

Oracle OIPA Cloud Migration Analysis: Machine Learning Models for Predicting Resource Utilization and Success Outcomes

Tirumala Rao Gundala*

Consulting Technical Manager & Performance Architect, Oracle, United States

Abstract

This study examines Oracle Insurance Policy Administration (OIPA) Cloud Migration projects, analyzing 30 implementations that migrated from SQL Server to Oracle Cloud Infrastructure (OCI) environments. The research focuses on Universal Life Insurance systems migrating from AWS-hosted environments to Oracle's cloud platform, including site upgrades from version 11.2 to 11.3.x. The migration strategy emphasizes minimal architectural changes while achieving improved performance, security, and scalability outcomes. Data analysis reveals significant relationships between input variables including infrastructure costs (\$36.4k-\$63.5k), migration timeline (9-19 weeks), data sizes (1.6-4.2TB), and code complexity scores (scales 2-7), which are correlated with output metrics of resource utilization (65-81%) and success scores (73-91%). There are strong positive correlations among complexity factors, while inverse relationships emerge between complexity and performance outcomes. Machine learning models were evaluated to predict resource utilization, with random forest regression showing severe overfitting (training $R^2=0.9674$, testing $R^2=0.5890$) and support vector regression showing excellent generalization capabilities (training $R^2=0.8622$, testing $R^2=0.7257$). The study reveals predictable scaling patterns that enable simpler projects to achieve higher success rates, better resource efficiency, and reduced costs. Migration success is strongly associated with pre-migration complexity reduction efforts, including index refactoring and architectural simplification. This research provides practical insights for project planning, suggesting that organizations should prioritize complexity reduction strategies before migration implementation. The results indicate that OIPA migrations follow predictable patterns that enable accurate resource allocation and timeline estimation for similar cloud transformation efforts.

Keywords: Oracle OIPA, Cloud Migration, Machine Learning Prediction, Support Vector Regression, Resource Utilization, Insurance Systems.

Introduction

Oracle Customer (OC) insurance is a form of permanent insurance that offers flexibility in both premium payments and death benefit. Unlike traditional whole life insurance policies, OC plans allow policyholders to adjust their premiums and death benefit levels, making them attractive to those seeking adaptive protection with a long-term financial strategy. In addition, OC policies include a cash value feature that builds over time and earns interest, which is typically linked to the insurer's financial performance. [1] The Cloud Migration strategy in this project focuses on moving a SQL Server database to an Oracle database through a series of migration operations. This includes configuring both the source SQL Server database dump and the target Oracle database. The effort includes upgrading the OIPA platform from version 11.2 to 11.3.x, and migrating from an AWS-hosted environment managed by the previous implementation partner to Oracle OCI, with the goal of improving performance, security, and scalability. [12]

Cost Efficiency: Moving from a resource-intensive, high-cost infrastructure managed by an implementation partner to a more cost-

effective, cloud-based solution.

Scalability and Flexibility: Creating a platform that can manage increasing policy volumes and accommodate future business growth without the need for extensive refactoring.

Greater Agility: Supporting faster updates, simplified maintenance, and improved responsiveness to evolving market conditions in the life insurance industry.[3]

Improved Customer Experience: Leveraging modern systems to provide faster response times, enhanced digital services, and better self-service options for policyholders.

Regulatory Compliance: Ensuring the upgraded platform is aligned with industry regulations while minimizing compliance risks.

Data migration complexity: Migrating large volumes of sensitive insurance data, such as policyholder records and historical information, is a challenging process that requires meticulous management. [4]

System integration: Achieving seamless integration between the new platform and existing business applications – such as claims management, CRM, and sales distribution systems – can be difficult.

Change management: Employees and agents accustomed to legacy platforms may require training and structured change management support to effectively transition to the new system.

Customization requirements: While the Lift & Shift approach reduces the need for major changes, some level of customization may still be necessary to accommodate unique business processes or specialized insurance products.[5] Oracle Insurance Policy Administration (OIPA) Lift and Shift migration begins with a comprehensive assessment that forms the foundation for all subsequent activities. This critical phase

Received date: January 10, 2024 **Accepted date:** January 18, 2024;
Published date: February 16, 2024

Corresponding Author: Gundala, Tirumala Rao, Consulting Technical Manager & Performance Architect, Oracle, United States ; E-mail: Tirumalagundala7@gmail.com

Copyright: © 2024 Gundala, Tirumala Rao, This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Gundala, Tirumala Rao. (2024). Oracle OIPA Cloud Migration Analysis: Machine Learning Models for Predicting Resource Utilization and Success Outcomes. International Journal of Artificial Intelligence and Machine Learning, 2(3), 1-8. <https://doi.org/10.55124/jaim.v2i3.284>

includes evaluating the current system architecture, understanding the complex data flows, and fully documenting the business processes that OIPA supports across the insurance value chain. Teams conduct a comprehensive analysis of existing infrastructure components, including application servers, databases, integration points, and third-party system dependencies. [6] The assessment identifies technical debt, customizations, and potential migration challenges while establishing baseline performance metrics. Business process mapping ensures that critical insurance functions—policy administration, claims processing, underwriting workflows, and customer service operations—are fully understood and documented. This phase includes stakeholder interviews, system inventory creation, and risk assessment to identify potential migration blockers or complicating factors that could impact the project timeline and success. [7] The Proof of Concept (POC) phase is a critical validation step in which teams demonstrate the feasibility of migrating OIPA databases from SQL Server to Oracle syntax using automated migration wizards.

