*A*

*Project Report*

*On*

## TOPIC

## Assessing Future Health Risks Based on Current Lifestyle

## A Focus on Type 2 Diabetes

*Field Project Report submitted for*

## Bachelors of Science (Computer Science)

*Submitted by*

| Roll. No. | Name | SAP No. |
|-----------|------|---------|
| **C029** | Hariom Jaiswal | 40721230017 |
| **C040** | Anshika Pangotra | 40721230011 |
| **C030** | Sangeeta Jogi | 40721230001 |
| **C032** | Sonum Katoch | 40315220072 |

*Under the guidance of*

## Mr. Omkar Mohite

## Year: 2024 - 2025

# CERTIFICATE

This is to certify that the field project titled *"Assessing Future Health Risks Based on Current Lifestyle: A Focus on Type 2 Diabetes"* is successfully done by <u>Hariom Jaiswal</u> in fulfilment of **Bachelor's of Science (Computer Science)** under the **University of Mumbai, Mumbai** through the **Mithibai College of Arts, Chauhan Institute of Science & Amrutben Jivanlal College of Commerce & Economics (Empowered Autonomous), Mumbai** carried out by him under our guidance and supervision.

| | | |
|---|---|---|
| **FP Teacher In-Charge** | **FP Coordinator** | **Principal** |

**Date: 08ᵗʰ March, 2025**

# 1. Objectives

**- Assessing Individual Health risk based on current lifestyle**
**- Particularly focusing on Type 2 Diabetes**

**Processed Followed:**

Assess Individual Health Risk:
Develop a scoring mechanism to evaluate an individual's health risk based on lifestyle factors such as exercise, sleep, and BMI.

Identify Diabetes Type 2 Risk:
Predict the likelihood of developing Type 2 Diabetes by analysing key health parameters, enabling early intervention.

Analyse the Impact of Lifestyle Choices:
Examine how lifestyle factors (e.g., physical activity, health condition, sleep duration, and stress) contribute to long-term health risks.

Categorize Risk Levels:
Classify individuals into low, moderate, and high-risk categories for long-term health risks and predicting the high risk and low risk for Type 2 Diabetes.

Develop Predictive Models:
Implement and compare machine learning models (Logistic Regression, Decision Tree, Random Forest) to determine the best approach for risk prediction.

# 2. Introduction

Diabetes, particularly Type 2 Diabetes, is a growing global health concern driven largely by lifestyle factors such as physical inactivity, poor diet, stress, and obesity. Unlike Type 1 Diabetes, which is primarily genetic and autoimmune-related, Type 2 Diabetes is preventable with early lifestyle interventions. Identifying individuals at high risk is crucial for preventive healthcare strategies.

This project aims to analyse various lifestyle, health, and demographic factors to assess an individual's risk of developing Type 2 Diabetes. By using machine learning models, we aim to build a predictive system that can help individuals and healthcare professionals identify high-risk individuals early and suggest preventive measures.

# 3. Methodology Overview

### 3.1. Data Collection:
Data was collected by conducting an offline survey and interviewing nearly 530 individuals from different age groups, 16, 17, 18, 19, 20, and 21+, and from multiple streams, such as Science, Commerce, Arts, and Others.

### 3.2. Survey Structure:
The dataset for this study was collected through a structured survey designed to capture key lifestyle and health-related factors influencing Health. The survey was divided into multiple sections to ensure a comprehensive analysis of an individual's daily routine, fitness habits, and medical condition.

### 3.3. Scoring System in the Survey:
To quantify lifestyle habits and health conditions, a scoring system was incorporated into the survey. Each response was assigned a numerical value to facilitate risk assessment and predictive model.

The scoring was designed as follows:
**Higher scores** indicate healthier habits and lower risk factors (e.g., regular exercise, sufficient sleep, strong fitness motivation).

**Lower scores** or zero values represent potential risk factors (e.g., sedentary lifestyle, poor sleep, existing health conditions, or lack of exercise).

**Cumulative scores** from different sections were later used to calculate risk categories, such as Health Risk Score and Diabetes Risk Score, helping in the classification of individuals into low, moderate, and high-risk groups.

### 3.4. Structure of Form:
Q1: Age

Q2: Gender

Q3: Stream

Q4: How often do you exercise?
    A. Rarely (0-1 times/week)        *[0]*
    B. Sometimes (2-3 times/week)     *[1]*
    C. Frequently (4+ times/week)     *[2]*

Q5: Describe your daily routine.
    A. Mostly sedentary (desk job, minimal movement) *[0]*
    B. Moderately active (some walking/movement)     *[1]*
    C. Very active (physical job, frequent exercise)     *[2]*

Q6: How long do you exercise daily?
    A. Less than 15 minutes    *[0]*
    B. 15-30 minutes    *[1]*
    C. More than 30 minutes    *[2]*

Q7: Your primary motivation for exercising?
    A. Health & longevity    *[2]*
    B. Weight loss / muscle gain    *[1]*
    C. Stress relief / mental well-being    *[1]*
    D. I don't exercise    *[0]*

Q8: Importance of Fitness for you.
    A. Very important    *[2]*
    B. Somewhat important    *[1]*
    C. Not important    *[0]*

Q9: Any pre-existing health conditions?
    A. Diabetes    *[1]*    *.....We have added the score*
    B. Hypertension    *[1]*    *in case of multiple selection*
    C. Heart disease    *[1]*
    D. Asthma or breathing problems    *[1]*
    E. No pre-existing conditions    *[0]*

