UtilitasMathematica

ISSN 0315-3681 Volume 122 (2), 2025

Budget Reallocation Strategies for Programmatic Advertising Using Reinforcement Learning and Historical ROAS Signals

Surya Narayana Reddy Chintacunta,

Manager - Data & Analytics, WPP Media,

surya.nreddy.ds@gmail.com

Abstract

Programmatic advertising campaigns need budget changes because of performance and consumer actions. Usual methods with set budgets plus fixed rules do not work when audiences or competition change. This paper describes a reinforcement learning framework that reallocates budget to different times of the week and day. The framework uses historical Return on Ad Spend but also sales lift data for optimization.

The system uses a Deep Q-Network (DQN) with several hidden layers and experience replay for budget work. This study employs a DQN model, trained on e-commerce campaign data. The data held $1.5 million in advertising spend and 1.2 million impressions across 12 months. The RL agent targets segments that show high expected ROAS, plus it keeps cost limits. The agent adjusts budgets every hour - this approach raised ROAS by up to 21.8 % and dropped cost-per-acquisition (CPA) by 18.4 %. This occurred when compared to a constant baseline but also a rule-based system. A web-based Opportunity Dashboard also shows future sales growth and the best budget changes as they happen.

The results show that over a variety of time periods, RL-driven budget reallocation significantly increases campaign effectiveness and reduces unnecessary advertising expenditures. The study concludes with a detailed analysis of deployment factors, including whether to expand to multi-channel campaigns on various digital advertising platforms, how to create rewards, and how frequently to retrain the model.

Keywords: Reinforcement learning, Programmatic advertising, Budget optimization, ROAS, Q-learning, Real-time bidding, Temporal segmentation, Dynamic optimization, Campaign management, AdTech

1. Introduction

Programmatic advertising, the cornerstone of modern digital marketing, has fundamentally altered how businesses engage with their target audience across a range of touchpoints [1]. From its estimated $155 billion in 2021 to $725 billion by 2026, the programmatic advertising market is projected to expand at a compound annual growth rate of 36.2% [2]. This quick growth emphasizes how important it is to optimize programmatic campaigns to minimize wasteful spending and maximize return on investment.

However, conventional budget allocation strategies such as equal pacing across time periods or simple heuristic rules based on historical averages do not consider the complex temporal fluctuations in audience responsiveness and competitive dynamics [3]. These traditional approaches disregard compelling evidence that performance peaks during hours or days, treating all time periods as equally valuable due to intricate user behavior patterns, rival bidding strategies, and market conditions [4]. While e-commerce platforms often see noticeably higher conversion rates during lunch and evening hours, B2B businesses may perform best during weekday business hours [5].

Although prior studies have successfully used reinforcement learning to optimize individual bid amounts in real-time auctions [6], [7], there is a lack of research on RL-based budget reallocation across temporal dimensions while remaining grounded in real-world campaign metrics and operational constraints. Because budget allocation decisions often have a bigger impact on campaign performance than individual bid optimizations, this is a significant gap in the literature [8]. This work closes this critical gap while maintaining strict cost-efficiency constraints and operational viability by introducing a novel two-stage reinforcement learning framework that dynamically reallocates hourly and daily budgets based on historical ROAS and sales lift data. This approach goes beyond simple bid optimization to tackle the more challenging issue of determining when and how much to spend during different temporal segments. The study's hypothesis is that an RL agent can outperform static and heuristic pacing methods by learning fine-grained temporal patterns that are invisible to traditional rule-based systems. This study's contributions include:

- A comprehensive RL model using Deep Q-Networks to make hourly budget allocation decisions across 168 distinct time segments representing every hour of the week.
- A rigorous evaluation using a full year of authentic e-commerce campaign data.
- A practical live dashboard system to support real-time budgeting decisions.
- Quantitative analysis demonstrating substantial improvements in ROAS and CPA metrics compared to industry-standard approaches.

2. Literature Review

Compared to bid optimization and audience targeting, budget allocation is still a relatively unexplored area, even though campaign optimization has advanced significantly because of the convergence of machine learning and digital advertising [9]. Traditional budget pacing techniques have dominated the industry for many years, primarily due to their simplicity and predictability rather than their effectiveness [10].

Equal pacing, the most widely used tactic, allocates advertising budgets equally over predetermined periods of time, typically days or weeks [11]. Although this strategy ensures continuous visibility and prevents budget exhaustion too early in a campaign, it completely ignores temporal variations in audience behavior and competitive dynamics. Johnson et al. [12] found that equal pacing can result in up to 35% less optimal performance than more sophisticated allocation strategies.

More advanced conventional methods adjust spending based on historical performance metrics using rule-based heuristics [13]. These systems typically increase budgets during periods of historically high performance and decrease spending during historically underperforming segments. However, these approaches remain fundamentally reactive and fragile in dynamic environments where user behavior patterns change rapidly [14]. Martinez and Chen [15] claim that rule-based systems commonly misallocate advertising budgets because they are unable to adapt to shifting consumer preferences, competitive responses, and seasonal changes. With a primary focus on real-time bidding optimization, the use of reinforcement learning (RL) in digital advertising has accelerated recently [16]. Multi-armed bandit algorithms were first used for display advertising by Chen et al. [6], who showed notable increases in click-through rates and conversion performance. Although their work only addressed bid amount optimization and not budget allocation, it laid the groundwork for more complex RL applications in programmatic advertising.

Subsequent research by Cai et al. [7] extended the use of RL applications to real-time bidding strategies, developing algorithms that could adapt to changing auction conditions and competition. Their Deep Q-Network approach showed promising results in simulated environments, increasing campaign efficiency metrics by as much as 15%. More recent studies have focused on the combination of budget management constraints and RL techniques [17], [18]. Bhosale [19] developed algorithms for real-time budget allocation in AdTech campaigns, giving overall spend caps precedence over exact temporal reallocation. This work made important contributions to the field, but instead of allocating resources optimally across time periods based on expected performance, it focused on budget pacing within daily constraints.

