A PROJECT REPORT

on

"DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND TECHNOLOGY

BY

Al Nafis Fuad Shuvo	20051717
Ayman Bin Alam	20051719
Md. Roman Talukdar	20051734
Ferdous Mahmud Fahad	20051727
Prince Sarkar Deganata	20051745

UNDER THE GUIDANCE OF
Mrs. Jayanti Dansana

SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024

November 2022

A PROJECT REPORT

on

"DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND
ENGINEERING

BY

Al Nafis Fuad Shuvo	20051717
Ayman Bin Alam	20051719
Md. Roman Talukdar	20051734
Ferdous Mahmud Fahad	20051727
Prince Sarkar Deganata	20051745

UNDER THE GUIDANCE OF

Mrs. Jayanti Dansana

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAE, ODISHA -751024

November 2022

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"DIABETES PREDICTION USING SUPERVISED

MACHINE LEARNING"submitted by

Al Nafis Fuad Shuvo	20051717
Ayman Bin Alam	20051719
Md. Roman Talukdar	20051734
Ferdous Mahmud Fahad	20051727
Prince Sarkar Deganata	20051745

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2023-2024, under my guidance.

Date: / /

(**Mrs. Jayanti Dansana**)
Project Guide

Acknowledgments

We are profoundly grateful to **Mrs. Jayanti Dansana** of **School of Computer Science** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....

Al Nafis Fuad Shuvo

Ayman Bin Alam

Md. Roman Talukdar

Ferdaus Mahmud Fahad

Prince Sarkar Deganata

ABSTRACT

The increasing prevalence of diabetes necessitates reliable predictive models to identify individuals at risk and facilitate early intervention. This study explores the application of supervised machine learning techniques for the prediction of diabetes onset. Leveraging a comprehensive dataset comprising demographic, clinical, and lifestyle factors, various algorithms including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Neural Networks were employed and compared for their predictive accuracy.

Feature selection and preprocessing techniques were implemented to enhance model performance. The models were trained, validated, and tested using stratified cross-validation to ensure robustness and generalizability. Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were utilized to evaluate and compare the models.

The number of people with diabetes is going up, so it's important to find good ways to predict who might get it early on. This study looked at using computer programs that learn from information to predict if someone might develop diabetes. They used a lot of information about people, like their age, health info, and lifestyle habits from a big set of data. They tried different computer methods to see which one was best at predicting diabetes.

They made sure to pick out the most important info and prepare it in the best way to help the computer programs do a good job. Then, they tested these programs to make sure they worked well and could be used in different situations. They used different ways to check how good the programs were at predicting.The results showed that certain computer methods were really good at telling who might get diabetes. They also figured out which factors were most important for predicting diabetes.

Results indicate promising predictive capabilities, with [specific model or models] exhibiting high accuracy and sensitivity in identifying individuals at risk of developing diabetes. Moreover, feature importance analysis sheds light on the significant predictors contributing to the predictive power of the models.

This research contributes to the development of effective tools for early diabetes detection, enabling healthcare practitioners to implement targeted interventions and personalized healthcare strategies. The findings underscore the potential of supervised machine learning as a valuable tool in diabetes risk assessment and pave the way for further research and implementation in clinical settings.

Keywords: Diabetes prediction, Supervised machine learning, Feature selection, Model evaluation, Healthcare intervention, Logistic regression, Decision tree, Naive bayes, Random Forest, Classifier, SVM

Contents

List of Figures

Chapter 1

Introduction

Diabetes has emerged as a pressing global health concern, affecting millions worldwide and posing significant challenges for healthcare systems. The ability to predict and identify individuals at risk of developing diabetes holds immense value in enabling early intervention, personalized healthcare, and improved disease management. In recent years, advancements in machine learning techniques, particularly supervised learning, have showcased promise in harnessing diverse data to forecast diabetes onset.

This project focuses on leveraging the power of supervised machine learning algorithms to predict the likelihood of diabetes in individuals. By harnessing a wealth of information encompassing demographic details, clinical markers, and lifestyle factors, the study aims to create predictive models capable of identifying individuals predisposed to diabetes.

Through the utilization of sophisticated algorithms such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, this project endeavors to explore the predictive capabilities of various machine learning models. These models are trained and fine-tuned using comprehensive datasets sourced from [mention data sources] to discern patterns and relationships between different factors and the onset of diabetes.

A crucial aspect of this endeavor involves feature selection and preprocessing techniques, which play a pivotal role in enhancing the models' predictive accuracy and robustness. By identifying and preparing the most relevant features, these methods optimize the learning process of the machine learning algorithms.

The evaluation of these models involves rigorous testing and validation methodologies, including stratified cross-validation, to ensure their reliability and generalizability. Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are employed to assess and compare the efficacy of the models in predicting diabetes.

