Policy Optimization for Financial Decision-Making

Done By: Lakshya Sharma **Date: 12/12/2025**

1. Exploratory Data Analysis (EDA)

A 200,000■row sample of the LendingClub dataset was analyzed due to its large size. Key insights:

- **Class imbalance:** Defaults are rare compared to fully paid loans.
- **Feature behavior:**
  - Loan Amount: Right■skewed, typical range \$5k–\$20k.
  - Interest Rate: Higher rates correlate strongly with default.
  - Annual Income: Highly skewed and weakly predictive alone.
  - DTI: Moderately correlated with default risk.

- **Categorical patterns:**
  - Grades/Subgrades show sharp increases in default from A→G.
  - Purposes like small_business and medical show higher risk.

- **Correlations:**
  - FICO strongly tied to repayment behavior.
  - Interest rate reflects risk pricing and correlates with default.

Overall, EDA shows the problem is imbalanced and influenced by multiple interacting financial features.

2. Predictive Deep Learning Model

A multilayer perceptron (MLP) classifier was trained with weighted binary cross■entropy to handle class imbalance.

Performance Metrics:
- AUC: **0.7065**
- F1-Score: **0.3925**
- Accuracy: **0.5975**

Interpretation:
- AUC ≈ 0.70 indicates moderate risk ranking capability.
- F1-score shows meaningful detection of default cases after rebalancing.
- Accuracy drops when the model stops predicting only the majority class, which is expected.

3. Offline Reinforcement Learning Agent

An RL environment was designed with:
- State = preprocessed borrower features
- Action = {0: deny, 1: approve}
- Reward = financial outcome from the approval decision

Due to environment constraints, a fallback threshold■based policy using the DL model’s probabilities was evaluated.

Estimated Policy Value (EPV): +275.6698

A positive EPV means the policy is expected to produce profit per applicant.

4. Analysis, Comparison, and Future Steps

4.1 Presenting the Results

The supervised model shows solid predictive performance but does not optimize profit. The RL evaluation shows a policy capable of generating positive financial value.

4.2 Why These Metrics?

AUC & F1 for DL:

- AUC captures ranking power across thresholds—ideal for imbalanced classification.
- F1 reflects the model's ability to detect rare defaults.

EPV for RL:

- RL’s goal is maximizing expected financial gain, not classification accuracy.
- EPV directly measures profitability, the true business objective.

4.3 Policy Comparison

DL Model Policy: Approve if predicted default probability < threshold.
RL Policy: Approve if expected reward > 0, considering profit vs. loss.

Example of Different Decisions:

A borrower may be high■risk numerically (DL denies), but RL may approve if the interest■earned reward outweighs the expected loss. RL evaluates value, not accuracy.

4.4 Future Steps

Model Enhancements:

- Train full offline RL (CQL/IQL) in a stable environment.
- Improve DL via focal loss, resampling, or boosting models.

Data Improvements:

- Add repayment histories, recovery rates, bureau scores, income stability.

Deployment Considerations:

- Tune thresholds for profit, not accuracy.
- Conduct fairness and bias evaluations.
- Validate RL policy on larger samples.

Conclusion:

DL provides reliable risk ranking, while RL optimizes business value directly. A hybrid approach combining both would offer stronger real■world performance.