Q10: On average, how many hours do you sleep per night?
    A. Less than 5 hours    *[0]*
    B. 5-7 hours    *[1]*
    C. More than 7 hours    *[2]*

Q11: What is the biggest challenge to staying consistent with exercise?
    A. Lack of motivation *[1]*
    B. Time constraints    *[1]*
    C. Health issues    *[2]*
    D. No challenges    *[0]*

Q12: Would you consider joining a gym or fitness class in the next 6 months?
    A. Yes *[1]*
    B. No  *[0]*

Q13: Dietary Habits?
    A. Junk food consumption    *[1]*
    B. Sugar intake    *[0]*

Q14: Stress Levels?
    A. Low    *[0]*
    B. Moderate    *[1]*
    C. High    *[2]*

## 3.5. Overview of Data Collection: *(CSV Format)*

Age, Gender, Stream, How often do you exercise?, Describe your daily routine, How long do you exercise daily?, Your primary motivation for exercising?, Importance of Fitness for you, Any pre-existing health conditions? , On average, how many hours do you sleep per night?, What is the biggest challenge to staying consistent with exercise?, Would you consider joining a gym or fitness class in the next 6 months?, Dietary Habits?, Stress Levels?

19, Male, Science, B. Sometimes (2-3 times/week), B. Moderately active (some walking/movement), B. 15-30 minutes, B. Weight loss / muscle gain, A. Very important, E. No pre-existing conditions, A. Less than 5 hours, B. Time constraints, A. Yes, Junk food consumption, Sugar intake, High

16, Male, Commerce, C. Frequently (4+ times/week), B. Moderately active (some walking/movement), C. More than 30 minutes, A. Health & longevity, B. Weight loss / muscle gain, C. Stress relief / mental well-being, A. Very important, C. Heart disease, A. Less than 5 hours, A. Lack of motivation, B. No, Junk food consumption, Sugar intake, Moderate

20, Female, Science, B. Sometimes (2-3 times/week), C. Very active (physical job, frequent exercise), A. Less than 15 minutes, C. Stress relief / mental well-being, B. Somewhat important, E. No pre-existing conditions, A. Less than 5 hours, A. Lack of motivation, B. No, Junk food consumption, Sugar intake, Moderate

19, Male, Science, A. Rarely (0-1 times/week), A. Mostly sedentary (desk job, minimal movement), A. Less than 15 minutes, C. Stress relief / mental well-being, A. Very important, D. Asthma or breathing problems, B. 5-7 hours, B. Time constraints, B. No, Junk food consumption, Sugar intake, Moderate

19, Female, Science, B. Sometimes (2-3 times/week), C. Very active (physical job, frequent exercise), A. Less than 15 minutes, A. Health & longevity, C. Stress relief / mental well-being, B. Somewhat important, E. No pre-existing conditions, B. 5-7 hours, D. No challenges, B. No, Junk food consumption, Sugar intake, High

20, Female, Science, A. Rarely (0-1 times/week), B. Moderately active (some walking/movement), A. Less than 15 minutes, B. Weight loss / muscle gain, B. Somewhat important, E. No pre-existing conditions, C. More than 7 hours, A. Lack of motivation, A. Yes, Junk food consumption, Sugar intake, Moderate

*....for full access to the data links related to the dataset, refer to the Source Code provided at the end.*

# 4. Feature Engineering

Feature engineering plays a crucial role in enhancing the dataset by transforming raw data into meaningful features that improve the model's predictive performance. The following steps were performed:

**4.1. Handling Missing Values:**
Missing values were imputed using the mode of the respective column to ensure consistency without distorting the dataset.

**4.2. Encoding Categorical Variables:**
Categorical responses (e.g., "Yes"/"No", "Low"/"High") were mapped to numerical values for machine learning compatibility.

Example:
Exercise Frequency:
Rarely → 0, Sometimes → 1, Frequently → 2

Activity Level:
Sedentary → 0, Moderately Active → 1, Very Active → 2

**4.3. Data Transformation:**
New numerical columns were created based on previous mappings.
The original categorical columns were removed to ensure the dataset was fully numerical for machine learning models.

**Column Structure: (Before)**
Age, Gender, Stream, How often do you exercise?, Describe your daily routine, How long do you exercise daily?, Your primary motivation for exercising?, Importance of Fitness for you, Any pre-existing health conditions? , On average, how many hours do you sleep per night?, What is the biggest challenge to staying consistent with exercise?, Would you consider joining a gym or fitness class in the next 6 months?, Dietary Habits?, Stress Levels?

