Proactive Diabetes Prediction: A Comprehensive Simulation Framework Incorporating Longitudinal Pre-Meal Parameters, Body Composition, and Dietary Habits

Rationale/ Introduction

Diabetes mellitus or most commonly known as Diabetes is a chronic metabolic condition characterized by high blood glucose (or blood sugar) levels, which over time cause catastrophic damage to the heart, blood vessels, eyes, kidneys, and nerves. There are two types of diabetes; type 1 diabetes and type 2 diabetes. The most common is type 2 diabetes, which mainly affects adults and arises when the body develops insulin resistance or fails to produce enough insulin. In the past 3 decades the prevalence of type 2 diabetes has risen dramatically in countries of all income levels. On the other hand, Type 1 diabetes, also known as juvenile diabetes or insulin-dependent diabetes, is a chronic illness in which the pancreas generates little to no insulin on its own (World Health Organization et al., 2023).

According to (PAHO, 2021) in 2019 alone, region wide diabetes mellitus (excluding kidney diseases due to diabetes) accounts for: 284,049 fatalities, including 139,651 men and 144,398 women. And the age-standardized death rate from diabetes was projected to be 20.9 per 100,000 people. Also, another research finding of (Statista Research Department, 2024) stated that according to preliminary data from January to September 2023, diabetes mellitus diseases accounted for 6.4% of all deaths in the Philippines. Deaths from such illnesses peaked in 2020, accounting for 6.5 percent of total deaths.Traditional diabetes care techniques frequently prioritize reactive measures upon diagnosis over proactive strategies for prevention and early intervention. However, recent advances in health technology and data analytics have created prospects for developing proactive methods to diabetes prediction and prevention.

The proposed research aims to address this gap by creating a complete simulation framework for proactive diabetes prediction using Logistic Regression, Decision Trees, Random Forests, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), Survival Analysis, or Neural Networks. This aims to give personalized risk assessments for people at different phases of diabetes risk by combining longitudinal pre-meal measurements, body composition data, and detailed eating habits. The simulation framework predicts the likelihood of diabetes onset using longitudinal data and advanced modeling approaches, allowing for prompt interventions to minimize risk variables and prevent disease progression. Furthermore, by incorporating multiple factors influencing diabetes risk, such as pre-meal parameters, body composition, and dietary habits, the simulation framework provides a comprehensive approach to diabetes prediction that takes into account the complex interaction of physiological, behavioral, and environmental factors. This method enables a more precise and personalized evaluation of diabetes risk, allowing for focused therapy based on individual needs.

Significance of the Study

The significance of this study lies in its potential to revolutionize diabetes prevention and management strategies through the development of a proactive simulation framework. By integrating longitudinal pre-meal parameters, body composition, and a comprehensive analysis of dietary habits, this research offers several notable contributions:

The simulation framework offers personalized predictions for diabetes risk, allowing individuals to understand their specific risk factors and take proactive measures to prevent or delay the onset of diabetes. Tailoring interventions to individual needs empowers people to make informed lifestyle choices and adopt healthier behaviors. By predicting diabetes risk early, the simulation framework creates opportunities for early intervention. This helps mitigate risk factors and prevent the progression of diabetes-related complications.

The development of a comprehensive simulation framework for proactive diabetes prediction has the potential to have a broader public health impact by reducing the incidence of diabetes and its associated complications, contributing to improved population health outcomes, and enhanced quality of life for individuals who may or may not yet be affected by diabetes.

Scope and Limitations of the Study

This study focuses on enhancing the accuracy and depth of sentiment analysis on Facebook post comments by integrating the Naïve Bayes Classifier and Semantic Role Labeling (SRL) techniques. It involves the development of a sentiment analysis system capable of classifying comments into positive, negative, or neutral sentiments. The system will be trained and validated using a dataset of Facebook post comments to ensure its effectiveness. Additionally, the study evaluates the system’s performance through key metrics such as accuracy, precision, recall, and F1-score while analyzing the impact of integrating Naïve Bayes and SRL on improving sentiment classification.

However, the study has several limitations that must be considered. First, the accuracy of the sentiment analysis system is dependent on the quality and diversity of the training dataset; biases in data collection may affect its generalizability to different contexts. Second, while the Naïve Bayes Classifier and SRL enhance sentiment classification, they may not fully capture complex linguistic nuances such as sarcasm, idiomatic expressions, or evolving online slang, which can lead to misclassification. Additionally, the study is limited to Facebook post comments, meaning that the findings may not be directly applicable to other social media platforms with different user behaviors and language patterns.

