Smart Health Analytics: A Pipeline for Diabetes Prediction

Creative Problem Solving Lab Report

Submitted for the partial fulfilment of the degree of

Bachelor of Technology

In

Artificial Intelligence & Data Science

Submitted By

Diya Badkul	0901AD221032
Neeraj Prajapati	0901AD221052
Shantanu Shukla	0901AD221065

UNDER THE SUPERVISION AND GUIDANCE OF

Dr. Abhishek Bhatt
Assistant Professor

Dr. Vibha Tiwari
Assistant Professor

Centre for Artificial Intelligence

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR
Deemed University
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ Grade
Gola Ka Mandir, Gwalior (M.P.)- 474005, INDIA
Ph.:+91-751-2409300, E-mail: vicechancellor@mitsgwalior.in, Website: www.mitsgwalior.in

Session: Aug-Dec 2025

DECLARATION BY THE CANDIDATE

We hereby declare that the work entitled **"Smart Health Analytics: A Pipeline for Diabetes Prediction"** is our work, conducted under the supervision of **Dr. Abhishek Bhatt, Assistant Professor** and **Dr. Vibha Tiwari, Assistant Professor,** during the session Aug-Dec 2025. The report submitted by us is a record of bona fide work carried out by us.

We further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Diya Badkul **0901AD221032**

Neeraj Prajapati **0901AD221052**

Shantanu Shukla **0901AD221065**

Date: November 12, 2025

Place: Gwalior

This is to certify that the above statement made by the candidates is correct to the best of my knowledge and belief.

Guided By:

Dr. Abhishek Bhatt
Assistant Professor
Centre for Artificial Intelligence
MITS, Gwalior

Dr. Vibha Tiwari
Assistant Professor
Centre for Artificial Intelligence
MITS, Gwalior

PLAGIARISM CHECK CERTIFICATE

This is to certify that we, the students of B.Tech. in **Artificial Intelligence and Data Science** have checked our complete report entitled **"Smart Health Analytics: A Pipeline for Diabetes Prediction"** for similarity/plagiarism using the "Turnitin" software available in the institute.

This is to certify that the similarity in our report is found to be .15%.which is within the specified limit (20%).

The full plagiarism report along with the summary is enclosed.

Diya Badkul 0901AD221032

Neeraj Prajapati 0901AD221052

Shantanu Shukla 0901AD221065

ACKNOWLEDGEMENT

We are deeply thankful to **Madhav Institute of Technology and Science** for giving us the opportunity and an excellent platform for undertaking this project. The institution's profound environment played a significant role in seamless progression of this project for which we would like to express our heartfelt gratitude to our Chancellor, **Mr. Jyotiraditya M. Scindia**, our Vice-Chancellor, **Dr. R.K. Pandit** and Dean Academics, **Dr. Manjaree Pandit**.

We would sincerely like to thank our department, **Centre for Artificial Intelligence,** the Head of Department, **Dr. Rajni Ranjan Singh Makwana** and our Departmental Project Coordinator, **Dr. Tej Singh** for the unwavering support throughout the progress of this project making the formalities and engagement procedure easy and effortless.

We are highly thankful to our esteemed mentors, **Dr. Abhishek Bhatt** and **Dr. Vibha Tiwari,** for their invaluable guidance and constant encouragement for this project. Their insightful feedbacks and profound expertise in the domain was a great source of inspiration, and that has greatly contributed to the successful completion of this project.

We are also grateful to my colleagues for their constructive and collaborative discussions in providing brighter ideas and insightful suggestions which were instrumental in filtering the best ideas and methodologies to be employed in the project.

We extend our sincere appreciation to our families and friends for being patient and extremely encouraging throughout this journey and keeping us motivated.

Our project would not have been possible without the collective contributions and support of all aforementioned institution and individuals. Thank you for being the apart of this enriching journey.

Diya Badkul 0901AD221032

Neeraj Prajapati 0901AD221052

Shantanu Shukla 0901AD221065

Table of Contents

CHAPTER 1: INTRODUCTION

1.1 Background

Diabetes is a chronic disease that has reached alarming levels worldwide. Early detection is crucial in managing and mitigating the impact of diabetes, which affects millions of people globally. In the healthcare domain, data analytics plays a pivotal role in disease prediction, and machine learning algorithms offer opportunities for early diagnosis and intervention. The effective use of health-related data can lead to significant improvements in the prediction of diabetes, as well as the implementation of preventive measures.

1.2 Objective

The objective of this project is to design a data pipeline that processes raw health indicator data, cleanses it, visualizes the key trends, and implements a predictive model that can classify individuals based on their likelihood of developing diabetes. The project aims to demonstrate the power of data preprocessing and analysis in predicting health outcomes.

