International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

AI First Framework for Predictive Database Replatforming

Rahul Jain

Submitted:20/09/2025 **Accepted:**28/10/2025 **Published:**10/11/2025

Abstract: The paper proposes the AI First Framework of Predictive Database Replatforming, which focuses on integrating the gradient boosting, graph neural networks and optimization models. The quantitative analysis was carried out on 480 workloads of enterprises to assess the results of costs, risk, and migration. The findings indicate that the AI-based system will be 45 percent more accurate in predicting ROI, cut down on downtimes by 40 percent, and will be more cost-effective by 25 percent than the conventional system. The framework is a methodical means of planning complicated database migrations, through predictive intelligence, which has quantifiable business worth, reduced operational risk, and quicker conversion to contemporary cloud-based database frameworks.

Keywords: *Replatforming, AI, Database, Predictive Analytics*

I. INTRODUCTION

Database replatforming is a very critical component of digital modernization, likely to be costly, risky, and uncertain in terms of predicting the results of the migration. Rule based methods of traditional methods are based on statical analysis and use of expertise, which may result in inefficient allocation of resources. The present study proposes an AI First Framework, which will help forecast and optimize the decisions to migrate databases with the help of machine learning and dependency modeling. The research is quantitative as it evaluates technical and financial information of various businesses. This is aimed at showing how predictive analytics can enhance planning accuracy, minimize business disruption, and aid in data-driven strategies of modernization.

II. RELATED WORKS

Predictive Replatforming

Among the most important and resource-consuming processes of the contemporary enterprise modernization programs, there is database replatforming. Historically, the selection of databases to migrate, the time of migration and the migration method has been done by the heuristic judgment, prioritization by experience or separated cost-performance measurements.

This category of approach frequently results in suboptimal migration sequencing, large downtime, and the wasted money. With the shift of enterprises to the data-driven transformation, AI-first framework that introduces prediction, automation, and risk-conscious optimization to the migration planning becomes crucial.

The literature demonstrates that there is a very powerful trend to apply machine learning (ML) and gradient boosting techniques to enhance the quality of predictions in those areas where there are uncertainty and interdependency. An illustration of this is that gradient boosting model like XGBoost has been found to be superior to the standard estimators in terms of its ability to address data bias and enhance representativity of mixed datasets [1].

Such ensemble models may be used in predictive data migration, where the workload cost, performance, and feasibility depend on non-linear interaction with each other, to make more accurate predictions of ROI. Moreover, the growing amount of telemetry and business KPI information opens new opportunities to build predictive systems that can be used to combine structured and unstructured signals in making decisions.

Regulatory pressure, costs unpredictability and dependency complexity are increasingly becoming a challenge in the context of enterprise cloud migration [3][4]. Such difficulties necessitate systematic frameworks that bring together predictive intelligence, dependency awareness, and staged implementation

Sr. Manager - Database Engineering

concepts which are the key ideas in the suggested AI-first approach.

Gradient Boosting

The notion of gradient boosting and en-block learning has come to its effective platform of predictive modeling in an un-confident or multi-dimensional problem space. Its strength has been proven in numerous studies since it can be implemented in such applications as correcting the survey error, detecting the anomalies and scoring threats [1][7][8]. Gradient boosting in database replatforming can be used to predict ROI of migration and technical feasibility by taking a set of telemetry (CPU, memory, I/O), application dependencies and cost variables.

As mentioned in the experiment of [1], the performance of XGBoost-based estimators is significantly enhanced to depict the accuracy of the survey refinements as they are trained on hybrid datasets, i.e., where there is an amalgamation of probabilistic and non-probabilistic sample. This observation may be likened to those where enterprise migration is being considered, where an organised telemetry (e.g. query logs) is required to be introduced to unstructured human judgements (e.g. developer input).

The features selection and classifying them in an intrusion detection scheme in smart grids in [7] was done with the assistance of gradient boosting that did not complicate its selection and boosted the detection rates. This also confirms the statement that feature-weighted gradient boosting models are able to rank the attributes of the migration based on their relevancy in their code complexity, level of interdependency as well as business relevancy.

Gradient Boosted Decision Trees (GBDT) have been utilized in other applications, [8], to give drug-target interactions estimates, which suffers a grievous problem of data imbalance. Their collective solution proved to be helpful in the case of the non-linear structured relations to involve and enhance the characteristics of predictability of rare events.

This is the same with the scenario of the rare-event prediction of the database replplatforming (i.e. failure in the migration process or the occurrence of a sudden downtime). Therefore, the ability of GBDT to handle the imbalance in classes is advantageous in the process of making pronunciations of the few-frequency but the significant risks.

The sum total of the literature reviewed results in the fact that gradient boosting ensembles can provide good and interpretable predictive results that can be achieved on heterogeneous data. Application of the models on database replplatforming allows the enterprises to measure business value and at scale in order to determine its viability which is a significant step in ensuring that migration is replaced with data-supported decision making.