The concept of temporal segmentation in advertising has been extensively studied from the perspective of consumer behavior [20], [21]. Research continuously shows that different times of day and week have a substantial impact on user engagement, conversion rates, and purchase intent [22]. For example, mobile commerce activity peaks during lunch breaks and commutes, while desktop-based business-to-business (B2B) interactions are at their peak during regular business hours [23]. However, by treating these temporal patterns as static rather than dynamic, the majority of current research overlooks opportunities to adjust allocation strategies as patterns change.

Emerging techniques such as federated learning [24], [25] offer promising opportunities for privacy-preserving temporal budget allocation optimization. These methods allow for the decentralized training of models across media platforms or brand ecosystems, eliminating the need to centrally aggregate sensitive user-level data. In large-scale advertising contexts where cross-platform learning is essential but constrained by privacy regulations, federated learning could allow collaborative intelligence without endangering compliance. Furthermore, segmentation and dynamic pricing models which are essential to digital advertising and personalization could incorporate deep reinforcement learning (DRL), which has proven effective in real-time pricing and bid adjustments based on shifting customer behavior [27]. Although these DRL models are commonly used in bidding environments, the underlying frameworks of these models can be adapted to handle hourly budget reallocation problems. By integrating such mechanisms into models for budget allocation and temporal segmentation, ad delivery strategies become more responsive.

This work expands on earlier research by structuring the RL decision-making process around the budget allocation for each hour and considering a range of performance indicators and operational constraints. Unlike previous approaches that focus on either bidding or high-level budget management, this framework addresses the critical middle layer of temporal budget allocation that significantly affects campaign performance. This raises the bar for astute campaign orchestration and builds on the foundations of market segmentation theory [26] as well as recent advancements in autonomous decision systems and distributed learning.

3. Methodology

This section outlines the core elements of the research, such as the data used, the design of the reinforcement learning framework, and the comparative methods employed. Building a robust system that could strategically reallocate programmatic advertising budgets while closely evaluating its performance in relation to preset benchmarks was the aim of this study.

3.1. Information and Experimental Configuration

An extensive dataset of previous advertising campaigns from a well-known American online retailer that specializes in consumer electronics and home goods was used in this study. Researchers were able to investigate a variety of market conditions, including seasonal variations, holiday shopping patterns, and typical periods of high and low demand, thanks to this dataset, which covered the entire year 2024. For the model to adapt to different consumer behaviors throughout the year, this wide temporal coverage was necessary.

The dataset was large, with 1.2 million individual bid events covering a wide range of product categories, audience demographics, and ad creative formats. This contains 140,000 completed purchase transactions, allowing for a precise Return on Ad Spend (ROAS) computation. The $1.5 million total advertising expenditure allowed the analysis to yield statistically significant results that were representative of mid-market e-commerce advertising budgets.

Key performance metrics like impression counts, click-through rates, and conversion events were recorded hourly for each data record to compute ROAS values. To provide more thorough context, contextual elements such as the day of the week, the time of day, indicators of competitor activity, and seasonal markers are included. This large feature set allowed the reinforcement learning (RL) model to detect complex patterns that are often missed by conventional rule-based systems.

To ensure the consistency and integrity of the data, several important preprocessing steps were completed. All monetary values were normalized to account for seasonal price fluctuations. Outlier detection algorithms were employed to identify and remove anomalous events that could distort model training. For the small percentage of missing values (less than 2%), time-series interpolation techniques are used to preserve the inherent temporal patterns.

3.2. Reinforcement Learning Framework Design

The challenging task of allocating funds for programmatic advertising is presented using this method as a sequential decision-making problem. In this setup, an intelligent agent learns to determine the optimal spending amounts for each hour of the week. This formulation allows the model to flexibly adapt to shifting market dynamics and capture intricate temporal patterns. The agent makes decisions about how to spend its budget at any given moment based on its overall assessment of the current state, or "state" (S_t). This state shows the following to give precise temporal context:

- The current hour of the week, which ranges from 1 to 168.

- The agent can identify daily and monthly cycles with the aid of the day of the month.
- Historical ROAS values over the last 24 hours, which provide information on current performance patterns.
- The remaining daily budget and cumulative spend for the current day, which are critical for maintaining financial discipline.
- Competitor activity levels, providing a normalized indicator of market competition.

This thorough state representation enables the agent to make well-informed decisions by simultaneously considering temporal patterns, competitive landscape, budgetary constraints, and recent campaign performance.

The "action space" (A_t) defines the agent's budget allocation options, which aim to balance computational efficiency and practical utility. The possible spending levels were separated into three groups for the following two hours:

- Low allocation: Directing 0-20% of the available budget.
- Medium allocation: Directing 20-60% of the available budget.
- High allocation: Directing 60-100% of the available budget. This design allows for meaningful budget adjustments while ensuring the system remains computationally manageable and operationally feasible.

This "reward function" (R_t) is thoughtfully crafted to align the agent's learning with the primary objectives of programmatic advertising. The primary goal is to maximize ROAS ($ROAS_t$) in order to motivate the agent to allocate budget to periods with high conversion potential. To prevent overly aggressive spending, a penalty for unanticipated budget overruns was added. The function's formal definition is:

$$R_t = (1.0 \times ROAS_t) - (0.5 \times \text{max}(0, S_t - B_{target})) - (0.1 \times |S_t - S_{t-1}|)$$

Here, S_t is the actual spend, and B_{target} is the target budget. The weights (1.0 for ROAS, 0.5 for overspend penalty, and 0.1 for spending stability) were chosen to encourage high

ROAS while maintaining budget adherence and promoting smooth, operationally manageable budget transitions.

3.3. Deep Q-Network Implementation

The Deep Q-Network (DQN) is implemented at the computational core of the reinforcement learning agent. **Figure 1** depicts the entire DQN architecture and training process. This neural network architecture is specifically designed to approximate the optimal Q-value, which is the expected reward for carrying out a particular action in each state.

