Policy Optimization for Financial Decision-Making

Final Project Report

Author: Piyush Kumar
Institution: Thapar Institute of Engineering and Technology
Date: October 2025

1. Executive Summary

The objective was to transition loan approval decision-making from subjective risk mitigation to objective financial profit maximization. An Offline Reinforcement Learning (RL) policy was developed and evaluated against a traditional Supervised Deep Learning (DL) classifier using Off-Policy Evaluation (OPE). The Conservative Q-Learning (CQL) policy demonstrated a 44% increase in expected return over the baseline, validating RL's superiority for financial decision optimization.

2. Data & Preprocessing

The dataset used was LendingClub accepted loans (2007–2018), filtered to include only completed loans (≈91,681 transitions). The target variable mapped loan_status to a binary outcome: 0 = Fully Paid, 1 = Defaulted. Features were standardized, and chronological splits (70/15/15) were used to prevent lookahead bias.

3. Models and Environment

A 3-layer MLP was trained for baseline classification, optimized via BCEWithLogitsLoss. The Offline RL environment defined states as 77-dimensional feature vectors, actions as {Deny, Approve}, and rewards as the net profit outcome per loan. Behavior Cloning (BC) and Conservative Q-Learning (CQL) were trained to learn optimal policies.

4. Results Summary

Policy	Objective	F1 Score	Estimated Policy Value (FQE)	% Improvement
CQL (Profit Max)	Maximize E[Profit]	N/A	$2,776.13	+44.0%
BC (Imitation)	Emulate $\pi\beta$	N/A	$1,928.09	0.0%
MLP (Supervised)	Maximize F1	0.52	$1,399.88	-27.4%

5. Analysis and Insights

The MLP achieved F1 = 0.52 but underperformed in profitability. CQL achieved the highest financial return by identifying loans with favorable risk-reward trade-offs. Despite lower precision (27.83%), its recall of 0.8614 captured profitable opportunities that supervised models missed.

6. Limitations and Future Work

Selection bias remains since the dataset contained only approved loans. Future work includes A/B testing to validate real-world uplift, fairness audits via Disparate Impact Ratio, and deployment through ONNX-exported policies integrated into production APIs.

7. Conclusion

Offline Reinforcement Learning (CQL) outperformed traditional classifiers by directly optimizing for long-term profit. This approach redefines financial decision-making by aligning model objectives with business outcomes, achieving a 44% improvement in expected profit per loan. The CQL policy is ready for deployment and live evaluation.