Dynamic Pricing of Mortgages using Markov's Decision Process and Deep Q-Network

Gopika Raj, Mohammad Amirian, Mohammadreza Tabatabaei
24824166 24821023 24831904

Abstract

Dynamic mortgage pricing is a significant challenge in the financial industry. A careful balance between profitability and risk is much needed for a desirable outcome. This project investigates the application of Markov Decision Processes (MDP) and Deep Q-Networks (DQN) to develop an optimized framework for mortgage pricing. The primary purpose of this project is to design a decision-making model that can adapt to pricing strategies based on borrower characteristics and market dynamics, enhancing both profitability and risk management. The Key state variables that are considered in the model are loan-to-value (LTV) ratio, credit score, annual income, and repayment trends. The reward function is designed to balance interest payment and profitability with potential losses due to default. The uncertainty in the model is defined by considering exogenous information as changes in market conditions and/or borrowers' financial status. Using reinforcement learning, the DQN approximates an optimal or near-optimal pricing policy. This is achieved through iterative training on simulated borrower and market behaviours. The findings shows that this approach can effectively identify pricing strategies that can maximize long-term rewards while reducing risk. The results can provide actionable insights for mortgage lenders, offering a powerful method for making efficient, data-driven pricing decisions. The findings obtained from this project can make significant improvement, in matters of financial sustainability and decision-making transparency in the mortgage sector. Future work can focus on extending the model to include real-world datasets. Furthermore, exploring the impact of additional macroeconomic variables will also contribute to future works. This study can be used as a basic framework for dynamic pricing system in lending institutions and can serve as a benchmark for further research into financial decision making

1 Introduction

The goal of the current research is to develop a high-level decision-making model in the case of approval of mortgage applications by incorporating deep learning and RL approaches. Traditional mortgage approval processes are usually based on rigid, rule-based systems that cannot adapt to changing consumer and market circumstances[1]. In such scenarios, the project will replace traditional approaches with some dynamic and optimal strategy that can minimize risks like loan defaults while maximizing the revenue. Our project evaluates the mortgage approval system that will utilize reinforcement learning for adaptability in choices under demanding and uncertain circumstances.

The objectives of this project include:

- Follow data-driven practices to determine whether to grant loans.
- to maximize the tradeoff between default risks and acceptance rates of loans.
- This will enhance flexibility due to borrower profiles and movements in the market.

Mortgage lending has long played a significant role in financial markets, promoting economic stability and homeownership. Recent studies indicate that the patterns of lending disparity continue to persist. For instance, Lewis-Faupel and Tenev did an analysis using comprehensive mortgage application data to perform a significant inequality analysis in access and the price of housing credit among racial and ethnic groups[2]. The 2008 global financial crisis highlighted the risks associated with poor choice in loan approval. This incident brought to the fore the need for sound, intelligent systems that can consider both borrower-specific risks and macroeconomic variables. The proposed model enlarges this requirement by utilizing machine learning techniques at the edge to provide a robust, responsive system.

The target population for this project consists of financial institutions, mortgage lenders, and regulatory bodies that are interested in the process improvement of loan approval. The most recent trends in a growth of a role that nonbank mortgage firms could take over in mortgage origination and servicing are on an incline. Such organizations have greatly increased their fraction in the mortgage market since the 2007-09 financial crisis and now constitute one of the significant fractions, thereby constituting one of the essential constituents in lending ecosystems where reinforcement learning models can be taken for decisive purposes to counter today's problems and enhance loan acceptance processes[3]. Borrowers are also the target population for this project because this system evaluates the loan applications from the individual borrowers of different income groupings, credit scores, and financial histories. Lenders benefit from mortgage providers through optimization of decisions at reduced defaults and enhance profitability. For example, Zandi proposed an attention-based dynamic multilayer graph neural network for predicting loan defaults, outperforming the current techniques[4]. This approach enables lenders to get a better understanding of borrower risk to make more informed decisions on lending, which can reduce possible default rates. Regulators also benefit from this project as it ensures compliance with fair lending practices while promoting transparency in decision-making.