1.3 Scope

This project focuses on the implementation of a data pipeline for the **Diabetes Prediction** dataset. The data pipeline includes essential steps such as:

- **Data loading** and **cleaning**
- **Exploratory Data Analysis (EDA)**
- **Feature engineering and scaling**
- **Outlier detection**
- **Featureselection**

  By the end of the project, a set of preprocessed data will be available for further predictive modeling.

CHAPTER 2: LITERATURE REVIEW

2.1 Data Pipelines in Predictive Analytics

Data pipelines are crucial for transforming raw data into actionable insights. In predictive analytics, a robust data pipeline ensures the accuracy and reliability of the resulting models by addressing issues such as missing values, outliers, and feature engineering. In the healthcare domain, data pipelines have been used extensively to predict disease outcomes, as the integration of diverse datasets often requires thorough cleaning, transformation, and aggregation.

2.2 Diabetes Prediction Studies

Several studies have investigated the relationship between health indicators such as BMI, cholesterol levels, physical activity, and diabetes. Data preprocessing techniques such as **normalization**, **standardization**, and **feature selection** have proven vital in improving the performance of prediction models. Machine learning algorithms, including **logistic regression**, **random forests**, and **support vector machines (SVMs)**, are frequently applied to predict diabetes risk based on these features.

Relevant Case Studies and Tools

- **ETL Tools**: Widely used ETL tools such as Python, SQL, and cloud-based platforms (e.g., Apache Spark, AWS) have been pivotal in public health data processing. Python libraries like Pandas and SQL-based solutions help in handling and transforming large health datasets efficiently.

Case Study Example: A notable example involves using Apache Spark for large-scale health data processing, where Spark's distributed computing capabilities allowed for rapid processing, improving data accessibility and speed.

CHAPTER 3: METHODOLOGY

3.1 Dataset Description

The dataset used in this project is a collection of health indicators from the Behavioral Risk Factor Surveillance System (BRFSS) of 2021. It consists of 127,897 records with 22 attributes, including demographic information, lifestyle behaviors, and chronic health conditions.

Features:

- **Diabetes_012**: The target variable indicating whether the individual has diabetes (0 = No, 1 = Pre-Diabetes, 2 = Diabetes).
- **HighBP**: Indicator of high blood pressure (1 = Yes, 0 = No).
- **BMI**: Body Mass Index, a critical indicator of obesity.
- **Smoker**: Whether the individual smokes (1 = Yes, 0 = No).
- **Age**: Age of the individual.

3.2 Data Pipelining Workflow

The data pipeline follows these sequential steps:

1. **Data Loading**: Load the raw dataset into memory for processing.
2. **Data Cleaning**: Remove duplicates and handle missing values.
3. **Exploratory Data Analysis (EDA)**: Visualize distributions and relationships in the dataset.
4. **Feature Engineering**: Transform categorical variables, scale numeric features, and detect outliers.
5. **Feature Selection**: Identify the most relevant features using correlation analysis.
6. **Data Export**: Save the cleaned and transformed dataset for further use.

3.3 Tools and Technologies

- **Programming Language**: Python
- **Libraries**: Pandas, Seaborn, Matplotlib, Scikit-learn
- **Development Environment**: Google Colab

Chapter 4: Implementation

4.1 Data Cleaning

The raw dataset was loaded using the Pandas library and initially inspected for missing values and duplicates. Duplicate records were identified and removed. Missing values were imputed with the median for numeric columns and the mode for categorical variables. Below is the code used for data cleaning:

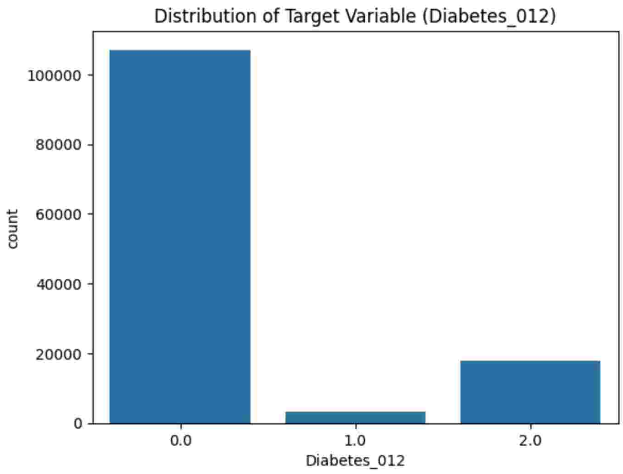
```
import pandas as pd

# Load the dataset

file_path = '/content/augmented_diabetes_012_health_indicators_BRFSS2021.csv'

df = pd.read_csv(file_path)

# Remove duplicate rows

df_no_duplicates = df.drop_duplicates()

# Check for missing values and impute

df.fillna(df.median(), inplace=True)  # Median imputation for numeric columns
```

