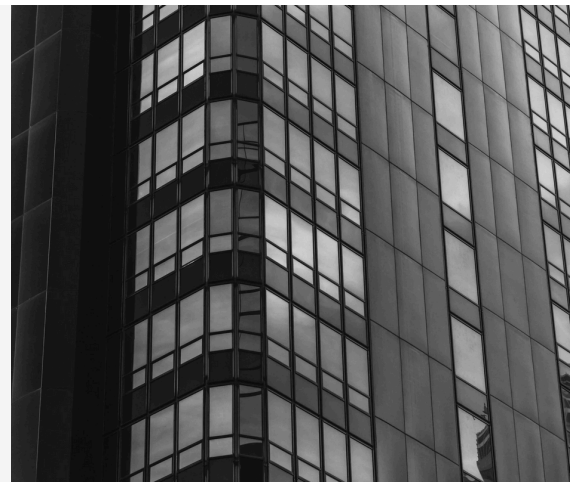


CodeAlpha Internship



Task Title: Credit Scoring model

Task Description: Develop a credit scoring model to predict the creditworthiness of individuals based on historical financial data. Utilize classification algorithms and assess the model's accuracy

Credit Scoring Model

A credit scoring model is a statistical tool used by lenders and financial institutions to evaluate the creditworthiness of a potential borrower. It predicts the likelihood that a borrower will default on a loan or make payments late.

Data Collection

Historical Data: The model is trained on historical data, which includes the credit history of past borrowers, their payment behaviors, loan characteristics, and outcomes.

Demographic Information: This may include age, income, employment status, and other personal details that can influence credit behavior.

Credit Bureau Data: Information from credit bureaus, such as credit scores, outstanding debts, and recent credit inquiries, is also incorporated.

Feature Selection

Financial Indicators: Income, current debt levels, credit utilization ratio, length of credit history, and types of credit used.

Behavioral Indicators: Payment history, number of late payments, defaults, or bankruptcies.

Demographic Features: Age, employment status, residential stability, etc., although some of these may be excluded due to regulatory or ethical considerations.

Model Training

Training: The model is trained with Random Forest Classifier on a subset of the data, learning to identify patterns and relationships between features and outcomes.

Validation: The model's performance is tested on a separate dataset to ensure it generalizes well to unseen data. Metrics like AUC-ROC, precision, recall, and F1-score are commonly used for evaluation.

Model Evaluation

Test the Model: Evaluate the model on a separate test dataset to measure its accuracy and generalization.

Confusion Matrix: Generate a confusion matrix to see how well the model performs on each class.

Accuracy and Loss Plots: Plot the training and validation accuracy/loss over epochs to visualize the model's learning process.