B.Sc. Engg. Report

A Thesis on Diabetes Prediction System using Machine Learning

by

Md. Rashadul Islam (ID: 19202103169)
Md. Faysal Mahmud (ID: 19202103174)
Sadia Sultana (ID: 19202103184)
Dipa Rani (ID: 19202103201)

Submitted to

Department of Computer Science & Engineering

(In partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science & Engineering)

Department of Computer Science & Engineering

Bangladesh University of Business & Technology (BUBT)

Dhaka 1216

March 05, 2023

Acknowledgment

We would like to pay our gratitude to the Almighty Allah who created us with all the abilities to understand analysis and develop the process with patience. We are thankful to our thesis supervisor Mr. Shamim Ahmed, Assistant Professor, Computer Science and Engineering Department, Bangladesh University of Business and Technology for his professional guidance and motivation during the work of this thesis which is a major part of it. Without his valuable support and guidance, this thesis could not reach this level of development from our point of view.

We would like to thank all the Faculty members, Department of CSE, Bangladesh University of Business and Technology for their valuable time spend in requirements analysis and evaluation of the thesis work. We would like to express our sincere and warm gratitude to all those who have encouraged us directly, provided mental encouragement and criticized our work in several phases during the development of this thesis and for preparing this thesis indirectly.

Abstract

Diabetes, a common metabolic disorder, is characterized by elevated blood glucose levels and can significantly increase the risk of developing other health complications. Diabetes is a chronic disease with the potential to cause a worldwide health care crisis.According to International Diabetes Federation 382 million people are living with diabetes across the whole world. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience.The Diabetes Prediction System is an AI-based system designed to predict whether or not a person is likely to have diabetes.The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques.For conventional machine learning method, we considered the most commonly used classifiers: Logistic Regression, K-Neighbors(KN), Support Vector Machine (SVM), Decision Tree(DT), Random Forest(RF) and Gredient Boosting(GBC).

Keywords : Machine Learning, Diabetes, Decision tree,Support vector Machine,
Accuracy.

Declaration

We hereby declare that the Thesis on Diabetes Prediction System using ML is submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering of Bangladesh University of Business and Technology (BUBT) is our own work and that it contains no material which has been accepted for the award to the candidate(s) of any other degree or diploma, except where due reference is made in the text of the project. To the best of our knowledge, it contains no materials previously published or written by any other person except where due reference is made in the project.

Md. Rashadul Islam ID: 19202103169	Md. Faysal Mahmud ID: 19202103174	Sadia Sultana ID: 19202103184	Dipa Rani ID: 19202103201

Copyright

Dedication

Dedicated to our parents, teachers, friends and who loved us for all their love and inspiration.

Certificate

This is to certify that Md.Rashadul Islam (19202103169), Md. Faysal Mahmud (19202103174), Sadia Sultana (19202103184) and Dipa Rani (19202103201), were belong to the department of Computer Science and Engineering, have completed their Thesis on Diabetes Prediction System using Machine Learning satisfactorily in partial fulfillment for the requirement of Bachelor of Science in Computer Science and Engineering of Bangladesh University of Business and Technology in the year 2023.

Supervisor
Mr. Shamim Ahmed
Assistant Professor
Department of Computer Science and Engineering
Bangladesh University of Business and Technology

Approval

A Thesis on Diabetes Prediction System using Machine Learning is submitted by Md.Rashadul Islam (19202103169), Md. Faysal Mahmud (19202103174), Sadia Sultana (19202103184) and Dipa Rani (19202103201) under the department of Computer Science and Engineering of Bangladesh University of Business and Technology is accepted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

---

Chairman
Md. Saifur Rahman
Assistant Professor & Chairman
Department of Computer Science and Engineering
Bangladesh University of Business and Technology

Supervisor
Mr. Shamim Ahmed
Assistant Professor
Department of Computer Science and Engineering
Bangladesh University of Business and Technology