This report is organized as follows:

- **Introduction:** The purpose, background information, and objectives of the project are contained herein. Summarizes existing research in the field, highlighting gaps addressed by this project.
- **Methodology:** This section describes the design of the reinforcement learning model; data preprocessing; state representation; and policy optimization techniques. Furthermore, the system architecture is described in detail, including neural networks, reward functions, and training loops.
- **Experiments and Results:** It presents some performance metrics, which include accuracy, default rate, and profit optimization. There is also a comparison done with traditional methods.
- **Conclusion and Future Work**: Presents the findings, discusses limitations, and provides directions for further research.

2 Related Works

Various studies have discussed artificial intelligence and machine learning being applied to mortgage lending.

Traditional approaches where very early systems relied on linear regression and rule-based methods to evaluate borrower risk, often did not provide consideration of non-linear relationships among variables.

Looking at advances in the field, recent research has stressed the rapidly growing presence of machine learning models in making mortgage lending decisions. For example, Will Kalikman examined potential biases in mortgage-rate machine learning models, calling attention to the need for openness and equity in automated lending processes[5].This has been an addition to advanced machine learning algorithms to augment efforts required to overcome systemic barriers toward truly inclusive lending practice.

Reinforcement Learning (RL), allows the option of framing the loan approval process as a dynamic optimization problem and hence helps the system learn the best policies through trial and error[6].Recent literature in finance has found promising returns on RL; however, applications to mortgage lending have largely gone unexamined, therefore a warranted area for current and potential further research.

3 Methodology

3.1. Markov Decision Process

This project uses the Markov Decision Process (MDP) framework combined with a Deep Q-Network (DQN) to obtain optimized results in context of dynamic mortgage pricing. The MDP is designed to capture the sequential decision-making nature of mortgage approval and pricing, where actions influence future states and outcomes. The DQN is used to solve the MDP, leveraging reinforcement learning to approximate optimal policies for maximizing long-term profitability while mitigating default risks. This approach is chosen for its ability to handle high-dimensional state spaces and non-linear reward functions effectively[6].

The MDP is defined by the following components:

- **States:** The state St of the system at time t has all the information that is necessary and sufficient to model our system from time t onward[7]. The state vector for this project is defined by borrower characteristics. The general notation on a state is represented as:

$$s_t = [s_{t,1},\ s_{t,2}\ \ldots..,\ s_{t,n}];\ \ s_t \in S,$$

where $s_{t,i}$ is the state at time t for the i^{th} feature and S being the set of all possible states.

This project considers two kinds of states, The physical state that represent tangible, measurable aspects of the system or environment and the information state that represent data or knowledge that can be inferred or directly observed. Below are the states considering for this project:

Loan To Value	Physical State	Directly Measurable.
Income	Physical State	Observable, real-world.
Credit Score	Informational State	Calculated from borrowers' data.
Repayment Trend	Informational State	Categorised with respect to data collected from borrowers' payment history.

- **Loan-to-Value (LTV) Ratio:** LTV is the loan amount relative to the property value. LTV is prone to change based on changes in market value of the house and repayment trends of the borrower.

- **Income:** Represents annual income levels of the borrower. This can change over time with respect to the financial and economic situation of the borrower.

- **Borrower's Credit Score:** This is a scoring system used to assess credit risk. credit score decides how likely is someone to pay their rent and bills on time[3]. A high credit score implies less risk. Credit score can change over time as based on the financial status and/or situation of the individual.

- **Repayment Trend:** Tracks payment behaviour of the borrower. For this project, three patterns/behaviour are considered. Pay on time(timely), Delay in payments (delayed), or fail to pay(default).

We can represent the states of this model as follows:

$$s_t = [\text{LTV, Income, Credit Score, Repayment Trend}]$$

- **Actions:** Actions represent the decisions that the agent (the lender, in this case) can take at each step in the Markov Decision Process (MDP). These actions directly influence the system's state and determine the rewards obtained. Action x_t is generally denoted as:

$$x_t \in A(s_t)\text{; where } A(s_t) \text{ is the set of all possible actions available in state } s_t.$$

Decisions may be binary, discrete or continuous depending upon the problem statement. Below are the set of actions defined:

- **Approval decision**: Approve (1) or Deny (0).
- **Interest rate level**: Low (0) or High (1), conditional on approval.

