4SSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 004.8

doi: 10.32620/reks.2024.4.20

Vladyslav DIDKIVSKYI, Dmytro ANTONIUK, Tetiana VAKALIUK, Yevhen OHINSKYI

Zhytomyr Polytechnic State University, Zhytomyr, Ukraine

USING THE PROXIMAL POLICY OPTIMIZATION AND PROSPECT THEORY TO TRAIN A DECISION-MAKING MODEL FOR MANAGING PERSONAL FINANCES

*The **subject of** this article is the development of a decision-making model that can, in the future, be incorporated into a personal finance simulator to improve personal finance literacy. The **goal** of this study is to develop decision-making models tailored to different investor profiles to provide personalized financial advice on asset allocation. This article employs reinforcement learning techniques and behavioral economics to achieve this objective, thereby contributing to the advancement of practical algorithms and approaches for financial decision-making. The **tasks** can be formulated as follows: 1) design a reinforcement learning environment featuring different investment options with varying average returns and volatility levels; 2) train the reinforcement learning agent using the Proximal Policy Optimization algorithm to learn recommended investment allocations; 3) implement a reward function based on Prospect Theory, incorporating parameters that reflect different investor risk profiles, such as loss aversion and diminishing sensitivity to gains and losses. The **results** reveal the development of distinct models for 3 investor profiles: risk-averse, rational, and wealth-maximizing. A graphical analysis of the recommended allocation percentages revealed significant patterns influenced by the value function parameters of Prospect Theory. The practical implications of this research extend to the development of simulation tools based on the model, which will enable individuals to practice and refine their financial strategies in a risk-free environment. These tools bridge the gap in personal finance education by providing experiential learning opportunities. **Conclusions**. The developed model effectively generates personalized financial advice that reflects individual risk preferences. Future work will focus on creating interactive simulation tools to enhance personal finance management skills. This study underscores the importance of integrating psychological and behavioral insights into financial decision-making models.*

Keywords: *personal finance; decision-making; reinforcement learning; prospect theory.*

1. Introduction

1.1. Motivation

Personal finance literacy is an essential skill in today's complex financial environment, enabling individuals to make informed decisions regarding their economic well-being. Effective financial management involves understanding and mastering various aspects, such as budgeting, saving, investing, and debt management [1, 2]. Despite its importance, many individuals, particularly those from low-income backgrounds, often lack the necessary knowledge and skills to make good financial decisions [3].

It is important to note that our research is not focused on trading or on individuals who are already engaged in trading activities. Instead, we target individuals without extensive experience in managing personal finances. Our long-term goal is to assist everyday individuals in improving their financial literacy and decision-making skills. We make financial education accessible to a broader audience, particularly those who may feel overwhelmed by the complexities of financial management.

To address this gap, this research explores the development of a decision-making model that leverages Proximal Policy Optimization and Prospect Theory to provide personalized financial advice. Proximal Policy Optimization (PPO), a cutting-edge reinforcement learning algorithm, is known for its high efficiency and stability when training complex models [4, 5]. Prospect Theory, on the other hand, offers a nuanced understanding of human decision-making by incorporating psychological factors into economic behavior [6]. Previously, we used Prospect Theory to analyze decision-making processes in personal finance. We investigated how individual risk perception influences the choice of financial instruments and investment strategies. Additionally, we developed a personal finance simulator although it currently lacks the feedback features necessary to evaluate the effectiveness of the decisions made [2]. By integrating PPO and Prospect Theory, we aim to create a model that allows interaction with a person using simulation technologies to develop a person's skills to make rational decisions in personal finance management.

1.2. State of the Art

Many scientists have examined the issue of personal financial literacy in their research. Lusardi et al. [7] proved that a lack of financial literacy and poor financial management can lead to irresponsible spending, problems with academic performance, and negative effects on mental and physical well-being [7]. T. Koskelainen's study, explores how digitalization affects financial literacy and capability, focusing on Fintech, digital financial behavior, and behavioral interventions. It proposes updates to financial literacy education and the development of digital tools and emphasizes public-private collaboration for a more inclusive economy [8].

PPO has been widely applied to various tasks, including trading activities. Lin, S.-Y. introduced a framework for optimal trade execution using PPO, which effectively handles time dependencies in market data [9]. Yang, H. proposed an ensemble strategy for stock trading that uses deep reinforcement learning to maximize investment returns. By integrating the strengths of a few algorithms, including PPO, the ensemble strategy effectively adapts to different market conditions [10].

Although prospect theory was developed many years ago, it remains highly relevant and continues to be applied in various research fields. Cabedo-Peris, J. explored the relationship between addictive behaviors and decision-making processes through the lens of prospect theory [11]. The study of Shrader, R.C. explored the relationship between the effort invested in developing financial forecasts and risk-taking behavior using prospect theory [12]. The study of Srivastava, S. presents a portfolio selection approach that integrates cumulative prospect theory with data envelopment analysis. By using a quadratic value function and assessing assets based on prospect theory value and long-term returns, this study investigates psychological factors in portfolio selection [13]. The study of Wang, X. introduces a three-way decision model based on cumulative prospect theory and outranking relations for portfolio selection. By incorporating a boundary region, the model reduces decision risk [14].

These studies underscore the importance of this research in the field of personal financial literacy and decision-making. The findings highlight the significant consequences of poor financial literacy and highlight the need for problem-solving skills to enhance financial capabilities. The application of advanced algorithms like PPO demonstrates its effectiveness in handling complex data and optimizing decision-making processes. This demonstrates the potential of PPO to improve financial decision-making strategies. Furthermore, the enduring relevance of prospect theory highlights its applicability to understanding and improving decision-making under risk and uncertainty.