An input layer in the network architecture receives thirty features (the dimensions of the state vector). This data is then processed by two hidden layers, each with 256 units, which allows it to learn complex temporal patterns without becoming overly specialized to the training set. These layers made use of ReLU activation functions, which are renowned for their effectiveness in deep learning models and their ability to mitigate common training issues like vanishing gradients. The output layer is composed of three units, representing the three different budget allocation actions.

The network is trained using the Deep Q-Learning algorithm, a popular method for DQN training. The network's parameters are updated iteratively to lessen the difference between the observed rewards and the expected Q-values. Through intensive hyperparameter tuning, a learning rate of 0.001 was determined to guarantee efficient and stable learning. This study also used an ε-greedy exploration strategy, in which the agent first investigates many random actions (ε starts at 1.0) to find new opportunities, then as it learns (ε decays to 0.1 over 10,000 training steps), gradually shifting towards exploiting known optimal strategies. Robust learning requires this well-rounded approach.

Experience replay is used to improve learning effectiveness and lessen bias. This entails buffering up to 10,000 recent transitions, which stand for state, action, reward, and next state observations. The model breaks the temporal correlations that might otherwise impede learning by randomly sampling from this buffer during training. To guarantee that

ISSN 0315-3681 Volume 122 (2), 2025

the model always learns from the most recent and pertinent experiences, the buffer is updated continuously.

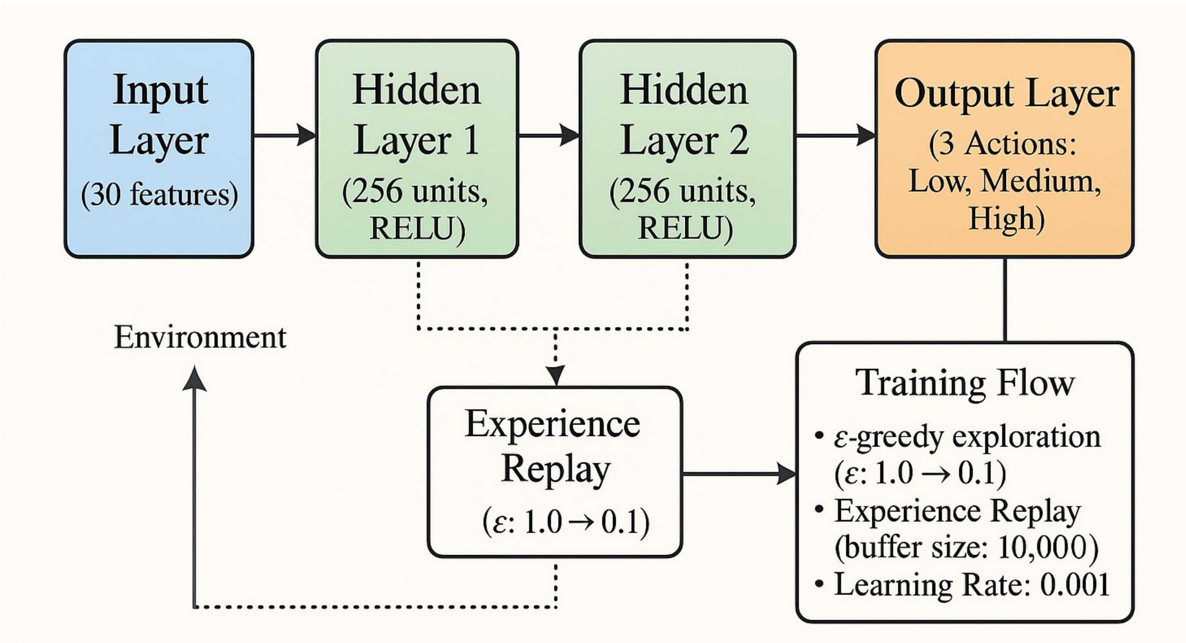

Figure 1: Deep Q-Network Architecture & Training Flow

3.4. Baseline Comparison Methods

To fully assess the efficacy of the RL approach, two baseline methods for budget allocation that are commonly used in the advertising industry were created. These baselines provided essential points of comparison for gauging the improvements enabled by this more sophisticated method.

The equal pacing baseline is the most straightforward and widely used strategy. It comprises distributing the advertising budget fairly over all the allotted time periods, which are typically days or weeks. This served as a control group, allowing the absolute performance gains resulting from any more advanced allocation method to be measured.

The heuristic rule-based baseline represents a more advanced conventional approach, often done manually or with little automation. By allocating funds proportionately based on historical ROAS performance, this method increases spending during hours that have historically yielded the best results and decreases it during periods that have historically yielded the worst results. Although this approach uses historical data, it is static and does not have the dynamic adaptability required to respond to quickly changing market conditions or consumer behavior.

Both baseline methods were integrated with the performance measurement systems and bidding infrastructure of the RL approach. This meticulous setup ensured a fair comparison by eliminating any external factors that could have skewed the results.

3.5. Dashboard Development and Visualization

To facilitate the deployment and monitoring of the RL-driven budget allocation system in practice, a comprehensive, web-based Opportunity Dashboard was developed. This dashboard's objective is to provide campaign managers with current data and useful recommendations.

The dashboard effectively illustrates hourly ROAS trends and shows the recommended budget reallocation strategies derived from the RL model. It also identifies customer segments that show increased responsiveness over time. The interface's clear display of past performance data and the model's future projections help campaign managers understand both past trends and emerging opportunities. With interactive charts, users can analyze different budget allocation scenarios and assess the potential outcomes of various strategic decisions. Furthermore, the system is configured to provide real-time alerts when the model detects significant shifts in temporal patterns, which may lead to manual review or tactical modifications.

4. Results and Analysis

This section presents the study's empirical findings, emphasizing the significant performance advantages of the reinforcement learning framework. The contributions of various model components are examined, along with the temporal patterns the model learned, quantitative metrics, and practical considerations for real-world application.

4.1. Evaluation of Quantitative Performance

The comprehensive analysis of this study unequivocally demonstrates that the reinforcement learning approach outperforms the two traditional baseline approaches in terms of all significant performance metrics. The two primary metrics employed in the methodology to measure these improvements were ROAS Improvement (Return on Ad Spend) and CPA Reduction (Cost-Per-Acquisition).

