**PREDICTIVE ANALYTICS PACKAGE**

* AHALYANJUNA S(22PD02)
* V S RITANYA(22PD38)
* VALINCIA JOHN(22PD39)

**ABSTRACT**

This project focuses on predictive modeling for diabetes using health indicators sourced from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) 2015 survey. Three datasets are analyzed:

1. Diabetes\_012\_health\_indicators\_BRFSS2015.csv: This dataset encompasses 253,680 survey responses and includes three target classes: no diabetes, prediabetes, and diabetes. Addressing class imbalance is a key consideration in this dataset.

2.Diabetes\_binary\_5050split\_health\_indicators\_BRFSS2015.csv: Comprising 70,692 survey responses, this dataset achieves a balanced 50-50 split between respondents with no diabetes and those with prediabetes or diabetes.

3. Diabetes\_binary\_health\_indicators\_BRFSS2015.csv: Similar to the first dataset, this one also contains 253,680 survey responses but with a binary target variable representing no diabetes versus prediabetes or diabetes. It is imbalanced.

**FEATURES**

The project employs diverse machine learning techniques:

* **Data Preprocessing**:

Null values are removed to ensure data integrity. The imbalanced datasets are balanced using SMOTE technique.

* **Modeling Approach:**

For Data 1: Multinomial Logistic Regression , one-vs-rest logistic regression and decision tree models are used to predict.

For Data 2 and Data 3 : Logistic regression model is used.

* **Visualization:**

Utilizing Streamlit for the frontend, the project offers interactive sliders, select boxes, checkboxes, and visualizations to input health metrics and view predictions and analyses generated by the machine learning models.

* **Insights**:

Correlation analysis, bar charts, pie charts, stacked column charts and box plots offer insights into the distribution of diabetes across various feature variables.

* **Libraries used:**

Pandas, scikit-learn, matplotlib, seaborn, and Streamlit for data processing, modeling, and visualization

**CONCLUSION**

Ultimately, this project aims to predict the probability of diabetes development based on health indicators and provide actionable insights for mitigating the risk through lifestyle adjustments and healthcare interventions.