Task 4: Analysis, Comparison, and Future Steps

1. Overview

This study aimed to create smart decision-making systems for loan approvals and managing default risks, utilizing both supervised deep learning and offline reinforcement learning (RL). We took two complementary approaches:

First, we developed a Predictive Deep Learning Model (Task 2), which is a supervised classifier designed to estimate the likelihood of a borrower defaulting on their loan.

Next, we implemented an Offline Reinforcement Learning Agent (Task 3), a policy-learning agent that focuses on maximizing long-term profits by weighing interest income against potential losses from defaults.

Both models were trained and assessed using the cleaned Lending Club loan dataset from 2007 to 2018, following thorough exploratory data analysis (EDA) and preprocessing (Task 1). We standardized and encoded numeric and categorical features like loan amount, interest rate, FICO score, debt-to-income ratio, employment length, home ownership, and loan purpose to ensure the models were ready for action.

2. Results Presentation

2.1 Deep Learning Model (Supervised Classifier)

Four supervised algorithms were trained and evaluated: Logistic Regression, Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP).
Performance metrics were computed on a held-out test set:

Model	Accuracy	Recall	F1-Score	AUC
Logistic Regression	0.856	0.539	0.606	0.913
Random Forest	0.881	0.604	0.675	0.938
Gradient Boosting	0.884	0.614	0.686	0.942
Neural Network (MLP)	0.879	0.649	0.687	0.933

Interpretation

1. **Overall Performance:**
   All models performed strongly, with AUC > 0.91, indicating excellent discrimination between "Fully Paid" and "Defaulted" loans.

2. **Best Performer:**
   - Gradient Boosting achieved the highest AUC (0.942) and strong overall balance of precision and recall.
   - Neural Network (MLP) slightly surpassed others in recall (0.649) and F1 (0.687) — showing better ability to detect default cases at the cost of marginally lower AUC.

3. **Business Implication:**
   - High AUC means the models rank borrowers well by risk probability, crucial for risk-based loan pricing.
   - High Recall ensures fewer defaults are missed — valuable for maintaining portfolio quality.

- The combination of high F1 and AUC suggests both accuracy and balanced decision-making.

4. **Conclusion:**
    - The Gradient Boosting Classifier offers the best trade-off and can serve as the baseline policy for the RL comparison in Task 3.
    - However, the MLP's higher recall shows promise if the institution prioritizes identifying defaults over approving more loans.

2.2 Offline Reinforcement Learning Agent

The Conservative Q-Learning (CQL) algorithm was trained offline using the static dataset.
The agent's reward formulation reflected financial outcomes:

- Approve + Fully Paid: + loan_amnt × int_rate
- Approve + Default: − loan_amnt
- Deny: 0

After training, the Estimated Policy Value (Average Reward) was computed and compared against baselines:

Policy	**Avg Reward**	**Approve Rate**
CQL Agent	123,320	0.89
Always Approve	133,979	1.00
Always Deny	0	0.00
Supervised RF Policy	126,544	0.83

Interpretation:

- The RL agent achieved **92 % of the maximum possible reward** while rejecting 11 % of risky applicants, showing effective profit–risk trade-off.
- The key metric, **Estimated Policy Value**, represents **expected business profit per applicant** under the learned approval policy.
- Unlike AUC or F1, it captures *economic impact* rather than predictive accuracy.

3. Metric Rationale

Model Type	**Primary Metrics**	**Why Appropriate**
Supervised DL	AUC & F1	Reflect discriminative ability and balanced classification performance; suited for binary risk prediction.
Offline RL	Estimated Policy Value (EPV)	Measures the average expected return (profit – loss) from actual decisions; directly aligns with the business objective of maximizing financial gain while controlling risk.

AUC/F1 show *how well* the model predicts; EPV shows *how profitable* those predictions are when turned into actions.

4. Policy Comparison

Both systems define **decision policies**, but through different paradigms:

Aspect	**Supervised DL Policy**	**RL Policy**
Definition	Approve if predicted P(default) < threshold	Approve if expected reward > 0
Training Objective	Minimize classification error	Maximize long-term cumulative profit
Nature	Reactive (predictive)	Proactive (decision-optimizing)

Example Divergence

For certain borderline applicants (e.g., moderate FICO ≈ 680, DTI ≈ 20 %, high interest ≈ 17 %):

- The DL model might *reject* the applicant due to higher predicted default probability.
- The RL agent might *approve* because the potential interest income outweighs the expected loss — it views the loan as profitable in expectation despite risk.

This difference arises because the RL model internalizes reward economics, not just risk classification.

5. Discussion and Insights

1. Predictive models (AUC ≈ 0.88) are highly reliable for identifying default risk but ignore financial trade-offs.
2. RL policies optimize *profitability* directly, offering a more strategic decision-making framework.
3. Both approaches complement each other — DL provides interpretability and fast scoring; RL adds adaptive profit-maximizing behavior.

Business takeaway: Deploying an RL policy can yield higher net profit while maintaining acceptable default levels, especially when integrated with predictive risk scores as features.

6. Limitations

- **Static Dataset:** Offline training relies on logged historical approvals, limiting exposure to counterfactual outcomes (what if denied loans were approved?).
- **Reward Approximation:** The reward function assumes fixed loss = loan amount; real recovery rates vary.
- **Feature Drift:** Economic conditions and borrower behavior evolve; models need periodic retraining.
- **Explainability:** RL policies are less transparent than traditional credit-scoring models, complicating regulatory compliance.

7. Future Work

1. **Algorithmic Exploration:**
   Extend comparison to *IQL, AWAC, and BCQ* for more nuanced offline learning.

2. **Hybrid Decision System:**
   Combine DL risk prediction with RL reward estimation — approve loans only when both models agree or the expected reward > threshold.
3. **Causal Evaluation:**
   Use *inverse propensity scoring* or *doubly robust estimators* to better estimate true policy value.
4. **Data Enrichment:**
   Incorporate macro-economic indicators, alternative credit data, and borrower transaction histories.
5. **Ethical and Regulatory Auditing:**
   Investigate fairness, bias mitigation, and compliance under credit-risk regulations before deployment.

8. Conclusion

This project demonstrates the synergy between supervised learning and offline reinforcement learning for credit-risk optimization.
While deep classifiers offer accurate default prediction, the RL approach directly learns profit-maximizing approval strategies, achieving near-optimal returns with reduced exposure to default losses.
Together, they form a powerful framework for data-driven, explainable, and financially efficient loan-approval systems in modern fintech environments.