον Random Forest με όλους τους δυνατούς συνδυασμούς. Μετά την εκπαίδευση και την αξιολόγηση (συχνά με cross-validation), επιλέγουμε τις ρυθμίσεις που δίνουν την καλύτερη απόδοση στο επιλεγμένο κριτήριο.

▪ **Grid Search για το XGBoost**

Το XGBoost (eXtreme Gradient Boosting) είναι ένα εξαιρετικά δημοφιλές πλαίσιο υλοποίησης του Gradient Boosted Decision Trees, σχεδιασμένο για υψηλή απόδοση και δυνατότητες παραλληλοποίησης. Οι βασικές υπερπαραμέτροι που μπορούν να ρυθμιστούν και να δοκιμαστούν μέσω Grid Search είναι:

1. **n_estimators**

- Περιγραφή: Ο αριθμός των δέντρων (γύρων boosting) που θα δημιουργηθούν.
- Τυπικές τιμές: 100, 200, 500
- Επίδραση: Όσο αυξάνεται, τόσο το μοντέλο μαθαίνει περισσότερα, αλλά κινδυνεύει να υπερπροσαρμοστεί (overfitting).

2. **learning_rate** (ή η)

- Περιγραφή: Ο ρυθμός μάθησης (learning rate). Όσο μικρότερος είναι, τόσο πιο αργά μαθαίνει το μοντέλο.
- Τυπικές τιμές: 0.01, 0.05, 0.1, 0.2.
- Επίδραση: Μικρότερη τιμή συνήθως αυξάνει την ακρίβεια αλλά χρειάζεται περισσότερα δέντρα (επομένως μεγαλύτερο χρόνο εκπαίδευσης).

3. **max_depth**

- Περιγραφή: Το μέγιστο βάθος των δέντρων.
- Τυπικές τιμές: 3, 5, 7, 10...
- Επίδραση: Μεγαλύτερο βάθος αυξάνει τη δυνατότητα του μοντέλου να μάθει σύνθετα μοτίβα, αλλά μπορεί να οδηγήσει σε υπερπροσαρμογή (overfitting).

4. **subsample**

- Περιγραφή: Το ποσοστό των δειγμάτων που χρησιμοποιούνται σε κάθε δέντρο.
- Τυπικές τιμές: 0.5, 0.7, 0.8, 1.0 (100% όλων των δειγμάτων).
- Επίδραση: Μικρότερο subsample γενικά αυξάνει τη σποραδικότητα και μπορεί να μειώσει το overfitting, αλλά πολύ χαμηλή τιμή μπορεί να οδηγήσει σε υποεκπαίδευση.

5. **colsample_bytree**

- Περιγραφή: Το ποσοστό των χαρακτηριστικών που χρησιμοποιούνται για την εκπαίδευση κάθε δέντρου.

- Τυπικές τιμές: 0.5, 0.7, 0.8, 1.0.
- Επίδραση: Παρόμοια με το subsample, αυξάνει την τυχαιότητα και μειώνει την πιθανότητα υπερπροσαρμογής.

6. **gamma** (ή γ)

- Περιγραφή: Η ελάχιστη μείωση στο loss που απαιτείται για την πραγματοποίηση ενός επιπλέον split σε έναν κόμβο του δέντρου.
- Τυπικές τιμές: 0, 0.1, 0.3, 0.5...
- Επίδραση: Μεγαλύτερη τιμή σημαίνει ότι ο αλγόριθμος είναι πιο «αυστηρός» στα split, άρα μπορεί να μειώσει την πολυπλοκότητα.

7. **reg_lambda** (ή λ) και **reg_alpha** (ή α)

- Περιγραφή: Οι παράμετροι κανονικοποίησης $L2$ (λ) και $L1$ (α) αντίστοιχα.
- Τυπικές τιμές: 0, 1, 10, 100... (ανάλογα το πρόβλημα).
- Επίδραση: Συμβάλλουν στον έλεγχο της πολυπλοκότητας των δέντρων και στη μείωση του overfitting.