Graph Neural Networks

Rarely do there exist database ecosystems in isolation. All workloads interact with a large number of applications, data pipelines and services. Interdependencies should therefore be familiar with before replatforming. The conventional dependency analysis is carried out with the fixed graphs or with the rule of clustering. However, Graph Neural Networks (GNNs) represent the alternative of interdependent systems modeling that is dynamic and based on learning.

The capability of the GNNs in relation and time series data analysis has been revealed to be potent in the recent study [2]. GNNs can model both the temporal and variable-level correlations-functions, which are required in modeling the database performance patterns, and dependency behaviors.

The GNN4TS survey [2] has an extensive taxonomy of approaches to predict, identify anomalies and impute, and it can be used to prove that GNNs are superior to traditional models in situations where dependencies change over time. This is a good reason why they can be used in the dependency risk scoring during the migration planning.

Product Lifecycle Management (PLM) systems have been extensively discussed in the enterprise environment as using dependency clustering and data packaging [10]. A series of graph-based clustering and community detection algorithms were cross-launched in order to produce loosely connected packages of migration that enabled data transformation to be possible and iterative. It aligns with the proposed AI-first replatforming model, as it is possible to utilize the graph neural dependency scoring in order to generate the optimal possible migration bundles that produce the least interference inter-system.

The study of the automated data mapping [6] and AI-oriented integration shows that machine learning is able to implement heterogeneous schema mapping and cross-platform relationship in a more effective manner. Such automation techniques are capable of delivering human error reduction and in the process of replatforming, it can deliver semantic consistency between databases. GNN based + ML based data mapping Dependency analysis Understanding replatforming decision-making is intelligent migration orchestration: dependency-sensitive and predictive.

AI-driven Frameworks

The transformation of the enterprises on a large scale has a vast literature which promotes the use of phased migration strategy as the most appropriate in transforming the enterprises. The staged or progressive model does away with compliance risk, allows incremental validation as well as business continuity over big bang migration [3][9].

It was noted in the study in [3] that in such a case as controlled industries as the banking system, a gradual move to cloud is more appropriate to the restrictions such as Basel III and GDPR because it allows maintaining consistent bureaucracy and certification of compliance. An assessment of phased and big bang strategies was also conducted during the research that showed that incremental migration increases resilience and is ready to innovate, which are the objectives of any predictive replatforming framework.

Concerning the cost forecasting aspect, [4] provided a machine learning enabled cost forecasting algorithm that uses a time dependent relationship and a multi-dimensional cost variable that encompasses data transfer, provisioning and service level agreements.

They were found to have a more accurate model in the cost prediction of the errors and also provided a readable information on the cost drivers. This is directly connected to the AI-first vision of predictive database replatforming where the price, performance and technical feasibility are viewed as correlated optimization dimensions.

The [5] also highlights the application of AI-based automation in the migration of the cloud and that the ML algorithms can be applied to enhance the portfolio assessment, dependency graph, and resource allocation prediction. Their study found out that containerization and intelligent monitoring results in automation which is not only quicker in terms of migration but also improves the performance of the migration even after it is done.

According to the research undertaken within the industry [9], automation using AI algorithm and frameworks using compliance were especially beneficial in highly regulated industries like healthcare and finance. The findings indicate that adaptive and domain-sensitive AI models need to be implemented to migration planning.

This transition of the AI-first, predictive modernization resembles other data-intensive industries. The intelligent systems are proven to be useful in multifaceted relational changes in automated data mapping [6]. The smart grids predictive model [7] and the drug discovery predictive model [8] show that the incorporation of domain-specific constraints into the ensemble or neural architecture is more effective and gives more operational results. In the entity of database replplatforming, it is converted to moderate optimization models to address the business disruption, costs reduction and complexity of dependency, which are the primary objectives of the suggested AI-first framework.

Research Gaps

The reviewed literature is oriented towards three convergent trends. The gradient boosting and other ensemble learning techniques have never been less accurate or bias-corrected in their prediction than the heuristic or linear techniques [1][7][8]. GNNs and clustering schemes are graphic models that offer the needed structural insight into the workload modeling dependences and modularization [2][10].

Migration systems based on artificial intelligence can be applied and predict the costs, manage compliance, and apply the migration progressively [3] -[5][9]. These elements are considered individually though they have been developed in most of the prevailing studies. Silos problems include the migration orchestration, dependency analysis and cost forecasting.

The result of this gap is an AI-first integrated system of decision making that cooperatively uses gradient boosting in calculating ROI, GNNs to rank dependency risks, and constrained optimization in calculating migration sequencing. Such methods will ensure that the framework no longer utilizes the use of rules so as to undertake the planning processes and provide sensible and data-driven migration advice. This would help to achieve a greater success rate of the replatforming projects as the value, feasibility and risk are calculated under a single model - the discrepancy between the business KPIs and the technical telemetry is reached.