1.3. Objectives and Approach

The primary **objective** of this research is to use PPO and Prospect Theory to train decision-making models tailored to different investor profiles that can provide personalized financial advice on asset allocation.

The **approach** in this research involves developing a reinforcement learning environment based on the PPO algorithm that includes different investment options with varying returns and volatilities. In this simulation, the agent is allowed to allocate a percentage of its incomes to these options. The action space is continuous, enabling flexible investment decisions. The reward function uses the Prospect Theory value function, which captures the asymmetry in how people perceive gains and losses. In summary, we will develop 3 distinct models tailored to different risk profiles (risk-averse, rational, and wealth-maximizing). These models can predict recommended investment percentages for each option based on the agent's risk tolerance.

2. Materials and methods of research

2.1. Reinforcement Learning

Reinforcement Learning (RL) is a machine learning paradigm that enables agents to learn effective behaviors through interactions with their environment. Unlike supervised learning, which relies on labeled data, RL focuses on learning from the consequences of actions and uses feedback in the form of rewards to refine decision-making policies over time [15].

In RL, the learning process is typically framed as a Markov Decision Process (MDP), which provides a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision-maker. An MDP is defined by a set of states, a set of actions, a transition function that describes the probability of moving from one state to another given an action, and a reward function that assigns a numerical value to each state-action pair [16].

The diagram in Figure 1 represents the interaction between the agent and environment in a reinforcement learning framework. The agent observes the current state S_t from the environment and selects an action A_t based on its policy. This action influences the environment, resulting in a new state S_{t+1} and a reward R_{t+1} given to the agent. The process continues iteratively, with the agent learning to optimize its actions based on the received rewards, with the goal of maximizing cumulative rewards over time.

4SSN 1814-4225 (print)
ISSN 2663-2012 (online)

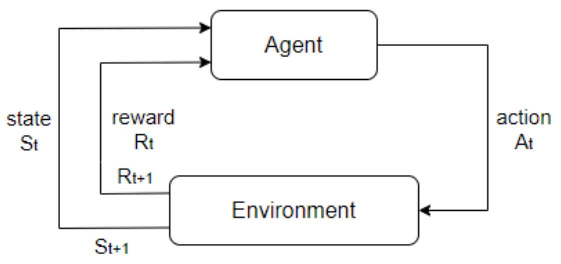

Fig. 1. The basic schema of the proposed RL algorithm

2.2. Proximal Policy Optimization

Proximal policy optimization (PPO) is a reinforcement learning algorithm that aims to find an optimal policy for decision-making. PPO uses an actor-critic approach, where the actor (policy) is responsible for selecting actions, and the critic (evaluator) assesses how well these actions align with expected rewards. The primary goal of PPO is to maximize the expected cumulative reward over time. This is achieved by adjusting the policy to increase the likelihood of actions that yield higher rewards.

Key principles of PPO:

1. PPO belongs to the class of policy gradient methods in which policies are iteratively improved based on the feedback received from the environment.

2. Unlike traditional policy gradient methods, PPO uses a clipped objective function that constrains the policy update to the neighborhood of the old policy. This ensures that the policy update does not significantly deviate, thereby maintaining stability during training.

Advantages of PPO:

1. PPO is relatively straightforward to implement compared to other reinforcement learning algorithms.

2. The proposed model effectively utilizes experience replay and batch updates to leverage data efficiently.

3. The use of a clipped surrogate objective helps stabilize the training process and prevents large policy updates that could lead to catastrophic performance reductions.

The core of PPO is the optimization of a surrogate objective function that balances the need for improving the policy while maintaining stability in the training process through a clipping mechanism. This mechanism involves clipping a coefficient in the PPO objective function (the clipped surrogate objective function) within a specific range. This clipping helps prevent large updates that can destabilize learning.

The objective of PPO is to maximize the expected advantage but with a constraint that ensures that updates are not too large. The surrogate objective function is defined as follows:

$$L^{CLIP}(\theta) = \widehat{E_t}\left[\min\left(r_t(\theta)\widehat{A_t}, \mathrm{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\widehat{A_t}\right)\right], \quad (1)$$

where θ represents the policy parameter;

$\widehat{E_t}$ denotes the expectation over time steps;

$r_t(\theta)$ is the ratio of the new and old policy;

$\widehat{A_t}$ is the estimated advantage at time step t;

clip is the clipping function that limits the value of $r_t(\theta)$ to the range $[1-\varepsilon, 1+\varepsilon]$.

The objective function in PPO allows the agent to perform multiple improvement epochs or iterations based on a batch of collected data. This process is like to practicing a skill repeatedly to achieve better performance. By iterating over the data multiple times, the agent can refine its policy more effectively.

The clipping mechanism prevents policy updates from being too large, thus maintaining stability and preventing drastic changes that could lead to suboptimal performance. By optimizing this clipped objective function, PPO balances exploration and exploitation, thereby allowing stable and efficient learning in complex decision-making environments.

2.3. Prospect Theory

Prospect Theory, developed by Daniel Kahneman and Amos Tversky, is a behavioral economic theory that describes how people make decisions between alternatives that involve risk and uncertainty. This theory deviates from the traditional expected utility theory by incorporating psychological insights into economic decision-making [17].

The value function in Prospect Theory is a core component that describes how individuals evaluate potential gains and losses relative to a reference point, typically their current state of wealth or a specific benchmark. This function captures the psychological nuances of how people perceive value, which differs significantly from the linear approach assumed in traditional economic theories [18].