8. **booster** (προαιρετικό)

- Περιγραφή: Καθορίζει τον τύπο της μηχανής boosting (π.χ. “gbtree”, “gblinear”, “dart”).
- Επίδραση: Συνήθως το “gbtree” είναι η προεπιλογή. Μπορεί να δοκιμαστεί και το “dart” ή “gblinear” ανάλογα με τα δεδομένα.

Και εδώ, στο Grid Search προετοιμάζουμε ένα πλέγμα από συνδυασμούς των παραπάνω τιμών. Στην πράξη, λόγω του μεγάλου πλήθους των υπερπαραμέτρων του XGBoost, επιλέγουμε συνήθως τις 2-3 παραμέτρους που περιμένουμε να έχουν τη μεγαλύτερη επίδραση (π.χ. learning_rate, max_depth, subsample) για αρχική αναζήτηση, και στη συνέχεια «βελτιστοποιούμε» πιο εξειδικευμένες ρυθμίσεις (π.χ. gamma, reg_lambda, reg_alpha).

Συνοψίζοντας, η μέθοδος Grid Search αποτελεί μια συστηματική και αξιόπιστη διαδικασία για την ανεύρεση των βέλτιστων ρυθμίσεων σε αλγόριθμους όπως ο Random Forest και το XGBoost. Παρότι μπορεί να απαιτήσει υψηλό υπολογιστικό κόστος, προσφέρει μια ολοκληρωμένη διερεύνηση του χώρου των παραμέτρων, εξασφαλίζοντας ότι δεν θα παραμεληθεί κάποιος σημαντικός συνδυασμός. Η επιλογή των παραμέτρων που θα τεθούν σε Grid Search, καθώς και η κλίμακα των πιθανών τιμών τους, είναι κρίσιμη, και συχνά βασίζεται σε προγενέστερη γνώση του προβλήματος και σε πειραματισμό. Με τον τρόπο

αυτό, μπορούμε να εκμεταλλευτούμε πλήρως τις δυνατότητες των αλγορίθμων Random Forest και XGBoost, καταλήγοντας σε μοντέλα με βελτιστοποιημένη απόδοση για την εκάστοτε εφαρμογή.

7.5 Μέθοδοι υπολογισμού δόσεων ινσουλίνης και επιπέδων γλυκόζης:

Υλοποίηση στον κώδικα

Στη συνέχεια, θα περάσουμε στις μεθόδους υπολογισμού δόσεων ινσουλίνης και επιπέδων γλυκόζης, όπως υλοποιούνται στον κώδικα της εφαρμογής.

Υπολογισμός Ινσουλίνης με Sliding Scale Protocol και Carbo Calculation

Η μέθοδος υπολογισμού “Sliding Scale + Carbo Calculation” βασίζεται στη θεωρητική ανάλυση της ενότητας 2.2.1 και 2.2.2. Συγκεκριμένα, το Sliding Scale Protocol καθορίζει τον αριθμό των μονάδων ινσουλίνης που πρέπει να χορηγηθούν, με βάση το πόσο υπερβαίνουν τα επίπεδα γλυκόζης το όριο των 100 mg/dL. Από την άλλη, το Carbo Calculation υπολογίζει την απαιτούμενη δόση ινσουλίνης ανάλογα με την ποσότητα των υδατανθράκων που καταναλώθηκαν, με αναλογία 1 μονάδα ινσουλίνης ανά 10 g υδατανθράκων. Ο υπολογισμός των δύο αυτών μεθόδων, όπου στο τέλος αθροίζονται για να εμφανίσουν την τελική δόση ινσουλίνης, απεικονίζεται στην παρακάτω εικόνα.

```
// Υλοποίηση του Sliding Scale + Carbo Calculation απευθείας στο Laravel
// χωρίς να καλείται το Flask API.
if ($algorithm === 'sliding_scale_carbo') {
    // Sliding Scale Calculation
    // Εάν το "glucose_old" > 100, υπολογίζουμε πόσα mg/dL πάνω από το 100 είναι
    $extra_units = 0;
    if ($glucose_old > 100) {
        $over_100 = $glucose_old - 100;
        // Για κάθε 50 mg/dL πάνω από το 100, προσθέτουμε 2 μονάδες ινσουλίνης.
        $extra_units = ($over_100 * 2) / 50;
    }

    // Carbo Calculation:
    // Υποθέτουμε λόγο ινσουλίνης προς υδατάνθρακες (ICR) 1:10
    // Δηλαδή 1 μονάδα ινσουλίνης ανά 10g υδατανθράκων.
    $insulin_for_carbs = $food_carbo / 10;

    // Το άθροισμα των δύο μεθόδων μας δίνει τη συνολική προτεινόμενη δόση
    $total_dose = $extra_units + $insulin_for_carbs;

    // Επιστρέφουμε τη δόση ως JSON απάντηση (σε AJAX κλήση)
    return response()->json(['message' => $total_dose]);
}
```

