Diabetes Prediction Analysis Report

I. Introduction

This report analyzes a diabetes prediction dataset containing patient health information, including age, gender, BMI, hypertension, HbA1c, blood glucose levels, heart disease, and smoking history. The objective is to explore data patterns, visualize key health indicators, and conduct hypothesis testing to determine statistical relationships related to diabetes status.

II. Methodology

The dataset was loaded using Pandas and explored through descriptive statistics, missing value analysis, and correlation checks. Seaborn and Matplotlib were used for visualization. Statistical hypothesis testing was conducted using Z-Test, T-Test, and Chi-Square test to evaluate relationships.

III. Results

Key findings from the dataset:

• Average BMI, Blood Glucose, and HbA1c levels are higher among diabetic patients compared to non-diabetic patients.

• Group means indicate diabetic patients tend to have higher risk factors overall.

• Z-Test revealed whether the population BMI significantly differs from 25.

• T-Test showed if Age distributions differ significantly between diabetic-positive and negative groups.

• Chi-Square test examined the association between Smoking History and Diabetes Status.

IV. Discussion

The analysis suggests that key indicators such as BMI, Glucose, and HbA1c play a significant role in predicting diabetes. The hypothesis testing provided statistical evidence to support differences between diabetic and non-diabetic groups. The correlation between glucose and HbA1c highlights medical relevance, while smoking history shows a potential association with diabetes.

V. Conclusion

In conclusion, this study highlights the importance of health indicators in predicting diabetes risk. Healthcare professionals can use these insights to design preventive strategies, focusing on patients with elevated BMI, glucose, and HbA1c levels. Statistical tests confirm significant differences between groups, reinforcing the predictive value of these indicators.