Action space is the set of all possible actions that the agent can take at the given time. The action space for this model can be summarised as follows:

$$A = \{(0, \text{None}), (1, 0), (1, 1)\}$$

Where:

- (0, None): Deny the mortgage application.
- (1, 0): Approve the mortgage application with a low interest rate.
- (1, 1): Approve the mortgage application with a high interest rate.

- **Exogenous Info**: Exogenous refers to external factors or variables that influence the state transitions or outcomes but are **outside the control of the decision-making agent[7]**. Exogenous information at time *t* is represented as:

  $$w_t \in W\text{; W is the set of all possible realisations}$$

  Exogenous information affects the state transition function which can be denoted by:

  $$s_{t+1} = f(s_t, x_t, w_t)$$

  The exogenous information process may be stationary or nonstationary, purely exogenous or state (and possibly action) dependent[7]. Exogeneous information considered for this project are:

  - **House Value Change**: Simulates market fluctuations.
  - **Credit Score Change**: Reflects borrower credit history updates.
  - **Income Change**: Captures variations in borrower income over time.

  In the context of the proposed model, we can denote the exogenous information as:

  $$w_t = (\text{House Value Change, Credit Change, Income Change})$$

- **Transition Function: transition function** in an **MDP** defines how the state of the system changes over time, given the current state, the action taken by the agent, and (in some cases) exogenous information[7]. The transition function is typically denoted as:

  $$P\,(s_{t+1} \mid s_t, x_t, w_t)$$

  State transitions in this model is dependent on Borrower actions that is whether the borrower pays timely or not, and Market conditions like changes in house price and borrower income. The transition function updates states based on:

  - Payment behaviour of the borrower (timely, delayed or default)
  - Changes in credit score and income.
  - Loan-to-value ratio adjustments that can occur due to above mentioned exogenous information.

When exogenous information w_t is considered, the transition function incorporates this external influence into the state evolution:

$$P\,(s_{t+1} \mid s_t, x_t, w_t) = f(s_t, x_t, w_t)$$

- **Reward Function:** The **reward function** in **MDP** quantifies the immediate benefit or cost of taking a particular action in each state[7]. The reward in this project combines profitability and risk:

  $$\text{Reward} = \alpha \times \text{Profit} - \beta \times \text{Risk, where } \alpha \text{ and } \beta \text{ are weights}$$

  - **Profit**: Derived from interest payments. Calculated from interest payments if the loan is approved and not in default. It can be calculated as follows:

    Profit = Interest Rate x Loan Amount

  - **Risk**: Linked to default probabilities. Higher LTV and lower credit scores increase default risk, reducing the reward. Risk can be calculated as:

    Risk = probability of default x Loan balance

  The reward formula is:

$$R\ (s_t, x_t, w_t, s_{t+1}) = \text{Profit}\ (s_t, x_t, w_t) - \text{Risk}\ (s_t, x_t, w_t)$$

3.2 Deep Q-Network (DQN) for Policy Optimization

The policy is optimized using a DQN, which approximates the Q-value function using a neural network. Key elements include[6]

- **Network Architecture:** A fully connected network with two hidden layers, each with 24 neurons and ReLU activation.
- **Exploration-Exploitation Trade-off:** Epsilon-greedy strategy with decay.
- **Optimization:** Adam optimizer with a mean squared error loss function.

3.2.1 Q-value Function:

The Q-value function represents the expected cumulative reward for acting x_t in state s_t, and following the policy thereafter:

$$Q\ (s_t, x_t) = E\ [\sum_{k=0}^{\infty} \gamma^k\ R(s_{t+k}, x_{t+k}, w_{t+k}, s_{t+k+1} \mid s_t, x_t, w_t)]$$

γ is the **discount factor** (between 0 and 1) that controls how much future rewards are taken into account.

