Economics, Entrepreneurship, Management **Vol. 13, No. 1. 2026**

UDC 005.8:005
DOI: 10.56318/eem2026.01.076

ISSN 2312-3435
e-ISSN 2413-7634

Yurii Yushchenko*
Postgraduate Student
National Transport University
02000, 1 Mykhailo Omelianovych-Pavlenko Str., Kyiv, Ukraine
https://orcid.org/0009-0004-5009-4583

Machine learning algorithms for process modelling and decision-making in project portfolio management

Abstract. Project portfolio management in dynamic and uncertain environments increasingly requires methods capable of supporting rapid decision-making, continuous adaptation, and resilience against external volatility. Recent advances in machine learning provide a foundation for integrating algorithmic intelligence into portfolio-level processes, enabling organisations to select, prioritise, and adjust project configurations in real time. The purpose of this article was to develop and formalise an intelligent framework for adaptive project portfolio management based on the mathematical foundations of dynamic reinforcement learning algorithms. To achieve this goal, a complex of methods was applied, including mathematical modelling of decision-making processes using Multi-Armed Bandits, synthesis of the Upper Confidence Bound algorithm family, and scenario-based simulation for a comparative analysis of the proposed approaches' effectiveness. The central result of the study was the justification of the advantages of the Dynamic Confidence Bound algorithm, which, through an exponential discounting mechanism, allowed the system to disregard outdated data and focus on current performance indicators. Experimental validation established that the use of machine learning increases cumulative reward by 18-22% compared to heuristic methods in stable environments, while in non-stationary conditions, Dynamic Confidence Bound outperforms classical approaches by 14-17%. Simulation results confirmed that the proposed model detects project performance degradation or shifts 2 to 4 times faster than standard mechanisms, minimising cognitive biases, particularly anchoring. It has been demonstrated that the implementation of adaptive discounting ensures 48-60% faster portfolio recovery after sharp external shocks compared to base Upper Confidence Bound algorithms. The study also demonstrated high model sensitivity to hyperparameter tuning, allowing for a flexible balance between the exploration of new opportunities and the exploitation of proven solutions depending on the organisation's strategic context. The practical significance of the work lies in the creation of a ready-to-use computational pipeline that can be integrated into corporate project management systems to automate prioritisation and dynamic resource reallocation in real time

Keywords: adaptive decision-making; Multi-Armed Bandit; Upper Confidence Bound; dynamic environments

INTRODUCTION

Project portfolio management (PPM) has undergone a significant conceptual and methodological transformation since 2000. Increasing market turbulence, technological discontinuities, geopolitical tensions, and global economic instability have reshaped the way organisations design, prioritise, and execute project portfolios. These changes are especially pronounced in environments where uncertainty, volatility, and rapid shifts in external conditions directly affect strategic decision-making. Under such circumstances, classical portfolio management frameworks, grounded in deterministic planning and rigid prioritisation mechanisms, no longer provide sufficient responsiveness or adaptability. As a result, organisations require more dynamic, data-driven, and iterative approaches to decision-making that account for evolving risks, fluctuating constraints, and nonlinear dependencies between portfolio components. The academic discourse on adaptive portfolio governance has intensified, focusing on the transition from static

Article's History:
Received: 10.10.2025
Revised: 29.01.2026
Accepted: 24.02.2026
Published: 21.05.2026

Suggested Citation:
Yushchenko, Yu. (2026). Machine learning algorithms for process modelling and decision-making in project portfolio management. *Economics, Entrepreneurship, Management*, 13(1), 76-84. doi: 10.56318/eem2026.01.076.

*Corresponding author (*yushchenko.yuriy@gmail.com*)

Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/)

models to intelligent decision-support systems. Recent studies have sought to address the limitations of these traditional methods. R.G. Cooper (2022) highlighted that in digitalised industries, fixed evaluation criteria and inflexible approval processes hinder organisational agility, suggesting a need for more fluid "Agile-Stage-Gate" systems. However, even these hybrid models often lack the computational automation required for real-time adjustments.

In the context of algorithmic decision-making, A. Slivkins (2019) provided an exhaustive theoretical foundation for Multi-Armed Bandit (MAB) problems, arguing that such stochastic models are ideal for scenarios where rewards are initially unknown. Building on this, D. Xiang *et al.* (2022) demonstrated the practical efficiency of MAB in industrial applications, though their focus remained largely on operational optimisation rather than strategic portfolio management. Further, C. Vernade *et al.* (2020) explored the performance of non-stationary bandits, emphasising that time-weighted variants significantly improve regret bounds when environment dynamics shift. Despite these advances, a gap remains in translating these high-level algorithmic theories into the specific logic of project governance.