The value function v(x) can be expressed mathematically as follows:

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\beta} & x < 0 \end{cases}, \quad (2)$$

where x represents the outcomes, which can be gains or losses;

α is the parameter that captures the diminishing sensitivity of gains;

λ is the loss aversion parameter, which captures the greater sensitivity to losses compared to gains;

β is the parameter that captures the diminishing sensitivity to losses.

The graph in Figure 2 illustrates a value function that is typically S-shaped and asymmetric. It is concave for gains and convex for losses. This shape reflects how people perceive gains and losses differently. The concavity for gains indicates diminishing sensitivity, meaning that the subjective value of gains decreases as the amount increases. Conversely, the convexity for losses shows that the subjective loss value increases steeply as the amount increases. The value function is steeper for losses than for gains, illustrating loss aversion. This means that losses are larger than gains of the same magnitude. For example, losing 100$ feels more painful than gaining 100$.

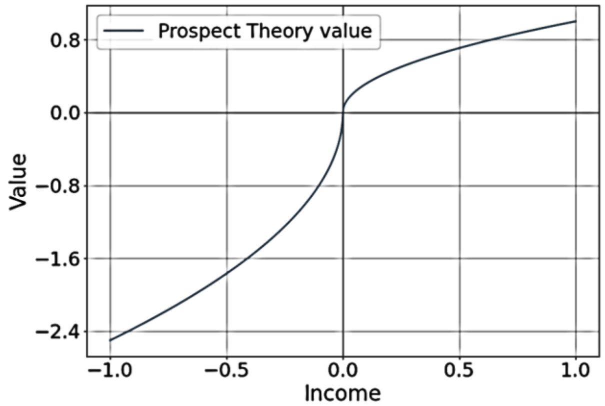

Fig. 2. An example of a graph of the prospect theory value function

Cumulative prospect theory further extends the original prospect theory by introducing the concept of decision weights, which models the nonlinear transformation of probabilities. This modification allows the model to better capture how individuals perceive and act on probabilities, particularly for low-probability, high-impact events that are often relevant in financial decision-making [19].

The weighting function w(p) can be expressed mathematically as:

$$w(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1 - p)^{\gamma})^{1/\gamma}}, \tag{3}$$

where p represents the probability;

γ is a parameter that determines the curvature of the weighting function.

The graph in Figure 3 illustrates the probability weighting function, which is typically inverse S-shaped. This function captures how people perceive probabilities in a nonlinear manner. Small probabilities are often overestimated, which makes rare events appear more likely than they are. Conversely, moderate to high probabilities are underestimated, making likely events seem less certain. This weighting affects how risky prospects are evaluated. For instance, people might overvalue lottery tickets (small probability of a large gain) and undervalue insurance (high probability of a small loss).

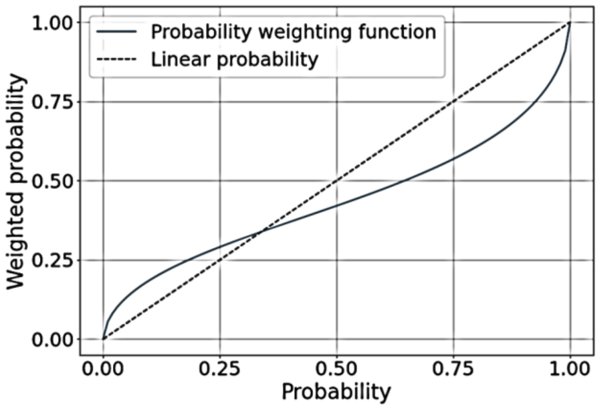

Fig. 3. An example of a graph of the prospect theory probability weighting function

For the experiments in this research, we utilize the original prospect theory value function. In future work, we will assess the feasibility and potential benefits of incorporating cumulative prospect theory.

3. Results and Analysis

3.1. Design and Implementation of the RL Environment

In this study, we developed a reinforcement learning environment using the OpenAI Gym toolkit to train and test the RL algorithms. We used the Proximal Policy Optimization algorithm from the Stable-Baselines3 library [20]. As a reward function, we used the Prospect Theory value function represented by equation (2).

Our environment featured 6 investment options, each characterized by different average returns and volatility.

Table 1

Environment investment options

№	Average return	Volatility
1	0.03	0.03
2	0.04	0.04
3	0.06	0.12
4	0.07	0.14
5	0.09	0.27
6	0.10	0.30

The simulation runs for 50 steps, where each step represents one year of human life. At each step, the agent can choose from the available investment options and allocate a percentage of their income to investments.

The action space is represented by the 1 continuous action within the range [0, 1], and the agent can choose how to allocate its available funds across different investments by specifying a percentage (between 0% and 100%) for each investment option. The total allocation of all investments must not exceed 100%.

Listing 1

Implementation of the action space

```
self.action_space = spaces.Box(
    low=0.0,
    high=1.0,
    shape=(n_investments,),
    dtype=np.float32)
```

where "n_investments" is number of investment options (6).

Listing 2

Implementation of the observation space

```
self.observation_space =
spaces.Dict({
    "invs": spaces.Box(
        low=0, high=1,
        shape=(self.n_investments,),
        dtype=np.float32),
    "step": spaces.Discrete(50)
})
```

where "invs" represents the total investment percentages for different options across all 50 steps, with values ranging from 0 (no investment at any step) to 1 (full investment at all steps). "step" represents the current step in the environment, with 50 possible discrete values. This configuration allows the agent to observe both its investment distribution and current steps in the environment.