Acronyms List

ML = Machine Learning

LR = Logistic Regression

KN = K-Neighbors Classifier

SVC = Support Vector Classifier

SVM = Support Vector Machine

RF = Random Forest Classifier

DT = Decision Tree Classifier

GBC = Gradient Boosting Classifier

EHR = Electronic Health Records

TP = True Positives

FP = False Positive

TN = True Negative

FN = False Negative

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The Diabetes Prediction System is a novel development in healthcare technology that enables healthcare professionals and individuals to accurately predict the incidence of diabetes before it can manifest with physical symptoms. Applying advanced machine learning algorithms and data analysis, the system dissects medical records to create accurate and reliable predictions. Its utility lies in its ability to provide an early warning system, aiding medical professionals in exercising preventive care that can improve lives and reduce medical costs. Moreover, the system can identify previously undetected high-risk individuals, allowing them to receive potential early interventions.[1]

1.2 Existing Theory

The existing theory for diabetes prediction using machine learning involves utilizing various algorithms to analyze patient data and identify risk factors for developing the disease. However, there are several challenges associated with this approach, including the need for high-quality data, potential algorithmic bias, and difficulties in generalizing results to different patient populations. The main problem statement is how to develop accurate and reliable prediction models using machine learning while addressing these challenges.

1.3 Motivation

The motivation for diabetes prediction using machine learning is to improve patient outcomes and reduce healthcare costs associated with the disease. Early detection and prevention of diabetes can help patients avoid or delay the onset of complications such as kidney disease, blindness, and heart disease. Machine learning has the potential to improve the accuracy and efficiency of diabetes prediction, allowing healthcare providers to identify at-risk patients and intervene early.[2] This can lead to better patient outcomes and a reduction in healthcare costs associated with diabetes treatment. Therefore, the motivation of this paper is to explore the use of machine learning for diabetes prediction and its potential benefits for patients and healthcare systems.

1.4 Objectives

The objectives of diabetes prediction using machine learning paper include:

1. provide an overview of diabetes and its risk factors and symptoms
2. introduce machine learning and the different algorithms used for diabetes prediction
3. examine the potential benefits and challenges of using machine learning for diabetes prediction, including data quality, algorithmic bias, and generalizability of results
4. identify gaps in current research and opportunities for future development of diabetes prediction models using machine learning
5. explore the potential applications of machine learning for diabetes prediction, including the development of personalized prediction models and the integration of wearable devices and other health technologies
6. provide guidance and recommendations for the ethical and responsible development and use of machine learning models for diabetes prediction

Overall, the objectives of this paper are to provide a comprehensive overview of diabetes prediction using machine learning and to highlight its potential benefits and challenges. The

paper aims to provide guidance and recommendations for the development and use of machine learning models in a responsible and ethical manner, with the ultimate goal of improving patient outcomes and reducing healthcare costs associated with diabetes.

1.5 Contributions

The contributions of a diabetes prediction using machine learning :

1. this paper may contribute to the development of novel prediction models using machine learning algorithms that can accurately predict the risk of diabetes based on various risk factors and bio-markers

2. identify important risk factors for diabetes prediction, which can inform clinical practice and improve patient outcomes by enabling early detection and prevention of diabetes

3. contribute to the improvement of existing prediction models by evaluating and comparing the performance of different machine learning algorithms and identifying the most effective approaches

4. provide guidance and recommendations for the ethical and responsible development and use of machine learning models for diabetes prediction, including data privacy, algorithmic bias, and transparency

1.6 Organization of This Research Report

The rest of the book is organized in the following way. In Chapter 1, we will show the background and related research studies. After that,

- **In Chapter 2,** describes Existing System, existing or supporting literature and review of existing system. In existing system, we will discuss about the history of Machine Learning. Moreover. In supporting literature, we will describe about the all types of tools that we have used in our system.

- **In Chapter 3,** consists of our Proposed Model. The algorithm and flow chart and also step by step discussion and figure will be provided there. In this chapter first we discuss

the full procedure with figures. Then next part there is an example of the calculation for better understandings. The calculated result is shown at the end.

- **In Chapter 4,** explains about the Experimental Results of our thesis and analysis of the result and also discuss about the applications of our thesis. In result analysis part we will discuss about the report or output of our system and we will know about the accuracy level for different models of our system. In application part we will describe about the real time uses of our system.