Σχήμα 32: Υπολογισμός Δόσης Ινσουλίνης με Sliding Scale και Carbo Calculation

Υπολογισμός ινσουλίνης με βάση το βάρος του ασθενούς (Weight-based Insulin Prediction)

Η μέθοδος Weight-based Insulin Prediction, όπως περιγράφηκε στην ενότητα 2.2.3, εφαρμόζεται αυτόματα σε περιπτώσεις όπου ο ασθενής έχει λιγότερες από τέσσερις εγγραφές.

```
// Μετράμε πόσες εγγραφές υπάρχουν ήδη καταχωρημένες για τον ασθενή
$countstats = PatientStatistic::where('patient_id', $patient_id)->count();

// Εάν υπάρχουν λιγότερες από 4 εγγραφές για τον ασθενή,
// επιστρέφουμε έναν απλό υπολογισμό: (βάρος / 2) / 3
// ως μία αρχική/προσεγγιστική δόση ινσουλίνης.
if ($countstats < 4) {
    $insulin_dose = ($patient_weight / 2) / 3;
    return response()->json(['message' => $insulin_dose]);
}
```

Σχήμα 33: Weight-based Insulin Prediction για Ασθενείς με Λιγότερες από 4 Εγγραφές

Στον συγκεκριμένο υπολογισμό, το αποτέλεσμα της διαίρεσης του βάρους του ασθενούς δια του 2 ($\$patient_weight / 2$) διαιρείται περαιτέρω με το 3, καθώς ο ασθενής μετρά τα επίπεδα γλυκόζης και λαμβάνει δόση ινσουλίνης τρεις φορές την ημέρα (πρωί, μεσημέρι, βράδυ).

Υπολογισμός Γλυκόζης Μετά το Γεύμα

Η μέθοδος υπολογισμού της γλυκόζης μετά το γεύμα εφαρμόζεται σύμφωνα με τη θεωρητικό υπόβαθρο της ενότητας 2.3 και απεικονίζεται στην παρακάτω εικόνα.

```

// 1ο σενάριο: Τοπικός υπολογισμός της νέας τιμής γλυκόζης
if ($calculation_mode === 'calculate_new_glucose') {
    $patient_weight = $patient->weight;
    if ($patient_weight <= 0) {
        return redirect()->back()->with('error', 'Το βάρος του ασθενούς πρέπει να είναι μεγαλύτερο από το μηδέν.')->withInput();
    }

    // Έλεγχος της διάγνωσης του ασθενούς
    $diagnosis = $patient->diagnosis;
    if ($diagnosis === 'Type I') {
        $TDD = $patient_weight * 0.5; // Type I: 0.5 μονάδες ανά κιλό
        $ISF = 1800 / $TDD; // Υπολογισμός ISF για Type I
    } elseif ($diagnosis === 'Type II') {
        $TDD = $patient_weight * 0.8; // Type II: 0.8 μονάδες ανά κιλό
        $ISF = 1500 / $TDD; // Υπολογισμός ISF για Type II
    } else {
        return redirect()->back()->with('error', 'Η διάγνωση του ασθενούς δεν είναι έγκυρη.')->withInput();
    }

    // Υπολογισμός του συντελεστή υδατανθράκων (Carbohydrate Ratio - CR)
    $carbohydrate_ratio = 500 / $TDD;

    // Υπολογισμός της απαιτούμενης ινσουλίνης για την κατανάλωση υδατανθράκων
    $insulin_for_carb = $food_carbo / $carbohydrate_ratio;

    // Υπολογισμός της συνολικής μείωσης στη γλυκόζη λόγω ινσουλίνης
    $glucose_reduction = $insulin_dose * $ISF;

    // Υπολογισμός της νέας τιμής γλυκόζης:
    // γλυκόζη πριν το γεύμα + αύξηση από υδατάνθρακες - μείωση από ινσουλίνη
    $predicted_glucose_new = $glucose_old + ($insulin_for_carb) - $glucose_reduction;

    // Αποφυγή αρνητικών τιμών στη γλυκόζη
    if ($predicted_glucose_new < 0) {
        $predicted_glucose_new = 0;
    }
}