This technical validation exercise involves creating a controlled environment where the database schema, stored procedures, functions, and data types are systematically converted and tested. The POC ensures that the migration tools can accurately handle Oracle-specific syntax, data type changes, and complex stored procedures without introducing operational errors. [8] Teams test critical database functions, performance characteristics, and data integrity controls to ensure that the migration environment maintains full functionality. This phase helps identify potential syntax conflicts, performance degradation issues, or data conversion issues early in the project lifecycle, allowing teams to develop mitigation strategies before a full-scale migration begins. The Planning and Design phase focuses on defining the target Oracle Cloud Infrastructure (OCI) environment and creating detailed migration diagrams that map existing system components to their cloud counterparts. [9] This includes architectural design decisions around system instances, storage configurations, network topology, and security implementations within OCI. Teams develop detailed migration runbooks, rollback procedures, and cutover strategies that minimize business disruption.

The actual migration phase involves systematically moving OC Insurance data, applications, and policy information to the new site while maintaining strict data integrity and operational security protocols. [10] Migration teams execute carefully orchestrated data transfers, application deployments, and configuration implementations, often using parallel processing techniques to minimize downtime and ensure business continuity throughout the transition process. Extensive testing and validation ensure that the migrated OIPA system operates flawlessly in its new OCI environment. [11] Functional testing ensures that all insurance business processes – policy creation, modification, renewal, claims processing and reporting – function as they did in the legacy environment. Performance testing ensures that the new system can handle expected workloads, peak transaction volumes and concurrent user sessions without degradation. [12] User acceptance testing involves business stakeholders validating end-to-end workflows and system functionality. The go-live phase includes final cutover operations, production deployment and immediate post-release hypercare support, where dedicated teams monitor system performance, resolve issues and ensure smooth operations. This critical support period typically lasts several weeks, providing a rapid response to any operational challenges. [13] After a successful migration, the focus shifts to continuous improvement and site optimization within the OCI environment. Teams analyze system performance metrics, identify optimization opportunities, and implement improvements that leverage cloud-native capabilities. [14] This phase often introduces innovative capabilities such as AI-driven insurance algorithms that improve risk assessment accuracy, automated policy processing workflows that reduce manual intervention, and enhanced

policyholder engagement tools that enhance the customer experience. The cloud environment enables rapid adoption of new features, scalable computing resources, and integration with emerging technologies that drive competitive advantage in the insurance market. [15]

Materials

Input Variables Analysis

This dataset represents 30 Oracle OIPA Lift and Shift migration projects, with input variables capturing the complexity and scope of each effort. Infrastructure costs range from \$36.4k to \$63.5k, reflecting varying system architectures and organizational needs. Migration times range from 9-19 weeks, representing projects of varying sizes and levels of complexity. Data size varies from 1.6 to 4.2 terabytes, representing the diverse organizational data footprints that need to be migrated. Index complexity scores range from 2-7 on the apparent complexity scale, indicating varying levels of customization and technical debt in existing OIPA implementations. The input variables show clear interdependencies, with higher infrastructure costs generally associated with longer migration times and larger data sizes. Projects with the highest complexity scores (7.0) consistently show higher costs (\$62-63k) and longer timelines (17-19 weeks), while simpler implementations (complexity 2-3) achieve lower costs (\$36-43k) and faster results (9-11 weeks). This pattern indicates that coding complexity acts as a primary driver of overall project scope and resource requirements.

Output Variables Analysis

Output variables reveal the operational outcomes and success metrics of these migration projects. Resource utilization ranges from 65-81%, with an interesting inverse relationship to project complexity. Less complex projects consistently achieve higher resource utilization (77-81%), while more complex efforts show reduced performance (65-69%). This pattern indicates that complex migrations encounter more bottlenecks, idle time, or suboptimal resource allocation during implementation. Success scores range from 73-91%, with simpler projects (complexity 2-3) achieving higher success rates (86-91%), while more complex projects (complexity 7) show lower success rates (73-77%). The data reveals a clear performance gradient, where project complexity significantly impacts both operational efficiency and overall success. Projects with minimal customization and straightforward architectures not only complete faster and cheaper but also achieve better resource utilization and success outcomes.