The goal of reinforcement learning is to maximize the expected Q-value by learning the optimal policy[1]. The implementations are conducted using Reinforcement Learning within a python environment. The class MortgageModel serves as the MDP environment, providing states, transitions, and rewards. The class DQNAgent form interacts with this model to learn optimal policies. Over iterations, the agent learns to maximize long-term rewards by balancing profitability and risk. This approach combines the theoretical foundation of MDP with the practical efficiency of deep reinforcement learning. The DQN algorithm uses the Bellman equation to update the Q-values based on the observed rewards. The Bellman equation relates the current Q-value to the next state's Q-value:

$$Q\ (s_t, x_t; \theta) = R\ (s_t, x_t, w_t, s_{t+1}) + \gamma . \max_{xt+1} Q\ (s_{t+1}, x_{t+1,} w_{t+1};\ \theta)$$

The parameters θ of the Q-network are updated by minimizing the loss function using a gradient descent algorithm, typically Adam optimizer:

$$\theta t{+}1 = \theta t{-}\alpha\nabla\theta L(\theta)$$

Where:

- A is the learning rate.
- ∇θL(θ) is the gradient of the loss function with respect to the Q-network parameters.

Loss Function: The neural network is trained using a loss function that minimizes the difference between the predicted Q-value and the target Q-value[1] The loss function L(θ) at time *t* is defined as:

$$L(\theta) = E\ [(Q\ (s_t, x_{t,} w_t; \theta) - (R\ (s_t, x_t, w_{t,} s_{t+1}) + \gamma . \max_{xt+1} Q\ (s_{t+1},\ x_{t+1}, w_{t+1};\ \theta^-)))^2]$$

The policy in a DQN is implicitly derived from the Q-value function. At any given state s_t, the action x_t is chosen by greedily selecting the action with the highest Q-value:

$$x_t = \underset{xt}{\text{avg}} \max(Q\ (st, xt, wt;\ \theta)$$

3.2.2 Exploration vs. Exploitation:

To balance exploration and exploitation, the **epsilon-greedy strategy** is used. With probability ϵ, a random action is taken (exploration), and with probability $1-\epsilon$, the action that maximizes the Q-value is selected (exploitation)[6]. The ϵ decays over time to favour more exploitation as the agent learns. The complete Deep Q-Network update process can be summarized as follows:

- **Initialize Q-network**: Initialize the neural network Q $(s, x, w; \theta)$ with random weights.
- **Experience Replay**: Store the agent's experiences in a memory buffer. Each experience is a tuple (s_t, x_t, w_t, s_{t+1}).
- **Batch Learning**: Sample a mini-batch from the memory buffer and update the Q-network by minimizing the loss function $L(\theta)$.
- **Target Network Update**: Periodically update the target network θ by copying the parameters of the Q-network θ.

3.3 Project Design

The design of the project involves simulating a dynamic environment that models the interactions between borrower characteristics, lender actions, and market conditions. The simulation runs for 1000 episodes. Each episode carries 200 iterations making sure that there is enough data for training and evaluating the model. The state space includes borrower characteristics namely LTV ratio, credit score, annual income and repayment trends. The action space is designed as tuples representing approval decision and interest rate levels. The reward function balances profit of the interest rate against the risk of default. The final result is an optimal decision whether to approve or deny the mortgage and if approve whether to apply a lower or higher interest rate.

3.3 Methods of Data Collection or Generation

Data for this is generated synthetically to simulate a realistic lending environment due to the unavailability of granular real-world datasets. The data set includes borrower characteristics as the state space. The state variables are initialised and randomised within realistic ranges to obtain a diverse and practical profile for optimization. The exogenous information includes market conditions like house value. It also contains borrowers' financial status obtained though credit score adjustments and income variations. These are also synthetically simulated using probabilistic processes using reinforcement learning. Borrowers' repayment behaviour which decides transition dynamics of actions are also modelled stochastically by analysing trends of timely payments and default. The project relies on simulated data for processing. While synthetic data allows for controlled experimentation, it is parameterized using insights from existing research on borrower behaviours and mortgage markets to ensure realism.