```

Σχήμα 34: Υπολογισμός Γλυκόζης Μετά το Γεύμα - Ροή Υλοποίησης

Με αυτόν τον τρόπο, οι μέθοδοι που περιγράφηκαν στις ενότητες 2.2 και 2.3 έχουν ενσωματωθεί απλά και κατανοητά στον κώδικα του “PredictionController.php”.

Η μέθοδος υπολογισμού γλυκόζης μετά το γεύμα βασίζεται σε δυναμικές παραμέτρους, όπως ο συντελεστής ευαισθησίας στην ινσουλίνη (ISF), ο συντελεστής υδατανθράκων (CR), και η αρχική τιμή γλυκόζης πριν το γεύμα. Αυτά τα δεδομένα, σε συνδυασμό με τις πληροφορίες που αντλούνται από τη βάση δεδομένων, επιτρέπουν την εξατομικευμένη πρόβλεψη για κάθε ασθενή.

Ο σχεδιασμός του κώδικα εξασφαλίζει ότι όλοι οι παράγοντες που επηρεάζουν τη γλυκόζη, όπως το βάρος, η διάγνωση, οι υδατάνθρακες του γεύματος και η χορηγούμενη ινσουλίνη, λαμβάνονται υπόψη. Επιπλέον, οι υπολογισμοί έχουν διαμορφωθεί με τέτοιο τρόπο ώστε να αποφεύγονται αρνητικές ή μη ρεαλιστικές τιμές, διασφαλίζοντας την ακρίβεια και την αξιοπιστία των αποτελεσμάτων.

7.6 Υλοποίηση Διαγραμμάτων στην εφαρμογή (Κώδικας)

Σε αυτό το κεφάλαιο παρουσιάζονται αναλυτικά τα τρία διαγράμματα που υλοποιήθηκαν στην εφαρμογή. Δίνεται έμφαση στον τρόπο με τον οποίο δομείται ο κώδικας τόσο στην πλευρά του Laravel (Blade templates, Controllers), όσο και στην πλευρά της Python (Flask/XGBoost) για την πρόβλεψη μελλοντικών τιμών γλυκόζης. Τα διαγράμματα που δημιουργήσαμε είναι:

1. Time In Range
2. Glucose-Insulin Correlation
3. Future Glucose Trend Prediction

Τα δύο πρώτα διαγράμματα ανήκουν στην ενότητα Diagrams, ενώ το τρίτο εντάσσεται στην ενότητα Glucose Trend.

1. Time In Range

Στόχος & Περιγραφή

Το διάγραμμα Time In Range απεικονίζει τα επίπεδα γλυκόζης ενός ασθενούς για ένα επιλεγμένο χρονικό διάστημα. Παράλληλα, εμφανίζει τις τιμές που καταγράφονται:

- Κάτω από 80 mg/dL (κίτρινη περιοχή)
- Μεταξύ 80 και 180 mg/dL (πράσινη περιοχή)
- Πάνω από 180 mg/dL (κόκκινη περιοχή)

Για κάθε ζώνη υπολογίζεται και εμφανίζεται το αντίστοιχο ποσοστό στο οποίο βρέθηκαν οι μετρήσεις του ασθενούς.