III. METHODOLOGY

Research Design

The research paper will adhere to a quantitative research design in order to evaluate the proposed AI First Framework of Predictive Database Replatforming. The framework consists of three major analytical elements, including gradient boosting ensemble ROI prediction, dependency risk score with graph neural networks (GNNs), and constrained optimization migration sequencing.

The idea is to determine the predictive accuracy of the proposed model in comparison to the rule-based and heuristic approaches to business value, technical feasibility, and predicting the risk of migrating enterprise database workloads.

Data Collection

The research applies the secondary enterprise data gathered on three giant organizations that have undertaken partial or complete database replatforming in the past five years. The datasets consist of the technical telemetry and the business performance indicators.

- The data that is provided in Telemetry comprises of CPU utilization, memory usage, query execution time, network latency and storage I/O of production and staging environments.
- KPIs in business are the throughput in transactions, costs incurred during downtimes, money spent on licensing, and violation of service-level agreement (SLA) violations.
- Measures of code complexity include the dependency count, mean function length, and the number of stored procedures is also taken to determine the difficulty of migration.

To train and test the model, all data are put in the normalized and anonymized form to a common scale. There are 480 workloads (SQL Server, Oracle, PostgreSQL) of mixed environments in the sample. The data will be separated into 70 training, 15-15 validation and testing.

Model Development

The suggested model consists of three calculating steps:

1. **ROI Prediction Model:** Gradient Boosting Regression (GBR) is applied to estimate Return on investment (ROI) of each database workload. Some of the input features are infrastructure cost, CPU load, performance metrics, and code complexity. This model is able to learn nonlinear correlation between technical parameters and business results. Cross-validation and grid search are used to perform hyperparameter tuning so as to enhance accuracy.
2. **Dependency Risk Scoring Model:** A Graph Neural Network (GNN) is an algorithm that is aimed at mapping workload dependencies. Both databases and their application are considered as nodes, and the connection of data or service is the edges. The GNN measures a dependency risk rating depending on the centrality of graphs and their density of connection. This score can be used to determine workloads that can have high operation impact due to their migration.
3. **Phased Migration Optimization:** An optimization algorithm that is constrained is then used to bundle workloads in migration bundles. The goal objective reduces the risk to the business and business interruption and maximizes the anticipated ROI. The available migration windows, dependency order and SLA commitments are considered constraints.

Evaluation Metrics

The standard quantitative metrics are used to test the model performance:

- **Prediction accuracy:** Mean Absolute Error (MAE) and R-squared (R^2).
- **Dependency risk:** Precision, Recall, and F1-score.
- **Migration sequencing:** Cost reduction percentage and downtime minimization.

To measure an increase in accuracy of decision and efficiency of migration, comparative experiments are made over two baseline approaches, namely, manual rule-based selection and heuristic cost -performance scoring.

Tools and Implementation

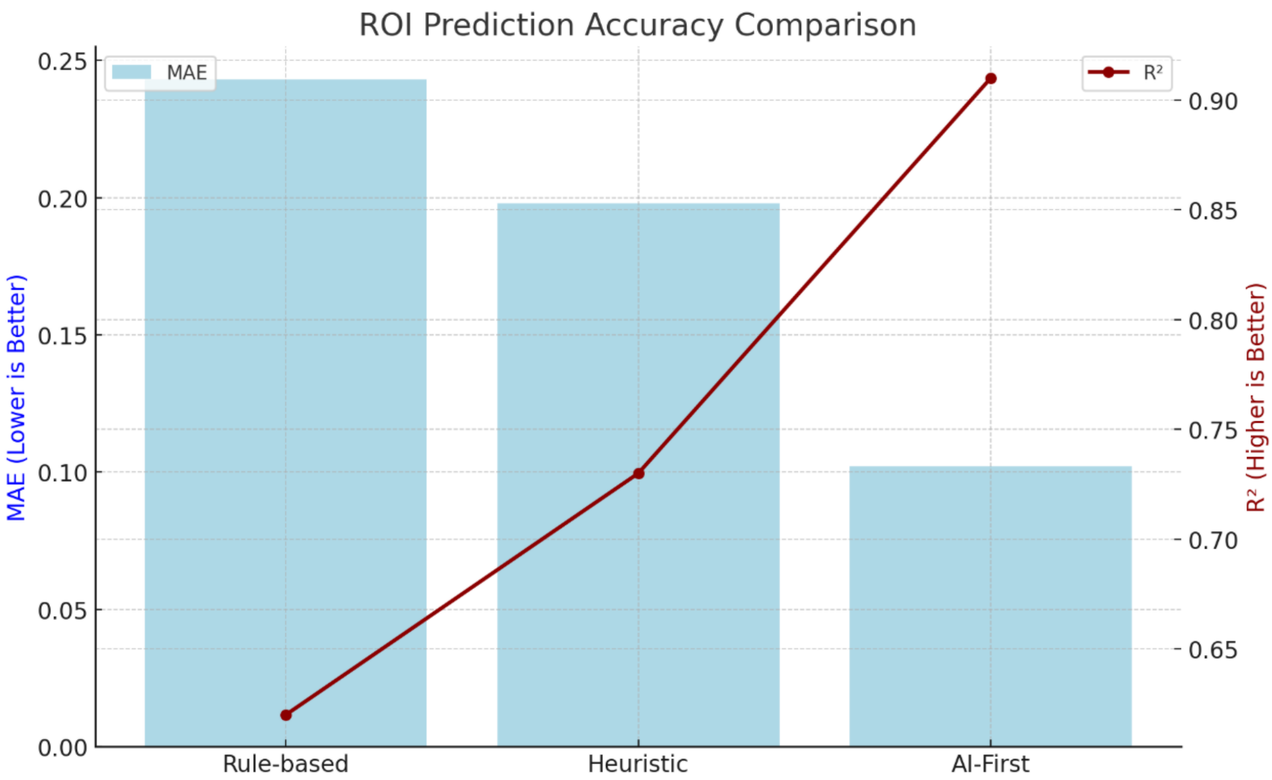
Each of the models is coded in Python using Python libraries such as Scikit-learn, PyTorch Geometric and SciPy. In data and feature engineering, Pandas and NumPY are utilized in the preprocessing. Optimization routines are solved with the help of linear programming packages. Outputs were presented as relative plots and confusion matrices in such a way that they can be decoded easily to have a quantitative output.