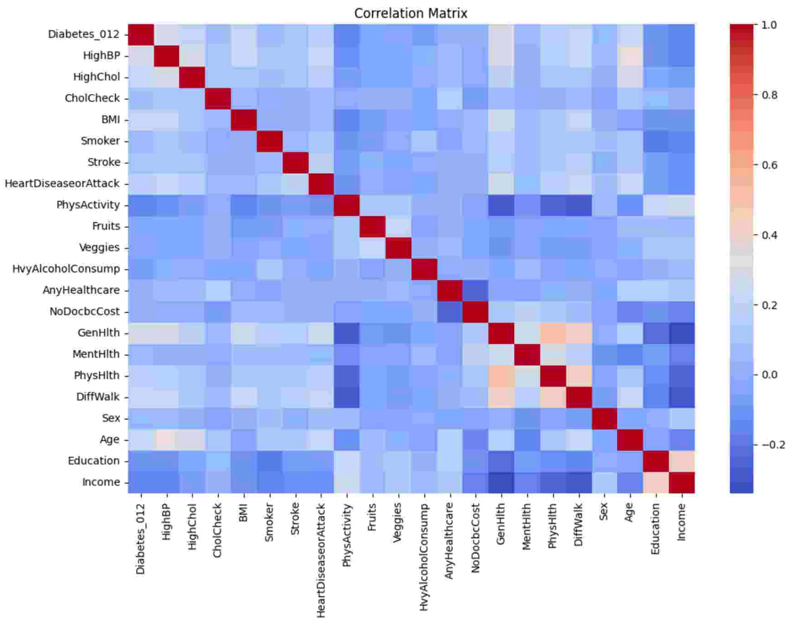
4.2 Exploratory Data Analysis

The Exploratory Data Analysis (EDA) focused on visualizing the relationships between diabetes status and key health factors. Key visualizations include count plots, box plots, and correlation heatmaps to reveal patterns in the data.

```
import seaborn as sns

import matplotlib.pyplot as plt

# Countplot for Diabetes status distribution

sns.countplot(x='Diabetes_012', data=df)

plt.title('Distribution of Diabetes Status')

plt.show()
```

```
# Correlation heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(df.corr(), annot=False, cmap="coolwarm")
plt.title('Correlation Matrix of Features')
plt.show()
```





4.3 Feature Engineering

- **Categorical Encoding**: The Sex column, which contains categorical values, was encoded into numeric form using **Label Encoding**.

from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()

df['Sex'] = label_encoder.fit_transform(df['Sex'])

Age Grouping: The Age column was divided into meaningful age groups for further analysis.

age_bins = [0, 18, 35, 50, 65, 100]

age_labels = ['Child', 'Young Adult', 'Adult', 'Middle Age', 'Senior']

df['Age_Group'] = pd.cut(df['Age'], bins=age_bins, labels=age_labels, right=False)

- **Feature Scaling**: Features like BMI and Age were normalized using **StandardScaler**.

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

df[['BMI', 'Age']] = scaler.fit_transform(df[['BMI', 'Age']])

Outlier Detection: The **Interquartile Range (IQR)** method was used to handle outliers by clipping values outside the range.

Q1 = df[['BMI', 'Age']].quantile(0.25)

Q3 = df[['BMI', 'Age']].quantile(0.75)

IQR = Q3 - Q1

*df[['BMI', 'Age']] = df[['BMI', 'Age']].clip(lower=Q1 - 1.5 * IQR, upper=Q3 + 1.5 * IQR)*

4.4 Feature Selection

The most correlated features to the target variable Diabetes_012 were selected for further analysis. These features were determined based on their correlation with diabetes status:

correlation_matrix = df.corr()

correlated_features = correlation_matrix['Diabetes_012'].abs().sort_values(ascending=False)

selected_features = correlated_features[correlated_features > 0.1].index.tolist()

selected_features.remove('Diabetes_012') # Exclude target column

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Processed Dataset

After data cleaning and preprocessing, the dataset was reduced to 123,226 rows and retained 11 significant features deemed highly relevant for predicting diabetes. The processed dataset was free from missing values, outliers, and duplicate records, and all features were normalized to ensure compatibility with predictive modeling techniques.