Υλοποίηση στο Blade Template

Στο αρχείο `diagrams.blade.php` υπάρχει φόρμα που επιτρέπει στον χρήστη να επιλέξει την περίοδο ανάλυσης (From-To) και τον τύπο διαγράμματος (π.χ. “Time in Range”). Όταν ο χρήστης υποβάλει τη φόρμα, το Laravel Controller ανακτά τα δεδομένα γλυκόζης από τη βάση και, αν επιλεγθεί το “Time in Range”, η σελίδα εμφανίζει ένα γραμμικό διάγραμμα `Chart.js`:

```

<!-- Time in Range Diagram -->
@if(request('diagram_type') == 'time_in_range')
    {{-- Εμφάνιση λεζάντας με ποσοστά για κάθε ζώνη γλυκόζης. --}}
    <div class="legend-box text-center mb-3">
        <span style="border: 2px solid yellow; padding: 5px 10px; color: black;">Under 80 mg/dL ({{ round($under80, 0) }}%)</span>
        <span style="border: 2px solid green; padding: 5px 10px; color: black;">80-180 mg/dL ({{ round($between80and180, 0) }}%)</span>
        <span style="border: 2px solid red; padding: 5px 10px; color: black;">Above 180 mg/dL ({{ round($above180, 0) }}%)</span>
    </div>

    <!-- minimum height για το chart container -->
    <div class="text-center" style="padding-right: 30px; min-height: 600px;">
        <canvas id="glucoseChart"></canvas>
    </div>

```

Σχήμα 35: Υλοποίηση Διαγράμματος Time In Range στο Blade Template

Στο Blade, χρησιμοποιείται JavaScript (Chart.js) για τη σχεδίαση. Παρατηρούμε ότι ορίζονται τρεις χρωματιστές “box” περιοχές (κίτρινη, πράσινη, κόκκινη) για τις ζώνες <80, 80–180 και >180 mg/dL. Επίσης, ένα custom plugin (highValueAnnotations) τοποθετεί ετικέτες (κόκκινο κείμενο) στα σημεία όπου η γλυκόζη ξεπερνά τα 180 mg/dL, ώστε να είναι ευδιάκριτες οι υψηλές τιμές.

Φόρτωση & Προετοιμασία Δεδομένων στον Controller

Στον DiagramController, η μέθοδος loadDiagram φιλτράρει τα ιστορικά δεδομένα του ασθενούς μέσω του μοντέλου PatientStatistic:

```

// Διαχείριση του διαγράμματος "Time in Range"
if ($diagramType == 'time_in_range') {
    // Ανάκτηση δεδομένων γλυκόζης (glucose_old) και χρονικού διαστήματος από τον πίνακα patients_statistics για τον συγκεκριμένο ασθενή.
    $patientStatistics = PatientStatistic::whereHas('patient', function ($query) use ($userId) {
        $query->where('user_id', $userId);
    })
        ->whereBetween('created_at', [$fromDate, $toDate])
        ->orderBy('created_at', 'ASC')
        ->get(['glucose_old', 'created_at']);
}

```

Σχήμα 36: Φόρτωση και Φιλτράρισμα Δεδομένων για το Διάγραμμα Time In Range στον Controller

Έπειτα, μετρά το ποσοστό τιμών που βρίσκονται σε κάθε ζώνη (<80, 80–180, >180) και επιστρέφει τα αποτελέσματα στο view, ώστε αυτά να «περαστούν» στη JavaScript και να αποτυπωθούν γραφικά.

2. Glucose-Insulin Correlation

Στόχος & Περιγραφή

Το διάγραμμα Glucose-Insulin Correlation αποτυπώνει, σε μορφή scatter plot, τη σχέση μεταξύ των επιπέδων γλυκόζης (x-άξονας) και της δόσης ινσουλίνης (y-άξονας). Στο ίδιο διάγραμμα προστίθεται μία γραμμή γραμμικής παλινδρόμησης, βοηθώντας να ελέγξουμε αν υπάρχει σαφής γραμμική συσχέτιση.