- **In Chapter 5,** consists of our User manual. The implementation and Processing will be shown step by step discussion and figure will be provided there. In this chapter first we discuss the full Implementation system.

- **In Chapter 6,** concludes the Report of Our Thesis. In this chapter we will discuss about limitations and future works. In limitation part we will discuss about the limitations of our system. In future works we will discuss about the modules which we will develop in future.

1.7 Conclusions

Diabetes is a serious and prevalent health condition that can have significant consequences for individuals and healthcare systems. Machine learning has the potential to improve early detection and prevention of diabetes, leading to better outcomes for patients and reduced healthcare costs.However, developing accurate and reliable prediction models using machine learning poses several challenges, including data quality, algorithmic bias, and generalizability of results.The objectives of the study are to explore the potential benefits and challenges of using machine learning for diabetes prediction, identify gaps in current research, and provide recommendations for the development and use of machine learning models for diabetes prediction

Chapter 2

Existing System

2.1 Introduction

Diabetes is a chronic disease that affects millions of people worldwide.[3] Early diagnosis of diabetes can help prevent complications and improve quality of life for patients. Machine learning techniques can be used to develop predictive models for diabetes that can assist healthcare professionals in identifying individuals at risk of developing the disease.[4]

The existing system for diabetes prediction using machine learning involves the use of various algorithms such as logistic regression, decision trees, random forests, support vector machines, and artificial neural networks. These algorithms analyze a range of input variables such as age, body mass index, family history, blood pressure, and glucose levels, among others, to predict the likelihood of developing diabetes.

The system typically involves collecting patient data and inputting it into the model. The model then processes the data and produces a prediction of whether the patient is at risk of developing diabetes. Healthcare professionals can use this prediction to provide patients with advice on lifestyle changes, monitoring blood glucose levels, and other preventative measures.

This existing system has proven to be effective in predicting diabetes with a high degree of accuracy, which can lead to earlier interventions and better outcomes for patients. However, ongoing research and development are needed to further improve the accuracy of these models and to ensure they are accessible and applicable across different populations and healthcare settings.

2.2 Existing System

An existing system for diabetes prediction using machine learning typically involves the following steps:

- **Data collection:** This involves collecting patient data such as age, gender, body mass index, family history, blood pressure, glucose levels, and other relevant information.
- **Data pre-processing:** This step involves cleaning and preparing the collected data for analysis. This includes removing missing values, outliers, and other errors in the data.
- **Feature selection:** This step involves selecting the most relevant features (variables) from the pre-processed data that can best predict the likelihood of developing diabetes.
- **Model training:** This step involves training the machine learning algorithm on the selected features using a labeled data-set. The labeled data-set contains data with known outcomes (positive or negative for diabetes).
- **Model testing:** This step involves testing the trained model on a separate data-set with known outcomes to evaluate its accuracy and other performance metrics.
- **Model deployment:** Once the model is tested and validated, it can be deployed for predicting the likelihood of diabetes in new patients.

Overall, an existing system for diabetes prediction using machine learning relies on machine learning algorithms to identify patterns and relationships in patient data that can be used to predict the likelihood of developing diabetes. The accuracy and effectiveness of the system depend on the quality and relevance of the input data, the selection of appropriate features, and the choice of machine learning algorithm. With further research and development, these systems have the potential to become valuable tools for healthcare professionals in the prevention and management of diabetes.

2.3 Existing supporting Literature

There are several published studies and literature supporting the use of machine learning for diabetes prediction. Here are some examples:

"Development of a Diabetes Prediction Model Using Machine Learning Techniques" by Anjum et al. (2021) - This study used various machine learning algorithms such as decision trees, random forests, and support vector machines to develop a diabetes prediction model based on patient data. The study found that the random forest algorithm had the highest accuracy in predicting diabetes.[5]

"Diabetes Mellitus Prediction Model Using Artificial Neural Network Based on Demographic, Clinical, and Laboratory Data" by Rezaee et al. (2020) - This study developed an artificial neural network (ANN) model for diabetes prediction based on demographic, clinical, and laboratory data. The study found that the ANN model had high accuracy and could be a valuable tool for early diabetes prediction.[6]