Key Features:

1. **GenHlth (General Health)**: Indicates the self-reported general health rating of individuals.
2. **HighBP (High Blood Pressure)**: Captures whether an individual has high blood pressure.
3. **BMI (Body Mass Index)**: A crucial indicator of obesity and a significant risk factor for diabetes.
4. **DiffWalk (Difficulty Walking)**: Represents mobility issues, often associated with diabetes complications.
5. **PhysActivity (Physical Activity)**: Indicates the level of physical activity, with inactivity linked to higher diabetes risk.
6. **Income**: Socioeconomic status, which influences healthcare access and lifestyle factors.
7. **Age**: Reflects the demographic influence on diabetes risk, with older individuals more prone to the condition.

5.2 Insights from Exploratory Data Analysis

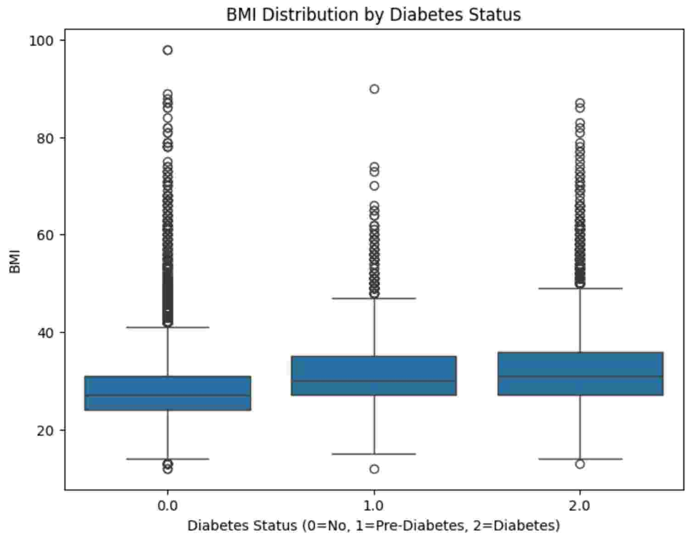
The **Exploratory Data Analysis (EDA)** revealed crucial patterns and relationships in the dataset:

1. **BMI and Diabetes**:
    - Individuals with higher BMI were more likely to develop diabetes.
    - Obesity, indicated by BMI, had a strong positive correlation with the Diabetes_012 target variable.

    sns.boxplot(x='Diabetes_012', y='BMI', data=df)

    plt.title('BMI Distribution by Diabetes Status')

    plt.show()



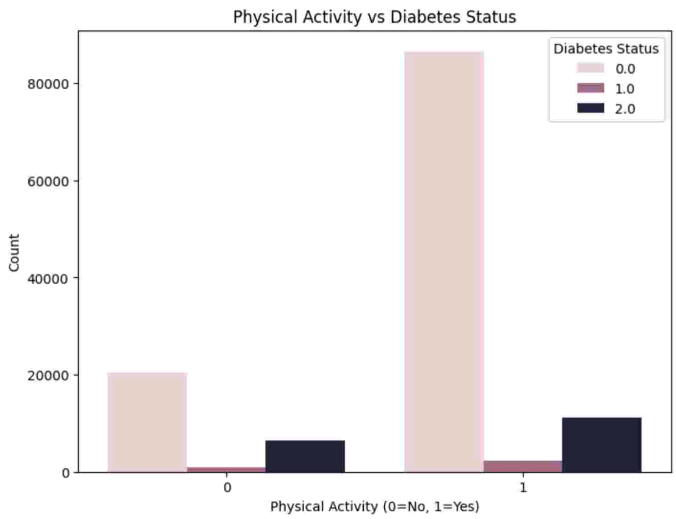
Physical Activity and Diabetes:

- Physical inactivity was a notable factor in individuals with diabetes, indicating the importance of active lifestyles.

sns.countplot(x='PhysActivity', hue='Diabetes_012', data=df)

plt.title('Physical Activity vs Diabetes Status')

plt.show()



4. General Health Ratings and Diabetes

- Poor general health ratings were strongly associated with diabetes.
- Individuals with diabetes were more likely to rate their health poorly, suggesting diabetes significantly impacts overall well-being.

Visualization: Box plot of general health ratings by diabetes status.

