Bank loan classification

Objective of this project : classify whether the personal loan was accepted or not .

Variables used in this analysis:

- ID: ID of the customer

- Age: Age of the customer
- Gender: M for Male, F for Female and O for Others
- Experience: Amount of work experience in years
- Income: Amount of annual income (in thousands)
- Home Ownership: Home Owner, Rent and Home Mortgage.
- Zip Code: Postal code in which the client lives
- Family: Number of family members
- CCAvg: Average monthly spending with the credit card (in thousands)
- Education: Education level (1: bachelor's degree, 2: master's degree, 3: advanced/professional degree)
- Mortgage: Value of home mortgage, if any (in thousands)
- Securities Account: Does the customer have a securities account with the bank?
- CD Account: Does the customer have a certificate of deposit account (CD) with the
bank?
- Online: Does the customer use the internet banking facilities?
- CreditCard: Does the customer use a credit card issued by the bank?
- Personal Loan: Did this customer accept the personal loan offered in the last campaign?

Task : to build a machine-learning model that can accurately classify whether the personal loan was accepted or not based on the information provided.

How is this project approached?

The dataset is converted from xlsx to csv for better understanding of data. The data is further cleaned , performed Exploratory Data Analysis(EDA) and Feature Engineering.

Missing values were present in the dataset so it is filled with mean.One-hot encoding is done to make string values into categorical values. Some unnecessary columns are dropped from the dataset and after making final clean data , Machine learning models like logistic regression, decision tree , gaussianNB are built from the data. Then those models are saved onto the pickle file so it can be easily loaded and used for prediction.

Key Findings:

1. The customers in the dataset have varying ages, ranging from young adults to potentially older individuals.
2. Customers in the dataset have diverse levels of work experience.
3. The income of the customers varies, indicating different financial backgrounds.
4. Family sizes among the customers differ, suggesting variations in household structures.
5. Customers exhibit different levels of average monthly spending with credit cards.
6. The dataset includes customers with various education levels.
7. Customers have different types of home ownership.
8. The presence of a personal loan indicates that customers accepted the loan offered in the last campaign.
9. The dataset captures information about whether customers have securities accounts, CD accounts, use online banking, and possess credit cards.

Insights:

1. Age, income, education, and home ownership are important factors to consider when analyzing customer behavior and loan acceptance.
2. The availability of personal loan acceptance data allows for potential analysis of factors influencing loan decisions.
3. The presence of securities accounts, CD accounts, online banking usage, and credit card ownership provides insights into customers' banking habits.
4. The distribution and range of values for each attribute indicate the diversity of customers in the dataset.

Thank You.