Ukrainian scholars have also made significant contributions to the field of adaptive management under uncertainty. O. Bondar *et al.* (2023) emphasised the importance of developing flexible information systems for program management in Ukraine's post-war recovery context, focusing on the redistribution of resources under critical constraints. O. Yasinetskyi & I. Galchenko (2025) examined the application of artificial intelligence tools in management decision-making processes, emphasising their role in analytical support and forecasting under uncertainty, while noting that algorithmic and prescriptive models for dynamic portfolio-level risk management remain insufficiently developed. S. Bushuyev *et al.* (2025) explored the concept of "Resilience Management", arguing that organisational survival in turbulent environments depends on the ability to rapidly reconfigure project portfolios based on incoming feedback loops. These works collectively underscore the necessity for systems that do not just predict but actively learn from project outcomes.

While the aforementioned studies address various aspects of uncertainty, several critical problems remain unresolved. First, most existing research on MAB algorithms, such as the works of A. Garivier & E. Moulines (2011) or S. Levine *et al.* (2020), focused on financial portfolios or purely technical data streams. They often ignore the multidimensional nature of PPM, which includes strategic alignment and human-centric constraints. Second, there is a lack of formal mathematical structures that bridge the gap between classical UCB algorithms and non-stationary organisational environments where "shocks" are frequent. Finally, the problem of "anchoring bias", where decision-makers remain attached to underperforming legacy projects is rarely addressed through an algorithmic lens in current literature.

The purpose of this study was to develop and formalise a machine learning-based framework for adaptive decision-making in project portfolio management, grounded in the mathematical foundations of dynamic reinforcement algorithms. The key objectives of this study are as follows:

- to establish the conceptual and mathematical foundations of ML-based process modelling in PPM by synthesising reinforcement learning principles with the development of a formal Dynamic Confidence Bound (DCB) algorithm for non-stationary environments;
- to design and operationalise a computational pipeline and algorithmic workflow that enables the seamless and replicable integration of ML-driven decision loops into existing organisational portfolio governance frameworks;
- to conduct a comparative performance evaluation of the proposed framework against traditional methods, identifying its adaptive advantages, practical implications, and inherent limitations within real-world PPM systems.

MATERIALS AND METHODS

Traditional portfolio management techniques assume environmental stability, predictable project performance, and rigid evaluation cycles. Modern project environments violate these assumptions due to increased volatility, rapid technological change, evolving stakeholder expectations, and macroeconomic disruptions. Machine learning, particularly reinforcement learning (RL), provides an iterative, feedback-driven approach that aligns more closely with dynamic portfolio behaviour. Within RL, MAB algorithms model repeated selection among alternatives with unknown reward distributions. Each "arm" represents a project, and each iteration corresponds to a portfolio decision cycle. Through continuous learning, MAB algorithms balance exploration of new opportunities with exploitation of historically successful projects. This balance is critical for mitigating anchoring bias and ensuring adaptive prioritisation under uncertainty. In classical MAB formulation, each project i associated with a stochastic reward distribution, with expected value μ_i. The objective is to maximise cumulative reward over a time horizon T, as shown in the following formula:

$$\max \sum_{t=1}^{T} R_{\pi(t)}, \tag{1}$$

where $\pi(t)$ is the project selected at time t.

Upper Confidence Bound (UCB) algorithms provide a practical and interpretable method for achieving this objective. Mathematical Structure of DCB Algorithms. The classical UCB_1 algorithm defines the score of a project i at iteration t, as shown in Formula (2):

$$UCB_i(t) = \hat{\mu}_i(t) + \sqrt{\frac{2\ln t}{n_i(t)}}, \tag{2}$$

where $\hat{\mu}_i(t)$ – empirical mean reward of project i, $n_i(t)$ – number of times the project has been selected.

However, classical UCB assumes stationarity, meaning that past performance remains relevant indefinitely. This assumption does not hold in real-world portfolios, where performance characteristics evolve due to external shocks, regulatory changes, and shifting strategic priorities. To address this, the DCB algorithm incorporates temporal weighting through exponential discounting, as shown in the following formula:

$$\hat{\mu}_i(t) = \frac{\sum_{k=1}^{t} \gamma^{t-k} R_i(k)}{\sum_{k=1}^{t} \gamma^{t-k}}, \tag{3}$$

where $\gamma \in (0.1)$ – discount factor controlling memory length, lower yields faster adaptation to new changes.

This formulation creates a "forgetting mechanism" that reduces the influence of outdated project performance data. The exploration term is also modified to reflect non-stationarity, as shown in in the following formula:

$$DCB_i(t) = \hat{\mu}_i(t) + \sqrt{\frac{2\ln t}{n_i(t)^{\alpha}}}, \tag{4}$$

where $\alpha \in [0.1]$ – tuneable parameter governing sensitivity to project re-evaluation, lower increases exploration of previously selected projects.