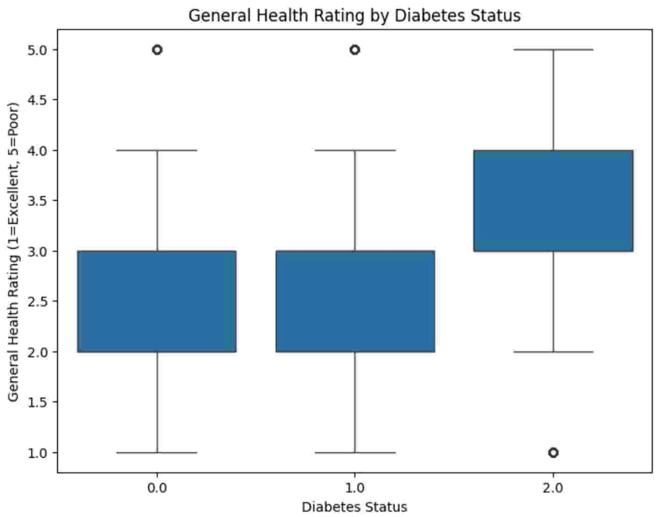
sns.boxplot(x='Diabetes_012', y='GenHlth', data=df, palette="coolwarm")

plt.title('General Health Rating by Diabetes Status')

plt.xlabel('Diabetes Status')

plt.ylabel('General Health Rating (1=Excellent, 5=Poor)')

plt.show()



5. Income and Diabetes

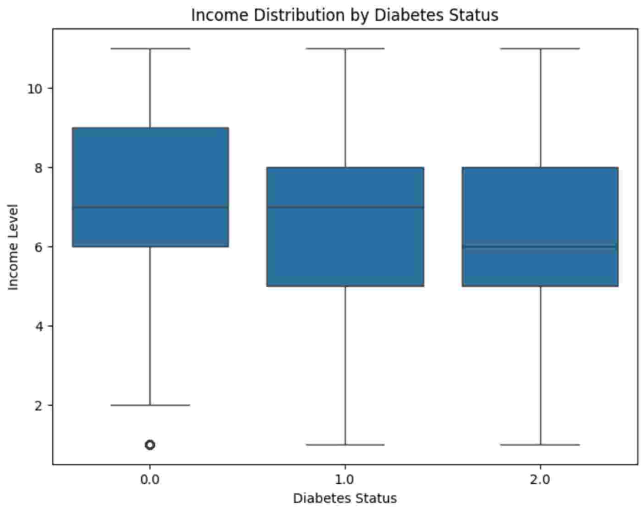
- Lower-income groups were more prone to diabetes, potentially due to limited access to healthcare and healthier lifestyle options.

Visualization: Income distribution by diabetes status.

sns.boxplot(x='Diabetes_012', y='Income', data=df)

plt.title('Income Distribution by Diabetes Status')

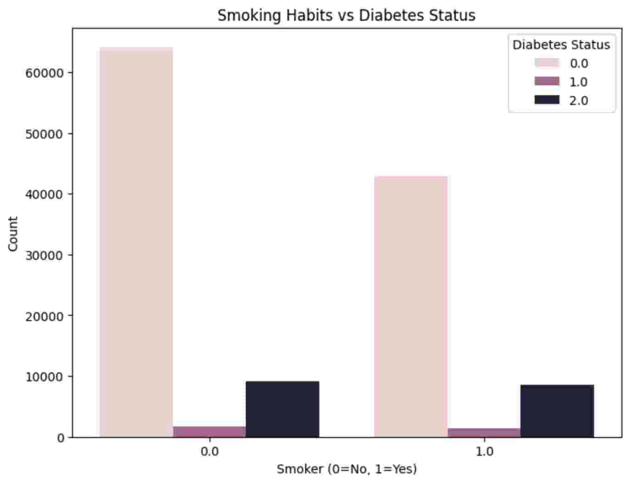
```
plt.xlabel('Diabetes Status')

plt.ylabel('Income Level (Categorical)')

plt.show()
```



6. Smoking and Diabetes

- Smoking habits showed a moderate correlation with diabetes, with a slightly higher prevalence in smokers.

Visualization: Count plot of smoking status by diabetes status.



7. Mental Health and Diabetes

- The number of days individuals reported poor mental health had a noticeable relationship with diabetes status.
- Those with diabetes tended to have more days of poor mental health, indicating a potential psychosocial impact of the condition.

Visualization: Box plot of mental health days by diabetes status.

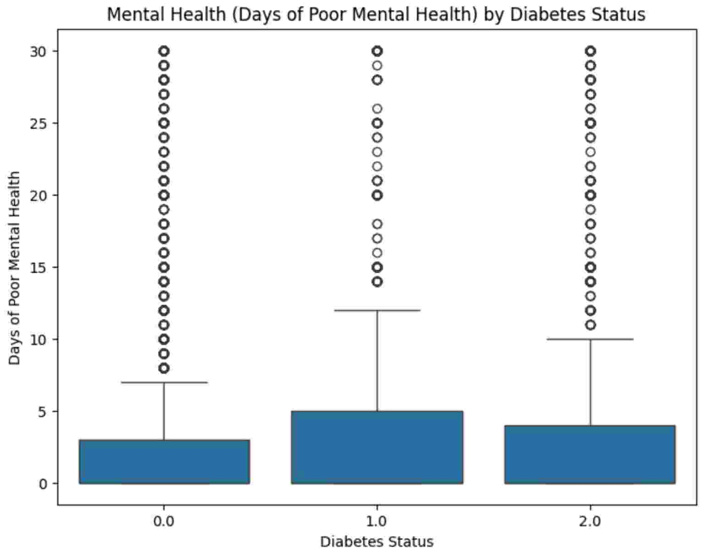
sns.boxplot(x='Diabetes_012', y='MentHlth', data=df, palette="pastel")