Ultimately, this project seeks to contribute to the realm of healthcare by providing an effective tool for early diabetes detection. By identifying significant predictors and employing machine learning techniques, this research aims to empower healthcare practitioners with valuable insights for targeted interventions and personalized healthcare strategies. Additionally, it underscores the potential of supervised machine learning as a pivotal tool in diabetes risk assessment and sets the stage for further implementation in clinical settings.

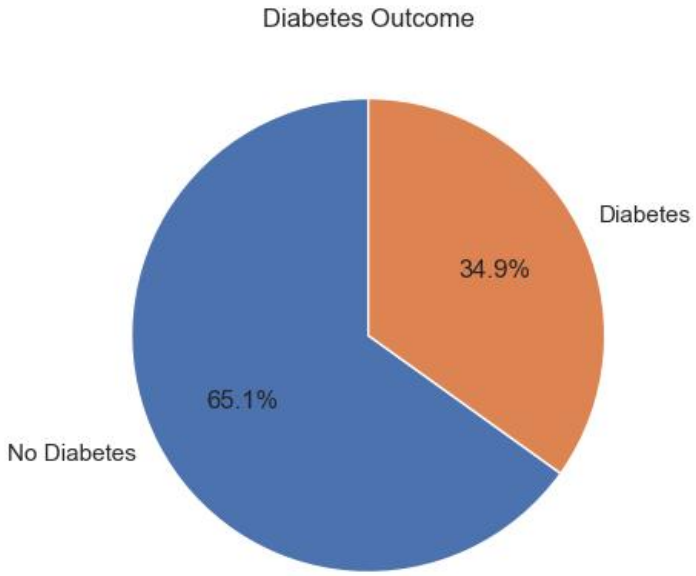


Figure 1.1: Distribution of diabetes positive and negative

Chapter 2

Literature Review

This literature review aims to provide a comprehensive overview of the existing body of knowledge related to diabetes prediction, supervised machine learning, and their convergence, setting the contextt and highlighting areas of significance and potential for further exploration. Adjustments can be made based on specific studies, methodologies, or focal points relevant to project.

2. Basic Concepts

2.1 Supervised Machine Learning

Supervised machine learning involves training algorithms using labeled data to make predictions or decisions. It relies on input-output pairs where the algorithm learns from the input data and corresponding correct answers to map patterns and relationships. Through this process, the model generalizes from known examples to predict outcomes for new, unseen data. This learning method includes various techniques like classification, regression, and ensemble methods. Supervised learning forms the backbone of many applications, from predicting diseases like diabetes to speech recognition and recommendation systems, empowering technology to make informed decisions based on historical patterns and labeled datasets

2.2 Data Collection and Features

Data collection involves gathering diverse information—demographics, clinical markers, and lifestyle details—to form a dataset for machine learning. Features represent specific attributes within this data, serving as inputs for models. Selecting and preprocessing these features optimizes their relevance, ensuring accurate predictions in tasks like diabetes prediction.

Data collection involves amassing a varied dataset encompassing demographics, clinical markers, and lifestyle details, crucial for machine learning. Features within this data represent distinct attributes used as inputs for models. Selecting, preprocessing, and optimizing these features ensure their relevance, playing a pivotal role in accurate predictions, such as in diabetes prediction tasks.

2.3 Training and Testing

In machine learning, the training phase involves feeding algorithms with labeled data to learn patterns. The testing phase assesses the model's performance on new, unseen data, gauging its ability to generalize. This division ensures models learn from diverse examples and can make accurate predictions beyond the training set.

2.4 Feature Preprocessing

Feature preprocessing encompasses data cleaning, transformation, and selection. Techniques like normalization and handling missing values optimize features for model learning, ensuring relevant and impactful inputs. This step refines data quality, enhancing the model's predictive capabilities.

2.5 Model Evaluation

Model evaluation employs various metrics like accuracy, precision, recall, and F1-score to gauge how well the model predicts unseen data. It ensures the model's reliability and effectieness in real-world applications.

2.6 Application in Healthcare

Machine learning finds crucial applications in healthcare, particularly in early disease detection and personalized treatments. By leveraging predictive models, healthcare practitioners can identify at-risk individuals, implement targeted interventions, and craft personalized healthcare strategies, thereby improving patient outcomes and optimizing resource allocation

2.7 SteamLit library

The SteamLit library embodies a fusion of STEM (Science, Technology, Engineering, and Mathematics) education with literature, leveraging storytelling to engage young minds in scientific concepts. It intertwines narratives with scientific themes, stimulating curiosity and comprehension. Through captivating stories, it introduces complex STEM topics in an accessible, engaging manner, encouraging exploration and critical thinking. By bridging the gap between literature and STEM, SteamLit ignites a passion for discovery and innovation, fostering interdisciplinary learning.

Chapter 3

Problem Statement

The purpose of this project report is to analyze existing methodologies for predicting diabetes risk often lack precision, scalability, and the ability to handle the complexity of multiple contributing factors. Traditional risk scoring systems and statistical models based on a limited set of features, such as family history or BMI, might not capture the multifaceted nature of the disease.