Υλοποίηση στο Blade Template

Και εδώ, στη σελίδα `diagrams.blade.php`, ενεργοποιείται η λογική μόνο αν το επιλεγμένο διάγραμμα είναι “Glucose-Insulin Correlation”:

```
<!-- Glucose-Insulin Correlation Chart -->
@elseif(request('diagram_type') == 'glucose_insulin_correlation')
    <!-- minimum height για το chart container -->
    <div class="text-center" style="padding-right: 30px; max-width: 1300px; min-height: 600px;">
        <canvas id="glucoseInsulinChart"></canvas>
    </div>
@endif
```

Σχήμα 37: Υλοποίηση Διαγράμματος Glucose-Insulin Correlation στο Blade Template

Στη JavaScript, χρησιμοποιούμε `type: 'scatter'` για να σχεδιάσουμε τα σημεία (glucose vs insulin). Κατόπιν, με το “`showLine: true`”, σχεδιάζουμε την διακεκομμένη γραμμή παλινδρόμησης. Επιπλέον, ορίζουμε δύο ξεχωριστά σημεία (intercept & slope point), ώστε να οπτικοποιούμε τις τιμές του «τεταγμένου στην αρχή» (intercept) και τη συμβολή της κλίσης (slope).

Προετοιμασία δεδομένων και Γραμμική Παλινδρόμηση

Στον Controller, αφού ανακτήσουμε τις μετρήσεις γλυκόζης και ινσουλίνης, καλείται η μέθοδος:

```
// Υπολογισμός γραμμικής παλινδρόμησης (Linear Regression)
$regressionData = $this->performLinearRegression($glucoseData, $insulinData);
```

Σχήμα 38: Υπολογισμός Γραμμικής Παλινδρόμησης για Συσχέτιση Γλυκόζης-Ινσουλίνης

Η `performLinearRegression` υπολογίζει:

- Την κλίση (*slope*)
- Την εκτομή (*intercept*)

- Τις προβλεπόμενες τιμές: $y = slope \times x + intercept$ και επιστρέφει τα αποτελέσματα στο Blade. Έτσι, η γραμμή παλινδρόμησης σχεδιάζεται ακριβώς πάνω από τα σημεία δεδομένων, υποδηλώνοντας τον βαθμό συσχέτισης γλυκόζης–ινσουλίνης.

3. Future Glucose Trend Prediction

Στόχος & Περιγραφή

Το Future Glucose Trend Prediction εμφανίζει τόσο τα ιστορικά επίπεδα γλυκόζης όσο και τις προβλεπόμενες μελλοντικές τιμές (π.χ. για τις επόμενες δύο μέρες, σε τρεις χρονικές στιγμές ανά ημέρα). Αποτελεί ένα διαφορετικό κομμάτι, το οποίο υλοποιείται στη σελίδα `glucose_trend_page.blade.php`. Τα δεδομένα αντλούνται από ένα Flask API σε Python, το οποίο αξιοποιεί το XGBoost για να εκτελέσει τη διαδικασία εκπαίδευσης και πρόβλεψης.

Υλοποίηση στο Blade Template

Το Laravel Controller (`showFutureGlucoseTrend`) καλεί την Python endpoint:

```
// Κλήση API Python, με το patient_id, για πρόβλεψη
$response = Http::get( url: 'http://127.0.0.1:5000/predict_future_glucose', [
    'patient_id' => $patientId
]);
```

Σχήμα 39: Κλήση Flask API για Πρόβλεψη Μελλοντικών Τιμών Γλυκόζης

Αν η κλήση επιτύχει, παίρνουμε πίσω έναν πίνακα με:

- `glucoseData` (ιστορικές τιμές)
- `timeData` (αντίστοιχα timestamps)
- `predictedFuture` (οι επόμενες έξι τιμές)
- `futureTimeData` (τα timestamps των μελλοντικών τιμών)