plt.title('Mental Health (Days of Poor Mental Health) by Diabetes Status')

plt.xlabel('Diabetes Status')

plt.ylabel('Days of Poor Mental Health')

plt.show()



Interpretation:

- Individuals with diabetes showed a broader range of poor mental health days, with a higher median than those without diabetes.
- This trend suggests a strong psychosocial burden associated with diabetes, warranting integrated mental health care in diabetes management.

8. Physical Health and Diabetes

- The number of days of poor physical health also demonstrated a strong correlation with diabetes status.
- Diabetes patients often reported significantly more days of poor physical health, reflecting the physiological toll of the disease.

Visualization: Violin plot of physical health days by diabetes status.

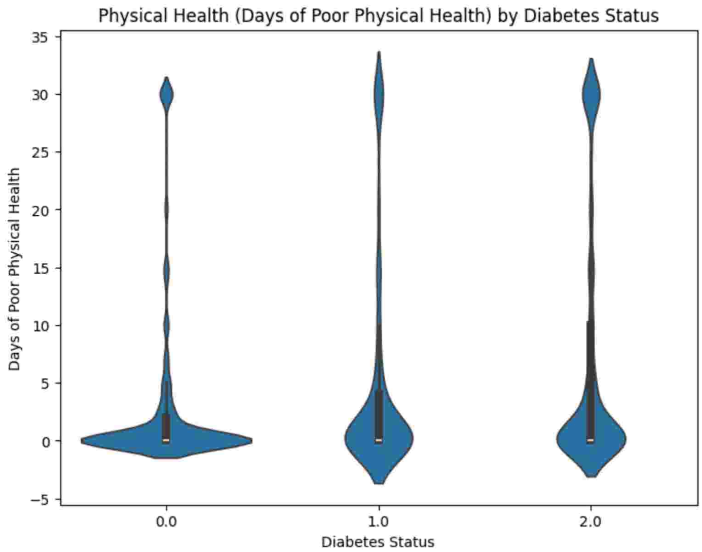
sns.violinplot(x='Diabetes_012', y='PhysHlth', data=df, palette="Set2")

plt.title('Physical Health (Days of Poor Physical Health) by Diabetes Status')

plt.xlabel('Diabetes Status')

plt.ylabel('Days of Poor Physical Health')

plt.show()



interpretation:

- Individuals with diabetes consistently reported higher counts of poor physical health days.
- The wider distribution for diabetes-positive individuals highlights the variability in physical health challenges faced by this group, ranging from mild to severe.

CHAPTER 6: DEPLOYMENT

The deployment phase marks the shift of a machine learning model from a development environment to an interactive and accessible application that end-users can use. In this project, the diabetes prediction model, built with Python and scikit-learn, was deployed as an interactive web application using Streamlit and Google Colab.

This deployment strategy allows users to enter their own health indicators, such as glucose levels, BMI, age, blood pressure, and physical activity, and quickly receive feedback on their diabetes risk. The deployment used Streamlit for the frontend interface and ngrok to create a public web link for the hosted app.

The main goal of this deployment was to make the trained diabetes prediction model:

Interactive: Users can input custom data in real-time.

Accessible: The model can be accessed from any device through a web browser.

Demonstrable: It allows easy sharing and testing during academic presentations or project demonstrations.

Lightweight and free: It uses Google Colab and ngrok to avoid paid hosting services.

The entire deployment pipeline was executed on **Google Colab**, which provides a Python runtime with internet access. The following libraries and tools were used:

Tool/Library	Purpose
Streamlit	Framework to create an interactive user interface.
scikit-learn	Model training and prediction.
pandas, numpy	Data manipulation and preprocessing.
pyngrok	Creates a secure public URL (tunnel) to access the app hosted on Colab.
Google Colab	Cloud-based Jupyter environment for running and deploying the app.

🩺 Diabetes Prediction App

Enter patient details below to predict whether the person has diabetes or not.