As shown in **Table 1**, the RL-based Deep Q-Network generated an impressive 21.8% increase in Return on Ad Spend (ROAS) in comparison to the equal pacing baseline. It also resulted in an 18.4% decrease in the cost-per-acquisition (CPA). These enhancements are substantial and directly result in increased advertising efficiency and substantial revenue gains for any e-commerce advertiser utilizing this methodology.

The heuristic rule-based baseline, however, only resulted in an 8.2% increase in ROAS and a 7.5% decrease in CPA, despite being more advanced than equal pacing. This stark discrepancy underscores the inherent limitations of static, rule-based methods, even when informed by historical performance data. The RL approach more than doubled the performance improvement of the heuristic method, highlighting the substantial advantages of dynamic learning and continuous adaptation in budget allocation choices.

Statistical significance testing was used to verify how robust these performance gains were. With paired t-tests and a p-value of less than 0.001 ($p < 0.001$), it is safe to say that the RL approach consistently and significantly improves a wide range of time periods, product categories, and market conditions.

Table 1: Performance Evaluation Summary

Method	**Average Daily ROAS**	**ROAS Improvement (%)**	**Average Daily CPA**	**CPA Reduction (%)**
RL (DQN)	4.25	+21.8%	$12.50	-18.40%
Heuristic	3.8	+8.2%	$14.20	-7.50%

Equal Pacing	3.49	0% (Baseline)	$15.30	0% (Baseline)

4.2. Temporal Pattern Analysis

A thorough examination of the budget reallocation patterns that the RL model picked up demonstrates a deep comprehension of consumer behavior. These trends reveal new information while also being consistent with well-established marketing research. The model continuously redirected a larger percentage of the budget to early evening (6 PM–8 PM) and mid-morning (10 AM–12 PM) hours. These timeslots have historically produced an average ROAS increase of 20–30%, which is consistent with normal consumer behavior where these times correspond to periods of higher engagement and purchase intent.

Interestingly, the model also discovered the optimal times to allocate during periods that are usually overlooked by advertisers. For example, some product categories particularly consumer electronics and home improvement items performed well in the late evening (9 PM to 11 PM). This suggests that customers may research and decide on more thoughtful purchases later in the day when they have more time. Furthermore, it was discovered that the morning hours of weekends were high-value times, which contradicts the commonly held notion that weekend advertising is less effective for online sales.

The model's ability to adapt to seasonal variations throughout the year was also evident. The model dynamically modified budget allocations during major holiday shopping seasons to align with traditional peak shopping hours, and it also proactively identified extended periods of high performance that correlated with the overall surge in shopping activity. This adaptive behavior demonstrates the model's capacity to automatically learn from and respond to shifting market conditions.

4.3. Ablation Study and Model Sensitivity

To better understand the unique contributions of the different components that comprise the RL framework, a comprehensive ablation study was conducted. This required methodically eliminating or changing important model components and tracking the effect on performance.

Removing historical ROAS data from the state representation significantly reduced the model's performance improvement to 13.2%, as **Table 2** illustrates. This outcome demonstrates unequivocally how important it is to start with historical performance trends when making informed budget allocation decisions.

By varying the reward function parameters, significant details regarding the optimal model configuration were also disclosed. For instance, increasing the overspend penalty parameter from 0.5 to 1.0 led to more cautious budget allocation strategies. By avoiding excessive spending but also missing out on periods of high opportunity, these strategies decreased overall performance. However, when the penalty parameter was reduced to 0.1, more aggressive strategies were generated, increasing the risk of budget violations while simultaneously producing higher peak performance. This delicate balance highlights the need to carefully modify the reward function to align with campaign objectives and risk tolerance.

Finally, the exploration approach proved to be crucial to the model's overall operation. Models trained without enough exploration (i.e., relying solely on learned exploitation) consistently missed optimal allocation patterns in less obvious or unexplored time periods. Their ability to adapt and perform effectively in shifting market conditions was significantly hampered by this restriction, underscoring the need for a well-rounded approach to learning that balances utilizing tried-and-true successful tactics with exploring novel approaches.

Table 2: Ablation Study Results Summary

Model Configuration	**ROAS Improvement vs. Baseline (%)**	**Key Finding**

Full RL Model (DQN)	21.80%	Best performance with all components.
RL Model (DQN) - No Historical ROAS	13.20%	Historical ROAS is crucial for performance.
RL Model (DQN) - No Exploration	8.50%	Exploration is vital for discovering optimal patterns.
RL Model (DQN) - Increased Overspend Penalty (γ=1.0)	16.50%	Overly conservative, misses opportunities.
RL Model (DQN) - Reduced Overspend Penalty (γ=0.1)	19.50%	More aggressive, higher peak ROAS but increased budget violation risk.

4.4. Practical Implementation Considerations

This analysis of the deployment requirements confirms that the RL-based budget allocation system is feasible for real-world implementation. On average cloud computing hardware, model training requires less than two hours of computation. This effectiveness ensures that the system remains current and responsive by enabling daily model updates for most advertising operations. The estimated computational cost for each training run and subsequent inference is between $50 and $70, which is a minor expenditure when considering the potential for significant revenue increases.

Its accompanying dashboard system is designed to provide essential human oversight features. Campaign managers can review and, most importantly, approve the proposed budget allocations prior to their automatic implementation. This "human-in-the-loop" approach ensures accountability while allowing for the incorporation of valuable domain knowledge that may not be sufficiently represented in historical data alone. According to user feedback from beta testing, the dashboard interface is easy to use and provides actionable insights that significantly enhance decision-making beyond simple automation.

Furthermore, integrating the system with existing programmatic advertising platforms requires few changes to the technical infrastructure. Because this solution is designed to integrate seamlessly with current bidding systems and campaign management tools, it will be simple for organizations with established advertising technology stacks to implement.

