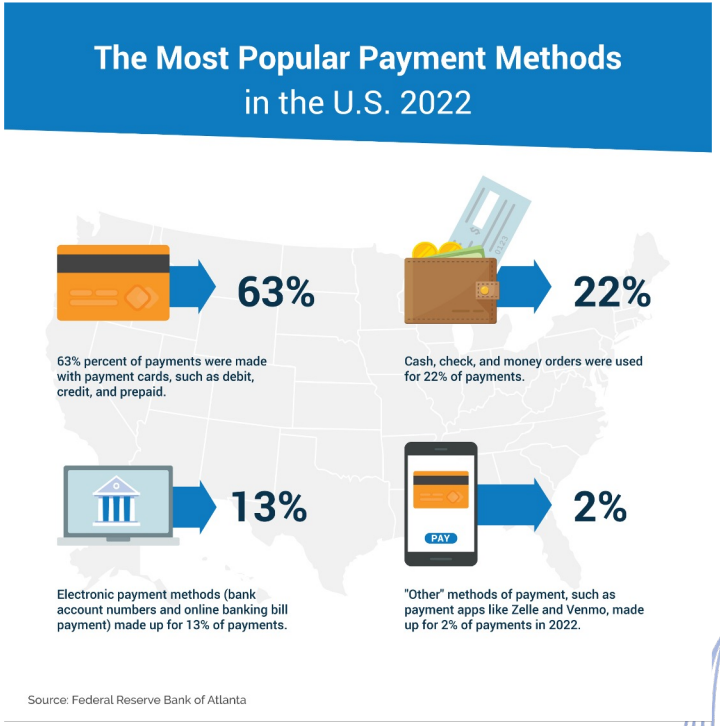
American Express - Default Prediction

Group 24- Sarthak Garg, Yeshwanth Karra, Adithya Mahesh, Tanmay Parulekar

Background

- Fundamental role of credit card in modern day
- Credit card contribute to the convenience of daily purchase
- Credit card reduce the need for carrying large amount of cash
- They offer lucrative rewards
- They facilitate the full amount payment over time

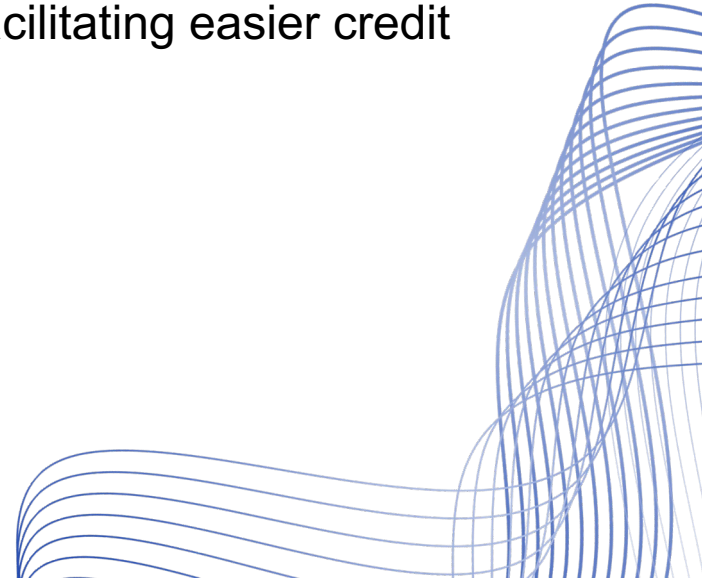


Problem Statement

- Develop a machine learning model to predict credit default
- Improving lending decisions & enhancing the customer experience

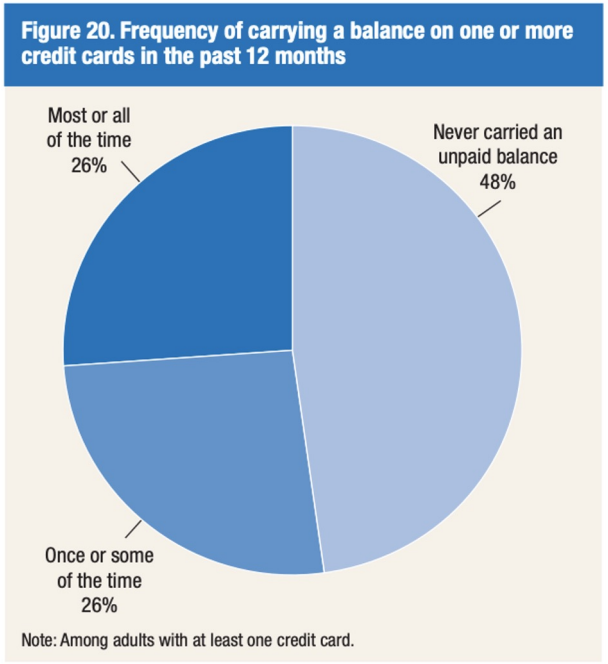
Problem Importance

- Mitigating Financial risks associated with default.
- Optimizing lending decisions for better business economics.
- Improving overall customer experience by facilitating easier credit card approval system for the applicants.
- Reduce Cost and Increase Efficiency