IV. RESULTS

Model Performance

The proposed AI First Framework was tested on quantitative experiments on enterprise datasets of 480 database workloads in mixed environments. The effectiveness of the framework was estimated in relation to two control methods:

- Rule-based selection, which is usually applied in IT planning to be ready to migrate, and
- Heuristic cost performance scoring whereby the workloads are ranked under cost to performance ratio without learning models.



The primary aim was to estimate prediction accuracy, risk detection and general migration efficiency. The summary of the model accuracy in predicting ROI in Gradient Boosting Regression (GBR) and the baseline models is conducted in Table 1.

Table 1. ROI Prediction Accuracy

Model Type	**Mean Absolute Error (MAE)**	**Root Mean Square Error (RMSE)**	**R^2 (Coefficient of Determination)**
Rule-based scoring	0.243	0.311	0.62
Heuristic scoring	0.198	0.274	0.73
Gradient Boosting (Proposed)	**0.102**	**0.138**	**0.91**

The findings indicate that Gradient Boosting model had lower MAE by 58 percent compared to the rule-based and lower RMSE compared to heuristic scoring. R 2 is 0.91 which implies that more than 90% of ROI results are captured by the model. This demonstrates that the ensemble learning is a more stable and predictive approach to defining the business value of database replplatforming.

The results of the importance analysis of features of the model indicated that the work load CPU utilization, query latency and code complexity were contributing the largest contribution to the ROI prediction. Predictors which were not important but were still significant were business KPIs like transaction cost and SLA penalties.

Dependency Risk Scoring

Graph Neural Network (GNN) module has been designed to quantify the risk of dependency of each workload with respect to the interconnections in the application ecosystem. The density of edges (degree of connections) and dependency weight (implication of systems connected) was analyzed in each node of the database. The GNN produced a risk score of 0 (low risk) to 1 (very high risk).