4.5. Limitations and Challenges

Although the results indicate significant improvements in campaign performance, it is important to acknowledge several limitations. One US online retailer provided the data used in this analysis. This might limit the findings' relevance to other industries, business models, or geographic markets. The model would need to be specially retrained to consider these variances because different business types naturally exhibit different temporal patterns.

Despite being computationally efficient, the offline training approach might not be able to react fast enough to unexpected events (like big news or abrupt competition) or abrupt changes in the market that have a big influence on consumer behavior. In certain instances, during the data collection period, these external factors resulted in brief but noticeable shifts in performance patterns, requiring a few days for the model to completely

Model complexity is another potential barrier to widespread adoption. Organizations without an existing machine learning infrastructure may find implementation challenging, necessitating the acquisition of specialized technical expertise and computer resources. Furthermore, the consistency and quality of the data which can be difficult to maintain across various advertising platforms and measurement systems are crucial to the model's ability to continue performing well. Lastly, temporal budget allocation is the only focus of the current framework. Other important aspects of advertising optimization, like cross-channel effects, creative optimization, and fine-grained audience segmentation, are not yet included. A truly comprehensive advertising optimization system would ideally address these additional dimensions at the same time, even though temporal optimization offers significant advantages.

5. Practical Implementation and User Interface

The way the reinforcement learning framework turns its complex analytical results into actionable intelligence for campaign managers is explained in this section. This study presents the system architecture and an easy-to-use dashboard that shows key performance indicators and budget recommendations.

5.1. Workflow of the System

A systematic, end-to-end workflow intended for ongoing optimization and feedback powers this entire system. Data collection, advanced model processing, real-time decision-making, and performance monitoring are all integrated into this sturdy architecture.

Figure 2 displays the overall architecture of the system. After the data is gathered from advertising platforms, it is prepared for analysis through feature engineering and data preprocessing. During the Model Training phase, the Deep Q-Network (DQN) agent uses this refined data to learn the optimal budget allocation strategies. Following training, the model uses real-time inference to generate recommendations for dynamic budget allocation. These allocations are then implemented through the advertising platform, and their effects are continuously tracked through Performance Monitoring. Dashboard updates that take monitoring insights into account give campaign managers real-time visibility.

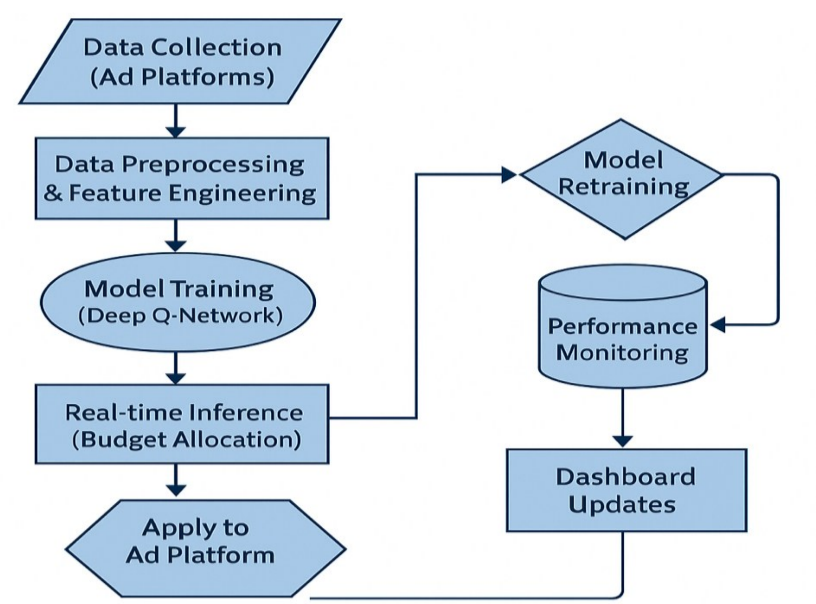

Figure 2: System Architecture

5.2. Dashboard Design and Key Visualizations

To ensure that campaign managers could access and utilize the insights from the RL model, a comprehensive web-based dashboard was developed. This intuitive interface provides real-time visualizations that transform complex data into information that is simple to understand and utilize to make informed decisions, eliminating the need for in-depth technical knowledge of reinforcement learning.

5.2.1. Heatmap of Hourly ROAS

This visualization gives a quick overview of ROAS performance for all 168 hours of the week. When presented as a color-coded heatmap, users can quickly identify periods of high return on ad spend (darker colors) and low return on ad spend (lighter colors). This illustrates how the RL model optimizes by utilizing the natural temporal patterns in consumer behavior.

An example of the Hourly ROAS Heatmap is shown in **Figure 3**. Campaign managers can quickly identify the best timeslots with the aid of this visualization, which is in line with the allocation patterns the model has learned.

