Final Report — Loan Approval Optimization Using Deep Learning and Offline Reinforcement Learning

Author: Jashandeep Singh | email id:jsingh13_be22@thapar.edu

Dataset: LendingClub Loan Dataset

Objective: Compare Supervised Deep Learning vs. Offline Reinforcement Learning for loan approval policy optimization.

1. Introduction

This project aims to develop and compare two intelligent models for optimizing loan approval decisions — a Deep Learning (DL) model for default prediction and an Offline Reinforcement Learning (RL) model for profit-maximizing policy learning. The motivation stems from a fintech scenario where maximizing profit and minimizing loan defaults must be balanced using historical data.

The overall workflow was divided into four major tasks:
1. Data Exploration and Preprocessing
2. Deep Learning Model (Default Risk Prediction)
3. Offline Reinforcement Learning (CQL)
4. Comparative Analysis, Interpretation, and Future Directions

This report focuses on Task 4, providing a complete analysis, metric comparison, policy interpretation, and recommendations for future work.

2. Presenting the Results

2.1 Deep Learning Model (Supervised Approach)

The DL model used a Multilayer Perceptron (PyTorch MLP) trained to predict default risk (0 = Fully Paid, 1 = Defaulted). After preprocessing (imputation, one-hot encoding, and scaling), the model achieved the following metrics:

Metric	Score
AUC	0.9283
F1-Score	0.7472

These metrics indicate a strong ability to distinguish between risky and safe applicants and a balanced precision-recall performance. The model implicitly defines a policy: 'Approve the loan if predicted default probability < threshold.'

2.2 Reinforcement Learning Model (Decision Policy Approach)

The RL model was trained using Conservative Q-Learning (CQL) via the d3rlpy library. The setup framed loan approval as a bandit-style MDP with two possible actions: {deny, approve}. Rewards were based on profit:
- Approved and repaid: positive reward (interest earned)

- Approved and defaulted: negative reward (loan loss)
- Denied: neutral (zero reward)

The trained RL agent was evaluated using Estimated Policy Value (EPV), which quantifies expected return under the learned policy.

3. Explaining the Metrics

Why AUC and F1 for the DL Model:

AUC measures the model's ranking ability — how well it separates defaulters from payers across all thresholds. F1-Score balances precision and recall, making it ideal for imbalanced datasets like credit defaults.

Why Estimated Policy Value (EPV) for RL:

EPV quantifies the expected cumulative reward achieved by the policy on historical data. It focuses on decision quality in terms of profit, not just prediction accuracy.

4. Policy Comparison and Insights

Both models define a loan approval policy but in different ways:
- DL model: Approves or rejects based on predicted default probability.
- RL agent: Learns to approve or deny based on expected long-term reward.

For example, a high-risk but high-interest applicant may be rejected by DL due to high default probability, while RL may approve due to positive expected profit.

Illustrative Diagram: Policy Decision Difference

The diagram below conceptually shows how the DL model and RL agent make different decisions for applicants across the risk spectrum.

Applicant Risk Level	DL Decision	RL Decision
Low Risk	Approve	Approve
Moderate Risk	Reject	Approve (High Reward Potential)

5. Future Steps and Recommendations

1. Deploy the RL model in a controlled environment for real-world feedback.
2. Combine DL risk prediction with RL decision-making (hybrid model).
3. Add customer lifetime value (CLV) and fairness metrics.
4. Explore model-based RL and Fitted Q Evaluation (FQE).
5. Collect richer datasets including behavioral and credit bureau data.

6. Conclusion

This project demonstrates a complete pipeline for loan approval optimization using Deep Learning and Offline Reinforcement Learning. The DL model provides accuracy-driven risk estimation, while the RL model optimizes for long-term profit. Together, they form the foundation for intelligent, data-driven lending strategies.

Final Metrics Summary

Model	Metric	Score
Deep Learning (MLP)	AUC	0.9283
Deep Learning (MLP)	F1-Score	0.7472
Reinforcement Learning (CQL)	Estimated Policy Value	High (Profit-Optimized)