Prior to each step in the environment, the "reset" function is invoked to restore the state to its initial values.

Listing 3

Implementation of the "reset" function

```
def reset(self, **kwargs):
    self.current_step = 0
    self.investment_totals = [0] *
self.n_investments
    self.update_state()
    return self.state, {}
```

where "current_step" is the current step index in a simulation run, "investment_totals" is the list of total investment percentages per each investment option.

The "update_state" function updates the state (observation space) to reflect the new investment percentages.

Listing 4

Implementation of the "update_state" function

```
def update_state(self):
    self.state = {
        "invs": [total / self.n_steps
for total in self.investment_totals],
        "step": self.current_step
    }
```

The "step" function simulates a single step in a reinforcement learning environment. It adjusts total investment percentages, calculates rewards, and updates the environment state.

Listing 5

Implementation of the "step" function

```
def step(self, action):
    action_sum = np.sum(action)
    if action_sum > 0:
        step_investments = action /
action_sum
        for i in range(self.n_invest-
ments):
            self.investment_totals[i]
+= step_investments[i]

        reward = self.calculate_re-
ward()
    else:
        reward = -100

    self.update_state()
    self.current_step += 1
    done = self.current_step ==
self.n_steps
    info = {
        "investment_totals": self.in-
vestment_totals
    }

    return self.state, reward, done,
False, info
```

The "step" function processes an action vector representing allocation percentages for each investment. The proposed method begins by summing the elements of the action vector. If the sum is positive, it normalizes the allocation percentages by dividing each element of the action vector by the total sum, ensuring that the normalized allocations sum to 1. Then, it updates the "investment_totals" with these normalized values. Next, the function calculates the reward using the "calculate_reward" method based on the updated investment totals. If the action sum is zero, the algorithm assigns a heavy penalty by setting the reward to -100. The function updates the environment's state to reflect the new investment values and increments the "current_step" counter. It checks whether the current step has reached the maximum number of steps, thereby indicating whether an episode has been completed. Finally, the function returns the updated state, the reward, a boolean indicating whether the episode has ended, and an information dictionary with the current investment totals.

The reward function defines the goal of the learning agent. The reward provides feedback to the agent, guiding it to learn the most effective actions to achieve its objective over time [21].

Listing 6

Implementation of the "calculate_reward" function

```
def calculate_reward(self):
    reward = 0
    for i in range(self.n_invest-
ments):
        inv = self.investments[i]
        inv_return = random.uniform(
            inv.average_return -
inv.volatility,
            inv.average_return +
inv.volatility)
        prospect_theory_value = pro-
spect_theory(
            inv_return,
            self.risk_pro-
file_params["alpha"],
            self.risk_pro-
file_params["lambda"])

        reward += self.investment_to-
tals[i] * prospect_theory_value

    return reward
```

The function iterates over all investment options and calculates return, a randomly sampled value within a specified volatility range. Then, it calculates the prospect theory value for the current investment return using the "prospect_theory" function implemented according to equation 2. Then, the function multiplies the agent's investment percentage in this asset by the prospect theory value, which ensures that the overall reward reflects both the investment's performance and the agent's allocation decisions and adds the resulting value to the reward.

The α, β and λ parameters of the Prospect Theory value function (equation 2) are defined based on the investor's risk profile. For training our models, we decided to use the same value for the β parameter, aligning it with the α parameter, to simplify the modeling process and focus on key aspects of Prospect Theory, such as loss aversion represented by the λ parameter. Empirical studies, including those by Kahneman and Tversky, have shown that α and β often take on similar values when estimating the value function for gains and losses [17].

The Prospect Theory value function parameters are defined as follows:

1. If the investor is risk-averse, $\alpha = 0.5$ reflecting diminishing sensitivity to both gains and losses, meaning the investor is cautious and values smaller gains and losses more than larger ones; $\lambda = 2.5$ indicates that losses are perceived much more significantly than equivalent gains.
2. If the investor is rational, $\alpha = 1$, reflecting linear sensitivity to gains and losses, indicating that the investor evaluates gains and losses proportionally without diminishing sensitivity; $\lambda = 1.5$ indicates that losses are perceived a bit more than equivalent gains.
3. If the investor is wealth-maximizing, $\alpha = 1.5$ reflecting increasing sensitivity to both gains and losses, meaning the investor places more value on larger gains and losses and is willing to take on more risk; $\lambda = 0.5$ indicates that losses are perceived less significantly than equivalent gains.

3.2. Hyperparameters Configuration

Table 2 outlines the PPO configurations used in the proposed reinforcement learning environment.

Table 2

PPO configurations

Parameter	Value
ent_coef	0.05
clip_range	0.15
n_epochs	15
n_steps	50
batch_size	50
learning_rate	3e-4
gamma	0.99
gae_lambda	0.95

A significant adjustment was made to the entropy coefficient "ent_coef" parameter, which was initially set to 0. Subsequently, it was increased to 0.02, 0.04, and 0.05. This change was intended to broaden the scope of exploration, which is crucial given the wide range of investments with varying volatilities. Without sufficient exploration, the model cannot learn optimal strategies.

4SSN 1814-4225 (print)
ISSN 2663-2012 (online)

We decreased the “clip_range” from 0.2 to 0.15. This parameter limits policy variations in updates to minimize sudden large fluctuations that may occur during the training phase. The clip range must be reduced to balance the increased entropy coefficient and ensure stable policy updates.