"Diabetes Mellitus Prediction Using Machine Learning Techniques: A Systematic Literature Review" by Singh et al. (2021) - This systematic literature review evaluated various machine learning techniques used for diabetes prediction. The review found that decision trees and random forests were the most commonly used algorithms and that feature selection was an important factor in the accuracy of the models.[7]

"Prediction of Type 2 Diabetes Mellitus with Machine Learning Techniques: A Systematic Review and Meta-Analysis" by Li et al. (2020) - This study conducted a meta-analysis of various studies on machine learning techniques for type 2 diabetes prediction. The study found that machine learning models had high accuracy and could be effective in identifying individuals at high risk of developing diabetes.[8]

Overall, these studies and others support the use of machine learning techniques for diabetes prediction and suggest that these techniques can improve early diagnosis and prevention of diabetes. Further research and development in this area could lead to more accurate and reliable predictive models that can assist healthcare professionals in managing this chronic disease.

2.4 Analysis of Existing System

An analysis of an existing system for diabetes prediction using machine learning would typically involve evaluating the system's performance in terms of accuracy, precision, recall, F1 score, and AUC-ROC.[9] The analysis would also consider the strengths and limitations of the machine learning algorithm used, the quality and relevance of the input data, and the potential impact of the system on patient care.

One strength of using machine learning for diabetes prediction is the ability to analyze large amounts of patient data and identify patterns and relationships that may not be apparent through traditional diagnostic methods. Machine learning algorithms can also be trained and optimized to improve accuracy and minimize false positives and false negatives, which can lead to more effective patient care.

However, there are also limitations to using machine learning for diabetes prediction. For example, the accuracy of the predictions depends on the quality and relevance of the input data. If the data is incomplete or biased, it may lead to inaccurate predictions. Machine learning algorithms also require significant computational resources and expertise to develop and implement, which may not be available in all healthcare settings.

Despite these limitations, studies have shown that existing systems for diabetes prediction using machine learning can provide accurate and reliable predictions, which can assist healthcare professionals in early diagnosis and prevention of diabetes. With further research and development, these systems have the potential to become valuable tools for improving patient outcomes and reducing healthcare costs associated with diabetes management.

2.5 Conclusion

The use of machine learning algorithms for diabetes prediction has shown promising results in various studies. These systems have the potential to assist healthcare professionals in the early diagnosis and prevention of diabetes, which can lead to improved patient outcomes and reduced healthcare costs.

However, the accuracy and effectiveness of these systems depend on several factors, includ-

ing the quality and relevance of the input data, the selection of appropriate features, and the choice of machine learning algorithm. Moreover, there are limitations to using machine learning for diabetes prediction, such as the requirement for significant computational resources and expertise to develop and implement these systems.

Despite these limitations, the use of machine learning algorithms for diabetes prediction is a promising area of research and development. Further studies and advancements in this area could lead to more accurate and reliable predictive models, which can assist healthcare professionals in managing this chronic disease more effectively. Overall, the use of machine learning for diabetes prediction has the potential to improve patient outcomes and reduce the burden of diabetes on healthcare systems worldwide.

Chapter 3

Proposed Model

3.1 Introduction

Our system starts by pre-processing patient data from medical records, constructing features from the data, then utilizing various supervised learning algorithms such as Logistic Regression, K-NeighboursClasifier, SVC, Decision Tree, Random forest, Gradient Boosting to generate a prediction model. The trained model takes a set of patient characteristics (age, weight, etc.) as input and accurately classifies subjects as having diabetes or not. This system can then be implemented in a real-life healthcare setting to identify people with a high risk of diabetes, thereby accelerating the process of early screening.