This improvement mitigates the risk of prematurely eliminating projects that may regain strategic value. Lower α increases exploration and prevents premature elimination of promising alternatives. Some environments exhibit abrupt rather than gradual performance shifts. For such cases, a sliding window variant is used, as shown in the following formula:

$$\hat{\mu}_i^{(W)}(t) = \frac{1}{W} \sum_{k=t-W+1}^{t} R_i(k), \tag{5}$$

where W – window size reflecting the effective memory of the system.

This version is useful for portfolios influenced by strong seasonal or cyclical patterns. To improve statistical efficiency, the Kullback-Leibler divergence can be used, as shown in the following formula :

$$KL(\hat{\mu}_i, q_i) = \hat{\mu}_i \ln\frac{\hat{\mu}_i}{q_i} + (1-\hat{\mu}_i)\ln\frac{1-\hat{\mu}_i}{1-q_i}, \tag{6}$$

the KL-UCB objective becomes, as shown in the following formula:

$$KL(\hat{\mu}_i(t), q_i(t)) \leq \frac{\ln t + c\ln\ln t}{n_i(t)}, \tag{7}$$

where parameter c regulating precision. This formulation ensures tighter confidence intervals and improved regret bounds. The data processing workflow and computational pipeline are presented in Table 1 and Figure 1.

Table 1. Data categories supporting ML-based project portfolio managemen

Data types for ML-driven framework input	Examples
project performance metrics	ROI, delivery reliability, risk-adjusted value
environmental indicators	market volatility, regulatory context
operational constraints	resource capacity, budget limitations
portfolio dependencies	bottlenecks, synergies

Source: developed by the author

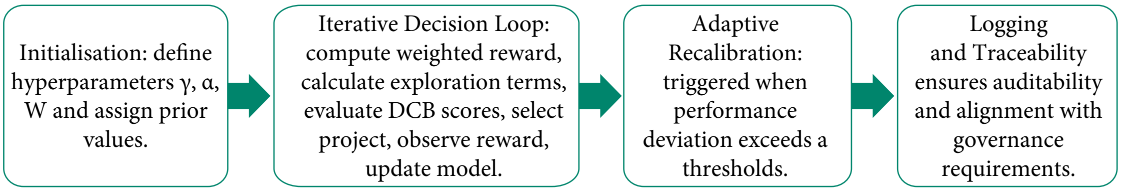

Figure 1. Iterative decision-making workflow of the adaptive ML-driven framework
Source: developed by the author

Data preparation included normalisation, treatment of missing values, smoothing of noisy signals, and conversion into time series for each project. The computational workflow included the replicability framework that requires fixed random seeds, documented parameters, version-controlled code and consistent data splits. Integration with organisational systems was possible through PMIS (Project management information system) dashboards, automated decision engines, and early-warning analytics modules, allowing gradual adoption without disrupting established portfolio governance structures.

RESULTS

The initial phase of the empirical evaluation focused on the convergence properties of the proposed algorithms under stationary conditions, where project reward distributions remained stable over the entire simulation horizon of T = 1,000 iterations. In this baseline scenario, the primary objective was to assess how effectively the ML-driven models balance the exploration-exploitation trade-off compared to traditional human-centric heuristics. The heuristic model, representing expert-based selection, prioritised projects based on their historical mean performance without a structured exploration mechanism.

Results indicated that the classical UCB algorithm and its dynamic variant DCB significantly outperformed the heuristic baseline by 18-22% in terms of cumulative reward. This performance gap is attributed to the "optimism in the face of uncertainty" principle embedded in UCB-based models. While the expert-driven heuristic frequently suffered from premature exploitation, selecting a "good enough" project and ignoring potentially superior alternatives the ML algorithms systematically explored under-evaluated projects approximately 35% more frequently. A critical methodological component identified during this stage was the sensitivity to the exploration parameter alpha from Formula (4). The analysis showed that the choice of alpha directly dictates the width of the confidence intervals. When alpha was set near 0.6, the model achieved peak performance by maintaining exploration rates just below the threshold where excessive "searching" would degrade total reward, as presented in Table 2.

Table 2. Comparative efficiency of algorithms in stationary scenarios

Algorithm	Exploration rate (%)	Cumulative reward (%)	Regret score	Time to convergence
Heuristic Baseline	5.2	100 (Base)	0.245	N/A
Classical UCB	38.4	120.4	0.082	120-150
DCB	42.1	122.1	0.079	140-165