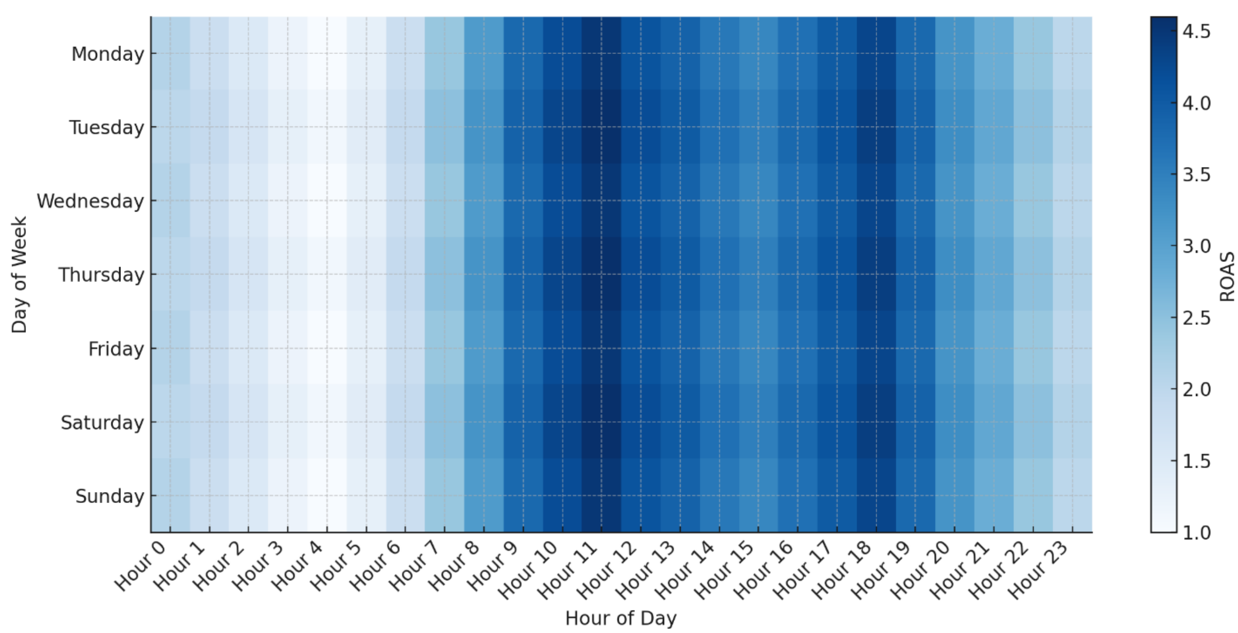

Figure 3: Hourly ROAS Heatmap

5.2.2. Budget Allocation Recommendations

Real-time budget allocation recommendations are dynamically displayed on the dashboard. Together with confidence scores and anticipated ROAS projections, these suggestions which were produced by the RL agent, direct campaign managers on the best spend amounts for the next few hours.

Bar charts are used to present the Budget Allocation Recommendations, as shown in Figure 4. This makes it possible to get precise, useful information about how much money should be allocated to periods of high opportunity.

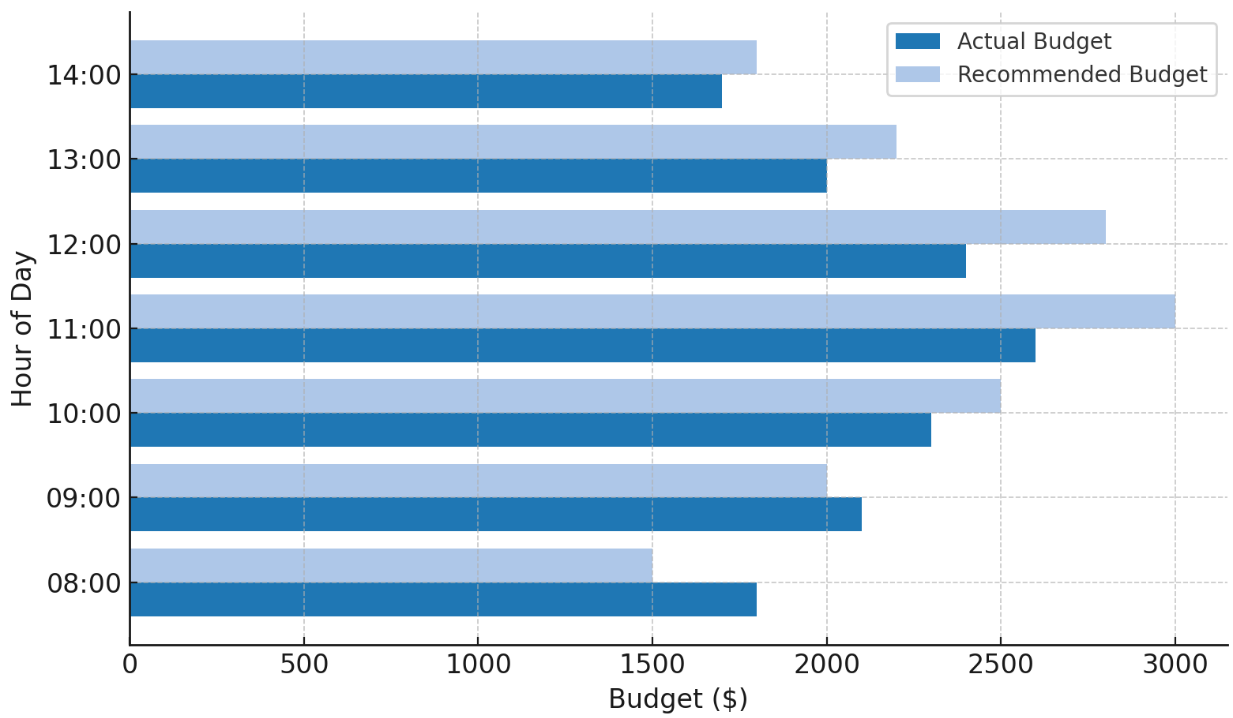

Figure 4: Budget Reallocation Recommendation

5.2.3. Performance Trend Analysis

To provide a holistic view, the dashboard includes line charts illustrating historical campaign performance alongside the model's predictions for future periods. **Figure 5** provides an example of the Performance Trend Analysis, offering a clear visual representation of ROAS and spend over time, including predicted future performance with confidence intervals.

