Initial Proposal
Data Science for Business
Feb/28/2024
Leon Zhang, Ishan Miglani, Keya Keya, Qing Shen

Proposal for Financial Risk Assessing Predictive Modeling

Background

Financial institutions face the inherent risk of customer loan defaults. To mitigate this risk, it's vital to accurately assess a potential borrower's creditworthiness. Traditional credit scoring systems often rely on a customer's established credit history, which might not be readily available for new customers. Moreover, identifying factors that increase the probability of default can be complex. To improve risk assessment and streamline the lending process, a predictive model leveraging a customer's recent account information and transactional data can provide valuable insights. By utilizing historical data, we can train a model capable of predicting future credit scores and default potential for new customers.

Dataset:

- **Title:** Credit Score Dataset (or a more specific name if you create your own)
- **Source:** https://www.kaggle.com/datasets/conorsully1/credit-score
- **Brief Description:** A dataset containing customer information with features related to income, spending, debt, and categorical attributes. Includes targets for credit score and default history.

Business Question of Interest:

- Can we develop accurate models to predict a customer's credit score and their likelihood of defaulting on a loan, using recent account and transactional data?

Type of DS Problem:

- **Credit Score Prediction:** Regression problem (predicting a numerical value)
- **Default Prediction:** Classification problem (predicting a yes/no category)

Features and Target Variables:

- **Features:**
  - **Descriptive Summaries:** Total income, savings, debt, etc.
  - **Transaction History:** Spending in different categories (clothing, education, gambling, etc.)
  - **Categorical Attributes:** Presence of debt, mortgage, dependents, etc.
- **Target Variables:**
  - **Credit Score:** Numerical value
  - **Default History:** Binary value (yes/no)

What do you expect to learn from the analysis and model:

- **Key Factors:** Which customer features (income, spending patterns, debt, etc.) are most strongly associated with high/low credit scores and likelihood of default.

- **Predictive Power:** How accurately can our models predict both the credit score and the potential for default for new customers.
- **Model Comparison:** Which modeling techniques are most effective for these prediction tasks.

What actions might be taken with a firm by the results:

- **Risk Assessment:** Identify high-risk customers likely to default on loans, enabling more selective lending.
- **Credit Line Decisions:** Determine appropriate credit limits for customers based on their predicted credit score and default risk.
- **Proactive Outreach:** Reach out to customers, likely to default, with financial advice or debt management options.
- **Marketing:** Tailor loan product offers or interest rates based on predicted risk profiles.