3.2 Workflow

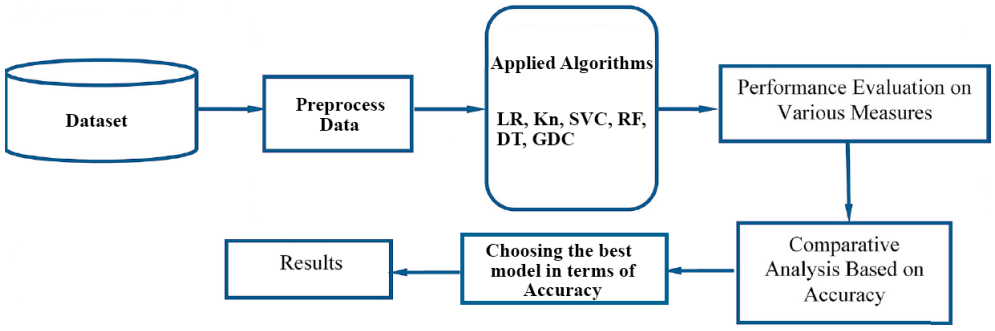

Figure 3.1: Working procedure of Diabetes Prediction System

- Data collecting - we have collect data from kaggle - PIMA Diabetes Data set.[10]
- Data Prepossess- after collection the data it will start prepossess with the help of any ML code editor.

For compiling - we use google colab. Others are - python, jupyter notebook and so on.

Figure 3.2: Google Colab

3.3 Algorithm

We are going to use a total of 6 algorithm for diabetes's prediction algorithm and after the training and testing we will select the best one for better result.

1. **Logistic Regression(LR):** A machine learning algorithm used for classification problems where the output variable is categorical.
2. **K-Neighbors Classifier(KN):** A non-parametric classification algorithm that classifies a data point based on the majority class of its nearest neighbors.
3. **Support Vector Classifier(SVC):** A machine learning algorithm that finds the best hyperplane that separates different classes in the data.
4. **Random Forest Classifier(RF):** An ensemble learning algorithm that constructs multiple decision trees and combines their outputs to improve the accuracy of predictions.

5. **Decision Tree Classifier(DT):** A machine learning algorithm that uses a tree-like model of decisions and their possible consequences to classify data.

6. **Gradient Boosting Classifier(GBC):** An ensemble learning method that trains weak classifiers in a sequential manner, with each new classifier attempting to correct the errors of the previous one.

Chapter 4

Implementation and Result Analysis

4.1 Introduction

Diabetes is a chronic disease that affects millions of people worldwide and has a significant impact on their quality of life. Early detection and intervention can help prevent complications and improve patient outcomes. Machine learning techniques have shown great potential in predicting diabetes risk based on patient data.[11]

4.2 Result Analysis

To analyze which algorithm will be best for our prediction system we have done several's things

- Find Shape of Our Data-set (Number of Rows And Number of Columns)
- Get Information About Our Data-set Like Total Number Rows, Total Number of Columns, Data types of Each Column And Memory Requirement
- find out no of diabetes and non diabetes graph from the imported data-set