**Column Structure: (After)**
Age , Gender_num, ~~Stream~~, Exercise Frequency, Activity Level, exercise_duration, motivation_score, Fitness Importance, health_condition_flag, Sleep Duration, challenge_score, gym_interest, Diet_score, Stress_score

*...the columns are aligned according to their previously mapped counterparts*
*...the strike 'Stream' column was dropped*

### 4.4. Creation of New Features:

**Estimate BMI**: Derived approximately based on fitness habits and activity levels.
*Code:*

```python
def estimate_bmi(row):
    if row["Exercise Frequency"] == 0 and row["Activity Level"] == 0:
        return "High (Overweight/Obese)"
    elif row["Exercise Frequency"] == 1 and row["Activity Level"] == 1:
        return "Moderate (Borderline Overweight)"
    else:
        return "Normal"

bmi_map = {"Normal": 0,
           "Moderate (Borderline Overweight)": 1,
           "High (Overweight/Obese)": 2}
```

*Figure 1: Code to calculate BMI*

**Risk Score**: Summed scores of key lifestyle and health factors to measure overall risk.
*Code:*

```python
df['risk_score'] = ((2 - df['Exercise Frequency']) +
                    (2 - df['Activity Level']) +
                    (2 - df['exercise_duration']) +
                    (2 - df['Fitness Importance']) +
                    df['health_condition_flag'] +
                    (np.where(df['Sleep Duration'] == 0, 1, 0)) +
                    df['challenge_score'] +
                    df['Diet_score']  +
                    df['Stress_score'])

def categorize_risk(score):
    if score <= 4:
        return 0  # Low Risk
    elif score <= 6:
        return 1  # Moderate Risk
    else:
        return 2  # High Risk
```

*Figure 2: Code to calculate Risk Score based on Lifestyle*

**Health Risk Category**: Created by categorizing the Risk Score into Low, Moderate, and High.
*Code:*

```python
df['health_risk'] = df['risk_score'].apply(categorize_risk)
```

**Diabetes Risk Score**: Computed based on factors such as age, BMI, health conditions (hypertension, heart issues), and lifestyle.
*Code:*

```python
def calculate_risk_score(row):
    score = 0

    # Age Factor
    if row["Age"] >= 40:
        score += 3  # Older age increases risk
    elif 30 <= row["Age"] < 40:
        score += 2

    # Lifestyle Factors
    if row["Exercise Frequency"] == 0: score += 2  # No exercise → Higher risk
    if row["Activity Level"] == 0: score += 2  # Sedentary lifestyle → Higher risk
    if row["Sleep Duration"] == 0: score += 2  # Poor sleep → Higher risk

    # Health Conditions
    if row["Diabetes"] == 1:
        score += 7  # Pre-existing diabetes is a strong predictor
    if row["Hypertension"] == 1:
        score += 3  # High BP increases diabetes risk
    if row["Heart"] == 1:
        score += 2  # Heart disease is linked to diabetes

    # BMI Factor (Estimated)
    if row["Estimated BMI"] == "High (Overweight/Obese)":
        score += 3  # Obesity is a major risk factor
    elif row["Estimated BMI"] == "Moderate (Borderline Overweight)":
        score += 1

    #Stress
    if row["Stress_score"] == 2:  # High stress
        score += 2
    elif row["Stress_score"] == 1:  # Moderate stress
        score += 1

    return score

# Apply the function
df["Diabetes Risk Score"] = df.apply(calculate_risk_score, axis=1)
```

*Figure 1: Code to calculate Diabetes Risk Score*

Diabetes Risk Classification: Converted Diabetes Risk Score into a binary variable (High Risk = 1, Low Risk = 0).
*Code:*

```python
df["Diabetes Risk"]=df["Diabetes Risk Score"].apply(lambda x:1 if x >= 7 else 0)
```

### 4.5. Feature Selection for Machine Learning:

Only the most relevant features were used for model training:
Redundant or highly correlated features were dropped to reduce multicollinearity and improve model efficiency.

By performing these feature engineering steps, we ensured that the dataset was clean, structured, and well-optimized for training machine learning models to predict health risks and Type 2 Diabetes susceptibility.

# 5. Machine Learning Model

After performing all necessary preprocessing steps, our dataset is now prepared for machine learning.
The goal is to predict future health risk based on lifestyle and health-related factors.

### 5.1. Features and Target Variable:

#### 5.1.1. Predicting Future Health Risk:

Features:
Age, Exercise Frequency, Activity Level, exercise_duration, motivation_score, Fitness Importance, health_condition_flag, Sleep Duration, challenge_score, gym_interest, Gender_num

Target:
health_risk (Categorized into Low, Moderate, and High Risk)

#### 5.1.2. Predicting Type 2 Diabetes Risk:

Features:
Exercise Frequency, Activity Level, Sleep Duration, Fitness Importance, Estimated BMI

Target:
Diabetes Risk (Categorized into Low and High Risk)

### 5.2. Model Selection:

To predict future health risk, we used three classification models:
1. Logistic Regression
2. Decision Tree
3. Random Forest

### 5.3. Model Training & Testing:

The dataset was split into training (80%) and testing (20%) using train_test_split. Each model was trained using the training data and evaluated on the train and test set.

### 5.4. Performance Metrics:

To assess the effectiveness of each model, we used:
**Accuracy**: Measures the overall correctness of predictions.

**Precision, Recall, F1-score**: Evaluates the balance between correctly identifying high-risk individuals and avoiding false positives.