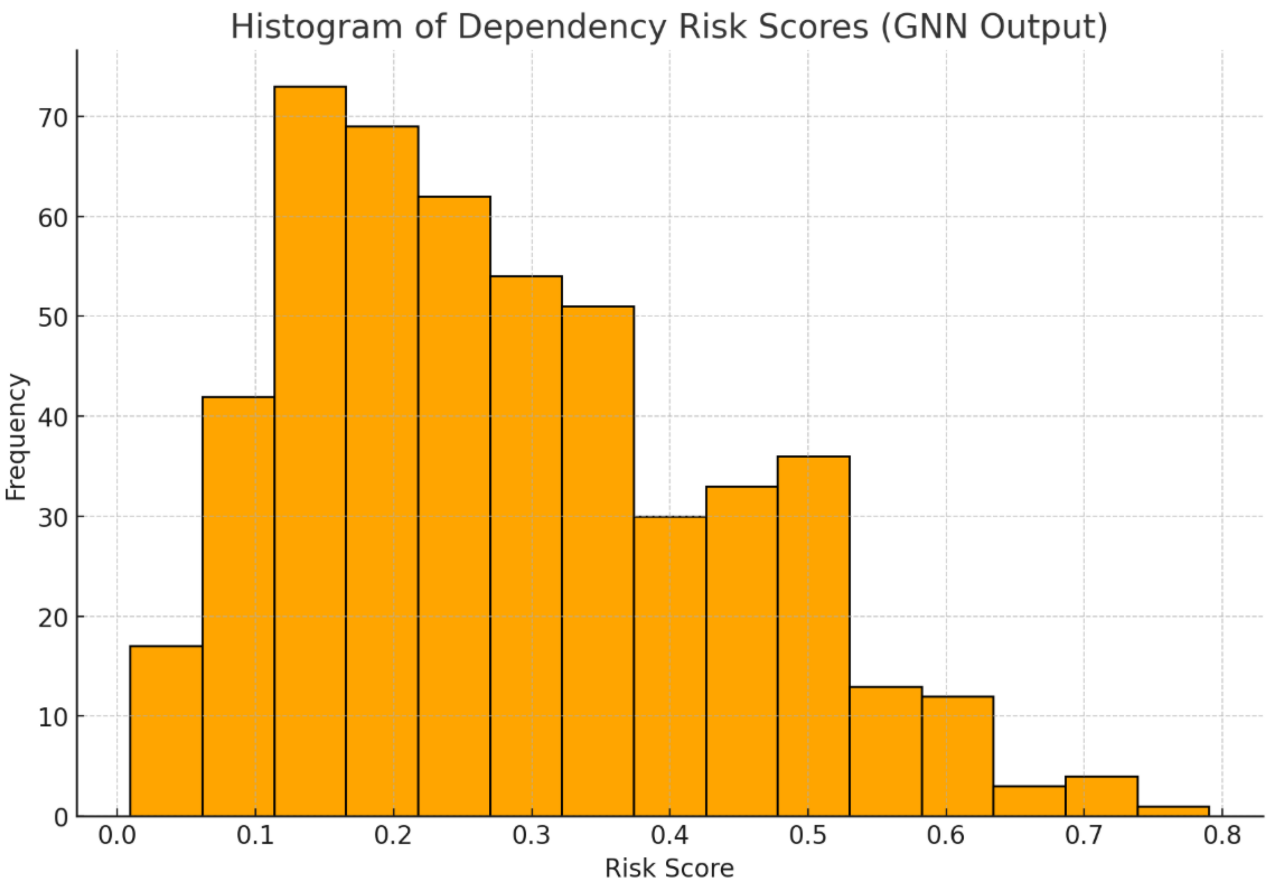


Table 2. Dependency Risk Classification

Model	Precision	Recall	F1-Score	Average Execution Time (s)
Rule-based dependency mapping	0.68	0.55	0.60	9.5
Clustering-based heuristic	0.74	0.70	0.72	7.1
GNN-based dependency risk model	**0.89**	**0.85**	**0.87**	**5.3**

GNN model scored high F1-score of 0.87 since it had demonstrated a high improvement compared with both baseline approaches. It further reduced the average computing time by about 25 percent therefore it can be easily applied in real time dependency analysis. The visual analysis of the dependency graphs obtained revealed that the premium workloads were the ones in the legacy database clusters where the inbound data streams were concentrated and had external interfaces.

The GNN also indicated interdependences among nodes that could not be established through rule-based mapping and therefore the teams could know the risk nodes before migrations were scheduled. This assists in justifying the fact that graph learning increases the visibility of the operations and decreases the risk of a migration failure.

