Machine Learning Model for Bank Loan Classification: A Brief Report

Introduction: In this report, I present an overview of my approach to building a machine-learning model to classify whether a personal loan was accepted or not based on a dataset containing 15 columns of information. The goal is to develop a robust model that accurately predicts loan acceptance.

Dataset Overview: The dataset comprises 15 features (columns) that provide information about loan applicants. These features could include attributes such as ID,age, income, Gender, Experience, Home Owenrship,Mortage,Securities Account, CD Account and more. Additionally, there is a target column indicating whether a personal loan was accepted (Personal Loan).

Approach:

1. **Data Preprocessing:** The initial step involved data preprocessing, which included handling missing values, handling duplicate values and null values. This step ensures the dataset is ready for training the machine-learning model.
2. **Feature Selection/Engineering:** Feature selection techniques were applied to identify the most relevant features for the classification task. Additionally, new features might be engineered if they were expected to contribute to the model's performance.
3. **Model Selection:** Several machine-learning algorithms were considered for the classification task, including but not limited to:
    - Logistic Regression
    - Decision Trees
    - Random Forest
    - Extra Tree Classifiers
4. **Model Evaluation:** The trained models were evaluated using metrics such as accuracy, precision. Cross-validation was employed to ensure the generalization performance of the models.

Key Findings and Insights:

1. **Feature Importance:** Through feature selection techniques, we identified the most influential features for loan acceptance prediction. This insight helps in understanding which aspects of an applicant's profile play a crucial role in their loan approval.
2. **Model Performance:** After thorough evaluation, we observed that certain algorithms may perform better than others for this specific task. The choice of the model heavily depends on the dataset's characteristics and the desired trade-off between precision and recall

Conclusion: Developing a machine-learning model for accurately classifying whether a personal loan is accepted involves a multi-step process encompassing data preprocessing, feature selection/engineering, model selection/training, and thorough evaluation. The final choice of the model depends on its performance metrics, interpretability, and the business's specific requirements. Additionally, insights gained from feature importance and model performance can guide decision-making in terms of lending policies and strategies. Continuous monitoring and updates to the model may be necessary to adapt to changing trends and data patterns. The model has been saved on pkl format and it seems like overfitting because of high accuarcy in model.