Project Report: Policy Optimization for Financial Decision-Making

BY Pushkar Jain: (102215170)

---

1. Introduction

The goal is to compare:

1. A **Deep Learning (DL) model** that predicts default probability.
2. An **Offline RL agent** that makes profit-aware approval decisions based on expected rewards.

---

2. Data Preparation and Problem Setup

The dataset used represents real-world lending scenarios, containing information such as **loan amount, interest rate, FICO score, DTI ratio, employment length,** and other applicant features.

Feature Engineering

- All numeric features were normalized.
- Categorical features were one-hot encoded.
- The target variable:
  - **1 → Defaulted (Bad Loan)**
  - **0 → Fully Paid (Good Loan)**

The dataset was divided into **training (80%)** and **testing (20%)** subsets.

---

3. Supervised Deep Learning Model

Model Architecture

A **Multi-Layer Perceptron (MLP)** was implemented in PyTorch with the following structure: - Hidden layers: [256, 128, 64, 32] - Activation: ReLU - Dropout: 0.3 - Loss: Weighted Binary Cross-Entropy (to handle class imbalance)

Training and Evaluation

The model was trained for 15 epochs with a learning rate of 5e-4 and a batch size of 256.

Performance Metrics:

AUC : **0.6938**

F1 (0.5 threshold) : **0.4127**

Best F1 : **0.4158 at threshold 0.58**

Interpretation

- **AUC (Area Under ROC Curve):** Measures how well the model separates defaulters from non-defaulters across thresholds.

- **F1-Score:** Balances precision (avoiding false approvals) and recall (approving genuine good applicants).

These metrics evaluate **classification accuracy**, not business profit.
While the model can identify risk well, it doesn't optimize financial returns — which is where the RL approach improves the decision-making process.

---

4. Offline RL Agent: Reward–Greedy Policy

The RL agent was designed to **learn a policy** that maximizes **expected profit per loan approval** rather than just minimizing classification error.

Reward Function

A reward was defined for each decision (approve or deny):

$$\text{reward} = \begin{cases} +\text{loan\_amount} \times \text{interest\_rate}, & \text{if approved and fully paid} \\ -\text{loan\_amount}, & \text{if approved and defaulted} \\ 0, & \text{if denied} \end{cases}$$

This explicitly models the business objective: maximize net gain per applicant.

Offline Dataset Construction

Each applicant's state (financial profile) was paired with: - Action: **approve (1)** or **deny (0)** - Observed reward - Terminal flag (single-step episode)

The agent was then evaluated using a **reward-greedy policy**, which approves only if expected reward > 0.

---

5. Comparative Results

Metric	Supervised (DL)	RL (Greedy Policy)
Expected Reward / Applicant (EPV)	−394.67	**+7.91**

Metric	Supervised (DL)	RL (Greedy Policy)
Total Reward on Test Set	−5.31M	**+106,394.20**
Approval Rate	68.05%	**2.15%**
Avg Observed Reward per Approved Loan	−579.95	**+600.27**

Interpretation

- The **supervised model** approves a large number of loans, many of which result in losses.
- The **RL policy** is conservative — approving only the most profitable loans — resulting in fewer approvals but positive total profit.

This demonstrates that **accuracy-based metrics (AUC/F1)** don't guarantee profit optimization, while **policy value metrics (EPV)** directly reflect financial performance.

---

6. Example of Policy Differences

Index	Greedy Action	Supervised Action	Prob(Default)	Loan Amt	Int Rate	True Default	Exp Reward	Obs Reward
5	0 (deny)	1 (approve)	0.547	13,000	0.0976	1	−6545.70	−13,000.00
8	0 (deny)	1 (approve)	0.422	10,000	0.0797	0	−3764.52	+797.00

Analysis:

- The **supervised model** approved these applicants because their predicted default probability was moderate.

- The **RL policy denied them**, realizing that even with low default risk, the **expected reward** (factoring potential loss) was negative.

- This highlights how the RL framework **aligns decision-making with economic objectives**, not just statistical risk.

---

7. Metric Comparison: Why Different Metrics Matter

AUC / F1 for DL

- Measure the model's ability to separate classes and balance recall vs. precision.
- Useful for **risk modeling** but **not sufficient for financial optimization**.

Estimated Policy Value (EPV) for RL

- Measures **average expected profit per decision**, integrating both reward and action probability.
- Represents **the true business metric** — how much the policy earns or loses on average per applicant.

In short: > AUC tells us *how well* we predict defaults.
> EPV tells us *how much* money we make when applying those predictions.

---

8. Visual Insights

- **ROC Curve (AUC = 0.694)** shows moderate discrimination between good and bad borrowers.
- **Policy EPV Plot** clearly demonstrates that the reward-greedy policy yields the highest net return per applicant.
  Confusion Matrix confirms that the RL policy makes fewer approvals but drastically reduces financial loss.

---

9. Future Steps and Recommendations

a. Next Steps

1. **Combine DL + RL:** Use the supervised model for default prediction, and the RL policy to approve only those loans with positive expected reward.
2. **Retrain RL Agent (CQL / TD3+BC):** Once version compatibility issues are resolved, train a full offline RL agent using `d3rlpy`.
3. **Expand Reward Modeling:** Incorporate repayment timelines, recovery rates, and transaction fees for more realistic profit estimates.

b. Limitations

- RL model currently relies on a **greedy policy approximation** (due to d3rlpy API limitations).
- The reward formulation is simplified and ignores **temporal dynamics** and **regulatory constraints**.
  No online feedback loop yet — the policy cannot adapt to changing borrower behaviors.

c. Data and Algorithm Improvements

- Collect more granular data (e.g., payment histories, income-to-loan ratio, credit utilization trends).
- Experiment with **offline policy gradient methods** or **conservative Q-learning (CQL)** for smoother decision boundaries.
- Explore **fairness-aware RL**, ensuring approvals remain equitable across demographic group

End of Report

Prepared by Pushkar Jain