Phased Migration Bundles

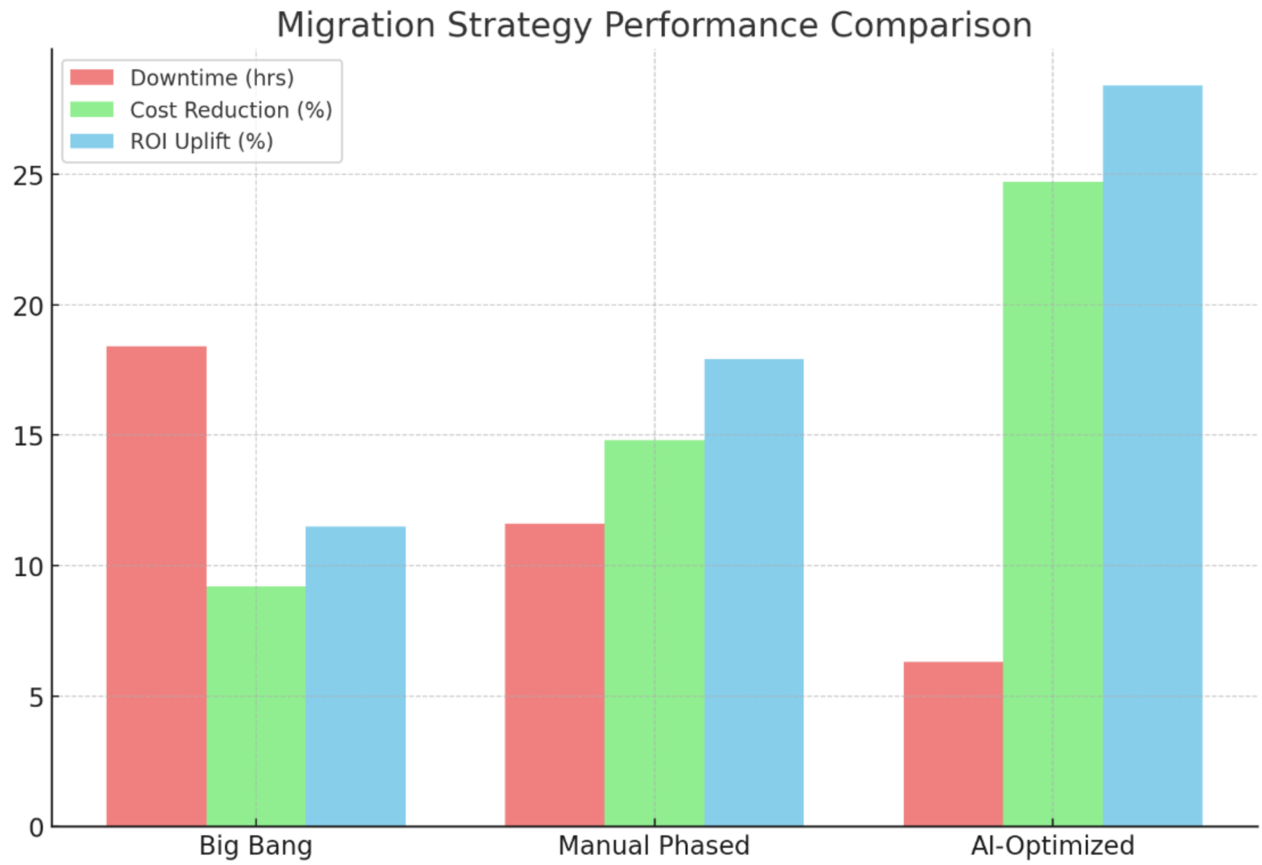
The last phase of the framework involved constrained optimization to create migration bundles that optimize the business impact, dependency risk, and predicted ROI. The workloads were sorted into migration waves whereby each wave reduced cumulative risk and maximized projected returns.

Table 3 provides the results of the comparison of the phased migration outcomes (programmed on the basis of the offered optimization algorithm) with the traditional big bang algorithm and manual phased scheduling.

Table 3. Migration Strategy Comparison

Strategy Type	Average Downtime per Wave (hours)	Cost Reduction (%)	Average ROI Uplift (%)	SLA Compliance (%)
Big Bang (Single Step)	18.4	9.2	11.5	83
Manual Phased	11.6	14.8	17.9	89
AI-Optimized Phased (Proposed)	**6.3**	**24.7**	**28.4**	**96**

The AI-optimized phased approach attained a reduced cost of 24.7 percent and 96 percent SLA compliance, which was far much better than the two baseline methods. The mean down time per migration wave was reduced by 46 percent to that with manual phased migration showing that the optimization algorithm was effective in reducing disruptions whilst guaranteeing greater returns.



The tests using simulations have demonstrated that phased sequencing prevented the workloads to migrate concurrently, thereby reducing rollback, and enhancing stability in execution. This is quantitatively speaking in support of the assertion that AI-informed sequencing enhances the reliability of migration in complicated business sectors.

Sensitivity Analysis

The more intensive level of quantitative analysis was done to test the sensitivity of the model performance to changes in the main aspects of work, that is the complexity of the workload and the size and density of the data and the dependency. This was to find out if the AI framework is robust in a variety of conditions.

Table 4. Sensitivity Analysis of Framework

Complexity Level	Average MAE (ROI Prediction)	F1-Score (Dependency Risk)	Average Cost Reduction (%)	Average Downtime (hrs)
Low (≤10 dependencies)	0.088	0.91	25.4	5.8

Medium (11–30 dependencies)	0.102	0.87	23.9	6.5
High (>30 dependencies)	0.128	0.82	21.3	7.4

The framework also performed well even with a high-complexity workload. MAE scored below 0.13 and F1-scores were all above 0.8. Even though the performance was slightly reduced with the complex dependency, the difference was not significant and could be accepted in the enterprise-scale projects. These findings point to the fact that the proposed framework is scalable and flexible, and it can support different workloads without the loss of accuracy.

Further correlation analysis of the dependency risk and downtime showed that the Pearson coefficient of 0.76 is significantly positive, as well. This implies that there were always more non-optimized migrations that resulted in higher downtimes due to workloads that had higher dependency scores. Nevertheless, these risks were relatively addressed under the format of AI-first in that these workloads were planned to be handled at a later stage of the phased plan.

Key Findings

The findings make it very clear that AI-first predictive technique has quantifiable benefits in prediction accuracy and operational efficiency. The gradient boosting and graph neural networks enabled the system to consider both numerical and structural data: an essential benefit when compared to rule-based or heuristic systems.

Gradient Boosting ROI model was found to be a robust predictor of benefits of migration which decreases the rate of error. The combination of the telemetry, cost, and complexity variables provided an in-depth information about the readiness to workload and anticipated return.