The “n_epochs” was increased to 15 to achieve more stable training by reducing the variance in policy updates. This method is particularly useful in environments with high reward variability.

The “n_steps” parameter was set to 50 to match the episode length, which was constant and always equal to 50.

The “batch_size” parameter was set to 50 to match the “n_steps” parameter because the episode length was relatively short, which made it unnecessary to split it into smaller batches.

The “learning_rate”, “gamma” and “gae_lambda” parameters are left at their default values.

3.3. Training Results and Analysis

We conducted 3 training runs, each dedicated to training a distinct model tailored to a specific investor profile based on risk preference [22]: risk-averse, rational, and wealth-maximizing. Each training run consisted of 1 million steps. As a result, we developed models capable of proposing allocation percentages for each investment option.

Figure 4 illustrates the model’s allocation predictions for different investor profiles. The only recommended investment option for a risk-averse investor profile is option 2, which offers an average return of 0.04 and a volatility percentage of 0.04, ensuring no losses. This type of investor prioritizes avoiding losses over achieving gains, and their investment decisions are driven by the desire for stability and minimal risk. In this context, Option 2 is the most suitable option due to its favorable risk-return profile. The low volatility of 0.04 aligns with a risk-averse investor's preference for minimizing potential fluctuations in their portfolio's value. The model’s recommendation to invest entirely in option 2 reflects the investor’s preference for safety and certainty. By allocating the entire investment to option 2, the investor avoids the higher risks associated with other options, which may offer higher returns but come with greater volatility. This conservative approach ensures that the investor’s capital is preserved, which aligns with their primary objective of avoiding financial losses.

According to the chart, the model’s recommended investment allocations for a rational investor profile are distributed as follows: 0.25 to option 3, 0.30 to option 4, 0.20 to option 5, and 0.25 to option 6. The allocation of funds across these options reveals a strategy that balances risk and return in alignment with rational investment principles. By distributing investments among these options, the profile seeks to optimize returns while managing risk. The presence of options 5 and 6, with their higher risk and return, indicates a willingness to accept some degree of risk to achieve higher potential returns. The diversified approach across options 3, 4, 5, and 6 reflects a rational investor strategy to mitigate risk through diversification while aiming for favorable returns based on individual risk tolerance and investment goals.

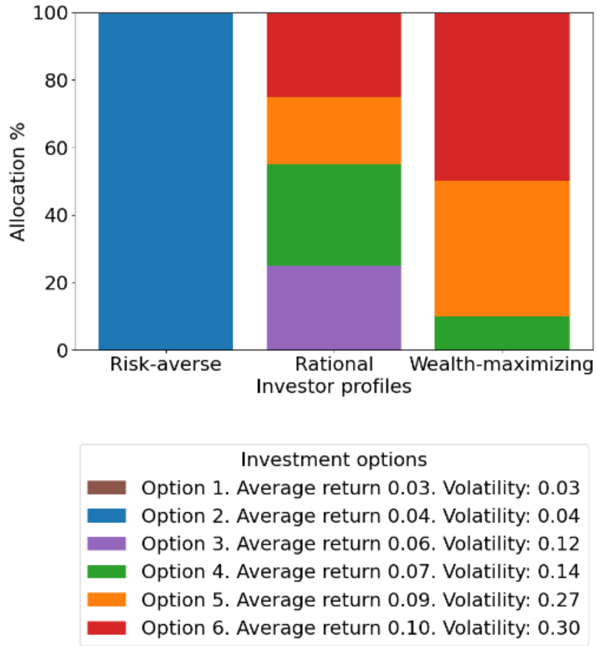

Fig. 4. Model allocation predictions based on the investor profile

The model’s allocation predictions for wealth-maximizing investor profiles are distributed as follows: 0.10 to option 4, 0.40 to option 5, and 0.50 to option 6. For wealth-maximizing investors, the primary objective is to achieve the highest possible return, even if that entails accepting higher risk levels. Accordingly, the model allocated significant portions of the investment to options 5 and 6, which offer higher returns despite their increased volatility. The high allocation to options 5 and 6 reflects the investor's willingness to take greater risks to achieve potentially higher returns. By focusing on these options, wealth-maximizing investors aim to capitalize on opportunities for significant financial growth. The diversified allocation approach ensures exposure to high-return options while maintaining some level of risk management through the inclusion of option 4.

The graphical analysis of recommended allocation percentages across various investment options shown in Figure 5, as influenced by the α and λ parameters from the Prospect Theory value function, reveals insightful patterns. Investment options with higher returns and volatility, such as options 5 (0.09 return, 0.27 volatility) and

6 (0.10 return, 0.30 volatility), exhibit significantly increased allocations as α and λ values rise. This trend underscores a critical observation: higher α values, reflecting increased sensitivity to gains and losses, prompt individuals to favor investments with higher returns, accepting the concomitant higher risks. Concurrently, increased λ values, indicating reduced loss aversion, further amplify this risk-taking behavior, encouraging allocation toward high-return, high-volatility investments. Consequently, our model demonstrates that as individuals' risk tolerance heightens, driven by elevated α and λ values, there is a pronounced preference for investment options that offer substantial returns despite their greater volatility. This behavior aligns with the core principles of Prospect Theory, affirming that decision-making in personal finance is profoundly influenced by individual attitudes toward gains, losses, and risk.