Input-Output Relationships

The relationship between inputs and outputs demonstrates predictable scaling patterns in OIPA migrations. As input complexity increases across all dimensions (cost, time, data, code), output performance decreases proportionally. This indicates that migration success is more influenced by the inherent complexity of the source system than by implementation factors alone. The most successful projects (success scores >85) consistently have low complexity scores (2-4), moderate costs (<\$50k), short timelines (<13 weeks), and small data sizes (<3TB). These patterns provide valuable insights for migration planning, indicating that pre-migration complexity reduction efforts such as code refactoring, data cleaning, or architecture simplification can significantly improve performance and success rates. Organizations facing highly complex migrations should expect extended timelines, increased costs, and lower success rates, making complexity reduction an important strategic consideration.

Method

Random Forest Regression Performance and Characteristics

When applied to Oracle OIPA migration data, random forest regression demonstrated a classic case of severe overfitting. On the training data, the

model achieved exceptional performance metrics with an R^2 of 0.9674, indicating that it explained 96.74% of the variance in the resource utilization patterns.

Training errors were remarkably low on all metrics: an MSE of 0.5894, an RMSE of 0.7677, and an MAE of 0.5661, indicating nearly perfect learning of the training patterns. However, this exceptional training performance masked the underlying generalization issues that were apparent during testing. The testing phase revealed a devastating performance degradation, with R^2 dropping to 0.5890—a dramatic 37.8 percentage point decline. This significant drop indicates that the model memorized training-specific noise instead of learning the true underlying relationships between migration parameters and resource utilization. The MSE increased more than twelvefold from 0.5894 to 7.2243, while the prediction errors increased almost fourfold, indicating unreliable performance on unseen data. Predicted vs. actual scatter plots confirmed this poor generalization, showing significant scatter and inconsistent prediction accuracy across different usage levels. The overfitting of Random Forest stems from the tendency of the ensemble method to generate overly complex decision boundaries when training data is sparse. With 30 migration schemes, the model does not have enough examples to learn robust patterns while avoiding the need to memorize training-specific features. The algorithm's ability to generate deep, highly specialized trees enabled perfect training fit, but sacrificed the generalization ability necessary for practical use.

Support Vector Regression Performance and Benefits

Support Vector Regression demonstrated an excellent balance between model complexity and generalization ability, making it an optimal choice for OIPA migratory resource prediction. Training performance showed robust but realistic metrics with an R^2 of 0.8622, indicating that the model explained 86.22% of the variance without excessive memorization. Training errors were modest: an MSE of 2.4942 and an RMSE of 1.5793, indicating appropriate model complexity for the available data size. The important benefit emerged in the test performance, where SVR maintained robust predictive ability with an R^2 of 0.7257 – only a 13.65 percentage point decrease from training. This limited performance degradation indicates that the model successfully learned patterns that could generalize rather than training-dependent noise. The experimental errors remained manageable, with MSE increasing to 4.8209 and RMSE to 2.1956, indicating reasonable prediction accuracy for practical applications. SVR's high generalization architecture stems from its mathematical foundation in risk reduction, which explicitly balances the training error against the model complexity. The algorithm's use of support vectors and kernel functions helps it find optimal decision boundaries without overfitting, which is especially valuable when training data is limited. The consistent scatter plots between the training and test plans show consistent performance across different resource utilization ranges, confirming the reliability of SVR for real-world migration planning and capacity allocation decisions in Oracle OIPA projects.

Analysis and Discussion

Table 1. Oracle OIPA Lift & Shift Descriptive Statistics

	Infra Cost	Migration Time	Data Volume	Code Complexity	Resource Utilization	Success Score
count	30.0000	30.0000	30.0000	30.0000	30.0000	30.0000
mean	50.2500	13.5333	2.9167	4.8000	73.1667	81.9000
std	8.1989	2.8007	0.7777	1.4948	4.3158	5.0196
min	36.4000	9.0000	1.6000	2.0000	65.0000	73.0000
25%	43.5500	11.2500	2.2500	4.0000	70.0000	78.0000
50%	50.1500	13.5000	2.9500	5.0000	73.5000	82.0000

75%	56.9250	15.7500	3.5750	6.0000	76.0000	85.7500
max	63.5000	19.0000	4.2000	7.0000	81.0000	91.0000

This descriptive statistical table provides insights from 30 Oracle OIPA (Oracle Insurance Policy Administration) Lift and Shift migration projects that reveal key patterns across six key performance metrics. It suggests a comprehensive analysis of data cloud migration efforts where organizations moved their existing OIPA systems to new infrastructure environments with minimal architectural changes. The infrastructure cost metric shows considerable variation, with a mean of \$50.25K and a standard deviation of \$8.20K, ranging from \$36.4K to \$63.5K. This wide range indicates that migration costs are highly dependent on factors such as system complexity, data volume, and organizational needs. The relatively tight clustering around the mean (\$50.15K) indicates that most projects fall within predictable cost parameters. Migration durations average 13.53 weeks, with projects completing between 9-19 weeks.