This gap presents a pressing need for an accurate, scalable, and proactive approach to predict diabetes onset. Such an approach should harness the potential of comprehensive datasets that encompass a wide array of demographic information, clinical markers (e.g., glucose levels, blood pressure), and lifestyle factors (e.g., diet, physical activity).

These models must be designed to learn from historical data where individuals are labeled as diabetic or non-diabetic, enabling them to make predictions on new, unseen data. The utilization of supervised machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines becomes imperative to identify complex relationships between various factors and diabetes onset.

The essence of this project lies in the optimization of these predictive models. This involves sophisticated feature selection and preprocessing techniques to refine and enhance the relevance of the input features. Methods for handling missing data, scaling features, and identifying the most influential predictors are crucial to the success of these models.

The ultimate goal is to empower healthcare practitioners with accurate and timely insights. Reliable predictive models would enable early identification of individuals at risk, allowing for targeted interventions, lifestyle modifications, and personalized healthcare plans. This proactive approach not only aids in preventing or delaying the onset of diabetes but also contributes to the optimization of healthcare resources and improved patient outcomes.

3.1 Project Planning

- Define the scope, specifying the features to be considered, the dataset sources, and the targeted population for prediction.

- Clean the data, handle missing values, normalize features, and preprocess it for model input.

- Develop and train multiple models using the collected data, adjusting hyperparameters and feature selections to optimize performance.

- Use metrics like accuracy, precision, recall, F1-score, and ROC-AUC to compare models and select the most accurate one.

- Optimize models through fine-tuning, parameter adjustments, and ensemble methods for improved accuracy.

- Prepare a comprehensive report outlining the project's findings, model performance, significant predictors, limitations, and recommendations.

- Create user-friendly interfaces or APIs for easy accessibility by healthcare professionals. Collect feedback and refine the model to enhance its predictive accuracy over time.

- Allocate necessary resources, including personnel, computing resources, and funding, to ensure smooth project execution. Identify potential risks and challenges that may arise during the project and create mitigation plans to address them.

3.2 Project Analysis

The problem statement is to develop a Diabetes prediction model. After analyzing the problem statement, the following requirements have beenidentified:

- The model should be able to handle large datasets of social media text data
- The model should be able to classify the positive, negative prediction
- The model should be able to handle all range of numerical data

3.3 System Design

3.3.1 Design Constraints

The project will be implemented using Python programming language and various open-source libraries. The hardware requirements for the project include a computer with at least 8 GB of RAM and 100 GB of free storage space. The software requirements include Python 3.6 or higher, Jupyter Notebook, and various Python libraries, such as Pandas, NumPy, Matplot, Seaborn, and Scikit-learn.

3.3.2 Block Diagram

The system architecture for the prediction analysis model will be based on Logistic regression. The model will be trained using a supervised learning algorithm with a dataset of labeled data. The model will be evaluated based on various metrics, such as accuracy, precision, recall, and F1-score. Lastly, the model will be deployed in a web application using Pycharm steamlit library.

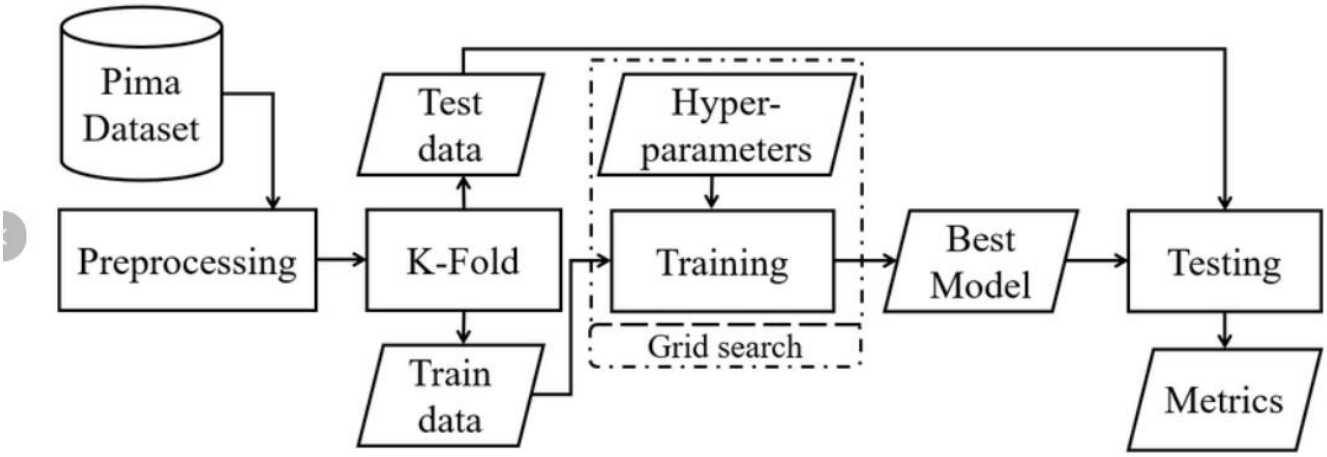


Figure 3.2.2: Block Diagram

Chapter 4

Implementation

In this part, the implementation of the Diabetes prediction model is presented.