**Confusion Matrix**: Provides insight into misclassifications across different risk levels.

# 6. Logistic Regression

## 6.1. Predicting Future Health Risk:
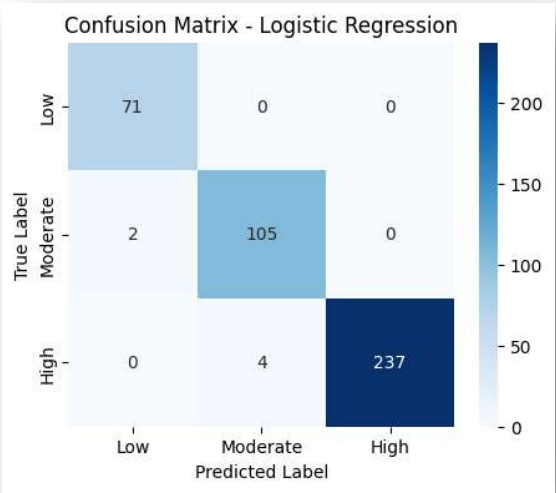
Confusion Matrix:



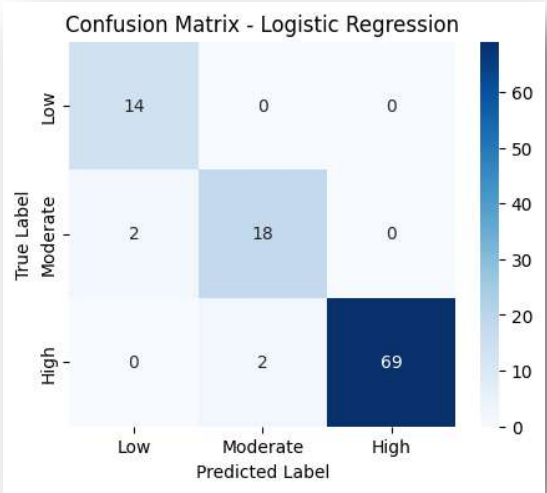*Figure 4: Logistic Regression (Training Phase) Confusion Matrix*



*Figure 5: Logistic Regression (Testing Phase) Confusion Matrix*

Classification Report:

Model Accuracy (Training Phase): **98.57%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.99        71
           1       0.96      0.98      0.97       107
           2       1.00      0.98      0.99       241

    accuracy                           0.99       419
   macro avg       0.98      0.99      0.98       419
weighted avg       0.99      0.99      0.99       419
```

*Figure 6: Logistic Regression (Training Phase) Classification Report*

Model Accuracy (Testing Phase): **96.19%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.88      1.00      0.93        14
           1       0.90      0.90      0.90        20
           2       1.00      0.97      0.99        71

    accuracy                           0.96       105
   macro avg       0.92      0.96      0.94       105
weighted avg       0.96      0.96      0.96       105
```

*Figure 7: Logistic Regression (Testing Phase)*
*Classification Report*

## 6.2. Predicting Type 2 Diabetes:

Classification Report:

Model Accuracy (Testing Phase): **93.33%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      1.00      0.96        78
           1       1.00      0.74      0.85        27

    accuracy                           0.93       105
   macro avg       0.96      0.87      0.90       105
weighted avg       0.94      0.93      0.93       105
```

*Figure 8: Logistic Regression (Testing Phase)*
*Classification Report*

# 7. Decision Tree

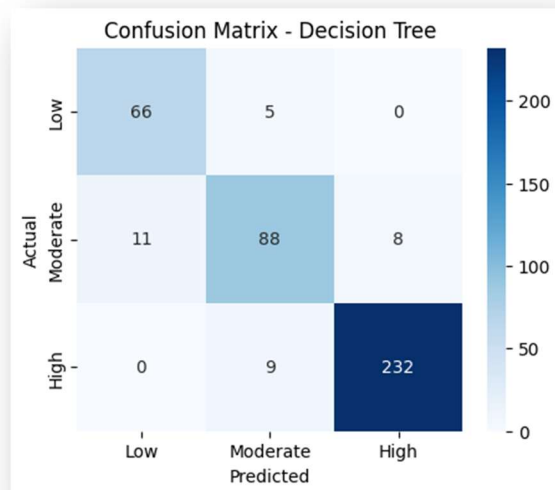## 7.1. Predicting Future Health Risk:

Confusion Matrix:



Figure 9: Decision Tree (Training Phase)
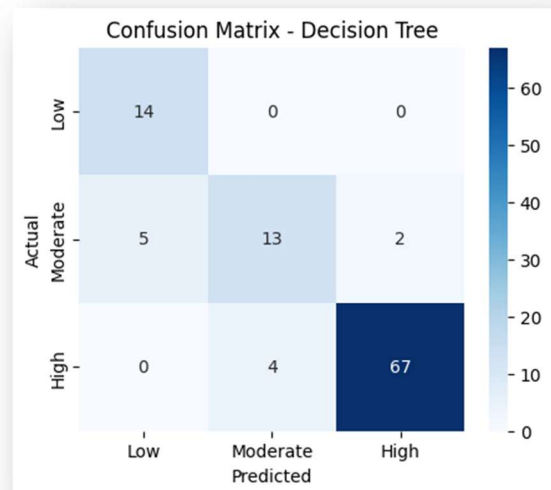Confusion Matrix



Figure 10: Decision Tree (Testing Phase)
Confusion Matrix

Classification Report:

Model Accuracy (Training Phase): **92.12%**



```
Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.93      0.89        71
           1       0.86      0.82      0.84       107
           2       0.97      0.96      0.96       241

    accuracy                           0.92       419
   macro avg       0.90      0.90      0.90       419
weighted avg       0.92      0.92      0.92       419
```

Figure 11: Decision Tree (Training Phase)
Classification Report

Model Accuracy (Testing Phase): **89.52%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.74      1.00      0.85        14
           1       0.76      0.65      0.70        20
           2       0.97      0.94      0.96        71

    accuracy                           0.90       105
   macro avg       0.82      0.86      0.84       105
weighted avg       0.90      0.90      0.89       105
```