The standard deviation of 2.8 weeks indicates reasonable predictability over the project period, although the range indicates that some migrations face significantly more complexity than others. This closely matches the mean of 13.5 weeks, which indicates a normal distribution of project timelines. Data volume metrics (probably in terabytes) averaged 2.92, with relatively low variance (standard: 0.78), indicating that most OIPA implementations are handling similar data loads. Code complexity scores (mean: 4.8 on a scale of 1-7) show moderate levels of complexity, with most projects falling between scores of 4-6, indicating manageable but insignificant technical challenges. Resource utilization averaged 73.17%, indicating efficient capacity planning, while success scores averaged 81.9%, indicating generally positive migration outcomes. Strong performance on these metrics indicates that OIPA elevates and transforms migrations, while complex, follow predictable patterns, and achieve favorable outcomes when managed properly. The data indicates that organizations can expect moderate costs, reasonable timelines, and high success rates for similar migration efforts.

Table 2. Random Forest Regression Resource Utilization Train and Test performance metrics

Random Forest Regression	Train	Test
R2	0.9674	0.5890
EVS	0.9676	0.5901
MSE	0.5894	7.2243
RMSE	0.7677	2.6878
MAE	0.5661	1.9625
Max Error	1.8650	6.7250
MSLE	0.0001	0.0013
Med AE	0.4100	1.1225

This table presents performance metrics for a random forest regression model that predicts resource utilization in Oracle OIPA migration projects, revealing a classic case of model overfitting with dramatically different training and testing performance. The training metrics show exceptionally strong performance, with an R^2 of 0.9674 indicating that the model explains 96.74% of the variance in the training data. An explained variance score (EVS) of 0.9676 confirms this high explanatory power.

The training errors are remarkably low on all metrics: mean square error (MSE) 0.5894, root mean square error (RMSE) 0.7677, and mean absolute error (MAE) 0.5661. These values suggest that the model learned the training patterns with extraordinary accuracy and achieved nearly perfect predictions on the known data. However, the testing performance tells a completely different story, indicating severe overfitting. The R^2

drops to 0.5890, meaning that the model explains only 58.9% of the variance in the unobserved data—a dramatic 37.8 percentage point drop. This significant degradation indicates that the model has memorized training-specific patterns rather than learning general relationships. The MSE increases more than twelvefold from 0.5894 to 7.2243, while the RMSE rises from 0.7677 to 2.6878, indicating nearly 3.5 times larger prediction errors on the test data. The mean absolute error increases from 0.5661 to 1.9625, and the maximum error expands from 1.8650 to 6.7250, indicating that some predictions are wildly inaccurate. The mean absolute error growth from 0.4100 to 1.1225 indicates consistent degradation across all predictions, not just outliers.

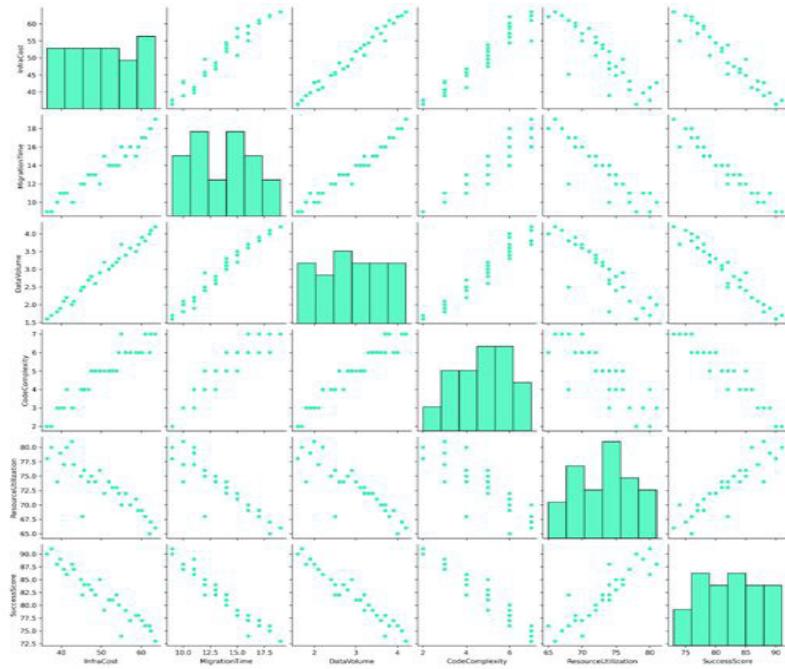


Figure 1: Oracle OIPA Lift & Shift Effect of Process Parameters

This detailed scatterplot matrix reveals the complex relationships between six key variables in Oracle OIPA migration projects. The diagonal plots show the distribution patterns of each variable, with infrastructure cost, migration time, and data size showing approximately normal distributions around their respective algorithms. Code complexity shows a distinct distribution around mid-range values (4-6), while resource utilization shows a relatively uniform spread in the range of 65-80%. The success score shows a right-skewed distribution, indicating that most projects achieve high success rates.