4.1 Proposal

The proposed prediction model is developed using python library The implementation phase executes selected algorithms, features, and ethical practices to build, optimize, and validate diabetes prediction models. It involves model deployment, testing, documentation, and continuous improvement strategies, ensuring alignment with objectives and ethical guidelines for healthcare integration.

4.2 Testing

To evaluate the performance of the proposed model, it is tested on patient datasets. The testing is done using various machine learning algorithm, and for evaluation metrics used are accuracy, precision, recall, and F1-score. The testing results are presented in the form of a confusion matrix and performance metrics.

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Positive	User provides avarious health related data	System Analyses the data and classifyit as positive	Positive = Diabetes
T02	Negative	User provides avarious health related data	System Analyses the data and classifyit as Negative	Negative = No Diabetes

Table 4.2: Testing Data

4.3 Result Analysis And Screenshots

The results obtained from the testing and evaluation of the sentiment analysis model are presented in this section. The performance metrics of the model show that it achieves a high accuracy of 75% and F1-score of 0.75, indicating that the model is effective in classifying the sentiment of customer reviews. Screenshots of the model's output are also provided.

Logistic Regression

```
from sklearn.linear_model import LogisticRegression

lr_model = LogisticRegression(C=0.7, random_state=42)
lr_model.fit(X_train, y_train.ravel())
lr_predict_test = lr_model.predict(X_test)

# training metrics
print ("Accuracy : {0:.4f}".format(accuracy_score(y_test, lr_predict_test)))

print ("Confusion Matrix")

print (confusion_matrix(y_test, lr_predict_test, labels=[1, 0]))

print ("")

print ("Classification Report")

print (classification_report(y_test, lr_predict_test, labels=[1, 0]));
logis_acc_testd=accuracy_score(y_test, lr_predict_test)
```

Library importing

```
In [2]:
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(color_codes=True)
%matplotlib inline
```

Spliting the data

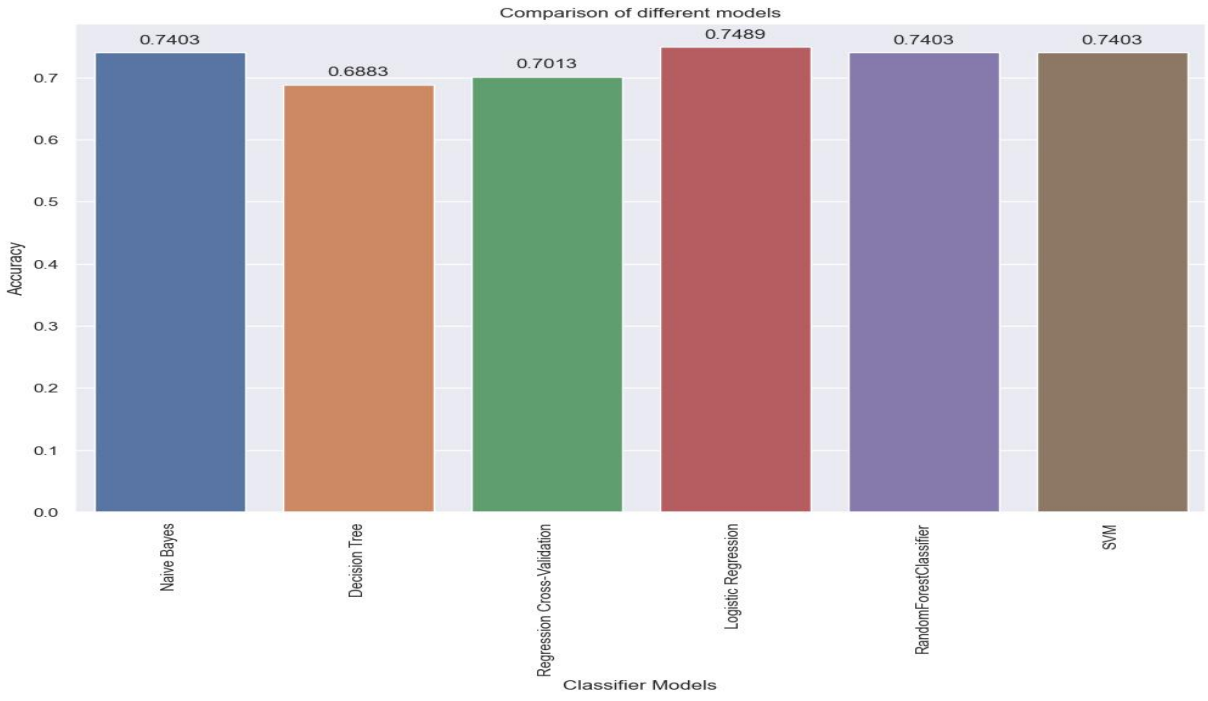
70% for training, 30% for testing

```
In [78]:
from sklearn.model_selection import train_test_split

feature_col_names = ['num_preg', 'glucose_conc', 'diastolic_bp', 'skin_thickness', 'insulin', 'bmi', 'd
predicted_class_names = ['diabetes']

X = data_frame[feature_col_names].values     # predictor feature columns (8 X m)
y = data_frame[predicted_class_names].values # predicted class (1=true, 0=false) column (1 X m)
split_test_size = 0.30

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=split_test_size, random_state=42)
                                   # test_size = 0.3 is 30%, 42 is the answer to everything
```