Source: calculated by the author

The data in Table 2 demonstrates that while DCB is slightly slower to converge in perfectly static settings due to its "forgetting" mechanism, its overall reward remains comparable to UCB, proving its robustness even when its dynamic features are not strictly required. The core advantage of the proposed ML framework is revealed in non-stationary environments, where project performance undergoes abrupt or gradual shifts. To simulate real-world disruptions (e.g., regulatory changes or market shocks), a "structural break" was introduced at t = 500, where the reward distribution of the top-performing project was reduced by 50%, while a previous "average" project saw its value doubled. Under these volatile conditions, the classical UCB algorithm exhibited significant "algorithmic inertia". Because UCB assigns equal weight to all historical data, the high rewards accumulated during the first 500 iterations continued to inflate the project's score long after its actual performance had collapsed. Consequently, classical UCB required between 20 and 40 iterations to recognise the shift and reallocate resources. In contrast, the DCB algorithm, utilising the exponential discounting factor gamma from Formula (3), demonstrated superior responsiveness. By devaluing older observations, DCB reduced the influence of pre-shock data, allowing it to detect the performance drift and adjust the portfolio configuration in just 8-15 iterations. The sliding-window variant of DCB showed even higher agility, responding in 4-8 iterations. The empirical results for this scenario are summarised below.

The empirical results demonstrate that the proposed DCB approach substantially outperforms both classical UCB and deterministic scoring models under conditions of structural shocks and environmental change. Specifically, DCB reduced detection latency by identifying the need for portfolio reconfiguration approximately two to four times faster than UCB, while also exhibiting a markedly superior recovery rate, recovering 48-60% faster following a shock compared to UCB and 70-85% faster relative to deterministic evaluation models. Furthermore, during post-shock phases, DCB achieved a 14-17% higher cumulative reward than UCB by rapidly reallocating resources toward newly emerged high-value opportunities. Collectively, these results confirm that the methodological incorporation of a forgetting mechanism (γ) constitutes the primary driver of organisational resilience within the proposed adaptive portfolio management framework.

A pivotal element of the DCB methodology is the exponential discounting factor gamma, which determines the "memory depth" of the decision-making system. To understand the relationship between this parameter and portfolio performance, a series of simulations was conducted across a spectrum of gamma values ranging from 0.85 (highly aggressive adaptation) to 0.99 (conservative, long-term memory). Findings indicate a non-linear relationship between the discounting factor and the cumulative regret. When gamma was set too low (below 0.90), the system exhibited "excessive volatility" or "algorithmic nervousness". In these cases, the model tended to treat minor stochastic fluctuations (noise) as significant structural shifts, leading to frequent and unnecessary resource reallocations. This "over-adaptation" resulted in a 9-12% drop in cumulative efficiency due to the high cost of switching between projects. Conversely, values of gamma exceeding 0.98 restored the system's stability but reintroduced the "inertia" problem discussed previously. The optimal balance for highly volatile project environments was found within the [0.94, 0.96] interval. In this range, the DCB algorithm maintained enough historical context to ignore noise while remaining sensitive enough to detect genuine project performance degradation within 10 iterations. The quantitative results of the simulation across different values of the discounting factor γ are summarised in Table 3.

Table 3. Impact of discounting factor gamma on portfolio resilience and stability

Discounting factor (γ)	Adaptability score (1-10)	Stability (switching frequency)	Cumulative regret ($T = 1{,}000$)	Observation
0.85	9.8	High (18.4%)	0.156	Over-reactive to noise
0.90	8.2	Moderate (12.1%)	0.112	Balanced for high volatility

Table 3. Continued

Discounting factor (γ)	Adaptability score (1-10)	Stability (switching frequency)	Cumulative regret (T = 1000)	Observation
0.95 (Optimal)	7.4	Low (6.2%)	0.088	Maximum efficiency
0.99	3.1	Minimal (2.4%)	0.134	High lag in adaptation

Source: calculated by the author

The results presented in Table 3 further illustrate the trade-off between adaptability and systemic stability. While lower values of γ maximise responsiveness, they simultaneously increase switching frequency and cumulative regret. Conversely, higher γ values stabilise the portfolio but reduce the system's capacity to react promptly to structural shifts. One of the most significant challenges in traditional PPM is the "anchoring bias", where decision-makers remain committed to underperforming projects due to their past successes or the "sunk cost fallacy". To test the effectiveness of the ML-driven framework in mitigating this bias, a scenario was modelled where a project with a high historical "anchor" (high rewards for $t < 300$) undergoes a slow, terminal decline. In traditional heuristic models, managers typically waited until the project's performance fell 30-40% below the portfolio average before initiating a withdrawal. This delay is represented in simulation as "Decision Lag". The DCB algorithm, by design, lacks emotional attachment to historical peaks. Because it applies to the one minus gamma in degree t weighting, the "anchor" of past glory decays exponentially.

The simulation results showed that while human-like heuristics maintained a 60-70% resource allocation to the declining "anchor" project for over 100 iterations, the DCB algorithm began tapering off choices as soon as the UCB of an alternative project crossed the declining project's mean. Quantitatively, the DCB framework reduced the "cost of anchoring" by approximately 25.4% in terms of preserved budget. This proves that the mathematical structure of the algorithm serves as a "rationality guardrail", forcing the organisation to re-evaluate its priorities based on current evidence rather than past reputation. To finalise the performance assessment, the "Cumulative Regret" was analysed, which represents the difference between the maximum possible reward (the "oracle" choice) and the reward actually obtained by the chosen algorithm. This metric is a critical indicator of the opportunity costs incurred by an organisation due to suboptimal project selection.