Despite these limitations, this research contributes to the advancement of sentiment analysis by demonstrating the effectiveness of integrating machine learning and semantic analysis techniques. The findings provide valuable insights into improving text classification models, which can be utilized in social media monitoring, brand analysis, and opinion mining. Future studies may address the limitations by incorporating deep learning models, expanding the dataset across multiple platforms, and exploring additional linguistic features for more comprehensive sentiment analysis.

Objectives of the Study

The objective of this study is to develop and validate a comprehensive simulation framework for proactive diabetes prediction. Specifically, the study aims to achieve the following objectives:

1. Design and implement a simulation framework that integrates longitudinal pre-meal parameters, body composition data, and detailed analysis of dietary habits to predict an individual's risk of developing diabetes over time.
2. Collect and incorporate longitudinal data on pre-meal parameters, including blood sugar levels, insulin sensitivity, and other relevant biomarkers, to capture dynamic changes in physiological indicators associated with diabetes risk.
3. Incorporate measurements of body composition, such as BMI, waist circumference, body fat percentage, and muscle mass, into the simulation framework to account for the influence of adiposity and lean body mass on diabetes risk.
4. Integrate detailed analysis of dietary habits, including caloric intake, macronutrient

composition, micronutrient intake, meal timing, and dietary patterns, to assess the impact of dietary factors on diabetes risk prediction.

5. Validate the accuracy and reliability of the simulation framework using real-world data sets and evaluate its performance in predicting diabetes risk compared to traditional risk assessment methods.

By achieving these objectives, the study aims to develop a novel approach to diabetes prediction that offers personalized risk assessments based on longitudinal pre-meal parameters, body composition dynamics, and dietary habits analysis, ultimately contributing to early detection, prevention, and management of diabetes.

Expected Outputs

The system begins by preparing the text data for analysis. This involves cleaning the text by removing unnecessary characters, punctuation, and symbols to ensure clarity and coherence. Then, the text will undergo tokenization, where it is broken down into individual words or tokens for further examination. Normalization follows by standardizing the text by converting it to lowercase to maintain consistency in analysis regardless of capitalization variations. Stop words such as common and insignificant words like "the" or "and," are then removed to focus on meaningful content. Lastly, lemmatization or stemming is applied to reduce words to their root forms. Once the text is preprocessed, the system will employ a Naive Bayes Classifier to determine the sentiment of the comments. This classifier learns from a labeled dataset of Facebook comments to understand the patterns associated with positive, negative, and neutral sentiments. By extracting relevant features from the training data, the classifier becomes proficient in predicting the sentiment of new comments based on the learned patterns.

Furthermore, the system integrates Semantic Role Labeling (SRL) to gain insights into the contextual relationships within the comments. Through dependency parsing, it analyzes the grammatical structure of sentences and identifies relationships between words, such as subject-verb-object relationships. Semantic roles including entities such as agents, patients, times, or locations are then assigned to words or phrases within the comments. By integrating this contextual information with sentiment analysis, the system will achieve a deeper understanding of the emotions expressed in the comments.

To implement the system effectively, several materials are required including a dataset of labeled Facebook post comments, Python libraries for natural language processing tasks, a machine learning library like scikit-learn, and an SRL toolkit such as Stanford CoreNLP or spaCy.

References

Concepcion, A., & Sy, C. (n.d.). Modeling the spread of fake news on social networking sites using the system dynamics approach. ASEAN Engineering Journal. https://journals.utm.my/aej/article/view/19251

Duzen, Z., Riveni, M., & Aktaş, M. S. (2023). Analyzing the spread of Misinformation on Social Networks: A process and software architecture for detection and analysis. Computers, 12(11), 232. https://doi.org/10.3390/computers12110232

Luo, H., Cai, M., & Cui, Y. (2021, December 16). Spread of misinformation in social networks: Analysis based on Weibo Tweets. Security and Communication Networks. https://www.hindawi.com/journals/scn/2021/7999760/

Simulation of social media networks with agent-based modeling (2019). https://www.theseus.fi/bitstream/handle/10024/174027/Frida-Gullichsen-thesis.pdf?sequence=2&isAllowed=y

Sulis, E., & Tambuscio, M. (2020). Simulation of misinformation spreading processes in social networks: an application with NetLogo. https://ieeexplore.ieee.org/document/9260064

T, S. M., & Mathew, S. K. (2022, February 15). The disaster of misinformation: A Review of Research in Social Media - International Journal of Data Science and Analytics. SpringerLink. https://link.springer.com/article/10.1007/s41060-022-0031