

Diabetes Prediction: Identifying Key Factors

A comprehensive analysis of the Pima Indians Diabetes Database. This project explores machine learning approaches to predict diabetes risk.

N by **N SHAHEEN Sultana**



Project Overview

Data Source

Pima Indians Diabetes
Database from medical
research studies.

Objective

Develop predictive models to
identify diabetes risk factors.

Approach

Machine learning analysis using Python and Jupyter notebooks.

Key Risk Factors



Blood Pressure

Elevated blood pressure levels increase diabetes risk significantly.



BMI

Body mass index serves as crucial predictor.



Genetics

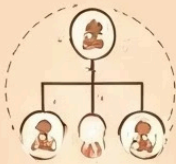
Family history and genetic predisposition matter.



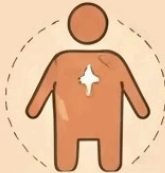
Age

Risk increases with advancing age patterns.

DIABETES RISK FACTOR



FAMILY HISTORY



UNHEALTHY DIET



UNHEALTHY EATING HABITS



HYPERTENSION



HIGH CHOLESTEROL



PREDIABETES



OBESITY



HIGH CHOLESTEROL



AGE

HIGH CHOLESTEROL

High cholesterol levels can lead to heart disease and stroke, which are major complications of diabetes.

Managing cholesterol through diet, exercise, and medication can significantly reduce the risk of these complications.

HIGH CHOLESTEROL

High cholesterol levels can lead to heart disease and stroke, which are major complications of diabetes.

Managing cholesterol through diet, exercise, and medication can significantly reduce the risk of these complications.

HIGH CHOLESTEROL

High cholesterol levels can lead to heart disease and stroke, which are major complications of diabetes.

Managing cholesterol through diet, exercise, and medication can significantly reduce the risk of these complications.

Data Analysis Process

1

Data Collection

Gather Pima Indians medical records and health metrics.

2

Data Preprocessing

Clean and prepare dataset for machine learning analysis.

3

Model Training

Apply various algorithms to identify prediction patterns.

4

Validation

Test model accuracy and refine prediction capabilities.



