Customer Credit Analysis

Blackwell Electronics / Credit One

UTDA C2T1 Submission - E. Jarrett

Project Introduction

Problem: Credit One customers (debtors) have begun defaulting at higher rates, posing significant risk to the operations/credibility of our credit-scoring business

Goals:

- *Achieve greater accuracy than existing models of classifying ‘at-risk’ customers*
- *Identify demographic/behavioral factors most predictive of future loan default*
- *Determine optimal credit limit to provide new (approved) loan customers*

The BADIR framework helps structure the data science process to derive actionable insights

- **B**usiness Question
  - "Who/What/Where/When/Why?"
  - Identify primary stakeholders and timing urgency
- **A**nalysis Plan
  - Set SMART Goals - Specific, Measurable, Achievable, Relevant, Time Bound
  - Formulate null hypotheses to prove or disprove
  - Methodologies - Aggregate, Correlation, Trend, Estimation
- **D**ata Collection
  - Source file and time period
  - Cleansing, validation, normalization, scaling, etc.
- **I**nsights
  - Quantified impact of hypotheses being proven/disproven
  - Visualizations aid comprehension
- **R**ecommendations - Most Important
  - Tailored to target audience/stakeholders
  - Must be actionable

Applying BA- Process to Credit One

- **B**usiness Questions – Urgent Timeline
  - How well do current models work and why has performance declined only recently?
  - What demographic or behavioral (charge/payment history) factors help identify “at-risk” customers?
  - If approved, what’s the revenue-maximizing credit limit that Credit One should grant a customer?
    - Presumably, most valuable customers are those not likely to fully default, but consistently make late payments and/or carry high monthly balances on revolving basis

- **A**nalysis Plan - Goals & Hypotheses
  - Exploratory Data Analysis
    - Variable relationship & summary statistics between currently defaulted / non-defaulted customers
    - Monthly performance for April - September 2005
    - Investigate other potential patterns as analysis proceeds on an iterative basis
  - ML Classification - Improve accuracy of existing models by +.05 OR achieve at-risk classification accuracy of 0.60 or higher
  - Hypotheses for traits contributing to increased likelihood of default:
    - Being SINGLE (!Prev_Relationship) or ‘less-educated’ (!College_Plus)
    - Customers with steadily increasing debt usage (as monthly % of limit)
    - Customers deviating from their typical payment pattern (no longer pay in full) and/or falling behind
    - Customers with lower LIMIT_BAL (fewer other resources to tap into)
    - There is no statistically significant correlation between age/gender and likelihood of default

Applying -DIR process to Credit One (contd.)

- **D**ata Collection
  - Raw CSV dataset contained 25 fields for over 30K customers, described in supplemental guide
  - Standard prep/cleaning practices already performed – Duplicate, null value, and incompatible data types removed, resulting in final dataset of exactly 30,000 records
  - All fields converted to numeric type (int64 or uint8) for later use in ML models
  - Cleaned dataset re-saved to disk as ‘progress checkpoint’
  - Derived additional fields to aid in analysis:
    - Pandas-standard ‘dummy’ variables created for education, marriage, and default status
    - Debt_Usage[1-6]: Monthly Bill as % of Total Limit for each month
    - Total_Bill: Total amount billed by customer throughout period
    - Total_Paid: Total amount customer paid throughout period
    - Net_Borrow: Total_Bill less Total_Paid
    - Debt_UsageAVG= Six month mean of Debt_Usage[1-6]
    - EvBehind: Boolean if ever paid late during the period
    - AvgPayTime - Mean of customers’ PAY_1:PAY_6 codes (higher values pay later)
- **I**nsights and **R**ecommendations - TBD
  - Initial exploratory analysis of dataset follows

Initial EDA: Defaults are Acute and Growing Worse

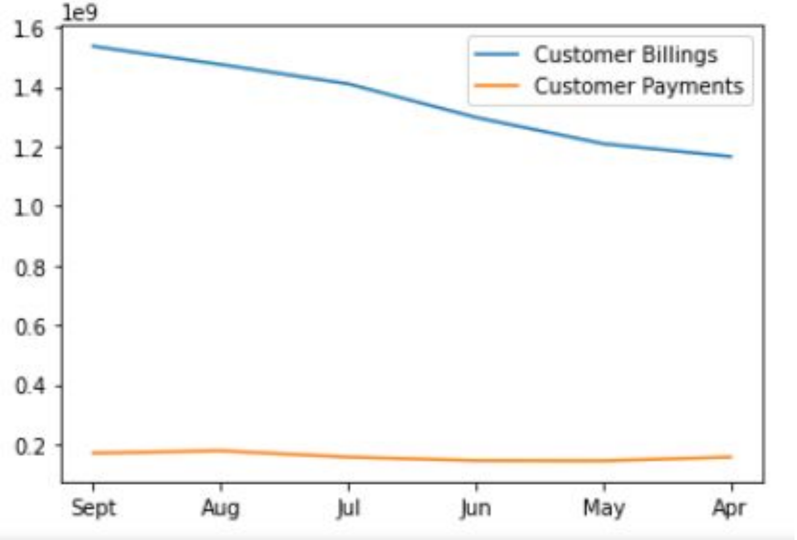
Over 6,000 (22%) of Credit One customers are currently in default