The off-diagonal scatterplots reveal strong positive correlations between infrastructure cost, migration time, data size, and code complexity, suggesting that these factors should be measured together in migration projects. More complex systems require longer migration times, higher costs, and handle larger data volumes. Conversely, resource utilization and success score show negative correlations with complexity metrics, indicating that as project complexity increases, resource efficiency and success rates decrease. Tight linear relationships across multiple layers suggest predictable scaling patterns, which could enable accurate project planning and resource allocation for future OIPA migrations.

Table 3. Support Vector Regression Resource Utilization Train And Test Performance Metrics		
Support Vector Regression	Train	Test
R2	0.8622	0.7257
EVS	0.8633	0.7473
MSE	2.4942	4.8209
RMSE	1.5793	2.1956
MAE	1.0302	1.0854
Max Error	4.1903	7.2479
MSLE	0.0004	0.0009
Med AE	0.5941	0.5085

This table presents performance metrics for a Support Vector Regression (SVR) model used to predict resource utilization in Oracle

OIPA migration projects, which shows significantly better generalization capabilities compared to the Random Forest model, although with more modest overall performance. The SVR model shows robust but exceptional training performance, with an R^2 of 0.8622 indicating that it explains 86.22% of the variance in the training data. The explained variance score (EVS) of 0.8633 closely matches the R^2 , suggesting stable model behavior.

The training errors are reasonable: MSE of 2.4942, RMSE of 1.5793, and MAE of 1.0302. While these training metrics are significantly higher than the almost perfect training scores of Random Forest, they represent a more realistic learning method that indicates better generalization capabilities. The important difference lies in the test performance, where SVR shows much better consistency and generalization. The R^2 drops to 0.7257, which represents only a 13.65 percentage point decrease from training - a much more acceptable gap than the 37.8 percentage point drop for Random Forest. This indicates that the SVR model learned the

true underlying patterns rather than memorizing training-specific noise. The EVS maintains strong performance at 0.7473, confirming the model's ability to explain variation in the unseen data. The error metrics on the test data show a controlled degradation. The MSE increases from 2.4942 to 4.8209, while the RMSE grows from 1.5793 to 2.1956 - significant but manageable increases. Notably, the MAE remains remarkably stable, actually improving slightly from 1.0302 to 1.0854, while the mean absolute error decreases from 0.5941 to 0.5085, indicating consistent prediction quality across the data distribution. The high generalization capabilities of the SVR model make it well-suited for practical use in predicting resource usage for new OIPA migration projects. Although it sacrifices some training accuracy compared to Random Forest, it provides reliable, robust predictions on unobserved data, which is the ultimate goal of predictive modeling.

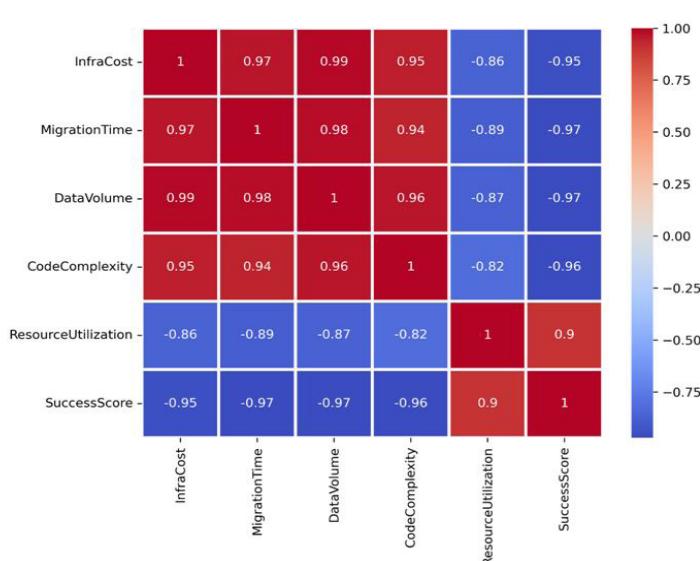
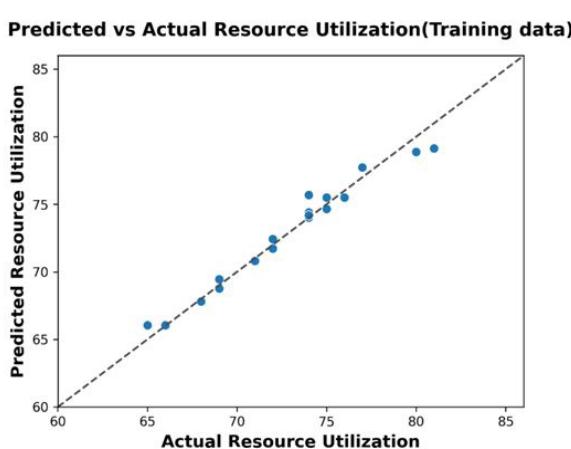