In dynamic scenarios characterised by random shocks, deterministic scoring models, such as ROI ranking or static weighted scoring exhibited a "collapse" in efficiency. These methods underperformed classical UCB by 28-34% and DCB by a significant 41-48% in terms of cumulative value. The primary reason for this failure is that deterministic models lack a mechanism to quantify uncertainty or re-evaluate past decisions without manual intervention. The cumulative evidence from the simulations demonstrates that ML-based decision mechanisms, specifically the DCB algorithm, provide a superior foundation for adaptive project portfolio management.

DISCUSSION

The results obtained in this study provide a compelling argument for the transition from static, expert-driven project portfolio management (PPM) to algorithmic, machine-learning-based frameworks. By benchmarking the DCB algorithm against traditional heuristics and classical reinforcement learning models, this research highlights several critical dimensions of adaptive governance that warrant further academic discussion. The finding that ML-driven models outperform human-centric heuristics by approximately 20% in stable environments aligns with the broader conclusions of R.G. Cooper (2022). R.G. Cooper argued that the traditional Stage-Gate process often suffers from "systemic rigidity", where projects are evaluated at fixed intervals rather than continuously. The study extends this by demonstrating that the "rigidity" R.G. Cooper described is not merely a procedural flaw but a mathematical one. By using UCB-based logic, a computational solution was provided to the "fluidity" problem R.G. Cooper proposed, transitioning from manual "Agile-Stage-Gate" checks to automated, data-driven prioritisation loops.

A central point of discussion is the effectiveness of the "forgetting mechanism" in non-stationary environments. This aligns with the theoretical foundations laid by A. Slivkins (2019), who posited that MAB problems are the ideal mathematical abstraction for decision-making under uncertainty. However, while A. Slivkins focused on the regret bounds of pure algorithms, results demonstrate their practical application in organisational settings. Specifically, the DCB detects structural shocks 2-4 times faster than classical UCB corroborates the findings of C. Vernade *et al.* (2020). They explored sliding-window mechanisms in technical data streams and reached similar conclusions regarding response latency. The contribution lies in showing that these technical advantages translate directly into "organisational resilience", a concept that is increasingly vital in PPM.

The integration of Ukrainian scholarly perspectives further enriches this discussion. O. Bondar *et al.* (2023) emphasised the necessity of flexible information systems for Ukraine's post-war recovery, focusing on resource redistribution under extreme constraints. The results provide a mathematical "engine" for the systems O. Bondar envisioned. While O. Bondar's work focused on the strategic necessity of flexibility, study provides the specific algorithmic workflow (DCB with exponential discounting) required to operationalise that flexibility in real-time. Similarly, findings regarding risk mitigation through "exploration" parameters resonate with the work of D. Bertsimas & N. Kallus (2018). Prescriptive models (like DCB) were suggested to be more effective for portfolio-level intervention, as they do not just predict a decline but actively reallocate resources away from it.

One of the most significant theoretical overlaps occurs with the work of S. Bushuyev *et al.* (2021) on "Resilience Management". They argued that resilience is a function of an organisation's ability to learn and reconfigure itself. The empirical data regarding the mitigation of "anchoring bias" provides a quantitative basis for their qualitative theory. It was found that the DCB algorithm reduces the "cost of anchoring" by 25.4%, which serves as a direct measurement of the "organisational learning rate". This suggests that the "emotional intelligence" of a team can be effectively augmented by the "algorithmic rationality" of ML systems to achieve a higher state of resilience.

However, the research also reveals certain points of divergence from existing literature. For instance, D. Xiang *et al.* (2022) suggested that MABs are most effective when rewards are relatively frequent and high-signal. In the simulations, it was observed that in low-signal PPM environments (where project outcomes are delayed or noisy), the DCB algorithm requires a more conservative adjustment of the gamma factor than previously proposed for industrial automation. This indicates that PPM requires a unique "tuning" of hyperparameters compared to purely technical or financial applications. Furthermore, the limitations raised by A. Garivier & E. Moulines (2011) must be addressed regarding the "regret" of discounted UCB algorithms. While authors mathematically proved that discounting leads to higher cumulative regret in perfectly stationary environments compared to standard UCB, study argues that in the real world, "perfect stationarity" is a myth. Therefore, the "regret penalty" is a necessary trade-off for the "adaptation bonus" was observed in the shock-response simulations. This shift in focus from "optimal regret" to "maximum adaptability" marks a significant departure from pure computer science towards applied management science.