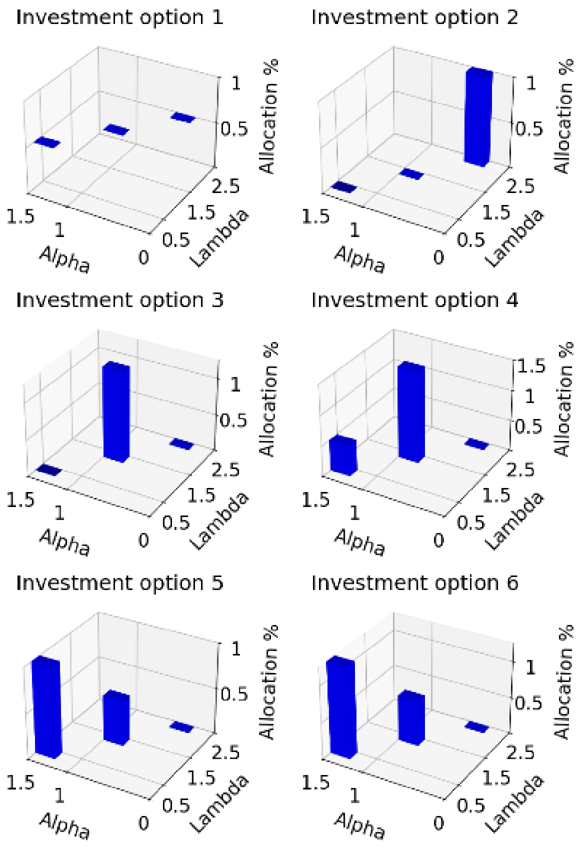

Fig. 5. Model allocation predictions based on the α and λ parameters of the value function of the Prospect Theory

4. Discussion

The findings of this study provide a strong foundation for creating interactive simulation tools that allow users to experiment with and refine their financial strategies in a risk-free environment. These tools are poised to address a significant gap in practical financial education by providing a safe platform for users to apply and test different investment strategies. Such simulations enable users to gain practical experience and build confidence in their financial decision-making skills, which is often limited in real-world scenarios where opportunities for hands-on learning are scarce.

The integration of psychological and behavioral insights into financial decision-making represents a significant advancement over traditional models, which typically rely on static assumptions about investor behavior. By incorporating Prospect Theory's parameters into our models, we provide a more nuanced understanding of how psychological factors influence investment choices. This approach not only enhances the realism of the models and empowers individuals with a deeper understanding of their financial behaviors and biases.

Furthermore, applying reinforcement learning techniques allows for dynamic and adaptable investment strategies that align with individual risk profiles. This flexibility is crucial for addressing investors' diverse needs and tailoring recommendations to their specific preferences and risk tolerances.

5. Conclusions

This research has demonstrated the successful development of a decision-making model that integrates proximal Policy Optimization and Prospect Theory. By leveraging these advanced techniques, we developed 3 models tailored to a specific investor profile based on risk preference: risk-averse, rational, and wealth-maximizing. The models' predictions for the recommended investment allocations highlight the nuanced ways in which individual risk preferences and psychological factors influence financial decision-making.

Our findings indicate that, for risk-averse investors, the model recommends conservative investment choices, focusing on options with lower volatility to minimize potential losses. In contrast, rational investors receive balanced recommendations that optimize returns while diversifying risk. Wealth-maximizing investors, on the other hand, are advised to concentrate their investments in high-return, high-volatility options, reflecting their greater tolerance to risk in pursuit of higher gains.

The detailed analysis of allocation patterns based on Prospect Theory parameters (α and λ) underscores the impact of psychological factors on investment behavior. Higher sensitivity to gains and losses (α) and reduced loss aversion (λ) drive investors toward riskier options with higher returns, which aligns with Prospect Theory's insights into human decision-making.

Overall, this research advances the field of personal finance by integrating reinforcement learning and behavioral economics into decision-making models.

The findings offer practical implications for developing personalized financial advice tools and highlight the importance of considering psychological factors when designing investment strategies.

Future research should focus on enhancing these models with additional behavioral insights and developing interactive simulation tools to further support financial literacy and decision-making. For example, incorporating additional behavioral biases, such as overconfidence or mental accounting, can further refine investment recommendations and provide a more comprehensive view of investor behavior. Additionally, expanding the range of investment options and market conditions in the simulation could enhance the model's robustness and applicability in real-world scenarios.

The development of interactive simulation tools based on these models will also be a key focus. Such tools can offer users practical, hands-on experience with financial decision-making, thus bridging the gap between theoretical knowledge and practical application. This experiential learning approach can significantly improve personal finance management skills and increase financial literacy.

Author Contributions

Conceptualization, methodology – **Vladyslav Didkivskyi**, **Dmytro Antoniuk**; formulation of tasks, analysis – **Vladyslav Didkivskyi, Yevhen Ohinskyi**; development of model, software, verification – **Vladyslav Didkivskyi, Tetiana Vakaliuk**; analysis of results, visualization – **Vladyslav Didkivskyi, Yevhen Ohinskyi**; writing, original draft preparation – **Dmytro Antoniuk**; writing review and editing – **Tetiana Vakaliuk**.

Conflict of Interest

The authors declare no conflict of interest.

Financing

This study was conducted without financial support.

Data Availability

Data will be made available upon reasonable request.

Use of Artificial Intelligence

During the preparation of this study, Microsoft Copilot was used to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

The authors thank the scientific and teaching staff of the Zhytomyr Polytechnic State University for their support.

All the authors have read and agreed to the publication of the final version of this manuscript.

References

1. Thomas, A., & Gupta, V. Social Capital Theory, Social Exchange Theory, Social Cognitive Theory, Financial Literacy, and the Role of Knowledge Sharing as a Moderator in Enhancing Financial Well-Being: From Bibliometric Analysis to a Conceptual Framework Model. *Frontiers in Psychology*, 2021, vol. 12. DOI: 10.3389/fpsyg.2021.664638.