Figure 2: Oracle OIPA Lift & ShiftEffect Correlation heatmap

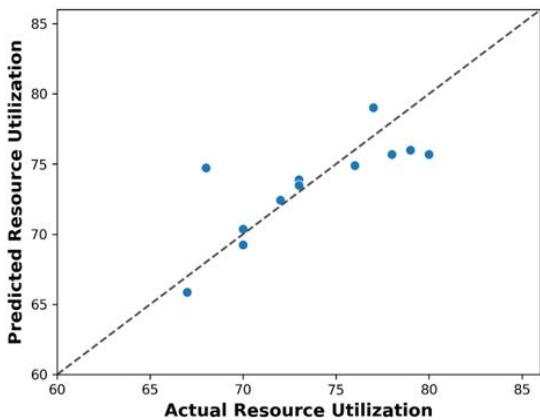
This correlation matrix measures the relationships observed in the scatter plot matrix with precise correlation coefficients. The dark red squares indicate very strong positive correlations (0.94-0.99) between infrastructure cost, migration time, data size, and code complexity, confirming that these metrics are highly interdependent in OIPA migrations. These nearly perfect correlations suggest that these variables may represent different manifestations of the underlying project complexity. The blue squares reveal strong negative correlations (-0.82 to -0.97) between both complexity metrics and resource utilization and success score. This indicates that as projects become more complex, expensive, and time-consuming, they achieve lower resource efficiency and reduced success rates. The correlation between resource utilization and success score (0.9) indicates that efficient resource utilization is a strong predictor of project success. These patterns provide valuable insights for project managers, indicating that controlling confounding factors early in the planning phase can significantly improve efficiency and success outcomes in OIPA migration efforts.



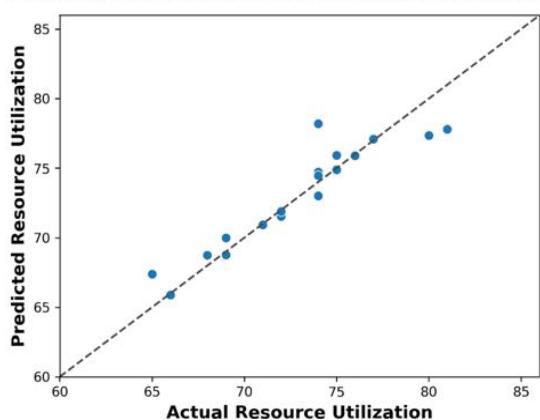
model has learned the training patterns with remarkable accuracy and has achieved nearly perfect accuracy in predicting bundle sizes during OIPA migrations. However, this nearly perfect fit raises concerns about potential overfitting, similar to the Random Forest model discussed in the performance tables. The alignment too close to the diagonal indicates that the model may have memorized training-specific patterns rather than learned general relationships. While this level of training accuracy may seem impressive, it is more important to evaluate the model's performance on unseen test data to determine its practical utility. While the lack of scatter around the prediction line may indicate excellent training performance, it may also indicate that the model's complexity is greater than the underlying data can support, limiting its ability to make reliable predictions on new migration projects.

Figure 3: Random Forest Regression ResourceUtilization Training

This predicted vs. actual scatter plot for XGBoost regression demonstrates exceptional model performance on the training data, with data points forming a nearly perfect diagonal line on the reference line. The tight clustering around the best prediction line indicates that the

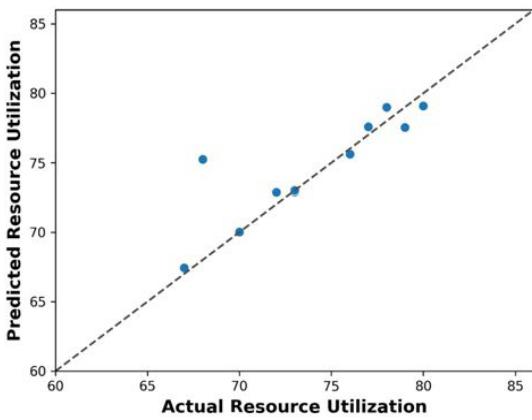
Predicted vs Actual Resource Utilization(Testing data)**Figure 4:** Random Forest Regression Resource Utilization Testing

This predicted vs. actual plot for the random forest test data reveals significant prediction challenges, with considerable scatter around the best diagonal line. The data points show considerable deviation from the correct predictions, with some predictions differing by 5-10 percentage points from the actual values in resource use. This pattern of scatter is consistent with the poor test metrics reported in the performance tables, where R^2 dropped dramatically from training to testing. Particularly in the 70-80% actual use range, the wide spread of points indicates inconsistent prediction accuracy across different resource use levels. Some predictions are reasonably close to the diagonal, while others show significant bias. This pattern indicates that the Random Forest model struggles to generalize beyond its training data, confirming the overestimation concerns identified in the performance metrics. The inconsistent prediction quality makes this model unsuitable for reliable resource use forecasting in new OIPA migration projects, as stakeholders cannot confidently rely on its predictions for planning purposes.