*Figure 12: Decision Tree (Testing Phase)*
*Classification Report*

## 7.2. Predicting Type 2 Diabetes:

Classification Report:

Model Accuracy (Testing Phase): **93.33%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      1.00      0.96        78
           1       1.00      0.74      0.85        27

    accuracy                           0.93       105
   macro avg       0.96      0.87      0.90       105
weighted avg       0.94      0.93      0.93       105
```

*Figure 13: Decision Tree (Testing Phase)*
*Classification Report*

# 8. Random Forest

## 8.1. Predicting Future Health Risk
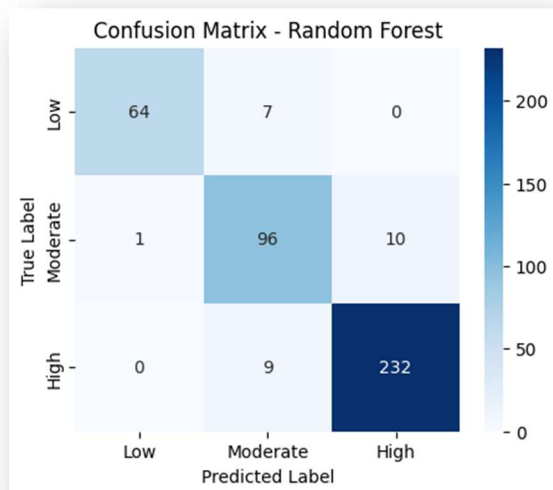
Confusion Matrix:



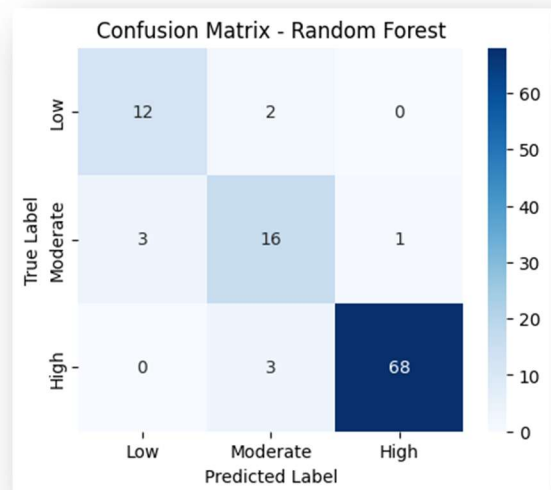Figure 12: Random Forest (Training Phase) Confusion Matrix



Figure 15: Random Forest (Testing Phase) Confusion Matrix

Classification Report:

Model Accuracy (Training Phase): **93.56%**



```
Classification Report:
             precision    recall  f1-score   support

          0       0.98      0.90      0.94        71
          1       0.86      0.90      0.88       107
          2       0.96      0.96      0.96       241

   accuracy                          0.94       419
  macro avg       0.93      0.92      0.93       419
weighted avg       0.94      0.94      0.94       419
```

Figure 16: Random Forest (Training Phase) Classification Report

Model Accuracy (Testing Phase): **91.43%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.80      0.86      0.83        14
           1       0.76      0.80      0.78        20
           2       0.99      0.96      0.97        71

    accuracy                           0.91       105
   macro avg       0.85      0.87      0.86       105
weighted avg       0.92      0.91      0.92       105
```

*Figure 17: Random Forest (Training Phase)*
*Classification Report*

## 8.2. Predicting Type 2 Diabetes:

Classification Report:

Model Accuracy (Testing Phase): **93.33%**

```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      1.00      0.96        78
           1       1.00      0.74      0.85        27

    accuracy                           0.93       105
   macro avg       0.96      0.87      0.90       105
weighted avg       0.94      0.93      0.93       105
```

*Figure 18: Random Forest (Training Phase)*
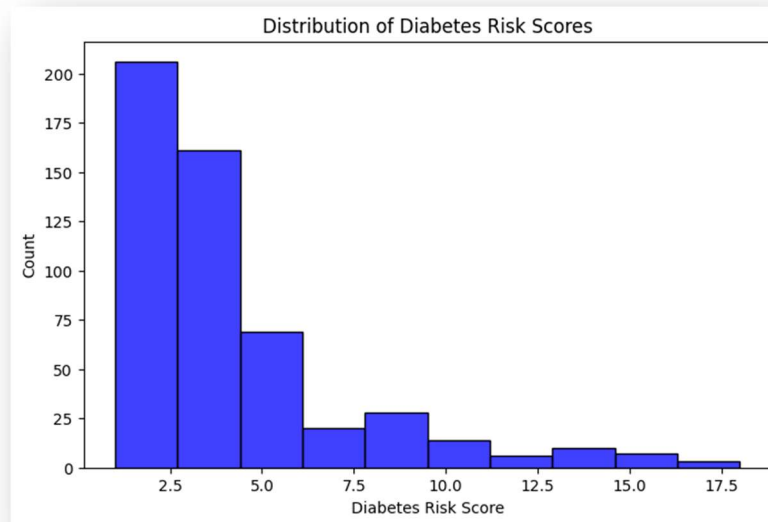*Classification Report*