2. Antoniuk, D. S., Vakaliuk, T. A., Didkivskyi, V. V., & Vizghalov, O. Development of a simulator to determine personal financial strategies using machine learning. In: *Proceedings of the 4th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2021)*, 2022, Virtual Event, Kryvyi Rih, Ukraine, December 18, 2021. Edited by Arnold E. Kiv, Serhiy O. Semerikov, Vladimir N. Soloviev, Andrii M. Striuk. CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073). Vol. 3077, pp. 12-26. Available at: http://ceur-ws.org/Vol-3077/paper02.pdf. (accessed 12.08.2024).

3. Xu, S., Yang, Z., Ali, S. T., Li, Y., & Cui, J. Does Financial Literacy Affect Household Financial Behavior? The Role of Limited Attention. *Frontiers in Psychology*, 2022, vol. 13. DOI: 10.3389/fpsyg.2022.906153.

4. Corecco, S., Adorni, G., & Gambardella, L. M. Proximal Policy Optimization-Based Reinforcement Learning and Hybrid Approaches to Explore the Cross Array Task Optimal Solution. *Machine Learning and Knowledge Extraction*, 2023, vol. 5, iss. 4, pp.1660–1679. DOI: 10.3390/make5040082.

5. Lobanchykova, N. M., Pilkevych, I. A., & Korchenko, O. Analysis and protection of IoT systems: Edge computing and decentralized decision-making. *Journal of Edge Computing*, 2022, vol. 1, iss. 1, pp. 55–67. DOI: 10.55056/jec.573.

6. Ruggeri, K., Alí, S., Berge, M. L., Bertoldo, G., Bjørndal, L. D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., Gibson, S. P., Jarke, H., Karakasheva, R., Khorrami, P. R., Kveder, J., Andersen, T. L., Lofthus, I. S., McGill, L., Nieto, A. E., & Pérez, J. Replicating patterns of prospect theory for decision under risk. *Nature Human Behaviour*, 2020, vol. 4, iss. 6, pp.622–633. DOI: 10.1038/s41562-020-0886-x.

7. Lusardi, A., & Mitchell, O. S. The Importance of Financial Literacy: Opening a New Field. *Journal of Economic Perspectives*, 2023, vol. 37, iss. 4, pp.137–154. DOI: 10.1257/jep.37.4.137.

8. Koskelainen, T., Kalmi, P., Scornavacca, E., & Vartiainen, T. Financial literacy in the digital age – A research agenda. *Journal of Consumer Affairs*, 2023, vol. 57, iss. 1, pp.507-528. DOI: 10.1111/joca.12510.

9. Lin, S. -Y., & Beling, P. A. An end-to-end optimal trade execution framework based on proximal policy optimization. *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 4548-4554. DOI: 10.24963/ijcai.2020/627.

10. Yang, H., Liu, X.-Y., Zhong, S., & Walid, A. Deep reinforcement learning for automated stock trading. *Proceedings of the First ACM International Conference on AI in Finance*, 2020, vol. 31, pp. 1-8. DOI: 10.1145/3383455.3422540.

11. Cabedo-Peris, J., González-Sala, F., Merino-Soto, C., Pablo, J. Á. C., & Toledano-Toledano, F. Decision Making in Addictive Behaviors Based on Prospect Theory: A Systematic Review. *Healthcare (Basel, Switzerland)*, 2022, vol. 10, iss. 9, article no. 1659. DOI: 10.3390/healthcare10091659.

12. Shrader, R. C., Simon, M., & Stanton, S. Financial forecasting and risky decisions: an experimental study grounded in Prospect theory. *International Entrepreneurship and Management Journal*, 2021, vol. 17, pp. 1827–1841. DOI: 10.1007/s11365-020-00697-4.

13. Srivastava, S., Aggarwal, A., & Bansal, P. Efficiency Evaluation of Assets and Optimal Portfolio Generation by Cross Efficiency and Cumulative Prospect Theory. *Computational Economics*, 2024, vol. 63, pp.129–158. DOI: 10.1007/s10614-022-10334-7.

14. Wang, X., Wang, B., Li, T., Li, H., & Watada, J. Multi-criteria fuzzy portfolio selection based on three-way decisions and cumulative prospect theory. *Applied Soft Computing*, 2023, vol. 134, article no. 110033. DOI: 10.1016/j.asoc.2023.110033.

15. Kumar Shakya, A., Pillai, G., & Chakrabarty, S. Reinforcement Learning Algorithms: A brief survey. *Expert Systems with Applications*, 2023, vol. 231, article no. 120495. DOI: 10.1016/j.eswa.2023.120495.

16. Brown, S., Sinha, S., & Schaefer, A. J. Markov decision process design: A framework for integrating strategic and operational decisions. *Operations Research Letters*, 2024, vol. 54, article no. 107090. DOI: 10.1016/j.orl.2024.107090.

17. Wang, J., Wu, C., & Zhong, X. Prospect theory and stock returns: Evidence from foreign share markets. *Pacific-Basin Finance Journal*, 2021, vol. 69, article no. 101644. DOI: 10.1016/j.pacfin.2021.101644.

18. Gisbert-Pérez, J., Martí-Vilar, M., & González-Sala, F. Prospect Theory: A Bibliometric and Systematic Review in the Categories of Psychology in Web of Science. *Healthcare*, 2022, vol. 10, iss. 10, article no. 2098. DOI: 10.3390/healthcare10102098.