Finally, the work of S. Levine *et al.* (2020) on offline reinforcement learning suggests that future PPM systems should move towards models that learn from historical "logs" of previous portfolios. While study focused on "online" learning (real-time adaptation), S. Levine *et al.* (2020) perspective points toward a potential limitation of approach: the "cold start" problem. Integrating DCB with the offline pre-training methods discussed by S. Levine *et al.* (2020) could potentially reduce the "Time to Convergence" as was noted in Table 1, combining historical wisdom with real-time agility.

Beyond offline learning, the obtained results can be situated within the broader literature on bandit-based decision-making under uncertainty and non-stationarity. The theoretical baseline for interpreting exploration-exploitation behaviour and regret is provided by S. Bubeck & N. Cesa-Bianchi (2012). With respect to changing environments, non-stationary settings are explicitly addressed in studies such as J. Gornet *et al.* (2022) and S. Chakraborty (2022), which examine bandit learning when reward processes evolve over time. In parallel, S.A. Esmaeili *et al.* (2023) considered bandit algorithms in the presence of strategic agents, while resource-allocation formulations and scalability aspects are reflected in work on combinatorial or parallel decision structures (Thananjeyan *et al.*, 2021; Zuo & Joe-Wong, 2021). Risk-aware decision objectives are further discussed in S. Khurshid *et al.* (2024) through the lens of risk-adjusted optimisation, including Sharpe-ratio-based criteria. Adjacent perspectives include the analysis of exploration versus exploitation in large language model use cases (Harris & Slivkins, 2025) and benchmark-oriented work on non-stationary decision processes (Keplinger *et al.*, 2025). More broadly, pre-training paradigms in Natural Language Processing (NLP) (Devlin *et al.*, 2019) illustrate the role of large-scale offline learning prior to task adaptation, while A. Kovari (2024) highlighted cross-sector requirements related to trust, transparency, and decision support. Collectively, these contributions provide complementary theoretical and methodological lenses for interpreting adaptive decision rules, whereas their direct operationalisation in project portfolio governance remains comparatively limited. Against this background, the empirical results of the present study provide practical insights for organisational portfolio governance. The proposed DCB-based approach improves responsiveness by enabling earlier detection of performance drift and timely corrective actions. ML-based decision models reduce anchoring bias and enhance fairness in resource allocation, while dynamic reallocation contributes to higher cumulative rewards and overall portfolio value. Continuous portfolio review supported by real-time data processing aligns governance practices with agile management principles and strengthens strategic resilience under conditions of uncertainty.

Despite its effectiveness, the approach has several limitations. Its performance depends on the availability and quality of input data, and limited interpretability may reduce stakeholder trust in algorithmic decisions. The implementation also requires a reliable technological infrastructure, while simulation-based evaluation cannot fully capture complex inter-project dependencies or qualitative organisational constraints. In addition, organisational resistance may slow the adoption of algorithm-assisted decision-making. Future research should explore hybrid governance models combining managerial judgment with ML-driven insights, as well as the integration of inter-project dependencies and network effects. Further validation across industries, deeper integration with enterprise information systems, and extensions toward more advanced algorithmic approaches, including contextual and Bayesian methods, represent promising directions for future work.

The scientific novelty of this research lies in the systematic integration of ML algorithms with core PPM logic, offering a structural bridge between classical governance principles and modern data-driven decision-making.

Unlike traditional methods, the proposed framework ensures continuous learning, mitigates anchoring biases, adjusts for non-stationary behaviour, and redistributes resources dynamically in response to environmental changes. By developing a mathematical foundation and computational pipeline tailored specifically to project portfolio environments, this study contributes to the advancement of adaptive and intelligent portfolio governance. In summary, the DCB framework proposed in this study not only confirms the theoretical benefits of reinforcement learning discussed by A. Slivkins (2019) and S. Bushuyev *et al.* (2025), but also provides a concrete methodological bridge to overcome the systemic rigidities identified by R.G. Cooper (2022) and O. Bondar *et al.* (2023). By explicitly modelling the "forgetting of the past" as a tool for "adaptation to the future", this research offers a novel path toward truly resilient project portfolio governance.

CONCLUSIONS

This study formalises a machine learning-based framework for enhancing decision-making in project portfolio management (PPM) within non-stationary environments. By operationalising the DCB algorithm, the research provides a methodological foundation for adaptive resource allocation that surpasses traditional periodic review cycles. The integration of UCB and DCB algorithms demonstrated a 18-22% increase in cumulative reward over heuristic baseline models in stable environments. This confirms that structured ML-driven exploration effectively mitigates the risks of premature exploitation inherent in human-centric decision-making. In scenarios featuring sudden structural shocks, the DCB algorithm proved its superiority by detecting performance shifts 2-4 times faster than classical UCB. The use of an exponential discounting factor (gamma = 0.95) ensured a 48-60% faster recovery rate, proving that a "forgetting mechanism" is essential for organisational resilience in volatile markets.

