

DIABETES PREDICTION

Step 1: Problem Definition

The objective of this project is to build a predictive model that can assess an individual's risk of developing diabetes based on various health-related features such as heart health, BMI, hypertension, and lifestyle factors. This could be a valuable tool for early intervention and preventive healthcare.

Step 2: Data Collection

Got dataset online from kaggle.

Step 3: Data Preprocessing

1. Checked for nulls
2. Removed Duplicates
3. Standardize numerical features
4. Did Encoding of categorical feature such as bmi and smoking history

Step 4: Exploratory Data Analysis (EDA)

1. Plotted bar chart for Distribution of Diabetes
2. Plotted bar chart of categorical features with distribution of diabetes
3. Plotted distribution of numerical features

Step 5: Data Modeling

1. Logistic Regression
2. Decision Tree
3. Random Forest

Step 6: Model Evaluation and Conclusion

The Random Forest model is the best because it predicts accurately 96% of the time while accuracy for Decision Tree and Logistic Regression was 94% and 88% respectively. It's also good at finding both people who have the condition and those who don't. This means it's balanced and does a great job overall.