# 9. Visualization and Inference



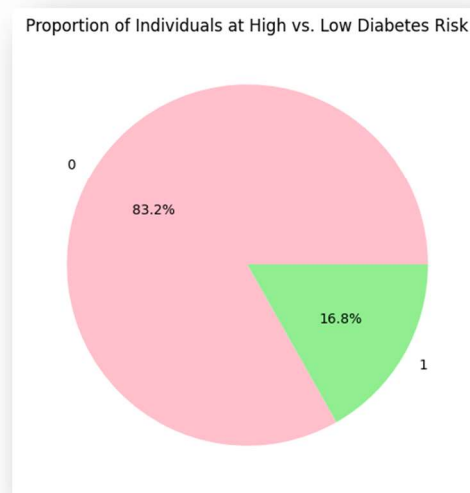*Figure 19: Distribution of Diabetes Risk Score*



*Figure 20: Proportion of Individuals at High VS Low Diabetes Risk*

*(Inference based on Figure. 19 & Fig. 20)*

**Majority at Low Risk** – The distribution is right-skewed, meaning most individuals have low diabetes risk scores (between 0 and 5). This suggests that a large portion of the surveyed population falls into the low-risk category.

**Moderate-Risk Group** – A smaller number of individuals fall within the moderate risk range (around 5 to 7). This group may have some lifestyle or health conditions that contribute to an increased risk.

**Few High-Risk Individuals** – Very few people have high diabetes risk scores (above 10), suggesting that only a small fraction of the population is at significant risk for Type 2 Diabetes. These individuals likely have multiple contributing factors like poor lifestyle habits or pre-existing conditions.
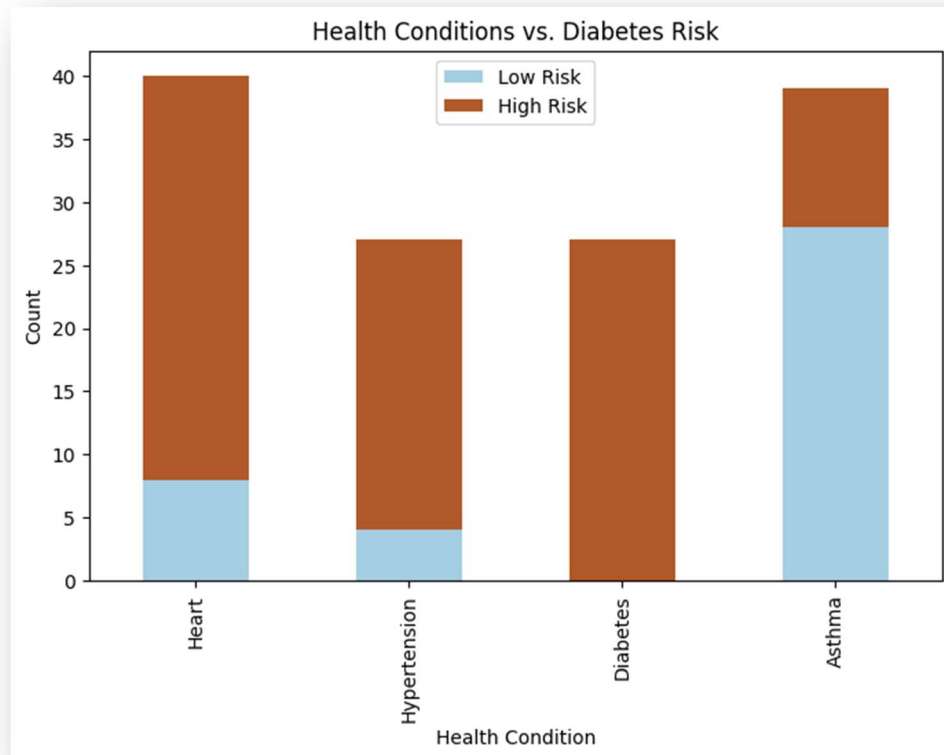
*Figure 21: Health Conditions VS Diabetes Risk*

*(Inference based on Figure. 21)*

**Heart Disease** Has the Highest High-Risk Association Among individuals with heart disease, the majority fall into the high-risk diabetes category (brown section).This suggests a strong correlation between heart disease and diabetes risk, possibly due to shared risk factors like poor diet, obesity, and sedentary lifestyle.

**Hypertension** and Diabetes Show High-Risk Trends Most individuals with hypertension and diabetes already have a high diabetes risk score. This reinforces the well-established medical link between high blood pressure and diabetes, as both conditions often co-exist due to metabolic syndrome and insulin resistance.