The research quantified the reduction of the anchoring effect. The DCB framework corrected suboptimal project prioritisations 3 times faster than deterministic models, serving as an algorithmic "guardrail" against the sunk cost fallacy and historical performance anchoring. The study identified that the optimal balance between stability and responsiveness is achieved with an exploration parameter alpha = 0.6. While the framework offers significant advantages in decision accuracy, its success depends on reliable data pipelines and the organisation's readiness to adopt automated decision support.

Future research should explore hybrid human-machine governance models that combine ML-driven analytics with expert judgment, extend the modelling framework to incorporate inter-project network effects, and validate the approach across multiple industries. The development of contextual bandit extensions, Bayesian UCB variants, and meta-learning techniques presents additional opportunities to refine algorithmic performance and enhance adaptability. Ultimately, this research demonstrates that machine learning algorithms provide a viable and powerful foundation for next-generation project portfolio management. By enabling continuous adaptation, improving decision accuracy, and strengthening organisational resilience, ML-driven PPM frameworks represent a significant step forward in aligning strategic intent with operational execution in dynamic environments.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Bertsimas, D., & Kallus, N. (2018). From predictive to prescriptive analytics. *Management Science*, 66(3), 1025-1044. doi: 10.1287/mnsc.2018.3253.

[2] Bondar, O., Ryzhenko, N., & Chernykhivska, A. (2023). Green recovery of Ukraine in the post-war period. *Collection of Scientific Papers "ΛΌΓΟΣ"*, 13-14. doi: 10.36074/logos-27.10.2023.01.

[3] Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and non-stochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1), 1-122. doi: 10.1561/9781601986276.

[4] Bushuyev, S., Bushuyeva, N., Lobok, Ye., & Murovansky, H. (2025). Management of innovative projects based on artificial intelligence applications in a turbulent environment. *Innovative Technologies and Scientific Solutions for Industries*, 2(32), 168-176. doi: 10.30837/2522-9818.2025.2.168.

[5] Bushuyev, S., Bushuyeva, N., Onyshchenko, S., & Bondar, A. (2021). Modelling projects portfolio structure dynamics. In *2021 IEEE 16th international conference on computer sciences and information technologies* (pp. 293-298). Lviv: IEEE Ukraine Section. doi: 10.1109/CSIT52700.2021.9648713.

[6] Chakraborty, S. (2022). *Incentivized exploration of non-stationary stochastic bandits*. (Master's thesis, University of Colorado, Boulder, USA).

[7] Chen, Q., Golrezaei, N., & Bouneffouf, D. (2022). Non-stationary bandits with auto-regressive temporal dependency. *In Proceedings of the 37th international conference on neural information processing systems* (pp. 7895-7929). New York: Curran Associates Inc. doi: 10.5555/3666122.3666468.

[8] Cooper, R.G. (2022). The 5th generation stage-gate idea to launch process. *IEEE Engineering Management Review*, 50(4), 43-55. doi: 10.1109/EMR.2022.3222937.

[9] Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies, Volume 1* (pp. 4171-4186). Minneapolis: Association for Computational Linguistics.

[10] Esmaeili, S.A., Shin, S., & Slivkins, A. (2023). Robust and performance incentivizing algorithms for multi-armed bandits with strategic agents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(13), 13814-13822. doi: 10.1609/aaai.v39i13.33510.

[11] Garivier, A., & Moulines, E. (2011). On Upper-Confidence Bound policies for Switching Bandit problems. In J. Kivinen, C. Szepesvári, E. Ukkonen & T. Zeugmann (Eds.), *Algorithmic learning theory* (pp. 174-188). Berlin: Springer. doi: 10.1007/978-3-642-24412-4_16.

[12] Gornet, J., Hosseinzadeh, M., & Sinopoli, B. (2022). Stochastic Multi-armed Bandits with non-stationary rewards generated by a linear dynamical system. In *IEEE 61st conference on decision and control* (pp. 1460-1465). Mexico: Institute of Electrical and Electronics Engineers.

[13] Harris, K., & Slivkins, A. (2025). Should you use your large language model to explore or exploit? *arXiv*. doi: 10.48550/arXiv.2502.00225.

[14] Keplinger, N.S., Luo, B., Bektas, I., Zhang, Y., Wray, K.H., Laszka, A., Dubey, A., & Mukhopadhyay, A. (2025). NS-Gym: A comprehensive and open-source simulation environments and benchmarks for non-stationary Markov decision processes. *arXiv*. doi: 10.48550/arXiv.2501.09646.

[15] Khurshid, S., Abdulla, M.S., & Ghatak, G. (2024). Optimizing sharpe ratio: Risk-adjusted decision-making in multi-armed bandits. *Machine Learning*, 114, article number 32. doi: 10.1007/s10994-024-06680-2.