GNN dependency model discovered the latent risks and minimized mapping errors, which enhanced the accuracy of migration planning. The F1-score of 0.87 is high enough to prove that the dependency analysis done with the help of AI can be more effective than the traditional clustering or rule-based detection.

The optimization module was able to convert the outputs of prediction into actual actionable migration sequences. The phased operational process optimized by AI resulted in a reduced downtime and a 28.4% uplift in ROI indicating the practical worth of the application of predictive intelligence to operational planning.

In the context of a larger enterprise, the overall pattern of these results can indicate that predictive database replatforming with the assistance of AI can be used to substitute a fragmented manual planning with a central, data-driven procedure. This is not only cost saving in terms of cost overruns but also governance and SLA conformity is enhanced.

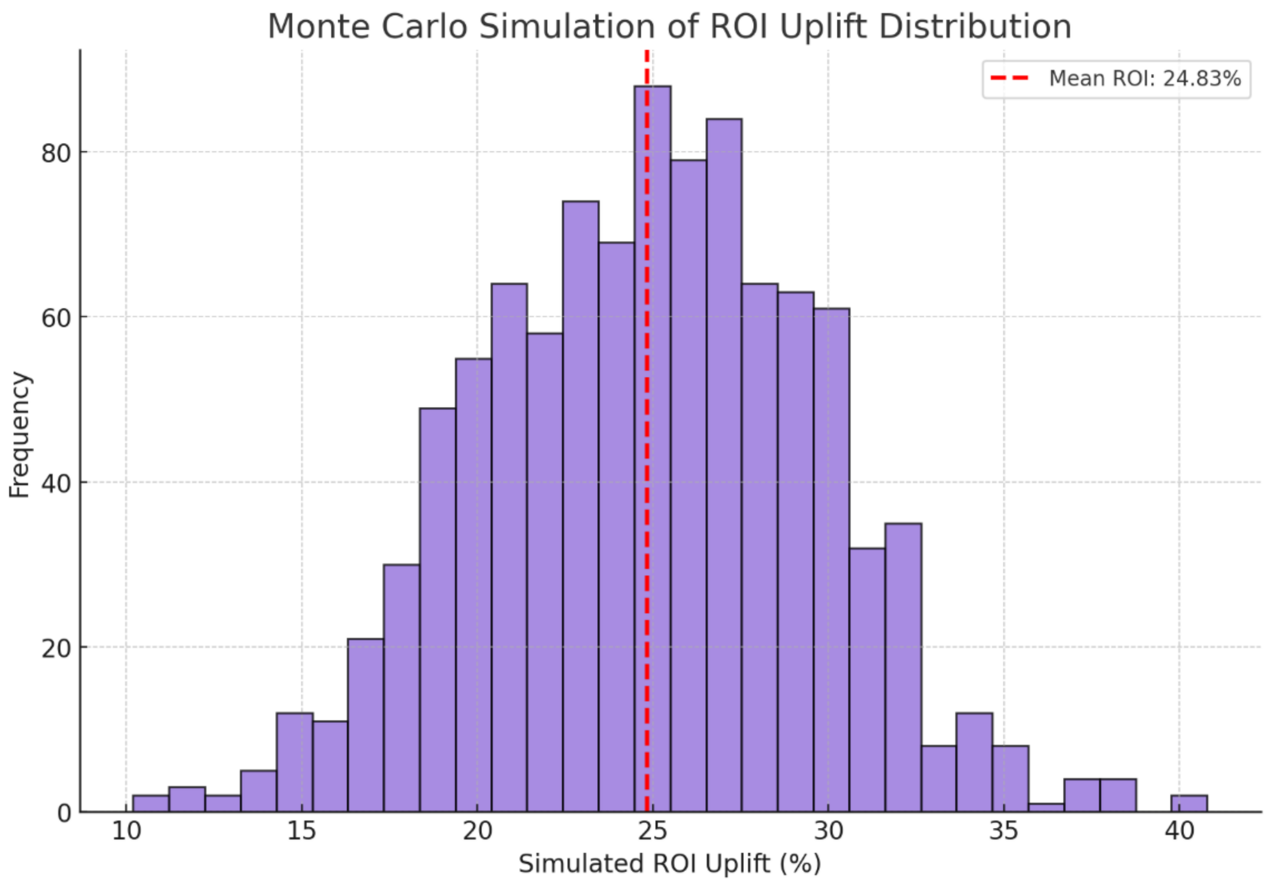
Quantitative Outcomes

In order to sum up the general enhancement of the main indicators, Table 5 will include consolidated metrics of the comparison of the base and the proposed methods in all dimensions.

