Credit Score Prediction Model Documentation

Introduction:

This document serves as documentation for a machine learning model developed to predict credit scores based on various features provided by individuals. The model utilizes a dataset containing information about individuals' financial attributes, such as age, income, loan history, payment behavior, and occupation, among others. The goal of the model is to assist in assessing the creditworthiness of individuals, which is crucial for financial institutions when making lending decisions.

Model Development:

The model development process involves several steps, including data preprocessing, algorithm selection, model training, evaluation, and deployment.

1. Data Preprocessing:

Data Cleaning: The initial dataset underwent cleaning procedures to handle missing values, outliers, and invalid entries. This included removing irrelevant columns, dropping rows with excessive missing data, and converting inconsistent data types.

Feature Engineering: Certain features were engineered or transformed to make them suitable for modeling. For example, categorical variables were encoded using label encoding or one-hot encoding, and numeric features were standardized to have a mean of 0 and a standard deviation of 1.

2. Algorithm Selection:

- Several machine learning algorithms were considered for credit score prediction, including Random Forest, Logistic Regression, Gradient Boosting and LSTM. These algorithms were chosen for their effectiveness in handling both numerical and categorical data, as well as their ability to handle classification tasks.

3. Model Training and Evaluation:

Random Forest Classifier: The Random Forest algorithm was trained on the preprocessed dataset. This ensemble learning method constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of individual trees.

Evaluation Metrics: The model's performance was evaluated using various classification metrics such as accuracy, precision, recall, and the F1-score. Additionally, confusion matrices were generated to visualize the model's predictions.

4. Model Deployment:

Flask Application: The trained Random Forest model was deployed using a Flask web application. The application provides an endpoint '/predict' where users can send JSON-formatted data containing individual financial attributes. The model processes the input data, predicts the credit score, and returns the prediction as a JSON response.

Conclusion

The developed credit score prediction model offers a reliable and efficient tool for financial institutions to assess the creditworthiness of individuals. By leveraging machine learning algorithms and appropriate data preprocessing techniques, the model provides accurate predictions that can aid in making informed lending decisions. Continuous monitoring and updates to the model can further enhance its performance and adaptability to changing financial landscapes.