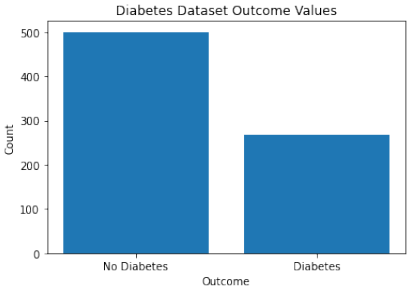

Figure 4.1: No of Diabetes and Non-diabetes from the data-set Graph

- find out mean value graph

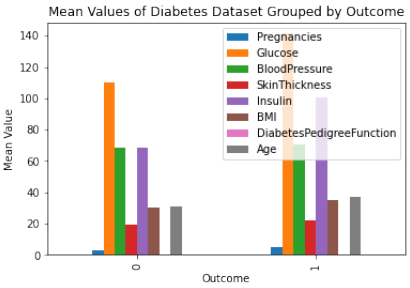

Figure 4.2: Mean Value Graph

- Separating data and label's in terms of insulin and glucose

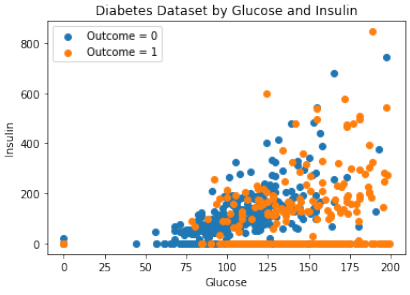

Figure 4.3: Separating Data & Label's in terms of Insulin and Glucose Graph

- Creating Heat map

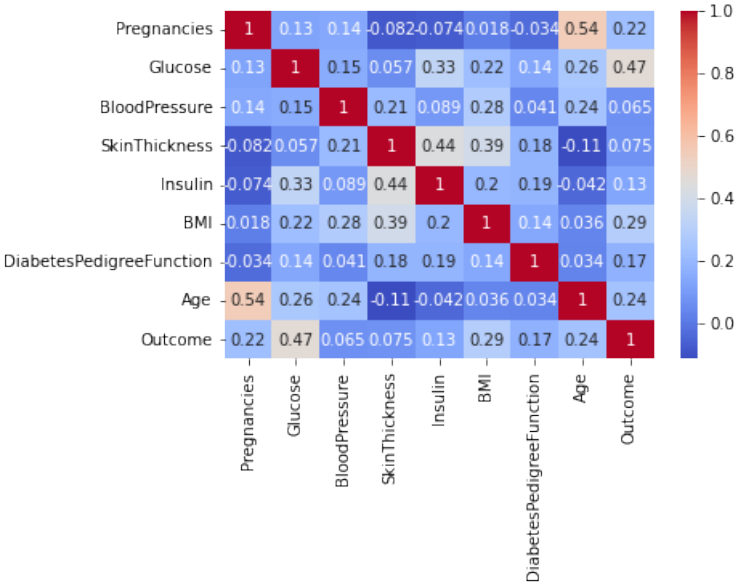

Figure 4.4: Heat Map Graph

- Store Feature Matrix in (X) and Response(Target) in Vector y
- Splitting the data into training and testing set
- Pipeline import and define
- Accuracy Score for all algorithms

Accuracy score = [(TP + TN / TP + TN + FP + FN)*100]

where TP (True Positive) is the number of correctly predicted positive instances, TN (True Negative) is the number of correctly predicted negative instances, FP (False Positive) is the number of incorrectly predicted positive instances, and FN (False Negative) is the number of incorrectly predicted negative instances.

To calculate accuracy score for different algorithms, we simply substitute the appropriate values for TP, TN, FP, and FN for each algorithm