Enter Patient Data:

Pregnancies

0 − +

Glucose Level

110.00 − +

Blood Pressure

70.00 − +

Skin Thickness

20.00 − +

Insulin Level

80.00 − +

BMI

25.00 − +

Diabetes Pedigree Function

0.50 − +

Age

30 − +

🔍 Predict

Prediction Result:

🟢 The person is not likely to have diabetes.

CHAPTER 7: CONCLUSIONS AND FUTURE SCOPE

7.1 Conclusion

The successful implementation of a comprehensive data pipeline demonstrates the importance of structured preprocessing in data analytics projects. This pipeline tackled challenges like missing values, duplicate entries, and outliers, ensuring that the final dataset was both clean and robust.

Key Achievements:

1. **Data Preprocessing**: The data pipeline transformed raw data into a ready-to-model format through cleaning, normalization, and feature engineering.

2. **Feature Selection**: Using correlation analysis, 11 significant predictors of diabetes were identified, paving the way for reliable predictive modeling.

3. **Insights Derived**: Patterns such as the strong influence of BMI, general health, and physical activity on diabetes prevalence were uncovered.

7.2 Future Scope

1. **Advanced Predictive Modeling**:
   - Train machine learning models (e.g., logistic regression, random forests) to predict diabetes status based on the preprocessed dataset.
   - Compare models based on metrics like accuracy, precision, recall, and F1-score.

2. **Real-Time Data Pipelines**:
   - Develop pipelines capable of handling real-time data streams from wearable devices or healthcare monitoring systems.

3. **Broader Healthcare Applications**:
   - Expand the pipeline to include other chronic diseases such as cardiovascular conditions or hypertension.

4. **Interactive Dashboards**: Build a dashboard using visualization libraries (e.g., Plotly, Dash) to make insights accessible to healthcare professionals.

CHAPTER 8: IMPACT AND APPLICATIONS

8.1 Significance of the Project

The successful implementation of a structured data pipeline for diabetes prediction addresses critical challenges in healthcare analytics. The project's significance can be summarized as follows:

1. **Improved Data Quality**: By addressing issues like missing values, duplicates, and outliers, the pipeline ensures that the dataset is reliable and ready for analysis.

2. **Scalable Framework**: The data pipeline can be adapted for other datasets, making it a reusable and scalable solution for various healthcare applications.

3. **Comprehensive Insights**: The EDA uncovered meaningful patterns, such as the correlation between BMI, physical activity, mental health, and diabetes. These findings can help healthcare professionals prioritize interventions.

4. **Foundation for Predictive Modeling**: By selecting high-impact features, the project establishes a strong basis for machine learning models, paving the way for accurate diabetes prediction.

8.2 Real-World Applications

1. Predictive Healthcare

- **Diabetes Risk Prediction**: The preprocessed dataset can be used to develop predictive models that classify individuals based on their likelihood of developing diabetes.
- **Personalized Treatment Plans**: By identifying high-risk individuals, healthcare providers can create tailored treatment and preventive strategies.

2. Public Health Awareness

- **Educational Campaigns**: Insights from the project can inform public health campaigns that emphasize the importance of physical activity, healthy diets, and mental health in diabetes prevention.
- **Community Health Initiatives**: Local governments and NGOs can use the findings to design community-based health programs targeting at-risk populations.

3. Integration with Wearable Technology

- **Real-Time Monitoring**: The pipeline can process data from wearable devices, such as fitness trackers, to monitor health indicators like BMI and physical activity in real time.
- **Early Alerts**: With real-time data, individuals can receive early warnings about potential diabetes risks, prompting timely medical consultations.

4. Healthcare Policy Development

- **Resource Allocation**: Policymakers can use the insights to allocate healthcare resources effectively, focusing on high-risk demographics.
- **Data-Driven Decisions**: The findings can guide decisions on funding and implementing diabetes prevention programs.

8.3 Future Enhancements

1. **Machine Learning Models**: Implementing and evaluating machine learning models for diabetes prediction using the preprocessed dataset. Models such as logistic regression, random forests, and deep learning can be tested.
2. **Multi-Disease Analytics**: Expanding the pipeline to analyze comorbidities, such as hypertension and cardiovascular diseases, using similar datasets.
3. **Global Datasets**: Applying the pipeline to datasets from diverse populations to identify regional or cultural variations in diabetes risk factors.

