Project Title

Bank Risk Controller Systems

Skills

Python, Analytics, statistics, Ploting, streamlit ,

machine learning

Domain

Banking

Problem Statement:

Please refer the given data and do the prediction of whether the customer will

get default or not. The column which you are going to predict in dataset is

“TARGET”.

Approach:

Methodology

**Data Collection**: Gather historical loan data from various sources.

**Data Preprocessing:** Clean and preprocess the data to handle missing values,

outliers, and categorical variables.

**Exploratory Data Analysis (EDA):** Perform EDA to understand data

distributions and relationships between variables.

**Feature Engineering**: Create new features that could enhance the predictive

power of the model.

**Model Selection:** Compare various machine learning models such as Logistic

Regression, Decision Trees, Random Forest, and Gradient Boosting, etc. to

find the best fit.

**Model Training:** Train the selected model(s) on the training dataset.

Model Evaluation: Evaluate the model performance using appropriate metrics.

**Hyperparameter Tuning**: Optimize the model parameters to improve

performance.

**Model Deployment**: Deploy the model for real-time prediction and integrate it

with business systems.

Steps followed:

EDA-Exploratory Data Analysis:

Plotted various plots to bring the insights and outliers in the real time data.

Model Selection:

Logistic Regression and Random Forest Algorithms are tried for the best fit.

Finally, Random forest algorithm gave the best fit.

Model Training:

Model is trained and tested for the prediction.

Model Deployment:

Model is deployed in the streamlit as per the Project document.

Pickle:

Model is pickled and saved.

Result:

Bank Defaulters are predicted using the Machine Learning Algorithm Random Forest and built it as a streamlit application for the real time business usage.