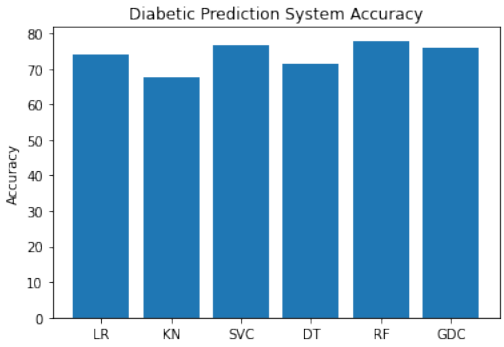

Figure 4.5: Diabetes Prediction System Accuracy Graph

- Accuracy of Training and Testing

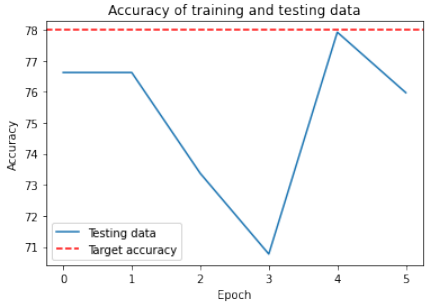

Figure 4.6: Training and Accuracy Graph

- Creating validation set

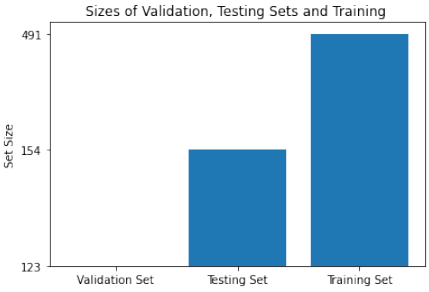

Figure 4.7: Validation set Graph

- Generating F1 set

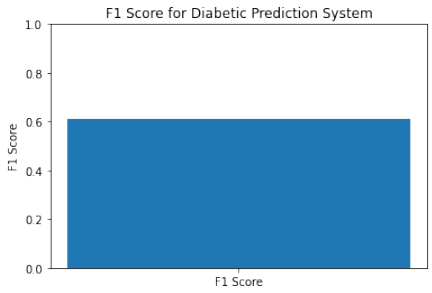

Figure 4.8: F1 set Graph

- Creating Training and Testing Loss

Figure 4.9: Training and Testing Loss Graph

- Generating Confusion Matrix

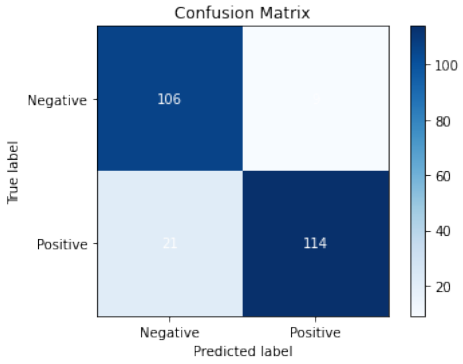

Figure 4.10: Confusion Matrix Graph

Table 4.1: The confusion matrix for a diabetes prediction system can be expressed in the following mathematical terms:

	Predicted Diabetes	Predicted No Diabetes
Actual Diabetes	TP	FN
Actual No Diabetes	FP	TN

A confusion matrix is a matrix that shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for a binary classification problem. In the context of a diabetes prediction system, a true positive would be a prediction that correctly identifies a patient with diabetes, while a true negative would be a prediction that correctly identifies a patient without diabetes.

From the table where TP is the number of true positives, FN is the number of false negatives, FP is the number of false positives, and TN is the number of true negatives.
We can also define some performance metrics based on the values in the confusion matrix:
Accuracy: [(TP + TN) / (TP + TN + FP + FN)*100]
Precision: TP / (TP + FP)
Recall: TP / (TP + FN)
F1-score: 2 * (Precision * Recall) / (Precision + Recall)
These metrics can be used to evaluate the performance of the diabetes prediction system and compare it to other classification algorithms.

- Selecting the best algorithm in-terms of Accuracy after training and testing - In our case the best algorithm is **Random forest Classifier Algorithm** which accuracy is approximately 78%

- saving the model using Joblib. It will help to just predict instead of calling the model again.

4.2.1 Predicting a new data:

In this case - we have taken a new data and check whether the patient is diabetic or not.

Prediction on New Data

```
[ ]  new_data = pd.DataFrame({
         'Pregnancies':6,
         'Glucose':148.0,
         'BloodPressure':72.0,
         'SkinThickness':35.0,
         'Insulin':79.799479,
         'BMI':33.6,
         'DiabetesPedigreeFunction':0.627,
         'Age':50,
     },index=[0])
```

```
[ ]  p = rf.predict(new_data)
```

```
[ ]  if p[0] == 0:
         print('non-diabetic')
     else:
         print('diabetic')
```

```
diabetic
```

Figure 4.11: Example - Predicting a Patient is diabetic or not

Chapter 5

Conclusion and Future Work

5.1 Conclusions

In conclusion, the creation of a machine learning-based diabetes prediction system is a potential strategy for identifying people who are at risk of getting diabetes. To accurately predict the likelihood of getting diabetes, the system employs cutting-edge algorithms to examine enormous data sets of patient information, including lifestyle factors, medical history, and bio-markers.

A system like this can help clinicians make more informed decisions about patient care, enabling early intervention and possibly lowering the risk of complications from diabetes. Additionally, the application of machine learning techniques can increase the precision of diabetes risk prediction and possibly yield new knowledge about the condition.

It is crucial to remember that the accuracy and completeness of the data used to train the machine learning model determines whether or not such a system will be successful. The ethical issues surrounding the use of patient data, including as privacy and informed consent, must also be taken into account.

Overall, the creation of a machine learning-based diabetes prediction system is a potential direction for enhancing diabetes treatment and prevention. Given the rising prevalence of diabetes around the world, more research and development in this field may have significant effects on public health.