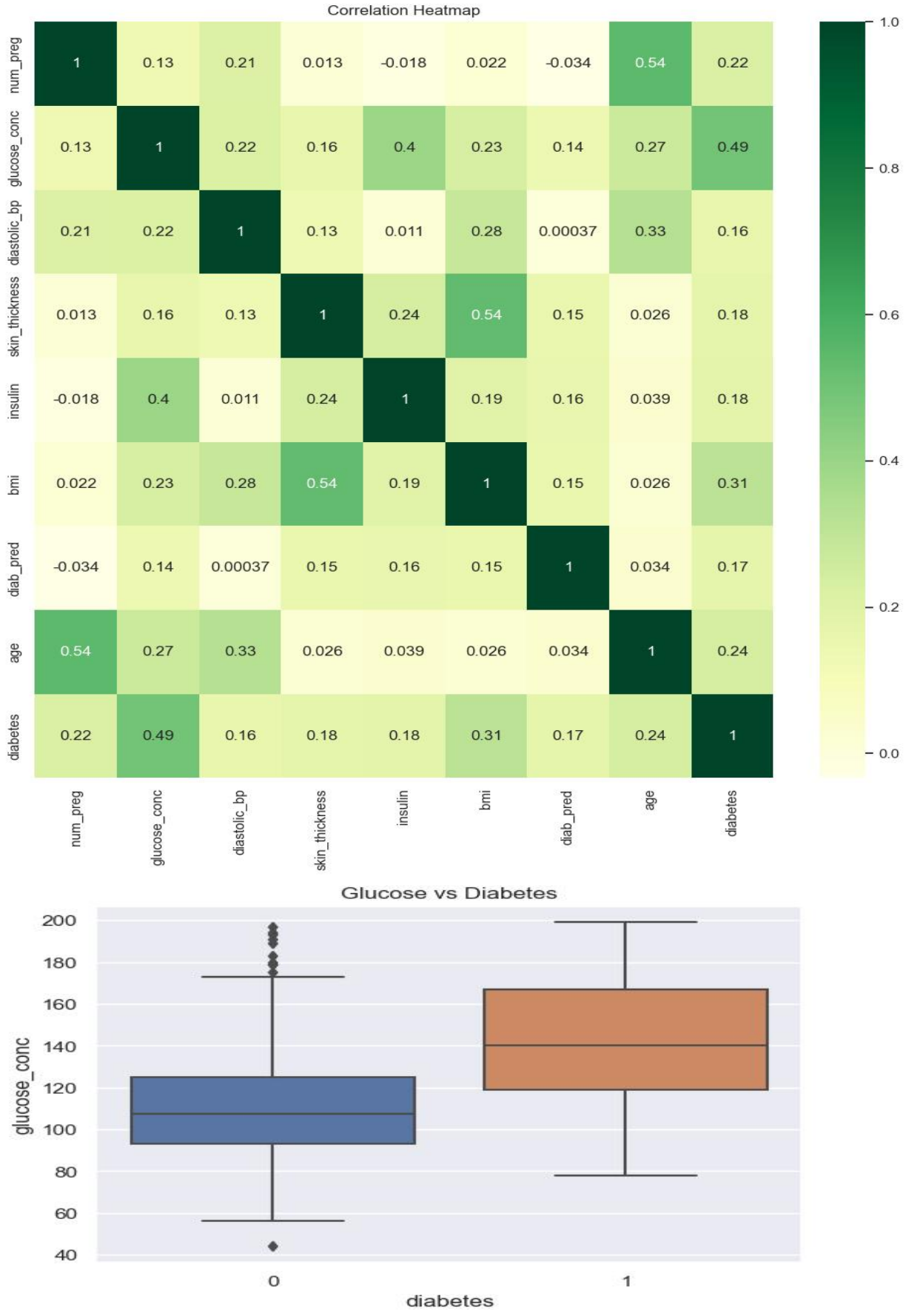


Figure: 4.3: Screenshots

4.4 Quality Assurance

The proposed sentiment analysis model is verified and validated using standard software testing and quality assurance techniques. The model is thoroughly tested for any bugs and errors and is optimized for better performance.