CO's exposure to default risk has steadily increased

- ➢ While customer billings grew ~32% April-Sept, respective payment
- ➢ For September, monthly spending reached 30.5% of the total cred

Customers' own financial challenges are evident

- ➢ Customers' average 'Debt Usage' (Monthly Bill as % of Limit) grew
- ➢ The share of CO customers at least 1 month behind in their payme



Default vs. 'Non-Default' Customers

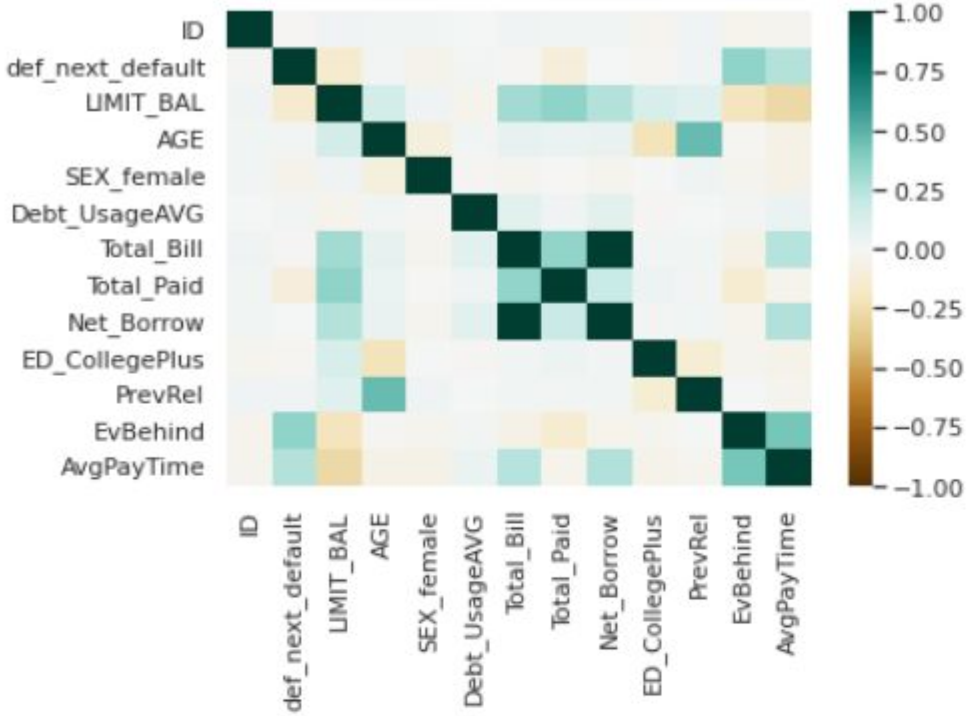
Customers who've defaulted exhibit few demographic differences from those who haven't, but their spending/payment patterns vary much more.

Mean Average	Default = 1	Default = 0
Age	35.7	35.4
SEX_female	.567	.614
ED_PlusCollege	.808	.824
LIMIT_BAL	130,109 NT	178,099 NT
Debt_UsageAVG	.382	.302
EVBehind	.648	.247
AVGPayTime	0.33	-0.33

Correlation Matrix: Default Likelihood & Predictive Features

Customer variables most strongly correlated (absolute value) with future default are:

1. ‘EvBehind’ 0.35
2. ‘AvgPayTime’ 0.25
3. ‘LIMIT_BAL’ -0.15
4. ‘Total_Paid’ -0.10



These variables pass basic ‘reality checks’, with their coefficients matching real world dynamics.

Recommendations

Based on preliminary analysis thus far, analyst suggests management:

1. **Enhance dataset's scope to include other demographic sources** likely useful 'pre-approval', as current Age/Gender/Education/Marriage relate minimally with defaults. Sources reflective of responsible mindsets (or lack thereof), such as employment history or criminal/driving record, may be most useful.
2. Customer billing/payment behavior, such as increased debt usage or delaying payment, appear more relevant for CO's purposes - **Expanding the dataset's time horizon further into the past** may enable trends to be identified sooner.
3. Barring new data improving CO's credit screening, the best strategy for minimizing defaults appears to act quickly toward 'at-risk' customers, starting outreach efforts upon the first missed/late payment.