19. Li, Y., Liu, P., & Wu, X. Failure mode and effect analysis approach considering risk attitude of dynamic reference point cumulative prospect theory in uncertainty contexts. *Artificial intelligence review*, 2023, vol. 56, iss. 12, pp.14557–14604. DOI: 10.1007/s10462-023-10501-8.

20. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. Stable-Baselines3: Reliable Reinforcement Learning Implementations. *Journal of Machine Learning Research*, 2021, vol. 22, iss. 268, pp. 1–8. Available at: https://jmlr.org/papers/v22/20-1364.html. (accessed 12.08.2024).

21. Fahad Mon, B., Wasfi, A., Hayajneh, M., Slim, A., & Abu Ali, N. Reinforcement Learning in Education: A Literature Review. *Informatics*, 2023, vol. 10, iss. 3, article no.74. DOI: 10.3390/informatics10030074.

22. van Dolder, D., & Vandenbroucke, J. Behavioral Risk Profiling: Measuring Loss Aversion of Individual Investors. *Journal of Banking and Finance*, 2024, article no. 107293. DOI: 10.2139/ssrn.4199169.

Received 17.08.2024, Accepted 18.11.2024

ВИКОРИСТАННЯ ПРОКСИМАЛЬНОЇ ОПТИМІЗАЦІЇ ПОЛІТИКИ ТА ТЕОРІЇ ПЕРСПЕКТИВ ДЛЯ НАВЧАННЯ МОДЕЛІ ПРИЙНЯТТЯ РІШЕНЬ ПО УПРАВЛІННЮ ПЕРСОНАЛЬНИМИ ФІНАНСАМИ

В. В. Дідківський, Д. С. Антонюк, Т. А. Вакалюк, Є. В. Огінський

Предметом цієї статті є розробка моделі прийняття рішень, яка в майбутньому може бути включена в симулятор особистих фінансів для підвищення грамотності в галузі персональних фінансів. **Мета** полягає в тому, щоб розробити моделі прийняття рішень, адаптовані до різних профілів інвесторів, які можуть надавати персоналізовані фінансові поради щодо розподілу активів. У цій статті використовуються методи навчання з підкріпленням і поведінкової економіки для досягнення цієї мети, сприяючи вдосконаленню практичних алгоритмів і підходів до прийняття фінансових рішень. **Завдання** можна сформулювати наступним чином: 1) розробити навчальне середовище з підкріпленням, що містить різні варіанти інвестування з різною середньою

4SSN 1814-4225 (print)
ISSN 2663-2012 (online)

прибутковістю та рівнями волатильності; 2) навчити агент навчання з підкріпленням за допомогою алгоритму проксимальної оптимізації політики визначати рекомендований розподіл інвестицій; 3) впровадити функцію винагороди на основі теорії перспектив, включаючи параметри, які відображають різні ризик-профілі інвесторів, такі як неприйняття втрат і зменшення чутливості до прибутків і втрат. **Результати** включають розробку окремих моделей для 3 ризик-профілів інвесторів: несхильний до ризику, раціональний і максимізуючий капітал. Графічний аналіз рекомендованих відсотків розподілу активів виявив суттєві закономірності впливу параметрів функції вартості теорії перспектив. Практичні наслідки цього дослідження поширюються на розробку інструментів моделювання, які дозволять людям практикувати та вдосконалювати свої фінансові навички в безпечному середовищі. Ці інструменти будуть спрямовані на подоланні розриву в освіті персональних фінансів, надаючи можливості для навчання на власному досвіді. **Висновки**. Розроблена модель ефективно генерує персоналізовані фінансові поради, що відображають індивідуальні рівні схильності до ризику. Майбутня робота буде зосереджена на створенні інтерактивних інструментів моделювання для подальшого вдосконалення навичок управління персональними фінансами. Дослідження підкреслює важливість інтеграції психологічних і поведінкових факторів у моделі прийняття фінансових рішень.

Ключові слова: персональні фінанси; прийняття рішень; навчання з підкріпленням; теорія перспектив.

Дідківський Владислав Валентинович – асп., Державний університет «Житомирська політехніка», Житомир, Україна.

Антонюк Дмитро Сергійович – канд. пед. наук, доц. каф. інженерії програмного забезпечення, Державний університет «Житомирська політехніка», Житомир, Україна.

Вакалюк Тетяна Анатоліївна – д-р пед. наук, проф., зав. каф. інженерії програмного забезпечення, Державний університет «Житомирська політехніка», Житомир, Україна.

Огінський Євген Володимирович –асп., Державний університет «Житомирська політехніка», Житомир, Україна.

Vladyslav Didkivskyi – PhD Student, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine,
e-mail: vladyslav.didkivskyi@gmail.com, ORCID: 0000-0002-4615-7578, Scopus Author ID: 57433774100.

Dmytro Antoniuk – Candidate of Pedagogical Sciences, Associate Professor at the Department of Software Engineering, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine,
e-mail: dmitry_antonyuk@yahoo.com, ORCID: 0000-0001-7496-3553, Scopus Author ID: 57216847072.

Tetiana Vakaliuk – Doctor of Pedagogical Sciences, Professor, Head of the Department of Software Engineering, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine,
e-mail: tetianavakaliuk@gmail.com, ORCID: 0000-0001-6825-4697, Scopus Author ID: 57211133927.

Yevhen Ohinskyi – PhD Student, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine,
e-mail: oginsky2@gmail.com, ORCID: 0000-0002-7777-8449.