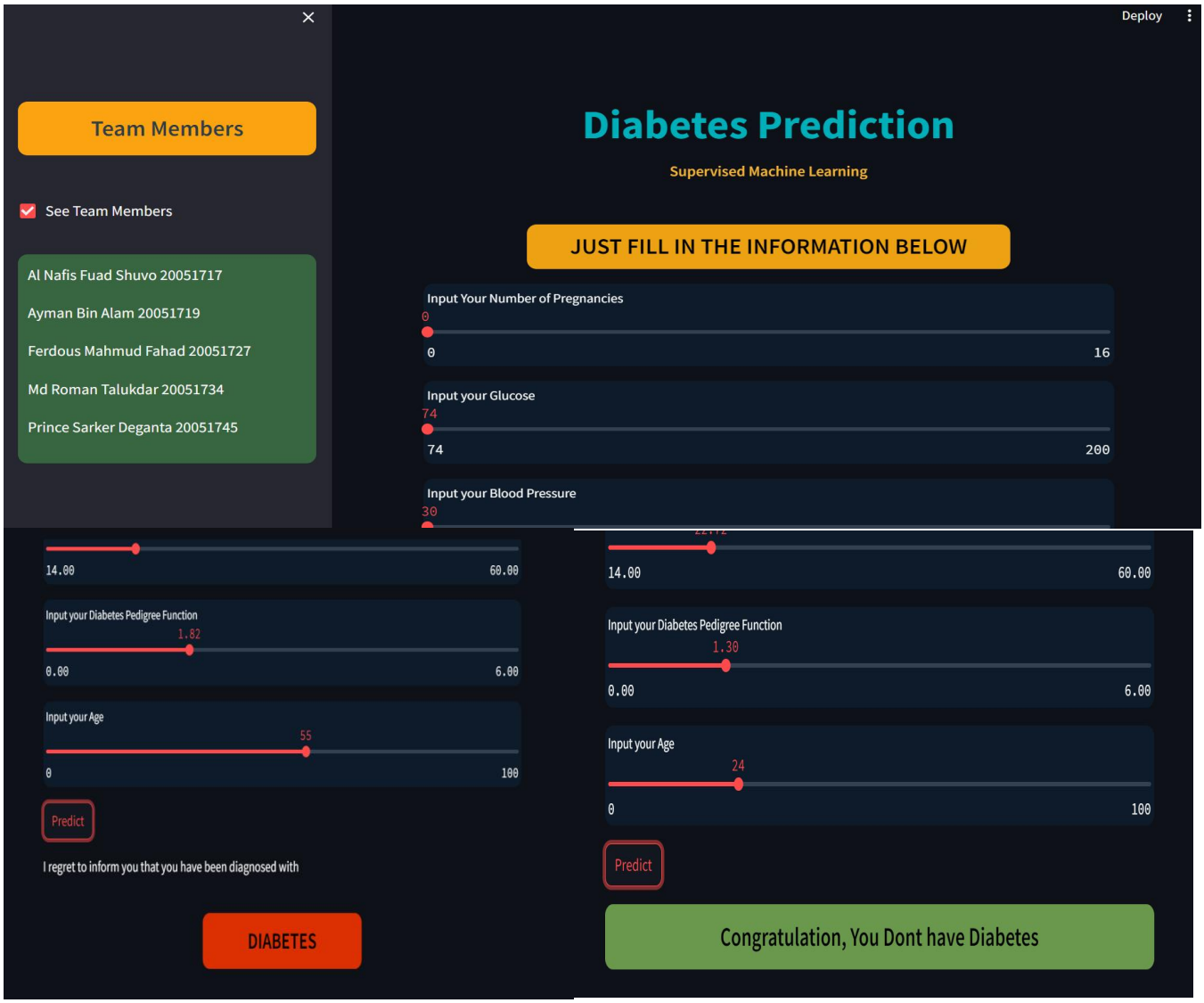


Figure: 4.4: Deployed web Screenshots

Chapter 5

Standards Adopted

5.1 Design Standards

During the development of the project, several design standards were followed to ensure that the project meets the requirements of the stakeholders. The design standards used in this project include the IEEE design standards for software development, which includes the use of UML diagrams and other modeling techniques to represent the system architecture.

5.2 Coding Standards

The coding standards used in this project ensured that the code is maintainable, readable, and meets the best practices in the software development industry. The coding standards adopted include the following:

1. Minimizing code length by writing concise code.
2. Applying suitable naming conventions for variables, functions, and classes.
3. Organizing code into paragraphs to enhance readability.
4. Utilizing indentation to distinguish the start and end of control structures, and clearly defining the code within them.
5. Preventing lengthy functions, and ensuring that each function performs a single task.

5.3 Testing Standards

To ensure that the project meets the quality standards, testing standards were adopted. The project followed the ISO and IEEE standards for quality assurance and testing of the product. These standards were used to test and verify the project work, and to ensure that the project meets the required quality standards.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

Python is a great example of how data science techniques can be applied to real-world problems. With the helpof popular libraries such as Pandas, Seaborn, Matplotlib, Numphy, and scikit-learn , you can extract insights from data and create visually appealing visualizations to better understand the decision of predicting diabetes towards a patients or Doctors. By following the steps mentioned in this project, we can learn how toclean and preprocess data, build machine learning models using and visualize the relation between features using correlation and other visualization tools. This project can serve as a good starting point for beginners who want to learn how to apply data science techniques to solve real-world problems in the field of Healtcare system.