**Asthma** Shows a Different Trend Unlike the other conditions, asthma has a relatively higher proportion of low-risk individuals (blue section).This suggests that asthma alone may not be a major factor in diabetes risk, though other lifestyle or genetic influences could play a role.
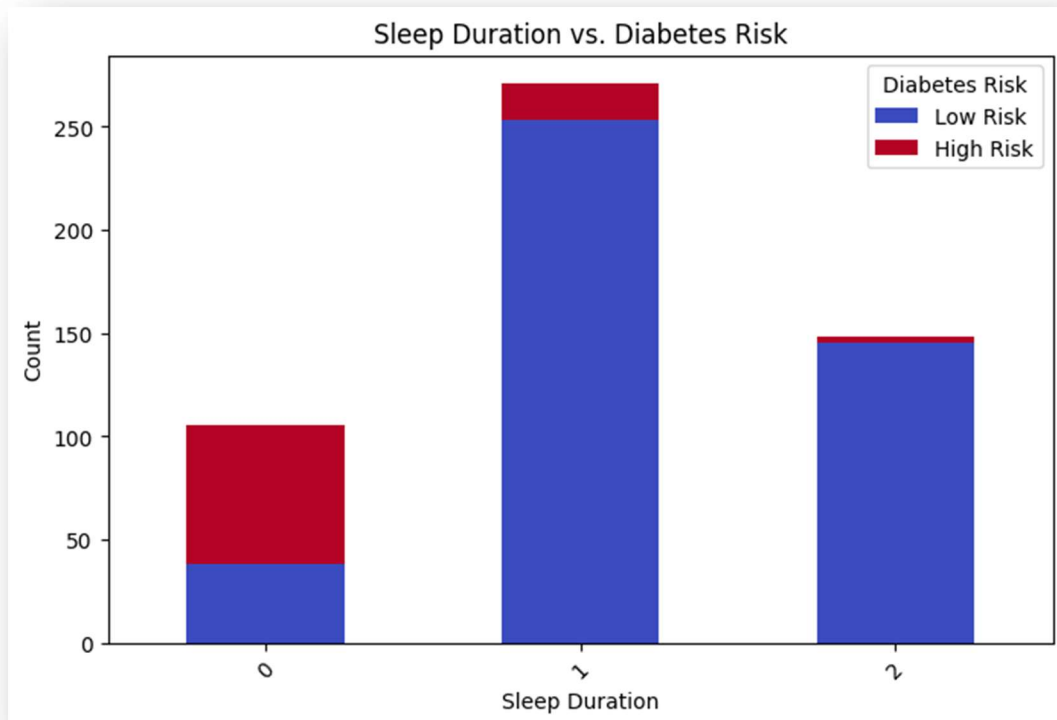
16

*Figure 22: Sleep Duration VS Diabetes Risk*

*(Inference based on Figure. 22)*

**Higher Diabetes Risk with Short Sleep Duration:**

Individuals with lower sleep duration (0 category) have a noticeable proportion of high-risk cases.

This aligns with research suggesting that insufficient sleep disrupts insulin sensitivity and glucose metabolism, increasing diabetes risk.

**Optimal Sleep Duration and Lower Risk:**

The majority of individuals in the mid-range sleep duration (1 category) have a predominantly low diabetes risk.

This supports existing studies indicating that around 7 hours of sleep is ideal for maintaining metabolic health.

**Slight Increase in Risk for Higher Sleep Durations:**

The group with the highest sleep duration (2 category) shows some increased diabetes risk cases.

This could be due to underlying health conditions, lower physical activity, or other metabolic disruptions associated with excessive sleep.
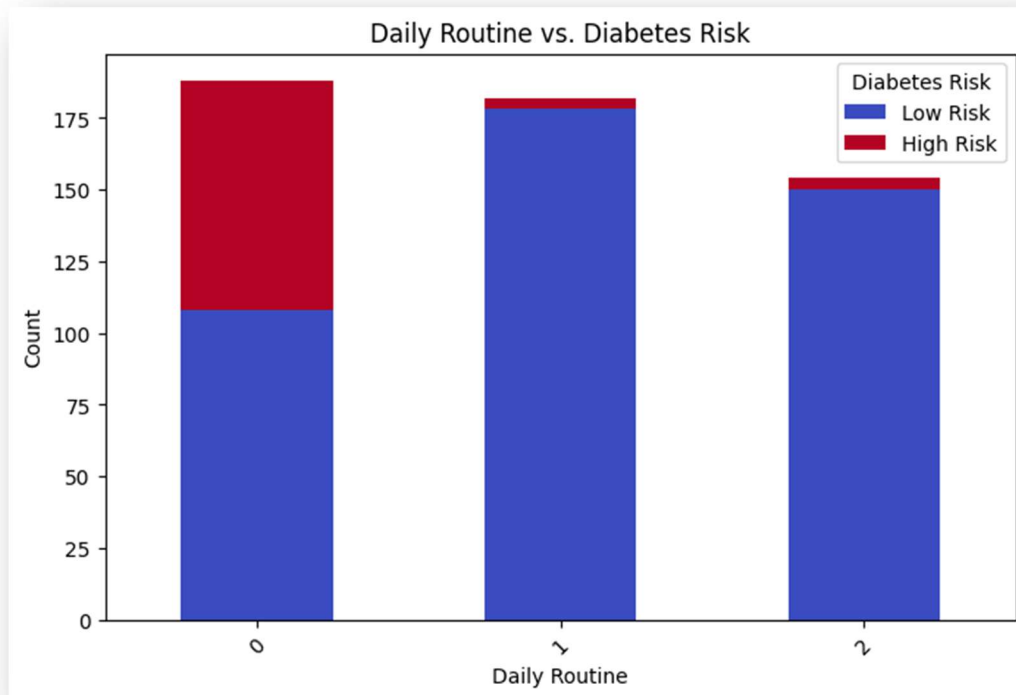
17

*Figure 23: Daily Routine VS Diabetes Risk*

*(Inference based on Figure. 23)*

**Sedentary Lifestyle Increases Risk:**

Individuals with a mostly sedentary routine (0) have the highest proportion of high-risk (red) individuals compared to other groups.

This suggests that lack of physical activity is a strong contributing factor to diabetes risk.

**Moderately Active Individuals (1) Show Lower Risk:**

The group with moderate activity (1) has a significantly lower proportion of high-risk individuals than the sedentary group.

This indicates that even moderate movement throughout the day can help reduce diabetes risk.

