Report on Bank Loan Classification

- The project began with importing necessary libraries and dataset to work with.
- Exploratory data analysis (EDA) was conducted to identify patterns and relationships between variables. There was strong correlation between Age and Experience, and moderate correlations like Income and CCAvg.
- Corrwith() function showed that there was high correlation of Income (0.44), CCAvg (0.37) and CD account (0.32) with the target column.
- Missing values were identified and according to the significance, the null values were either dropped or imputed. For categorical data, manual encoding was done.
- The categorical values were encoded using most frequent data. The numerical values were imputed using the strategy ‘mean’.
- The dataset was then split into features (X) and target (y) for model training.
- The Random Forest model was chosen for its robustness and ability to handle high-dimensional data. The model evaluation showed strong performance, with an accuracy of 98%, precision of 99%, recall of 82%, and an f1 score of 0.89, indicating its effectiveness in predicting personal loan acceptance.
- Finally, a chatbot was designed to assist users with inquiries about the bank's services, including personal loan applications built using the framework Streamlit.