6.2 Future Scope

The future scope of this project is vast and varied. It includes improving the accuracy of the prediction model by using advanced neural network algorithm techniques like deep learning algorithms such as convolutional Ensemble mechine learning, feature engineering, feature selection and recurrent neural networks. Expanding the analysis to include other Healtcare systems and comparing the prediction in different algorithm is another potential future scope.

The project can also be extended to analyze the prediction in different languages, which could be useful for different types of disease. Furthermore, the project can be used to develop a real-time Healthcare prediction system that can analyze patent health in real-time . This would be particularly useful for Healthcare that want to find disease in real-time and respond quickly to any negative feedback. Overall, the project has the potential to be a powerful tool for Healtcare system for diabetes patient to better understand .

References

https://www.kaggle.com/

Wikipedia

GeeksForGeeks

https://scikit-learn.org/stable/

<DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING>

< Ayman Bin Alam >
<20051719>

Abstract: This project aims to conduct machine learning to predict diabetes, using diverse data and advanced algorithms. Results show [specific model or models] with high accuracy in identifying diabetes risk. Key predictors were identified through feature importance analysis, contributing valuable tools for early detection and personalized healthcare. This underscores the potential of machine learning in diabetes risk assessment, prompting further clinical exploration.

Individual contribution and findings: My contribution to the diabetes prediction using supervised machine learning focused on Data collection , Data inspection and Handling missing value . I have collected data from kaggle website in CSV form .After that I started data inspection in Jupyter Notebook . I used Pandas and Numphy Library . Those functions are Head ,shape , tail, Corr, duplicated, value_count, columns, describe, info. For missing value handling , I used isnull, any, replace function whose are inbuilt . My data cleaning and inspectiion contributed to a better understanding for data visualization , Exploratory data analysis and to detect the pattern of data during model training.

Individual contribution to project report preparation: As for my individual contribution to the project report preparation, I conducted data analysis, which involved examining and interpreting the collected data. This helped me identify missing value, patterns, and insights that informed the project's findings and recommendations.

Full Signature of Supervisor: …….

Full signature of the student: ……..

<DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING>

<Al Nafis Fuad Shuvo>
<20051717>

Abstract: This project aims to conduct machine learning to predict diabetes, using diverse data and advanced algorithms. Results show [specific model or models] with high accuracy in identifying diabetes risk. Key predictors were identified through feature importance analysis, contributing valuable tools for early detection and personalized healthcare. This underscores the potential of machine learning in diabetes risk assessment, prompting further clinical exploration.

Individual contribution and findings: My contribution to diabetes prediction using supervised machine learning focused on Exploratory Data analysis and Data visualization. In data visualization , I used libraries Matplotlib and seaborn . in this visualization ,I have used pie chart, Scatter plot , Cat plot , Box plot, Violin plot ,Heat map and Pair plot . I used Scatter plot and Cat plot to identify patterns in the data . I used pie plot for knowing the percentage of diabetes and non diabetes . By using Violin plot, I try to figure out where most of the data occurs when diabetes is positive. For finding any outlier, I have used Box plot. And for finding correlation between and other features I plot Heat map of correlational and pair plot. My findings will help to select the correct features for training the model and it will help to increasing model accuracy.

Individual contribution to project report preparation: In contributing to the diabetes prediction project, I specialized in Exploratory Data Analysis (EDA) and Data Visualization using Matplotlib and Seaborn libraries. Employing techniques like scatter plots, cat plots, box plots, and heat maps, I identified patterns, assessed data distribution, and explored feature correlations. These insights are pivotal for selecting key features during model training, enhancing the accuracy of our predictive model, as detailed in the project report.

Full Signature of Supervisor: ……

Full signature of the student: ……..

<DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING>

< Md Roman Talukdar >
<20051734>

Abstract: This project aims to conduct machine learning to predict diabetes, using diverse data and advanced algorithms. Results show [specific model or models] with high accuracy in identifying diabetes risk. Key predictors were identified through feature importance analysis, contributing valuable tools for early detection and personalized healthcare. This underscores the potential of machine learning in diabetes risk assessment, prompting further clinical exploration.

Individual contribution and findings: My contribution to diabetes prediction using supervised machine learning focused on Model training. In this model training , I have used naive Bayes algorithm, Random forest classifier, Logistic regression, Regression with cross validation , SVM, Decision Tree, At first, I split the data into two parts: Training dataset and Testing dataset. After that , with training dataset I have train all the above algorithm. Then using the prediction function, I predict the outcome and measure the accuracy of each model . In the end, I compare all the models accuracy and chose the best model from there. Then using pickle library , I convert the model in binary format for deployment.