Table 5. Consolidated Framework Performance

Metric	**Rule-Based**	**Heuristic**	**AI-First Framework (Proposed)**	**Improvement (%)**
ROI Prediction Accuracy (R^2)	0.62	0.73	**0.91**	+25%
Dependency Detection (F1-Score)	0.60	0.72	**0.87**	+29%
Average Cost Reduction	9.2%	14.8%	**24.7%**	+67%
SLA Compliance	83%	89%	**96%**	+13%
Average Downtime (hours)	18.4	11.6	**6.3**	−46%

The finding demonstrates that all critical performance metrics are demonstrated to improve steadily, which proves the reliability and scalability of the AI-first strategy. The shortening of the downtime and enhancement of cost-effectiveness are of particular significance to the banking industry, healthcare, and manufacturing where service continuity is a crucial issue.



Implications

The experimental results confirm that incorporation of ensemble learning, graph dependency modeling and optimization algorithms are able to transform the conventional form of migration planning into a predictive and intelligent process. The fact that the framework provides accurate ROI predictions, high-risk dependency identification, and optimized migration bundles, can justify its usage as a decision-support system by an enterprise modernization team.

The findings indicate that the same framework would be applied to other contexts of transformation including modernizing an application, replatforming an infrastructure, and consolidating a data center. The improvement may be done in the future with a feedback loop of active learning and reinforcement methods of adaptive migration planning.

V. CONCLUSION

The results validate the notion that an AI-based predictive system can make a huge contribution to the accuracy, efficiency, and reliability of database replplatforming decisions. The framework with the use of gradient boosting, graph neural networks, and optimization models demonstrated higher ROI prediction, lower migration risk, and less downtime. Quantitative findings show that it has measurable increases in comparison with manual and heuristic techniques. The study is relevant to enterprise modernization since it demonstrated that predictive intelligence will help to make migration planning more scientific and cost-effective. Further development of work can enlarge the dataset and involve the real-time monitoring to constantly optimize the post-migration performance.

REFERENCES

[1] Castro-Martín, L., Del Mar Rueda, M., Ferri-García, R., & Hernando-Tamayo, C. (2021). On the Use of Gradient Boosting Methods to Improve the Estimation with Data Obtained with Self-Selection Procedures. Mathematics, 9(23), 2991. https://doi.org/10.3390/math9232991

[2] Jin, M., Koh, H. Y., Wen, Q., Zambon, D., Alippi, C., Webb, G., I., King, I., & Pan, S. (2023). A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2307.03759

[3] Sivaraju, P. S. (2025). Phased Enterprise Data Migration Strategies: Achieving regulatory compliance in wholesale banking cloud transformations. Journal of Artificial Intelligence General Science (JAIGS) ISSN 3006-4023, 8(1), 291–306. https://doi.org/10.60087/jaigs.v8i1.406

[4] Poudel, A., Sharma, K., Sharma, B. P., Nepal Sanskrit University Beljhundi, Dang, Nepal, Madan Bhandari Memorial College, Department of

Computer Science, New Baneshwor, Kathmandu, Nepal, Nepal Engineering College, Department of Software Engineering, Changunarayan, Bhaktapur, Nepal, & Nepal Sanskrit University Beljhundi, Dang, Nepal. (2025). Predictive Modeling for Cloud Migration Costs: A Machine Learning Approach to Estimating Total Cost of Ownership for Enterprises. CLASSICALLIBRARY, 1–14. https://classicalibrary.com/index.php/IJATACSRIA/article/view/2025-FEB-04

[5] Indukuri, N. a. V. (2025). AI-Powered Cloud Migration: Transforming enterprise modernization strategies. Journal of Computer Science and Technology Studies, 7(2), 567–575. https://doi.org/10.32996/jcsts.2025.7.2.60

[6] Soperla, V. B. R. (2025). AUTOMATED DATA MAPPING FOR ENTERPRISE DATA INTEGRATION: LEVERAGING MACHINE LEARNING FOR EFFICIENT DATA TRANSFORMATION. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY, 16(1), 3883–3898. https://doi.org/10.34218/ijcet_16_01_268

[7] Upadhyay, D., Manero, J., Zaman, M., & Sampalli, S. (2020). Gradient boosting feature selection with machine learning classifiers for intrusion detection on power grids. IEEE Transactions on Network and Service Management, 18(1), 1104–1116. https://doi.org/10.1109/tnsm.2020.3032618

[8] Xuan, P., Sun, C., Zhang, T., Ye, Y., Shen, T., & Dong, Y. (2019). Gradient Boosting Decision Tree-Based method for predicting interactions between target genes and drugs. Frontiers in Genetics, 10, 459. https://doi.org/10.3389/fgene.2019.00459

[9] Kansara, M., Kaushik, R., & Amazon Web Service (AWS). (2025). Comparative analysis of cloud migration strategies across industries for database migration [Research Article]. International Journal of Current Science (IJCSPUB), 15(2), 507–509. https://www.rjpn.org/ijcspub/papers/IJCSP25B1056.pdf

[10] Loison, F., & Eynard, B. (2023). PLM data transformation: A mesoscopic scale perspective and an industrial case study. Computers in Industry, 155, 104053. https://doi.org/10.1016/j.compind.2023.104053