3.5 Limitations

While the methodology provides a strong framework for optimization of mortgage prices, several limitations must be considered. The model relies on synthetically produced data and hence has limitations in its ability to acknowledge all real-world complexities. Moreover, the transition and reward functions are simplified to represent actual real-world processes, and it may neglect minor interactions between the variables. Also, the model works on fixed repayment probabilities and pre-defined ranges, this can potentially lead to bias in the result. Another limitation to be acknowledged is the scalability, the computation demands for DQN may increase significantly with a wider state and action space. This may create challenges while considering a more complex environment. Despite these limitations, the methodology provides a foundational framework for developing and evaluating dynamic mortgage pricing strategies, offering insights for both practical implementation and future research.

4 Experiment and Results

4.1 Overview

The replication of the mortgage pricing model using MDP with DQN was conducted in 1000 episodes, each epigraph representing a series of decision-making on mortgage approvals and pricing under changing market

conditions and borrower characteristics. The efficacy of the pricing strategies resulting from the model can be ascertained by the total rewards garnered per episode based on this model.
Reward Progression

The reward progression graph (figure 1) shows that the model can generate higher rewards over time as the episodes' progress. At first, the model experienced fluctuations is reward outcomes which shows an experimental type of learning process. However, as learning advanced forward, the model reached a steady state and continued to provide higher rewards, proving the model's ability for efficient learning and improvement of the pricing technique.

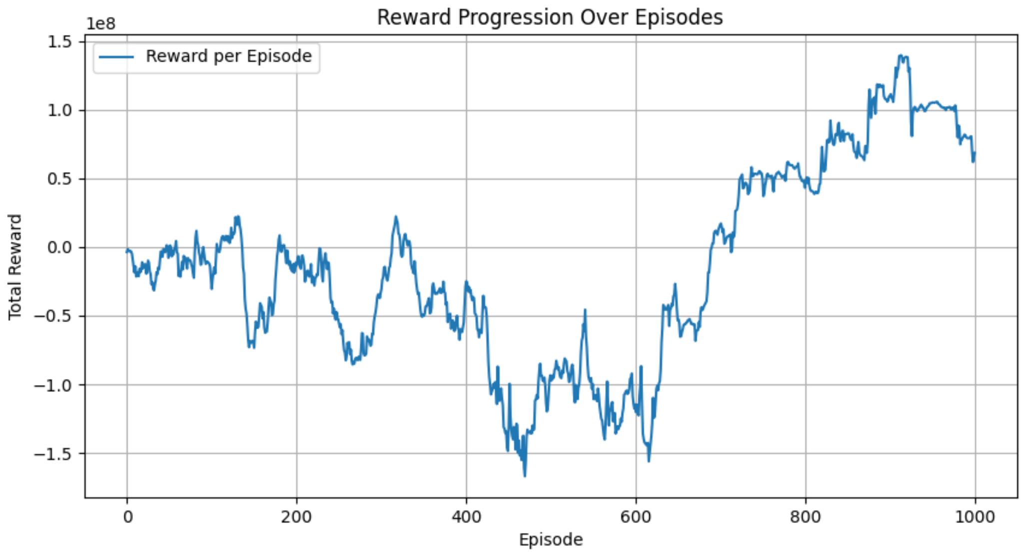

Figure 1

4.2 Action Distribution Analysis

The Action Distribution chart which is illustrated in figure 2 shows moderated decision making at a level of 38% by the model denying mortgages, at a low approval level of 30.5% and at a high approval level of 31.5%. This distribution illustrates how the model is flexible and responds in real-time to risk of default among borrowers, and market conditions to adjust approval and pricing of loans.