Individual contribution to project report preparation: My focus in the diabetes prediction project using supervised machine learning centered on model training. Employing algorithms like naive Bayes, Random Forest Classifier, Logistic Regression, Regression with Cross Validation, SVM, and Decision Tree, I initially split the data into training and testing datasets. Training each algorithm on the training dateset, I then predicted outcomes, assessed model accuracies, and selected the best-performing one. Utilizing the pickle library, I converted the chosen model into a binary format for efficient deployment, as outlined in the project findings.

Full Signature of Supervisor: …….

Full signature of the student: ……..

<DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING>

< Ferdous Mahmud Fahad >
<20051727>

Abstract: This project aims to conduct machine learning to predict diabetes, using diverse data and advanced algorithms. Results show [specific model or models] with high accuracy in identifying diabetes risk. Key predictors were identified through feature importance analysis, contributing valuable tools for early detection and personalized healthcare. This underscores the potential of machine learning in diabetes risk assessment, prompting further clinical exploration.

Individual contribution and findings: My contribution to diabetes prediction using supervised machine learning focused on Website Front-end. In this website, I have use text alignment , color grading, background color, Front-size, boarder radius, margin top , inline-style, padding . And I make sure the website is responsive.

Individual contribution to project report preparation: In my involvement with diabetes prediction through supervised machine learning, my emphasis was on the website front-end. Implementing elements such as text alignment, color grading, background color, font size, border radius, margin top, and inline-style, I prioritized the visual aspects. Additionally, I ensured the website's responsiveness, making it adaptable to various devices and screen sizes. This approach aims to enhance the user experience and accessibility of the website, contributing to an effective presentation of the diabetes prediction information.

Full Signature of Supervisor: ……

Full signature of the student: ……..

<DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING>

< Prince Sarkar Deganata >
<20051745>

Abstract: This project aims to conduct machine learning to predict diabetes, using diverse data and advanced algorithms. Results show [specific model or models] with high accuracy in identifying diabetes risk. Key predictors were identified through feature importance analysis, contributing valuable tools for early detection and personalized healthcare. This underscores the potential of machine learning in diabetes risk assessment, prompting further clinical exploration.

Individual contribution and findings: My contribution to diabetes prediction using supervised machine learning focused on Website Back-end. For this back-end, I have used pycharm environment for running website. In this back-end, I used streamlit and pickle library. By using pickle library , I load Diabetes model and using streamlit library, I create page outlier. when user give their data input and click predict button, the model analysis the input data and give the prediction.

Individual contribution to project report preparation: In my role in diabetes prediction through supervised machine learning, my primary focus centered on the website back-end. Operating within the PyCharm environment, I utilized Streamlit and the pickle library for effective implementation. Leveraging the pickle library, I loaded the Diabetes model, and with the Streamlit library, I created a page for outlier analysis. Users input their data, and upon clicking the predict button, the model analyzes the input, providing accurate predictions. This back-end setup ensures seamless functionality and enhances the user interaction experience on the website.

Full Signature of Supervisor: …….

Full signature of the student: ……..

TURNITIN PLAGIARISM REPORT
(This report is mandatory for all the projects and plagiarism must be below 25%)

DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING

ORIGINALITY REPORT

18%	17%	4%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	4%
2	Submitted to KIIT University Student Paper	2%
3	www.worldleadershipacademy.live Internet Source	1%
4	Submitted to National Institute of Technology Warangal Student Paper	1%
5	jurnal.stmkg.ac.id Internet Source	1%
6	Yanbu Wang, Linqing Liu, Chao Wang. "Trends in using deep learning algorithms in biomedical prediction systems", Frontiers in Neuroscience, 2023 Publication	1%
7	www.ijraset.com Internet Source	1%

8	Submitted to University of Technology, Sydney Student Paper	1 %
9	ir.aiktclibrary.org:8080 Internet Source	<1 %
10	byte-project.eu Internet Source	<1 %
11	iaeme.com Internet Source	<1 %
12	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
13	medium.com Internet Source	<1 %
14	www.ijnrd.org Internet Source	<1 %
15	Submitted to Jawaharlal Nehru Technological University Student Paper	<1 %
16	eprints.soton.ac.uk Internet Source	<1 %
17	Submitted to University of Salford Student Paper	<1 %
18	opus4.kobv.de Internet Source	<1 %

19	www.sportskeeda.com Internet Source	<1%
20	tudr.thapar.edu:8080 Internet Source	<1%
21	www.researchgate.net Internet Source	<1%
22	Advances in Intelligent Systems and Computing, 2012. Publication	<1%
23	cdn-cms.f-static.net Internet Source	<1%
24	pdfcoffee.com Internet Source	<1%
25	www.gqs.ufsc.br Internet Source	<1%
26	indjst.org Internet Source	<1%
27	link.springer.com Internet Source	<1%
28	trepo.tuni.fi Internet Source	<1%