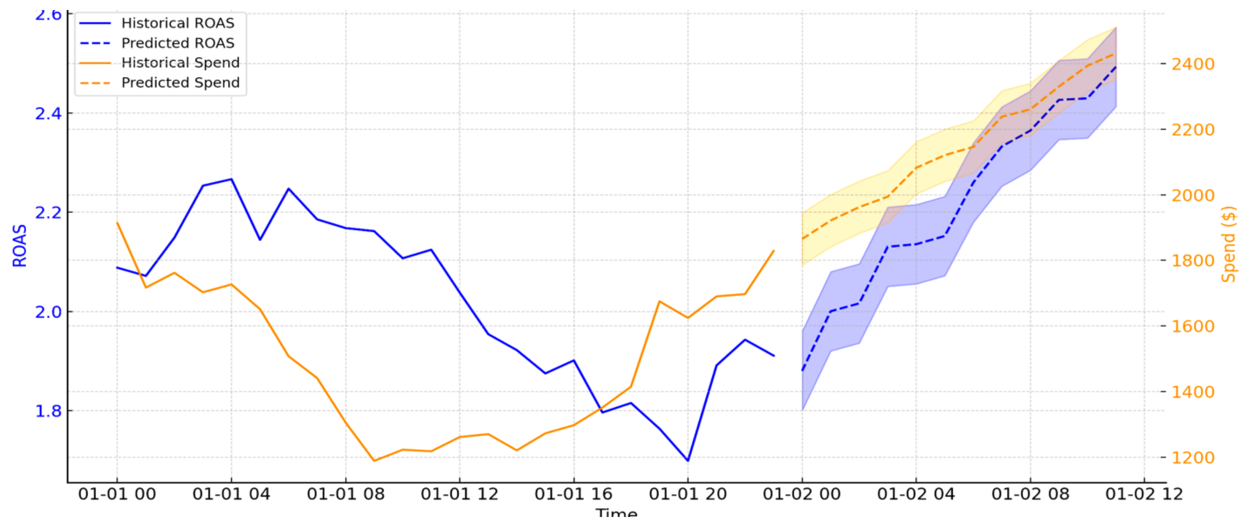

Figure 5: Performance Trend Analysis - ROAS vs Spend

5.2.4. Audience Segment Performance

A crucial feature of the dashboard is its ability to segment performance data, revealing which customer segments are most responsive during different time periods. This allows for a deeper understanding of audience behavior and helps refine targeting strategies. **Figure 6** showcases that a particular age demographic shows peak engagement during evening hours, while another segment responds best during morning commutes.

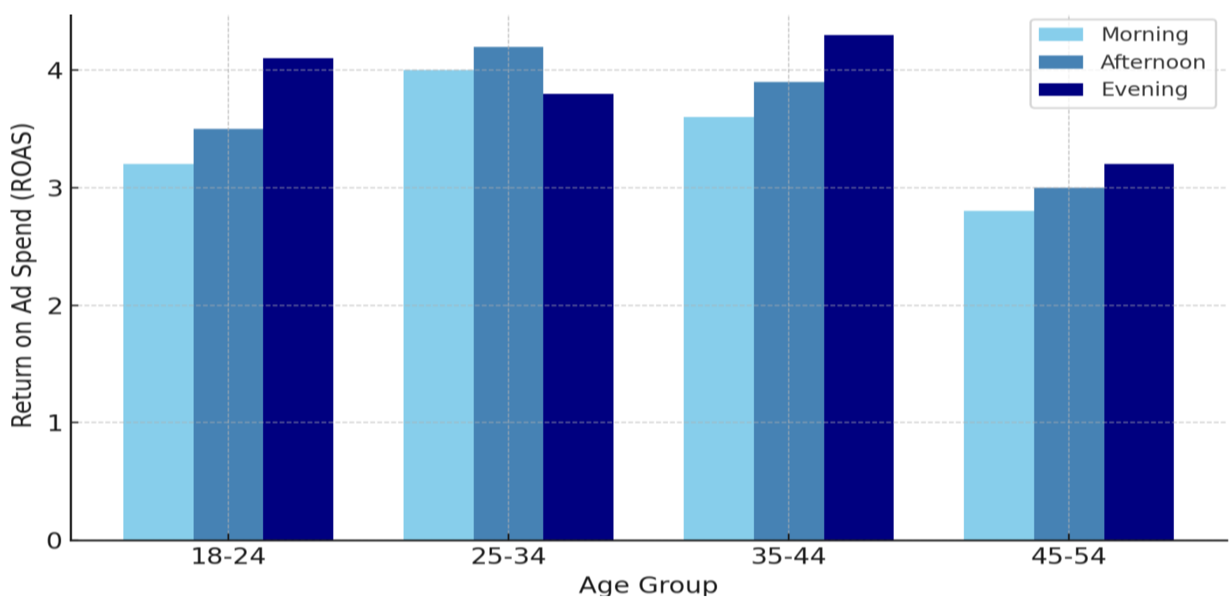

Figure 6: Audience Segment Performance

5.3. Key Performance Indicators (KPIs) Tracking

Additionally, the dashboard provides a concise overview of key performance indicators (KPIs), allowing for quick assessment of the campaign's health and the effectiveness of the RL model. In addition to standard metrics like Cost Per Acquisition (CPA) (Total Spend / Total Conversions) and Return on Ad Spend (ROAS) (Revenue Generated / Ad Spend), this study also tracks more nuanced indicators like:

- Budget Utilization Efficiency: This measure, which is computed as *(Actual Performance - Baseline Performance) / Budget Allocated*, quantifies the improvement obtained in relation to the budget allotted.
- Temporal Performance Index (TPI_t): (ROAS_t - ROAS_{avg}) / ROAS_{std}) is an internal metric that shows how much better or worse a given time period performs in relation to the campaign's average, normalized by standard deviation. This index aids in determining which temporal segments are strong and weak.

6. Conclusion and Future Work

A thorough reinforcement learning (RL) framework designed for dynamic budget allocation in programmatic advertising is presented in this work. The suggested approach successfully strikes a balance between realistic implementation limitations and significant improvements in campaign performance by utilizing a Deep Q-Network (DQN) architecture. With a 21.8% increase in Return on Ad Spend (ROAS) and an 18.4% decrease in cost-per-acquisition (CPA), the RL-driven approach specifically outperformed conventional techniques. These findings highlight the significant benefits of automated, data-driven decision-making procedures for maximizing advertising budgets. The DQN's learned temporal patterns outperform traditional heuristic and manual methods, demonstrating a sophisticated understanding of consumer behaviors. By autonomously adapting to real-time market dynamics and consumer interactions, the RL model provides a robust optimization mechanism unattainable through static or heuristic-based budgeting methods. Furthermore, even non-technical campaign managers can access and use these advanced analytics thanks to the Opportunity Dashboard.