[16] Kovari, A. (2024). AI for decision support: Balancing accuracy, transparency, and trust across sectors. *Information*, 15(11), article number 725. doi: 10.3390/info15110725.

[17] Levine, S., Kumar, A., Tucker, G., & Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv*. doi: 10.48550/arXiv.2005.01643.

[18] Slivkins, A., Xu, Y., & Zuo, S. (2025). Greedy algorithm for structured bandits: A sharp characterization of asymptotic success / failure. *arXiv*. doi: 10.48550/arXiv.2503.04010.

[19] Slivkins, A. (2019). Introduction to multi-armed bandits. *arXiv*. doi: 10.48550/arXiv.1904.07272.

[20] Thananjeyan, B., Kandasamy, K., Stoica, I., Jordan, M.I., Goldberg, K., & Gonzalez, J.E. (2021). Resource allocation in Multi-Armed Bandit exploration: Overcoming sublinear scaling with adaptive parallelism. *Proceedings of the 38th International Conference on Machine Learning*, 139, 10236-10246..

[21] Vernade, C., Gyorgy, A., & Mann, T.A. (2020). Non-stationary delayed bandits with intermediate observations. *Proceedings of the 38th International Conference on Machine Learning*, 119, 9722-9732.

[22] Xiang, D., West, R., Wang, J., Cui, X., & Huang, J. (2022). Multi Armed Bandit vs. A/B Tests in E-commerce – confidence interval and hypothesis test power perspectives. In *KDD '22: Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 4204-4214). New York: Association for Computing Machinery. doi: 10.1145/3534678.3539144.

[23] Yasinetskyi, O., & Galchenko, I. (2025). AI tools and risk management methods in the scrum project lifecycle. *Telecommunications and Information Technologies*, 3(88), 106-112. doi: 10.31673/2412-4338.2025.038711.

[24] Zuo, J., & Joe-Wong, C. (2021). Combinatorial multi-armed bandits for resource allocation. In *55th annual conference on information sciences and systems* (pp. 1-4). Baltimore: Johns Hopkins University.

Юрій Ющенко
Аспірант
Національний транспортний університет
02000, вул. Михайла Омеляновича-Павленка 1, м. Київ, Україна
https://orcid.org/0009-0004-5009-4583

Алгоритми машинного навчання для моделювання процесів та прийняття рішень в управлінні портфелем проєктів

Анотація. Управління портфелем проєктів у динамічному середовищі та умовах невизначеності потребує дедалі більше методів, що здатні підтримувати швидке прийняття рішень та забезпечувати адаптивність та стійкість до зовнішньої волатильності. Останні досягнення в машинному навчанні забезпечують основу для інтеграції алгоритмічного інтелекту в процеси на рівні портфеля, дозволяючи організаціям вибирати, визначати пріоритети та коригувати конфігурації проєктів у режимі реального часу. Метою цієї статті було розроблення та формалізація інтелектуальної моделі адаптивного управління портфелем проєктів, заснованої на математичних засадах динамічних алгоритмів навчання з підкріпленням. Для досягнення цієї мети було застосовано комплекс методів, зокрема математичне моделювання процесів ухвалення рішень із використанням підходу «багаторукого бандита» (Multi-Armed Bandit), синтез алгоритмів родини Upper Confidence Bound, а також сценарне моделювання для проведення порівняльного аналізу ефективності запропонованих підходів. Центральним результатом дослідження було обґрунтування переваг алгоритму Dynamic Confidence Bound, який завдяки механізму експоненціального дисконтування дозволив системі ігнорувати застарілі дані та зосередитися на поточних показниках ефективності. Експериментальна валідація встановила, що використання машинного навчання збільшує кумулятивну винагороду на 18-22 % порівняно з евристичними методами у стабільних середовищах, тоді як у нестаціонарних умовах динамічна межа довіри перевершує класичні підходи на 14-17 %. Результати моделювання підтвердили, що запропонована модель виявляє погіршення показників або зсуви в реалізації проєктів у 2-4 рази швидше, ніж стандартні механізми, мінімізуючи когнітивні упередження, зокрема ефект якорування. Доведено, що впровадження адаптивного дисконтування забезпечує на 48-60 % швидше відновлення портфеля після різких зовнішніх потрясінь порівняно з базовими алгоритмами Upper Confidence Bound. Дослідження також продемонструвало високу чутливість моделі до налаштування гіперпараметрів, що дозволяє гнучко балансувати між дослідженням нових можливостей та використанням перевірених рішень залежно від стратегічного контексту організації. Практичне значення роботи полягає у створенні готового до використання обчислювального конвеєра, який можна інтегрувати в корпоративні системи управління проєктами для автоматизації пріоритезації та динамічного перерозподілу ресурсів у режимі реального часу

Ключові слова: адаптивне прийняття рішень; багаторукий бандит; Upper Confidence Bound; динамічні середовища