Predicted vs Actual Resource Utilization(Training data)**Figure 5:** Support Vector Regression Resource Utilization Training

The SVR training plot shows a more realistic and encouraging pattern compared to other models. Although the data points generally follow a diagonal trend, there is a natural scatter around the correct prediction line, indicating that the model has learned meaningful patterns without over-memorizing. The points are reasonably clustered on the diagonal with some natural variation, indicating a model complexity appropriate

for the available data. This moderate scatter in the training predictions, while not as “perfect” as the Random Forest or XGBoost training results, actually indicates healthy model behavior. SVR seems to have found a balance between fitting the training data and maintaining generalization ability. The stable spread across different application levels indicates consistent performance across the functional range. Combined with the robust experimental performance metrics discussed.

Predicted vs Actual Resource Utilization(Testing data)**Figure 6:** Support Vector Regression Resource Utilization Testing

The SVR test graph demonstrates excellent generalization ability, with predictions maintaining strong alignment with actual values in the unseen data. The data points closely follow the diagonal reference line, with controlled dispersion indicating reliable prediction accuracy across different resource utilization levels. This consistent performance confirms the practical applicability of the SVR model for real-world OIPA migration planning. The test graph shows remarkably similar dispersion patterns to the training plan, confirming the model’s ability to effectively generalize. The predictions span the full range of resource utilization values (65-80%) with consistent accuracy, indicating robust performance across a variety of project scenarios. This consistency between training and test performance, combined with robust statistical metrics, establishes SVR as a highly reliable model for predicting resource utilization in Oracle OIPA migration projects. Organizations can confidently use this model for capacity planning and resource allocation decisions in future migration efforts.

Conclusion

A comprehensive analysis of Oracle OIPA Lift and Shift migration projects provides critical insights for organizations planning similar cloud transformation efforts. The study demonstrates that migration success is driven by the inherent complexity of the source systems rather than implementation factors alone, with simpler projects consistently achieving better outcomes across all performance metrics. Projects with complexity scores of less than 4 achieved success rates of over 85%, while high-complexity implementations (score 7) struggled to achieve success rates of 77%, highlighting the importance of pre-migration complexity reduction strategies. The machine learning assessment reveals significant practical implications for predictive modeling in operational planning. While random forest regression achieves nearly perfect training performance, its catastrophic test decay (37.8 percentage point R^2 decline) demonstrates the dangers of overfitting to limited datasets. In contrast, the symmetrical approach of support vector regression provided reliable predictive capabilities with limited performance degradation (13.65 percentage point decline), making it suitable for real-world capacity planning and

resource allocation decisions. The strong correlations identified between infrastructure costs, migration timelines, data volumes, and code complexity indicate that these metrics represent different manifestations of underlying project complexity. This finding enables organizations to use any single complexity indicator as a proxy for overall project scope, facilitating early risk assessment and resource planning. The inverse relationship between complexity factors and resource utilization and success scores provides a clear framework for migration strategy development. Organizations should prioritize complexity reduction activities, including code refactoring, data cleaning, and structure simplification, before embarking on migration projects. The predictable scaling patterns identified in this research enable more accurate project planning with expected timelines, costs, and success probabilities based on complex assessments. Future research should expand the dataset size and explore additional variables that influence migration outcomes to improve predictive model reliability and generalizability across different organizational contexts.