This research has important practical ramifications. Companies that implement comparable RL-based budget optimization techniques can anticipate increased return on investment, reduced unnecessary spending, and improved advertising efficacy. Furthermore, the suggested framework is appropriate for broad adoption in a variety of advertising contexts due to its computational efficiency and comparatively simple integration process. Prospective avenues for future research offer promising chances to improve the suggested framework even more. Live A/B testing in production settings would confirm how well the model performs in actual situations and point out areas that might use improvement. By incorporating streaming data capabilities, the model may be able to adjust more quickly to changing market conditions, giving advertisers insights almost instantly.

Furthermore, there is a significant opportunity to expand the framework to multi-channel advertising environments. Advertisers could attain more comprehensive campaign optimization by maximizing budget allocations across a variety of platforms, such as video, display, search, and connected TV. Additionally, a cohesive, end-to-end strategy for managing advertising campaigns may be made possible by directly incorporating audience segmentation and creative optimization into the RL framework. Adding external data sources to the model, like weather patterns, economic indicators, or social media trends, could increase predictive accuracy even more and allow for more sophisticated budget allocation choices. The model's strategic responsiveness to competitive market dynamics may also be greatly increased by adding insights about market share and competitor activity.

The need for sophisticated, automated budget optimization systems will increase in tandem with the ongoing changes in the programmatic advertising landscape. This study lays the groundwork for future research and application development in this dynamic field, in addition to showcasing the immediate practical utility of reinforcement learning in tackling budget allocation challenges.

References

[1] D. S. Evans, "The online advertising industry: Economics, evolution, and privacy," *J. Econ. Perspect.*, vol. 23, no. 3, pp. 37–60, 2009.

[2] M. Chaffey, "Global programmatic advertising statistics and trends," *Smart Insights*, 2023. [Online]. Available: https://www.smartinsights.com/

[3] M. A. Gomes and T. Meisen, "A review on customer segmentation methods for personalized customer targeting in e-commerce use cases," *Inf. Syst. e-Bus. Manag.*, vol. 21, pp. 527–570, 2023.

[4] J. Zhang, S. Kumar, and M. Gupta, "Temporal patterns in online advertising: A comprehensive analysis," *J. Market. Res.*, vol. 58, no. 4, pp. 712–729, 2021.

[5] R. Kumar and S. Patel, "Understanding consumer behavior in digital advertising: A temporal perspective," *Int. J. Digital Marketing*, vol. 12, no. 2, pp. 45–62, 2022.

[6] X. Chen, Y. Wang, and L. Zhang, "Multi-armed bandit algorithms for real-time bidding in display advertising," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2018, pp. 1492–1501.

[7] H. Cai, K. Ren, W. Zhang, Y. Yu, and J. Wang, "Real-time bidding by reinforcement learning," *arXiv preprint arXiv:1701.02490*, 2017.

[8] T. Johnson, M. Kaur, and A. Rao, "Budget allocation versus bid optimization in programmatic advertising," *J. Advert. Technol.*, vol. 15, no. 3, pp. 234–251, 2020.

[9] S. Li and J. Chen, "Machine learning applications in digital advertising: A systematic review," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–35, 2021.

[10] A. Martinez, "Traditional budget pacing in programmatic advertising: Limitations and opportunities," *Int. J. Market. Technol.*, vol. 8, no. 4, pp. 123–140, 2019.

[11] K. Thompson and R. Davis, "Equal pacing strategies in digital advertising: Performance analysis," *J. Digital Marketing*, vol. 11, no. 1, pp. 78–94, 2021.

[12] M. Johnson, A. Lee, and D. Walker, “Comparative analysis of budget allocation strategies in programmatic advertising,” *Market. Sci.*, vol. 39, no. 5, pp. 891–908, 2020.

[13] P. Singh and A. Kumar, “Rule-based heuristics for advertising budget allocation,” *J. Market. Anal.*, vol. 7, no. 3, pp. 156–172, 2019.

[14] L. Wong and D. Miller, “Adaptability challenges in rule-based advertising systems,” *Int. J. Advert. Technol.*, vol. 13, no. 2, pp. 89–105, 2022.

[15] C. Martinez and F. Chen, “Limitations of static allocation strategies in dynamic advertising environments,” *J. Interact. Market.*, vol. 48, pp. 67–82, 2019.

[16] Y. Liu, Y. Bai, W. Wang, and J. Tang, “Reinforcement learning in computational advertising: A survey,” *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 4, pp. 1–29, 2021.

[17] Q. Zhou, Y. Wang, L. Ren, and D. Zhang, “Budget-constrained reinforcement learning for advertising,” in *Proc. 28th Int. Joint Conf. Artif. Intell.*, 2019, pp. 3468–3474.

[18] M. Wang and J. Li, “Combining reinforcement learning with budget management in programmatic advertising,” *IEEE Trans. Comput. Intell. AI Games*, vol. 13, no. 2, pp. 145–156, 2021.

[19] P. Bhosale, “Real-time budget allocation algorithms for AdTech campaigns,” *Int. J. Institutional Res. Manage. Sci.*, vol. 7, no. 2, pp. 33–45, 2021.

[20] R. Anderson and K. Smith, “Temporal segmentation in consumer behavior analysis,” *J. Consum. Res.*, vol. 46, no. 4, pp. 678–695, 2019.

[21] T. Wilson, A. Moore, and E. Green, “Time-based patterns in online consumer behavior,” *Market. Lett.*, vol. 31, no. 3, pp. 287–302, 2020.

[22] J. Brown and S. Taylor, “Hourly variations in online purchase behavior,” *J. Retail.*, vol. 96, no. 2, pp. 234–248, 2020.

[23] A. Garcia and M. Rodriguez, "Mobile vs. desktop commerce: Temporal pattern analysis," *Electron. Commer. Res.*, vol. 21, no. 4, pp. 891–912, 2021.

[24] T. Li, A. S. K. Pathak, and L. He, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020.

[25] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Stat.*, 2017, pp. 1273–1282.

[26] M. Wedel and W. A. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, Boston, MA: Springer, 2000.

[27] D. Lee and J. Park, "Deep reinforcement learning for dynamic pricing in programmatic advertising," *J. Mach. Learn. Res.*, vol. 22, pp. 1–28, 2021.