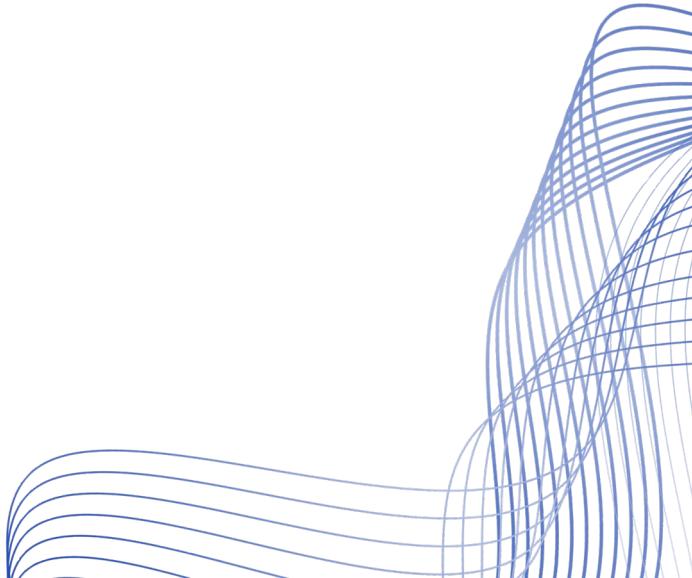
Problem Challenges

- Diverse Customer Profiles
- Evolving Economic Conditions
- Behavioral Data
- Data Anonymization



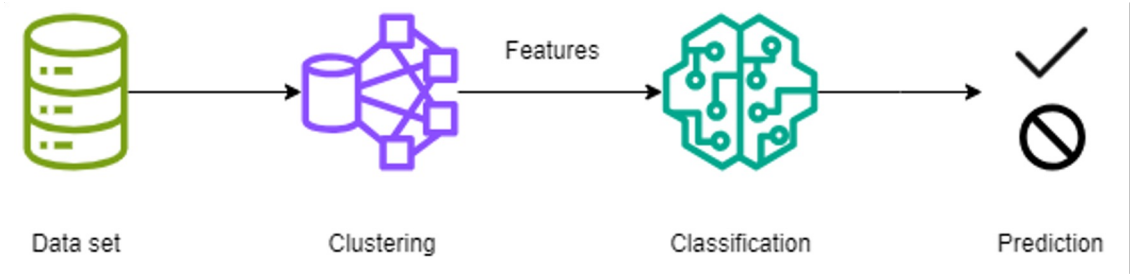
Data Challenges

- Size of the dataset
- Data Schema
- High Dimensionality

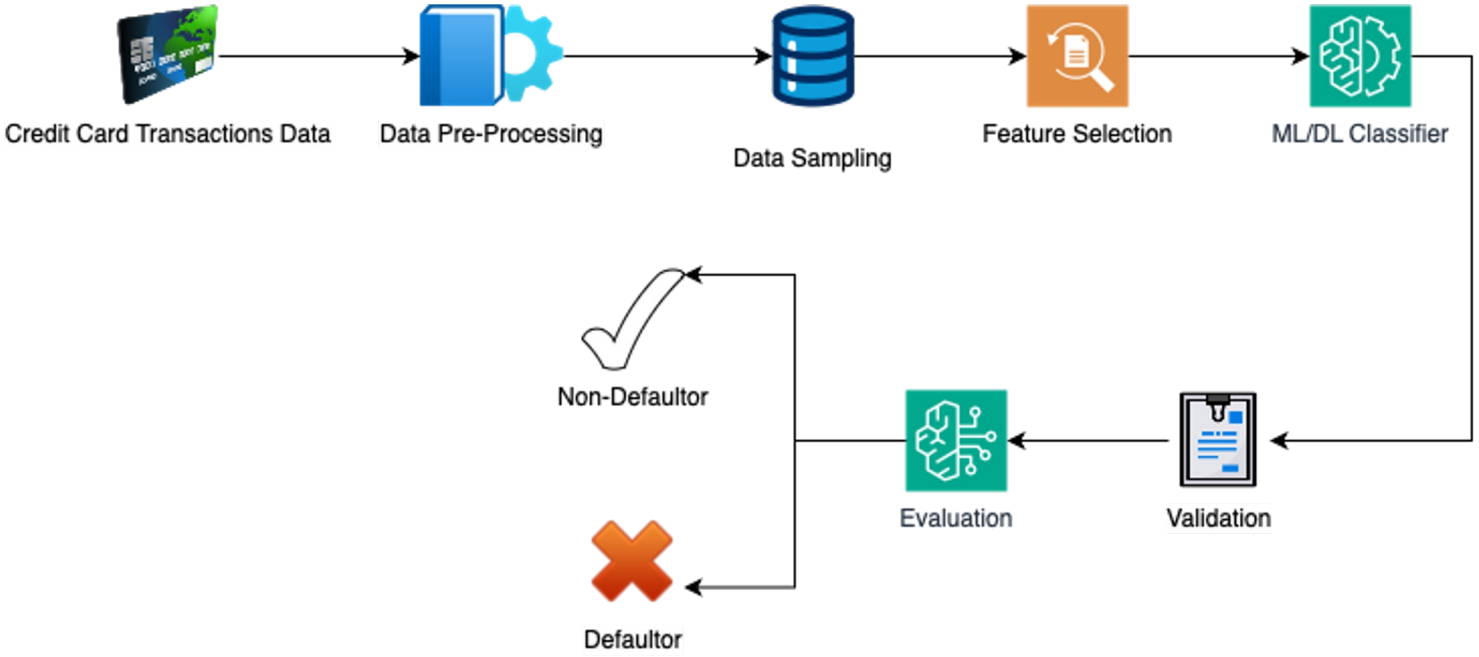


Essential Task

- Clustering
- Classification



Data Mining Pipeline



Initial Exploratory Data Analysis

- The raw data requires preprocessing before it can be fed to a model.
- Data Preprocessing steps:
  - There are multiple entries for a single ‘Customer_ID’ which need to be grouped
  - Remove irrelevant columns
  - Handle missing values in the remaining columns

Initial Exploratory Data Analysis

Null values in each column

customer_ID	0
S_2	0
P_2	1147
D_39	0
B_1	0
...	
D_140	1065
D_142	124272
D_144	1079
D_145	2632
target	0

Train data: (5531451, 190)

Is Data Collection Needed?

Thank You