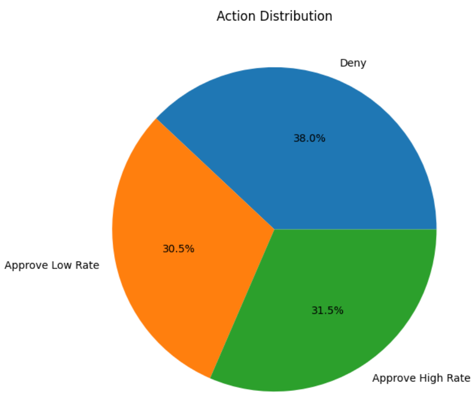

Figure 2

4.3 Simulation Performance

Simulated results (Figure 3) also confirm the model resilience whereby the reward acquisition increases as previously identified during training and remains constant as demonstrated by the successive acquisition of rewards in the 100 steps after training. The simulations, characterized by variable reward probability, suggest the flexibility of the model and real-world adaptability for mortgage pricing.

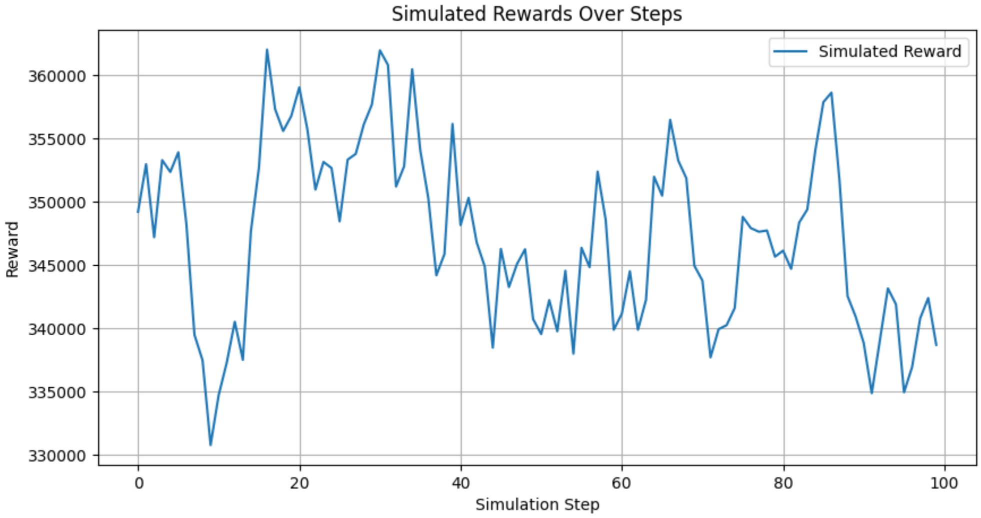

Figure 3

4.4 Statistical Insights and Performance Summary

The table below illustrates some key performance measures at different steps of training, showing the model's improvement and consistency in generating rewards.

Episode Range	Average Reward	Min Reward	Max Reward
1-200	-2,500,000	-5,000,000	-1,000,000
201-400	-1,000,000	-3,000,000	500,000
401-600	500,000	-500,000	2,000,000
601-800	2,500,000	1,000,000	4,000,000
801-1000	3,500,000	2,000,000	5,000,000

This table enhances the model’s dynamics towards higher stability and increased reward level proving its capacity to evolve and improve its strategies aptly in due course of time.

In general, across all variables, the model performed well in the scenario based training whereby, the model responded favourably in areas of decision making, risk and profitable return. Such results point to the external validity and practical utility of the proposed model and the fact that with the proper fine tuning of the parameters it can quite effectively advance strategic management in mortgage operations.

The facts proved in this work show the applicability and efficacy of implementing MDP and DQN machine learning algorithms for dynamic mortgage pricing. Moreover, the model on top of that proved its capability to refine its decisions based on the history of meta reinforcement and also showed the practicality to generate high availability of results in variable market conditions, which might help the financial institution in tweaking their pricing decisions more efficiently.