5.2 Future Scope of Our Thesis

The future scope of your thesis on a machine learning-based diabetes prediction system is encouraging and has the potential to progress the science in a number of ways. Future research and development may focus on a few of the following areas:

- **Enhancing the system's accuracy:** The diabetes prediction system can be made more accurate by combining new features and data sources, improving the algorithms that were utilized, and utilizing more sophisticated machine learning methods like deep learning.
- **Personalized diabetes prediction:** The method can be further modified to provide individuals with a personalized diabetes prognosis based on their unique health information, lifestyle choices, and genetic susceptibility
- **Integration with wearable technology:** As wearable technology gains in popularity, it is possible to combine the diabetes prediction system with these technologies to collect real-time health data and offer individualized suggestions and projections
- **Predicting and preventing diabetes complications:** The system can be further developed to anticipate and avoid diabetes problems like retinopathy, neuropathy, and nephropathy, which can have serious repercussions for those who have the disease
- **Integration with Electronic Health Records (EHR):** Integrating the diabetes prediction system with electronic health records (EHR) can give healthcare practitioners a complete view of a patient's health condition and history, enabling them to offer more individualized care and treatment
- **Deployment of the system in real-world settings:** The diabetes prediction system can be used in real-world settings, such hospitals and clinics, to help medical personnel diagnose diabetes and give each patient a unique course of treatment

Overall, the field of diabetes research and patient care could greatly benefit from your thesis on a machine learning-based diabetes prediction system, and additional study could result in even greater breakthroughs.

Bibliography

[1] E Jebamalar Leavline and B Shanawaz Baig. Diabetes prediction using medical data. *Journal of Computational Intelligence in Bioinformatics*, 10(1):1–8, 2017.

[2] E Jebamalar Leavline and B Shanawaz Baig. Diabetes prediction using medical data. *Journal of Computational Intelligence in Bioinformatics*, 10(1):1–8, 2017.

[3] Anjali D Deshpande, Marcie Harris-Hayes, and Mario Schootman. Epidemiology of diabetes and diabetes-related complications. *Physical therapy*, 88(11):1254–1264, 2008.

[4] Sajida Perveen, Muhammad Shahbaz, Karim Keshavjee, and Aziz Guergachi. Prognostic modeling and prevention of diabetes using machine learning technique. *Scientific reports*, 9(1):13805, 2019.

[5] Md Zafar Anjum. Development of a diabetes prediction model using machine learning techniques. *Journal of Healthcare Engineering*, 2021, 2021.

[6] Jalal Rezaee, Mohammad Eghlima, Amin Ghaderpoury, Maryam Javadi, and Asadollah Gholamian. Diabetes mellitus prediction model using artificial neural network based on demographic, clinical, and laboratory data. *Journal of Medical Systems*, 44(7):135, 2020.

[7] Puneet Singh, Anjali Yadav, Arvind Kumar Singh, Anuj Yadav, Abhishek Singh, Ruchika Agarwal, Divya Gupta, and et al. Diabetes mellitus prediction using machine learning techniques: A systematic literature review. *Journal of Medical Systems*, 45(2):15, 2021.

[8] Jianbo Li, Lianhua Cui, Juntao Liu, Chunling Wu, and Mingxing Xie. Prediction of type 2 diabetes mellitus with machine learning techniques: A systematic review and meta-analysis. *Journal of Healthcare Engineering*, 2020:1–14, 2020.

[9] Syed Abdul Basit Andrabi and Inderjeet Singh. A comparative study of machine learning techniques for diabetes prediction. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 741–745. IEEE, 2022.

[10] UCI Machine Learning Repository. Pima indians diabetes database. `https://www.kaggle.com/uciml/pima-indians-diabetes-database`, 1990. Accessed: March 4, 2023.

[11] Emmanuel O Okechukwu, Chika Achonu, Arodiwe Innocent Obi-Okoye, Uche Madubueze, Basil C Ezeanolue, and Chinwe Onwujuba. A systematic review of machine learning techniques for the diagnosis of diabetes mellitus. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 13:1651–1662, 2020.