Final Report — Policy Optimization for Loan Approvals Using Deep and Reinforcement Learning

Author: Anjan Gupta

Dataset: LendingClub Loan Dataset

1. Introduction

This study develops and contrasts two intelligent models aimed at improving loan approval decisions. The first model—a supervised Deep Learning (DL) classifier—predicts the likelihood of loan default. The second—an Offline Reinforcement Learning (RL) agent—learns a policy that directly maximizes financial return. The work simulates a real fintech context, where institutions must balance risk management and profitability by leveraging historical data. The complete workflow consisted of four phases: 1. Data analysis and preprocessing 2. Deep learning model for default prediction 3. Offline RL model for policy optimization 4. Evaluation, interpretation, and proposed future work. This report focuses primarily on the analysis and comparative evaluation of these two approaches.

2. Experimental Results

The experimental study compares both the supervised and reinforcement learning approaches using the LendingClub dataset. Metrics, training configurations, and evaluation methods are discussed below.

2.1 Deep Learning Model (Supervised Framework)

A multilayer perceptron (MLP) built using PyTorch was trained to estimate the probability that a borrower will default (1) or fully repay (0). After preprocessing (imputation, encoding, and scaling), the model achieved strong predictive performance.

Metric	Value
AUC	0.9283
F1-Score	0.7472

2.2 Offline Reinforcement Learning Model (Decision Policy Framework)

The reinforcement learning agent was trained using Conservative Q-Learning (CQL) implemented in d3rlpy. The decision setup used two possible actions: deny (0) or approve (1). Rewards were based on loan profitability: approved and repaid loans yield positive rewards, while defaults incur negative rewards. Denied loans have zero reward. Policy performance was assessed using Estimated Policy Value (EPV), reflecting expected return under the learned policy.

3. Understanding the Metrics

• AUC and F1 for the supervised model: AUC measures ranking quality; F1 balances precision and recall for imbalanced datasets. • EPV for RL: EPV measures the expected financial return of the decision policy, aligning directly with business objectives.

4. Comparative Insights

Both models generate actionable decisions but optimize different objectives. The DL classifier minimizes prediction error, while the RL agent maximizes expected financial return. For instance, an RL model may approve a moderately risky but high-interest loan if the expected profit outweighs the risk, whereas the DL classifier may reject it due to higher default probability.

5. Future Work and Recommendations

1. Evaluate RL model performance in a real or simulated deployment. 2. Develop hybrid DL–RL frameworks combining risk prediction and reward optimization. 3. Include fairness, regulatory, and customer lifetime value metrics. 4. Explore Fitted Q Evaluation (FQE) and model-based RL approaches. 5. Augment datasets with behavioral and bureau data for richer modeling.

6. Conclusion

This project demonstrates an end-to-end approach to optimizing loan approval processes using Deep Learning and Offline Reinforcement Learning. The DL model offers accurate risk estimation, while the RL policy directly targets profit maximization. Together, these approaches lay the groundwork for data-driven, intelligent, and financially sound lending strategies.

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