**Highly Active Individuals (2) Have the Lowest Risk:**

The most active group (2) has the smallest proportion of high-risk individuals.

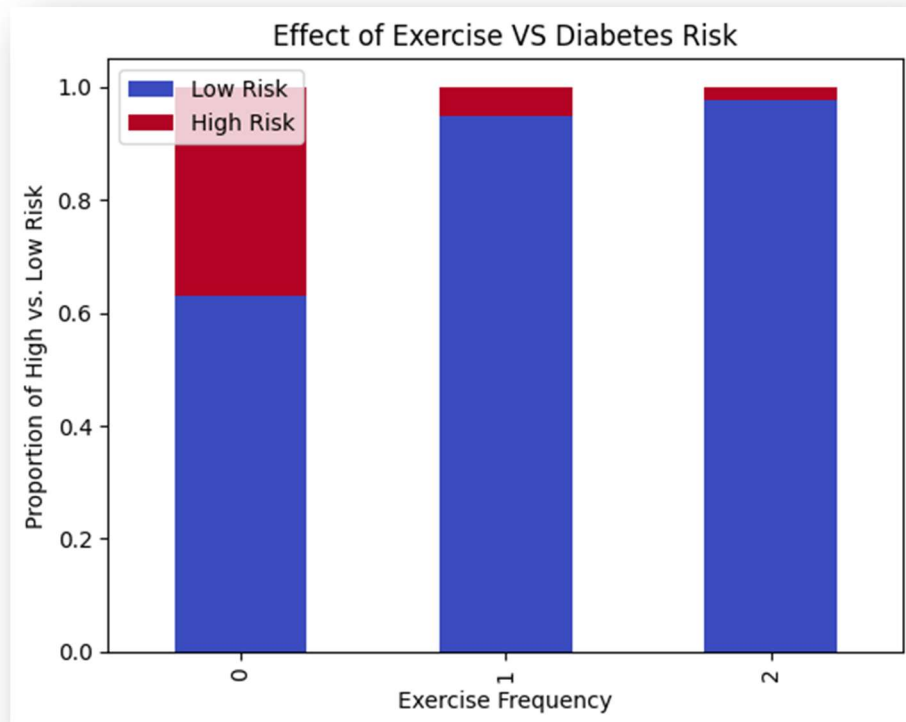This suggests that an active lifestyle is highly effective in mitigating diabetes risk.

*Figure 24: Proportion of High VS Low Risk*

*(Inference based on Figure. 24)*

**Higher Exercise Frequency Correlates with Lower Diabetes Risk**

Individuals who exercise frequently (categories 1 and 2) have a significantly lower proportion of high-risk cases compared to those who exercise rarely (category 0).

This suggests a strong inverse relationship between exercise and diabetes risk.

**Sedentary Individuals Are at Higher Risk**

The highest proportion of high-risk individuals is in the group with little to no exercise (category 0).

This aligns with existing research showing that a lack of physical activity increases the likelihood of insulin resistance and Type 2 Diabetes.

**Exercise as a Preventive Measure**

Since high-risk cases are almost absent in individuals who exercise regularly, promoting physical activity could be a key intervention to reduce diabetes risk.

Encouraging even moderate exercise may have substantial health benefits.

# 10. Observations

The analysis of various lifestyle factors, including exercise frequency, daily routine, stress levels, and sleep duration, reveals clear patterns in their relationship with diabetes risk. Key findings from our visualisations suggest that:

*(Inference based on Figure. 24)*
**Physical Activity Plays a Critical Role** – Regular exercise is strongly associated with a lower risk of diabetes. Individuals who rarely exercise are more likely to fall into the high-risk category, while those who engage in frequent physical activity show a significantly lower risk.

*(Inference based on Figure. 23)*
**Daily Routine and Lifestyle Choices Matter** – A sedentary daily routine correlates with a higher diabetes risk, reinforcing the importance of an active lifestyle.

*(Inference based on Figure. 22)*
**Sleep Duration and Stress Management Are Essential** – Both inadequate sleep and high-stress levels show a strong association with increased diabetes risk. This underscores the need for a balanced lifestyle that includes proper rest and stress reduction strategies.

*(Inference based on Figure. 19)*
**Preventive Measures Can Reduce Risk** – Since a significant portion of individuals falls into the moderate-risk category, early interventions such as dietary improvements, exercise programs, and stress management techniques can prevent progression to high risk.

Diabetes is a growing health concern, but the insights from our data analysis highlight that it is largely preventable through lifestyle modifications. By prioritizing physical activity, maintaining a structured daily routine, managing stress, and ensuring adequate sleep, individuals can significantly reduce their risk of developing diabetes.
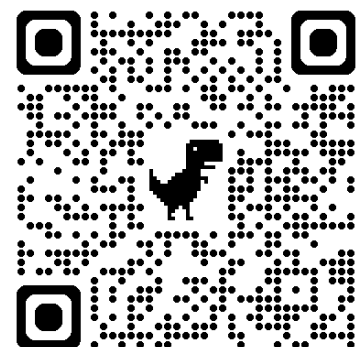
# 11. References

https://www.healthline.com/health/difference-between-type-1-and-type-2-diabetes
https://uvahealth.com/services/diabetes-care/types
https://www.sciencedirect.com/science/article/pii/S1877050920308024

# 12. Source Code

**GitHub Link: https://github.com/Hariom-Jaiswal/Predict-Lifestyle-Diabetes-Risk.git**

*Scan For Source Code*