

Leveraging Economic and Transactional Data for Predictive Analytics and Synthetic Consumer Behavior Modeling

By: Ethan Pollock, Omkar Pardeshi, Ajay Kumar





Dataset

- 1.3 million rows of transaction data
- Dates range from December 31, 2018 to June 21, 2020
- Variables include:
 - Credit card number
 - Merchant/store
 - Transaction category
 - Transaction amount
 - First name, Last name
 - Gender of card holder
 - Address details of cardholder



API Usage

- From FRED Economic Data (Federal Reserve Bank of St Louis)
 - Gross domestic product
 - Real GDP
 - Nominal GDP
 - Consumer price index
 - Inflation
 - Unemployment rate
- Data is directly from the government
 - Specifically the Federal Reserve



Models Selected to Test

- Random Forest
 - Provides modeling for both classification and regression analysis
 - Capable of capturing nonlinear dependencies in spending behavior
- Support Vector Machines
 - Excels in high-precision tasks
 - May not be optimal for large datasets, designed smaller high-dimensional analysis
- Deep Learning Neural Networks
 - Detects nonlinear or sequential patterns that traditional models might miss
 - Can be effective in analyzing time-based spending data



Models Selected to Test (cont'd)

- Gradient Boosted Trees (specifically XGBoost)
 - Performs superior with large tabular datasets
 - Designed to handle nonlinear relationships between spending and economic variables
 - Results in high predictive accuracy
 - F1 scores above 0.92
 - Area Under Receiver Operating Characteristic curve near 0.99
- Clustering and Unsupervised Learning
 - Specifically K-means clustering and Hierarchical clustering
 - Can segment spenders into behavior clusters
 - May reveal patterns across different economic environments



Project Timeline and Progress

- Model Evaluation
 - F1 scores
 - AUC
 - Root Mean Squared Error
- Addition of stock market elements
 - Provides insight into an additional market of public spending
 - Strong and weak stock markets typically support their respective strong and weak economic periods
- Generative AI methods to simulate predictive outcomes with possible variable changes
 - Synthetic data will be generated
 - Will enhance scenario analysis
- Possible map generation for areas with most credit card transactions
 - Could provide insight into best location to open business (category dependent)



Significance of Produced Results

- Consumer credit card debt is worth \$1.2 billion as of Q2 2025
 - This microeconomic amount can compound into macroeconomic impacts
- Ability to charge to credit cards and repay balances is crucial to credit companies business models and success
 - Cash back/points are offered as a reward to good credit users
 - High interest rates help to pay for cost of debt and discourage credit usage
- Debt is not free, so being able to generate credit card usage based on economic variables will help credit banks generate appropriate business plans
 - Generally speaking, the Great Recession is a prime example of what happens when too much debt is defaulted at once