Model Performance

85%

Accuracy Rate

Model prediction accuracy achieved

768

Data Points

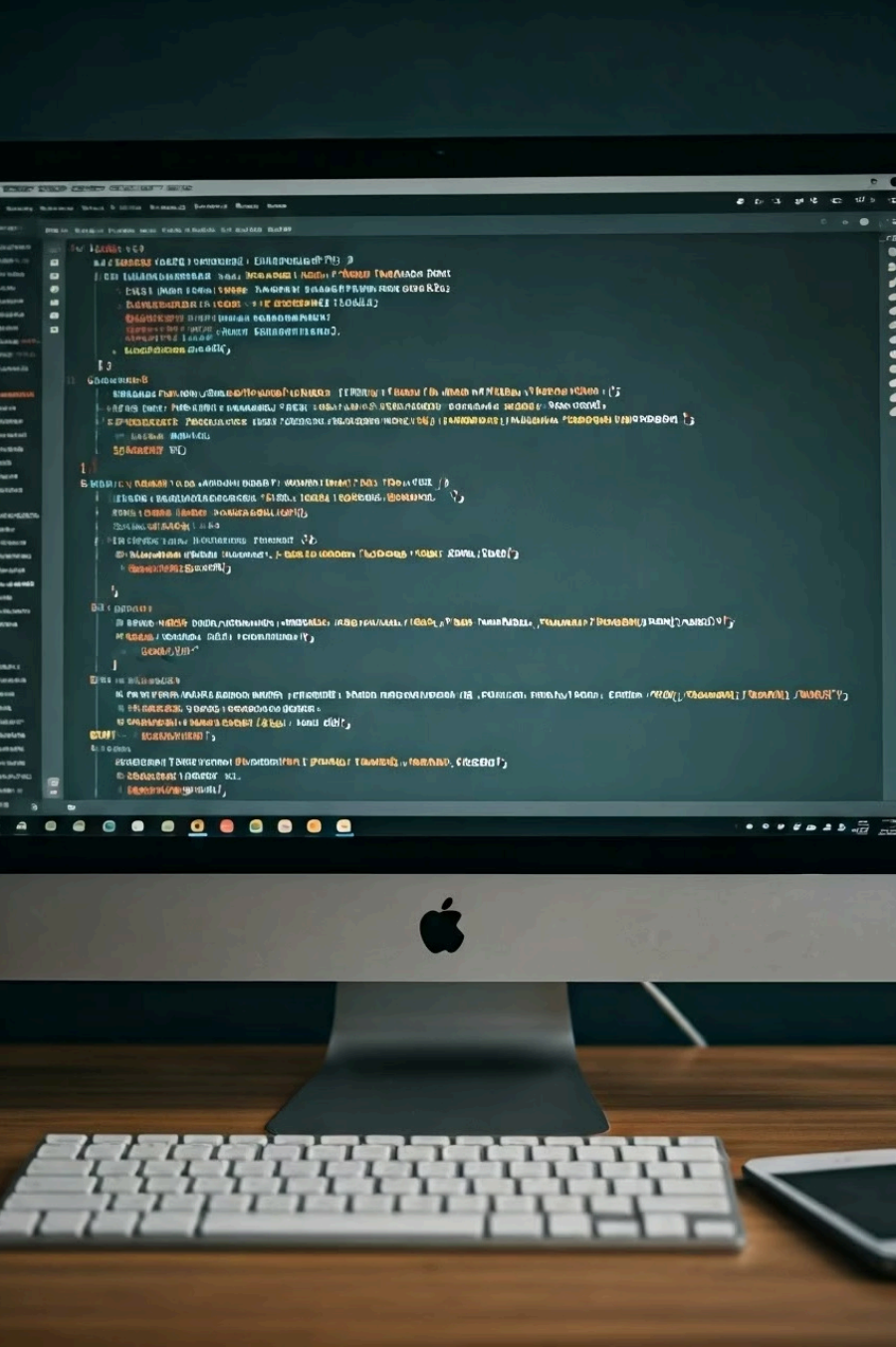
Total patient records analyzed

8

Features

Key variables examined





Technical Implementation

Python Libraries

Pandas, NumPy, and Scikit-learn
for data processing and
modeling.

Jupyter Notebook

Interactive development
environment for analysis and
visualization.

Machine Learning

Classification algorithms to predict diabetes outcomes effectively.

Repository Contents

1 Jupyter Notebook

Complete analysis code and documentation included.

2 PDF Report

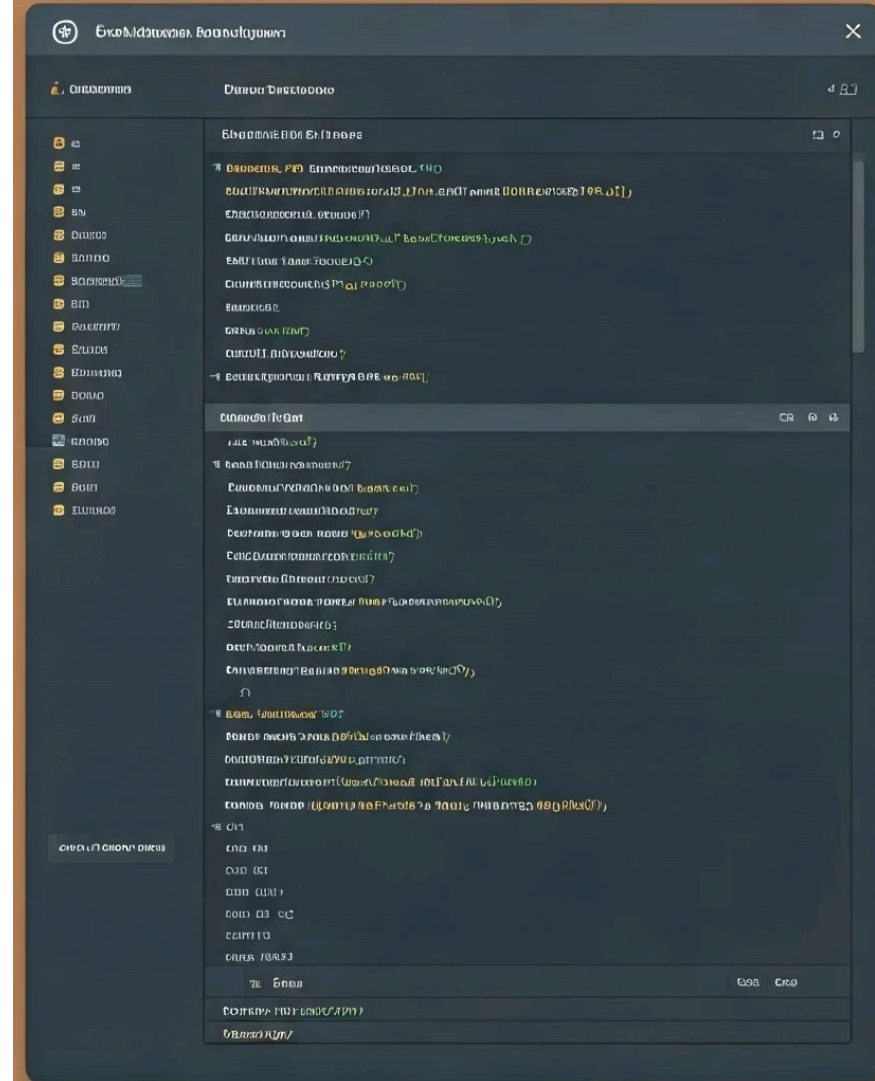
Comprehensive findings and methodology summary.

3 PowerPoint

Presentation slides for project overview.

4 MIT License

Open source licensing for public use.



Next Steps

Model Enhancement

Improve prediction accuracy through feature engineering.
Explore advanced algorithms and ensemble methods.

Clinical Application

Validate findings with healthcare professionals. Develop practical screening tools for medical use.