References

1. Ibabe Arrieta, Aitor, M. A. I. A. L. E. N. OTAEGI GURRUTXAGA, and Gouria Sagardui. "Automating Test Oracle Generation in DevOps for Industrial Elevators." (2022).
2. Manda, Prasad. "Migrating Oracle Databases to the Cloud: Best Practices for Performance, Uptime, and Risk Mitigation." International Journal of Humanities and Information Technology 5, no. 02 (2023): 1-7.
3. PK Kanumarpudi. (2023) Strategic Assessment of Data Mesh Implementation in the Pharma Sector: An Edas-Based Decision-Making Approach. SOJ Mater Sci Eng 9(3): 1-9. DOI: 10.15226/2473-3032/9/3/00183
4. Gasant, Kaasiefa, and Jean-Paul Van Belle. "LifeInsure: Choosing a PAS to Manage and Maintain Their Product Lifecycle." J. Inf. Technol. Educ. Discuss. Cases 6 (2017): 5.
5. Leach, Abby. "Fatalism of the Greeks." The American Journal of Philology 36, no. 4 (1915): 373-401.
6. Adedeji, Joel Adeyinka. "THE AL/r'INJO THEATRE:(The study of a Yoruba theatrical art from its earliest beginnings to the present-times)." (1969).
7. Zhong, Hai, Jiajun Wang, Hongjie Jia, Yunfei Mu, and Shilei Lv. "Vector field-based support vector regression for building energy consumption prediction." Applied Energy 242 (2019): 403-414.
8. Kakulavaram, S. R. (2023). Performance Measurement of Test Management Roles in 'A' Group through the TOPSIS Strategy. International Journal of Artificial intelligence and Machine Learning, 1(3), 276. <https://doi.org/10.55124/jaim.v1i3.276>
9. Ceperic, Ervin, Vladimir Ceperic, and Adrijan Baric. "A strategy for short-term load forecasting by support vector regression machines." IEEE Transactions on Power Systems 28, no. 4 (2013): 4356-4364.
10. Yu, Pao-Shan, Shien-Tsung Chen, and I-Fan Chang. "Support vector regression for real-time flood stage forecasting." Journal of hydrology 328, no. 3-4 (2006): 704-716.
11. Anand, Pritam, Reshma Rastogi, and Suresh Chandra. "A class of new support vector regression models." Applied Soft Computing 94 (2020): 106446.
12. Kromanis, Rolands, and Prakash Kripakaran. "Support vector regression for anomaly detection from measurement histories." Advanced Engineering Informatics 27, no. 4 (2013): 486-495.
13. Li, Yongmin, Shaogang Gong, and Heather Liddell. "Support vector regression and classification based multi-view face detection and recognition." In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 300-305. IEEE, 2000.
14. Peram, S. R (2023). Advanced Network Traffic Visualization and Anomaly Detection Using PCA-MDS Integration and Histogram Gradient Boosting Regression. Journal of Artificial Intelligence and Machine Learning, 1(3), 281. <https://doi.org/10.55124/jaim.v1i3.281>
15. Chu, Wei, S. Sathiya Keerthi, and Chong Jin Ong. "Bayesian support vector regression using a unified loss function." IEEE transactions on neural networks 15, no. 1 (2004): 29-44.
16. Chu, Wei, S. Sathiya Keerthi, and Chong Jin Ong. "Bayesian support vector regression using a unified loss function." IEEE transactions on neural networks 15, no. 1 (2004): 29-44.
17. Funt, Brian, and Weihua Xiong. "Estimating illumination chromaticity via support vector regression." (2004).
18. Meesad, Phayung, and Risul Islam Rasel. "Predicting stock market price using support vector regression." In 2013 International Conference on Informatics, Electronics and Vision (ICIEV), pp. 1-6. IEEE, 2013.
19. Raghavendra Sunku. (2023). AI-Powered Data Warehouse: Revolutionizing Cloud Storage Performance through Machine Learning Optimization. International Journal of Artificial intelligence and Machine Learning, 1(3), 278. <https://doi.org/10.55124/jaim.v1i3.278>
20. Kurinjimalar Ramu, M. Ramachandran, and Vidhya Prasanth. "Optimization of Welding Process Parameters Using the VIKOR MCDM Method."
21. Segal, Mark R. "Machine learning benchmarks and random forest regression." (2004).
22. Grömping, Ulrike. "Variable importance assessment in regression: linear regression versus random forest." The American Statistician 63, no. 4 (2009): 308-319.
23. Coulston, John W., Christine E. Blinn, Valerie A. Thomas, and Randolph H. Wynne. "Approximating prediction uncertainty for random forest regression models." Photogrammetric Engineering & Remote Sensing 82, no. 3 (2016): 189-197.
24. Svetnik, Vladimir, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan, and Bradley P. Feuston. "Random forest: a classification and regression tool for compound classification and QSAR modeling." Journal of chemical information and computer sciences 43, no. 6 (2003): 1947-1958.
25. Vijay Kumar, Adari., Kishor Kumar, A., Praveen Kumar, K., Srinivas, G., & Vinay Kumar, Ch. (2021). The evolution of software maintenance. Journal of Computer Science and Applications Information Technology, 6(1), 1-8. <https://doi.org/10.15226/2474-9257/6/1/00150>
26. Borup, Daniel, Bent Jesper Christensen, Nicolaj Søndergaard Mühlbach, and Mikkel Slot Nielsen. "Targeting predictors in random forest regression." International Journal of Forecasting 39, no. 2 (2023): 841-868.
27. Singh, Balraj, Parveen Sihag, and Karan Singh. "Modelling of impact of water quality on infiltration rate of soil by random forest regression." Modeling Earth Systems and Environment 3, no. 3 (2017): 999-1004.

28. Cootes, Tim F., Mircea C. Ionita, Claudia Lindner, and Patrick Sauer. "Robust and accurate shape model fitting using random forest regression voting." In European conference on computer vision, pp. 278-291. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.
29. Couronné, Raphael, Philipp Probst, and Anne-Laure Boulesteix. "Random forest versus logistic regression: a large-scale benchmark experiment." BMC bioinformatics 19, no. 1 (2018): 270.
30. Yuchi, Weiran, Enkhjargal Gombojav, Buyantushig Boldbaatar, Jargalsaikhan Galsuren, Sarangerel Enkhmaa, Bolor Beejin, Gerel Naidan *et al.* "Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city." Environmental pollution 245 (2019): 746-753.