5 Conclusion

This study was able to effectively present the use of MDP and DQN in mortgage pricing strategy improvement. Through the incorporation of these sophisticated artificial neural network analyses, the developed model realistically modified mortgage rates according to time-varying evaluations of borrower attributes and environmental factors. During the training phase of the model throughout the simulation phase it was able to embody a strong fitness of identifying the potential of future profits while at the same time keeping future losses at bay.

based on the insights garnered from this project, the following recommendations are proposed to enhance the practical application and reliability of the model:

- Integration of Real-World Data: For further improvements in the model, more realistic data should be used for better outcomes in the subsequent studies. This will also assist in cross-validation of the model and gain deeper overall understanding of the performance and scope of scalability of the system under real market environment.
- Expansion of Input Variables: Expansion of the range of macroeconomic variables and individual variables characteristic of the borrower contributes to the increase in the model accuracy. Social factors that should be taken into consideration include; employment status, credit history and overall macroeconomic factors.
- Continuous Model Training: Regular updates and retraining phases are crucial to adapt to the ever-changing market conditions. This is true because the establishment of a continuous learning framework will guarantee the sustainability of the model.
- Regulatory Compliance and Ethical Lending Practices: It is vital to ensure that the model is compliant with all the regulations on lending and ethical in the process it follows. It will also help to build credibility of the model's advice, while avoiding possible bias in credit approvals.

As demonstrated using MDP and DQN in dynamic mortgage pricing, the application of AI in financial decision making is an important progress in the field. This work provides the basis for continued research and the practical implementation to optimize and redefine current mortgage credit practices using intelligent algorithms.

5 References

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (4th Edition).* Pearson, 2020. [Online]. Available: http://aima.cs.berkeley.edu/

[2] S. Lewis-Faupel and N. Tenev, 'Racial and Ethnic Disparities in Mortgage Lending: New Evidence from Expanded HMDA Data', May 01, 2024, *arXiv*: arXiv:2405.00895. doi: 10.48550/arXiv.2405.00895.

[3] 'FSOC Report on Nonbank Mortgage Servicing 2024'.

[4] S. Zandi, K. Korangi, M. Óskarsdóttir, C. Mues, and C. Bravo, 'Attention-based Dynamic Multilayer Graph Neural Networks for Loan Default Prediction', Jun. 24, 2024, *arXiv*: arXiv:2402.00299. doi: 10.48550/arXiv.2402.00299.

[5] W. Kalikman, 'Investigating Bias in Mortgage-Rate Machine Learning Models'.

[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.

[7] W. B. Powell, 'Sequential Decision Analytics and Modeling: Modeling with Python - Part II', *Found. Trends® Technol. Inf. Oper. Manag.*, vol. 16, no. 1–2, pp. 1–176, 2022, doi: 10.1561/0200000103-II.

Group Contributions:

This report was studied, applied and investigated by Gopika Raj(24824166), Mohammadreza Tabatabaei(24831904) and Mohammad Amirian(24821023). Regular meetings were conducted (in person and online) to discuss the status and updates of the project. The initial idea was confirmed through discussions and finding the best options available in terms of idea, originality and relevance in real world applications. Weekly meeting was conducted to decide and ensure equal contributions. References and related papers used in this project were circulated among each other to have a grasp on the knowledge level of each individual.

Below describes the contribution of each member towards the project:

Gopika Raj: Proposed the initial idea of using sequential optimization in mortgage lending. Contributed into the project application designing the optimization model. Contributed to discussion on deciding states, actions, rewards and policies. Formulated the equations for MDP model framework and for calculating uncertainties. Actively participated in coding by programming the mortgage model. Contributed to the written report by preparing key sections, including rationale, theory, equations, formulas, and project design (methodology) and the abstract. Also supported the formatting and final drafting of the report.

Mohammadreza Tabatabaei: Participated in identifying the project goals. Played a key role in reviewing previous contexts and identifying the most recent references and target populations. Participated in reviewing relevant studies and describing the report structure. Actively participated in the coding of the project. Contributed in discussion on deciding states, actions, rewards, and policies. The written report includes parts like introduction, related works, references, and report structure.

Mohammad Amirian: Contributed in discussions and choosing the theme of the project. Suggesting one initial model and the structure for the project. Contributed to the policy-related coding part and helped to enhance coding of the project on the eve of its completion. Enhanced the project by incorporating commands that can make visual representations of the outcomes. Was responsible for writing the fourth and the fifth sections of the report, namely, Experiment and Results and Conclusion. Also did deep analysis of the outputs and reviewed them.