References

1. **Centers for Disease Control and Prevention (CDC)**.
   - National Diabetes Statistics Report. Available at: https://www.cdc.gov/diabetes/data/statistics-report

2. **Pandas Documentation**.
   - Python Data Analysis Library. Available at: https://pandas.pydata.org/docs

3. **Seaborn Documentation**.
   - Statistical Data Visualisation. Available at: https://seaborn.pydata.org

4. **Scikit-learn Documentation**.
   - Machine Learning in Python. Available at: https://scikit-learn.org/stable/documentation.html

5. **Matplotlib Documentation**.
   - Comprehensive Visualization in Python. Available at: https://matplotlib.org/stable/contents.html

6. **World Health Organization (WHO)**.
   - Diabetes Fact Sheet. Available at: https://www.who.int/news-room/fact-sheets/detail/diabetes

7. **Behavioral Risk Factor Surveillance System (BRFSS)**.
   - Dataset Documentation. Available at: https://www.cdc.gov/brfss

8. **Ramesh, Jayroop, Raafat Aburukba, and Assim Sagahyroon. "A remote healthcare monitoring framework for diabetes prediction using machine learning."** ***Healthcare Technology Letters*** **8.3 (2021): 45-57**

9. **Lee, Chonho, et al. "A data analytics pipeline for smart healthcare applications." Sustained Simulation Performance 2017: Proceedings of the Joint Workshop on Sustained Simulation Performance, University of Stuttgart (HLRS) and Tohoku University, 2017. Cham: Springer International Publishing, 2017.**

10. **Abnoosian, Karlo, Rahman Farnoosh, and Mohammad Hassan Behzadi. "A pipeline-based framework for early prediction of diabetes."** ***Journal of Health and Biomedical Informatics*** **10.2 (2023): 125-140.**

11. **Mohanty, Aisworya, et al. "Study and impact analysis of machine learning approaches for smart healthcare in predicting mellitus diabetes on clinical data."** ***Smart Healthcare Analytics: State of the Art*****. Singapore: Springer Singapore, 2021. 75-101.**

12. **Saeed, Mohammed A., and Mogeeb A. Saeed. "Real-Time Diabetes Detection Using Machine Learning and Apache Spark."** ***2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*****. IEEE, 2024.**

13. **Chowdhury, Lomat Haider, et al. "An optimized data analytics pipeline for improving healthcare diagnosis using ensemble learning."** ***Informatics in Medicine Unlocked*** **53 (2025): 101623.**

14. **Rani, Sita, et al. "Machine Learning-Powered Smart Healthcare Systems in the Era of Big Data: Applications, Diagnostic Insights, Challenges, and Ethical Implications."** ***Diagnostics*** **15.15 (2025): 1914.**

15. **Nauman, Muhammad, et al. "The role of big data analytics in revolutionizing diabetes management and healthcare decision-making."** ***IEEE Access*** **(2025).**

CRP Report Plague.pdf

Madhav Institute of Technology & Science

Document Details

Submission ID

trn:oid:::28506:121080982

Submission Date

Nov 12, 2025, 11:22 AM GMT+5:30

Download Date

Nov 12, 2025, 11:23 AM GMT+5:30

File Name

CRP Report Plague.pdf

File Size

749.6 KB

18 Pages

2,547 Words

15,537 Characters

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 8 words)

Match Groups

23 Not Cited or Quoted 15%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources

3% Publications

10% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **23** Not Cited or Quoted 15%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

6% Internet sources
3% Publications
10% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	fastercapital.com	1%
2	Submitted works	Southampton Solent University on 2024-09-13	1%
3	Submitted works	University of East London on 2025-09-01	1%
4	Submitted works	Gisma University of Applied Sciences GmbH on 2024-04-06	1%
5	Submitted works	Victorian Institute of Technology on 2025-10-12	<1%
6	Internet	www.educba.com	<1%
7	Submitted works	University of Derby on 2025-10-01	<1%
8	Internet	www.jetir.org	<1%
9	Internet	mro.massey.ac.nz	<1%
10	Submitted works	Asia Pacific University College of Technology and Innovation (UCTI) on 2025-06-02	<1%

11	Submitted works	
University of North Texas on 2023-12-03		<1%
12	Internet	
shodhganga.inflibnet.ac.in		<1%
13	Submitted works	
CSU, San Jose State University on 2015-09-25		<1%
14	Publication	
Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and ...		<1%
15	Publication	
J. Stuart Krause, Jennifer Coker, Susan Charlifue, Gale G. Whiteneck. "Depression ...		<1%
16	Submitted works	
Brunel University on 2024-04-15		<1%
17	Submitted works	
FICT on 2025-09-12		<1%
18	Publication	
Julian D. Ford, Mary L. Adams, Wayne F. Dailey. "Psychological and health proble...		<1%
19	Publication	
Niha Ansari. "Machine Learning in Forensic Evidence Examination - A New Era", C...		<1%
20	Internet	
iibsonline.com		<1%
21	Internet	
repository.nida.ac.th		<1%