Έπειτα τα δεδομένα περνούν στο `glucose_trend_page.blade.php`, όπου σχεδιάζονται με `Chart.js`. Συγκεκριμένα, στο `data`: ορίζουμε ένα συνδυαστικό σύνολο ιστορικών + προβλεπόμενων labels (χρονικών στιγμών), και για το `dataset` της πρόβλεψης δημιουργούμε null τιμές για τη διάρκεια των ιστορικών, ώστε η διακεκομμένη γραμμή να ξεκινήσει ακριβώς μετά τα πραγματικά δεδομένα.

Κώδικας Python με XGBoost

Ακολουθεί ένα χαρακτηριστικό κομμάτι του Python script (predict_future_glucose) που επιστρέφει τις έξι μελλοντικές τιμές:

```
# Εκπαίδευση μοντέλου XGBoost
model = XGBRegressor(n_estimators=80, random_state=0)
model.fit(X, y)

# Υπολογισμός του MAE (Μέσο Απόλυτο Σφάλμα)
y_pred_train = model.predict(X) # Πρόβλεψη στις εκπαιδευτικές τιμές
mae = mean_absolute_error(y, y_pred_train) # Υπολογισμός του MAE
print(f"Training MAE: {mae:.4f}") # Εκτύπωση του MAE στο terminal

# Καθορισμός της τελευταίας χρονικής στιγμής των δεδομένων
last_time = df.index[-1]
predicted_future = [] # Λίστα για τις μελλοντικές προβλέψεις
last_window = values[-window_size:].copy() # Τελευταίο παράθυρο δεδομένων

# Καθορισμός χρονικών στιγμών: Πρωί (08:00), Μεσημέρι (13:00), Βράδυ (20:00)
time_slots = [8, 13, 20]

# Εύρεση της επόμενης χρονικής στιγμής με βάση την τελευταία εγγραφή
last_hour = last_time.hour
next_time_slot = None

for slot in time_slots:
    if last_hour < slot:
        next_time_slot = datetime(last_time.year, last_time.month, last_time.day, slot, 0)
        break

# Αν η ώρα είναι μετά τις 20:00, μεταβαίνουμε στην επόμενη ημέρα στις 08:00
if not next_time_slot:
    next_time_slot = datetime(last_time.year, last_time.month, last_time.day, 8, 0) + timedelta(days=1)

# Δημιουργία 6 μελλοντικών χρονικών στιγμών
prediction_times = []
for _ in range(6):
    prediction_times.append(next_time_slot) # Προσθήκη χρονικής στιγμής
    next_slot_index = (time_slots.index(next_time_slot.hour) + 1) % len(time_slots) # Εύρεση επόμενης ώρας
    next_time_slot = datetime(
        next_time_slot.year, next_time_slot.month, next_time_slot.day, time_slots[next_slot_index], 0
    )

# Ενημέρωση ημέρας αν η ώρα μεταβαίνει από βράδυ (20:00) σε πρωί (08:00)
if next_time_slot.hour == 8:
    next_time_slot += timedelta(days=1)

# Υπολογισμός 6 μελλοντικών προβλέψεων
for _ in prediction_times:
    pred = model.predict(last_window.reshape(1, -1))[0] # Πρόβλεψη
    rounded_pred = round(float(pred)) # Στρογγυλοποίηση στην πλησιέστερη ακέραια τιμή
    predicted_future.append(rounded_pred) # Αποθήκευση πρόβλεψης
    last_window = np.append(last_window[1:], rounded_pred) # Ενημέρωση παραθύρου
```

Σχήμα 40: Υλοποίηση Πρόβλεψης Μελλοντικών Τιμών Γλυκόζης με XGBoost

Εδώ φαίνεται ξεκάθαρα πως:

1. Εκπαιδεύουμε το μοντέλο με τα ιστορικά δεδομένα γλυκόζης του ασθενούς, χρησιμοποιώντας ένα παράθυρο 12 σημείων (window_size) ως χαρακτηριστικά.
2. Παίρνουμε το τελευταίο διαθέσιμο “πακέτο” 12 μετρήσεων και προβλέπουμε την επόμενη τιμή.
3. Ενημερώνουμε το “παράθυρο” (αφαιρούμε την παλαιότερη τιμή, προσθέτουμε την καινούρια πρόβλεψη) και το επαναχρησιμοποιούμε για να παράξουμε την επόμενη πρόβλεψη. Κάνουμε αυτό το βήμα έξι φορές, μία για κάθε μελλοντική στιγμή.

4. Επιστρέφουμε τα αποτελέσματα στην εφαρμογή Laravel σε μορφή JSON, τα οποία έπειτα απεικονίζονται ως “Predicted Glucose” στο Chart.js διάγραμμα.

8 Βιβλιογραφία

- [1] Εφαρμογή GlucoLog RapidCalc. *App software for diabetes management*. Ανακτήθηκε στις 01/01/2025 από: <https://glucomenday.com/newplatform/app-software-for-diabetes-management/gluccolog-rapid-calc/>
- [2] Εφαρμογή Diabetes:M. *Diabetes Management*. Ανακτήθηκε στις 01/01/2025 από: <https://diabetes-m.com/>
- [3] Dzida, G., & Συνεργάτες. (2019). *Expert Opinion: Recommendation of diabetes experts on the use of FreeStyle Libre in diabetic patients in Poland*. Ανακτήθηκε από: <https://ruj.uj.edu.pl/server/api/core/bitstreams/bdc6855a-0baa-497e-aad7-aa6afa5049ce/content>
- [4] Εφαρμογή Hedia Diabetes Assistant. Ανακτήθηκε στις 01/01/2025 από: <https://www.hedia.com/>
- [5] Εφαρμογή Sugarmate. Ανακτήθηκε στις 01/01/2025 από: <https://www.sugarmate.io/features>
- [6] Qu, K. (2024). Research on Linear Regression Algorithm. *MATEC Web of Conferences*, 395, άρθρο 01046. Ανακτήθηκε στις 01/01/2025 από: https://www.matec-conferences.org/articles/matecconf/pdf/2024/07/matecconf_icpcm2023_01046.pdf
- [7] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning* (3rd ed.). Springer Texts in Statistics. Ανακτήθηκε στις 01/01/2025 από: https://doi.org/10.1007/978-3-031-38747-0_3
- [8] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- [9] Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- [10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
- [11] Serokell. Random Forest image. Ανακτήθηκε στις 01/01/2025 από: <https://serokell.io/blog/random-forest-classification>
- [12] Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media. Ανακτήθηκε στις 01/01/2025
- [13] Keen, B. A. (2017). Feature Scaling with scikit-learn. Ανακτήθηκε στις 01/01/2025 από: <https://benalexkeen.com/feature-scaling-with-scikit-learn/>

- [14] GeeksforGeeks. (2024). Linear Regression in Machine learning. Ανακτήθηκε στις 01/01/2025 από: <https://www.geeksforgeeks.org/ml-linear-regression/>
- [15] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 785–794. Ανακτήθηκε στις 01/01/2025
- [16] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- [17] Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378.
- [18] Laravel Documentation. *Laravel 11.x Documentation*. Ανακτήθηκε στις 01/01/2025 από: <https://laravel.com/docs/11.x>.
- [19] GeeksforGeeks. Laravel import/export Excel file. Ανακτήθηκε στις 01/01/2025 από: <https://www.geeksforgeeks.org/laravel-import-export-excel-file/>
- [20] Baranwal, A., Bagwe, B. R., & Vanitha, M. (2020). Machine learning in Python: Diabetes prediction using machine learning. Ανακτήθηκε στις 01/01/2025
- [21] Tatachar, A. V. (2021). [Comparative assessment of regression models based on model evaluation metrics](#). *International Research Journal of Engineering and Technology (IRJET)*. Ανακτήθηκε στις 01/01/2025