Shodh AI - Financial Policy Optimization Report

Applicant: Vansh Goel

1. Project Summary & Key Results

This project evaluated two machine learning approaches to optimize loan approvals. The goal was to shift the decision-making process from simple risk prediction to **financial profit maximization**.

Key Results Comparison

Metric	**Model 1: Deep Learning (DL)**	**Model 2: Offline RL (CQL)**
Primary Goal	Predict Default Risk (Probability)	Maximize Financial Return (Policy)
Key Metric	**ROC AUC Score**	**Estimated Policy Value**
Metric Value	**0.7329**	**$212.50** (per loan)
F1-Score (Class 1)	0.2348	N/A
Baseline (Historical)	N/A	**$-1806.30** (per loan loss)

The RL agent successfully learned a policy that converts the historical **-1806.30 loss** into an estimated **212.50 profit**.

2. Analysis of Models and Metrics

2.1 Deep Learning (DL) Model (MLP Classifier)

The DL model was trained to output the *probability* of default.

- **ROC AUC Score (0.7329):** This metric is used because it's robust to **class imbalance**. A score of 0.7329 shows the model is competent at *ranking* applicants by risk (better than a random guess of 0.5), validating its **predictive power**.
- **F1-Score (0.2348):** The low F1-Score (for the Default class) confirms the difficulty of making a hard "Yes/No" decision. The model lacks the financial context needed to balance false positives (denying good loans) against false negatives (approving bad loans).

2.2 Offline Reinforcement Learning (RL) Agent (Discrete CQL)

The RL agent was trained to maximize the dollar value of the decision.

- **Estimated Policy Value:** This is the most crucial **business metric**. It quantifies the expected monetary gain/loss *per loan* if the agent's policy is deployed. The shift from a loss to a profit confirms the RL agent learned to be a profit-maximizing **decision-maker**.
- **Reward Justification:** The reward function $\text{Profit} = (\text{Loan Amount} \times \text{Interest Rate} / 100)$ was chosen to force the agent to value loans based on **potential profit**, not just low risk. The penalty, $\text{Loss} = - \text{Loan Amount}$, creates an **asymmetric risk signal** (one big loss wipes out many small profits).

3. Policy Disagreement and Limitations

Policy Disagreement: The High-Risk/Low-Loss Case

The RL agent demonstrates superior decision-making in cases where the financial outcome outweighs the statistical probability of risk.

Profile	**DL Model's Decision**	**RL Agent's Decision**	**Reason for Disagreement**
Low Principal, High Interest (e.g., $2,000 at 24% with low FICO/high DTI)	**DENY**	**APPROVE**	The **DL Model** sees high default risk (e.g., 70% probability). The **RL Agent** calculates that the potential loss of $\text{-\$2,000}$ is a **small, acceptable gamble** that is statistically outweighed by the high-interest profit, ultimately favoring the decision that increases the *average expected dollar return*.

Long-Term Business Risk

The most dangerous, incorrect conclusion a non-technical stakeholder could draw is that the $\mathbf{+\$212.50}$ is a **guaranteed future profit**.

- **The Caveat:** This number is a **statistical estimate** derived from a model trained on *past data* (specifically, only *approved* loans). It does not account for changes in the economy, and the policy has never been validated on real-world denied applicants.
- **The Risk:** The reward function ignores the **loan duration (term)**. The policy, therefore, systematically **underestimates the time-based risk** of 60-month versus 36-month loans. Deploying this policy would likely lead to a portfolio over-indexed on riskier, long-term products, causing un-modeled losses.

4. Future Steps

1. **Deployment:** Conduct a controlled **A/B test** of the RL policy (e.g., 1% of new applicants) to validate the estimated profit value with real-world financial results.
2. **Refinement:** Integrate the **Time Value of Money (TVM)** and **risk-based penalties** into the reward function to force the agent to prioritize short-term, low-risk loans.
3. **Data Strategy:** Acquire or synthesize data on **denied applications** to eliminate the crucial selection bias that limits the policy's real-world accuracy.