Open Access   **Review Article**

AI-Powered Data Warehouse: Revolutionizing Cloud Storage Performance through Machine Learning Optimization

Raghavendra Sunku*

Data Engineer, Central Mutual insurance., United States

Abstract

AI-powered data warehousing represents a significant evolution in cloud storage technologies, reshaping the way organizations handle, analyze, and extract insights from their growing data volumes. As data generation accelerates across a variety of industries, traditional on-premises storage solutions are proving inadequate to meet the growing demands for scalability, high-speed processing, and intelligent data management. Combining Cloud-based machine learning (ML) and artificial intelligence (AI) warehousing addresses these challenges by unlocking powerful capabilities such as real-time analytics, predictive insights, and automated data manipulation. This study explores the transformative role of AI in cloud storage, focusing on key performance indicators such as data volume, query latency, AI optimization, redundancy strategies, and compression performance. Using advanced Machine learning techniques, especially random forest, decision tree regression, and XG boost regression– we evaluated the performance of cloud storage systems under various configurations.

Our analysis confirms that AI-integrated solutions outperform conventional methods, especially in optimizing queries, ensuring data quality and dynamically allocating resources. The results highlight that data compression rates and AI optimization levels are the most influential factors affecting cloud storage performance. Correlation analysis reveals strong positive correlations between these variables and overall performance metrics. Among the models tested, Random Forest Regression demonstrated the best accuracy, achieving R^2 scores of 0.9965 on the training data and 0.9520 on the test data, indicating excellent generalization. XGBoost, while accurate on the training data, showed signs of overfitting. In contrast, decision tree regression struggled with the complexity of the data and revealed its limitations. Furthermore, the study highlights how AI enables real-time data processing, supports intelligent automation through robotic process automation (RPA), and improves cyber security through continuous operational monitoring and anomaly detection. Server less architecture benefit from AI-driven optimization for resource management and cost control, while hybrid and multi-cloud deployments benefit from adaptive workload distribution and seamless data integration. While barriers such as talent shortages and administrative requirements are important considerations, these advances empower organizations to make faster decisions, increase efficiency, and gain a competitive advantage.

Key words: Artificial Intelligence, Cloud Data Warehouse, Machine Learning Optimization, Real-time Analytics, Predictive Modeling, Automated Data Management, Cloud Storage Capacity, XGBoost Regression, Random Forest Algorithm, Data Compression Technology.

Introduction

AI-powered data warehousing is playing a key role in revolutionizing cloud storage, significantly changing How businesses manage, analyze, and derive value from their data. As data volumes grow rapidly across a variety of industries, traditional on-premises data warehouses are struggling to meet the growing demands for scale, speed, and intelligent processing. The shift to cloud-based data warehousing Solutions powered by machine learning (ML) and artificial intelligence (AI) represents a major shift. This advancement not only overcomes the limitations of legacy systems, but also enables faster decision-making, improved operational efficiency, and better business adaptability. [1] The key to this transformation lies in AI's ability to streamline and optimize critical data warehousing operations. AI and ML are increasingly being used in areas such as data ingestion, transformation, intelligent query processing, and predictive analytics. For example, AI can study past query behavior and automatically optimize queries for better speed and accuracy. This leads to faster and more accurate data access, allowing companies to respond quickly with informed strategies. In addition, AI-powered predictive tools help companies predict trends, understand customer behavior, and assess potential risks - increasing their competitiveness in dynamic markets.[2] A unique feature of an AI-enhanced cloud data warehouse is its ability to provide real-time analytics. Unlike traditional systems that are often hampered by infrastructure issues, latency issues, and the need for manual processes, modern AI-integrated cloud platforms can analyze streaming data as it arrives.

This real-time capability is particularly beneficial Rapid insights can be productive in sectors like manufacturing, retail, and finance. better customer service, streamlined inventory control, and faster operational responses.[3] The benefits of AI extend beyond analytics. In the field of data management, AI significantly improves data quality by automating tasks such as data cleansing, de duplication, classification, and outlier detection. These processes, which were once Tasks that were previously difficult and prone to human error are now handled by intelligent systems in a timely and accurate manner. This automation is not only workload for data engineers, but also ensures that incoming data meets high standards of accuracy and consistency. Furthermore, AI facilitates efficient metadata management, improving data sequence tracking and

discovery – key components for effective data governance and regulatory compliance.[4]A key benefit of AI-driven cloud data warehouses is scalability. Unlike traditional systems that require manual upgrades and significant hardware investments, cloud-based platforms offer flexible, on-demand scaling. Both Storage and processing power can be automatically adapted to changing workloads and fluctuations in data volume without compromising performance. AI further enhances this adaptability by predicting usage patterns and proactively allocating resources.

This intelligent scaling ensures optimal performance while keeping costs efficient, allowing the organization to be responsive and efficient under changing operational demands.[5] Equally important is the contribution of AI to security and regulatory compliance within the cloud data warehouse. AI-driven security solutions provide continuous monitoring of user behavior, data access, and system operations to quickly detect and respond to suspicious actions or breaches. These tools automate the enforcement of data security policies and ensure compliance with industry regulations including PCI DSS, GDPR, and HIPAA. By identifying risks in real-time and providing recommended mitigation strategies, AI strengthens the overall resilience of the system, helping organizations stay protected against emerging cyber security threats.[6] Intelligent automation is made possible by integrating AI with cloud data warehousing, particularly through the use of Robotic Process Automation (RPA). With the help of machine learning, modern RPA systems can perform complex data tasks such as ETL (extract, transform, load) with minimal human intervention.

These intelligent algorithms can learn from human interactions, adapt to changing data structures, and perform better over time. As a result, reduced operational costs and faster access to insights. For example, organizations that use AI-powered RPA within their cloud data warehouse can automate the creation of dashboards and reports, which enables real-time analytics to support strategic decision-making.[7]Another significant innovation in this space is the adoption of server less architectures within AI-enhanced cloud data warehousing platforms. Server less architectures allow users to run queries and run applications without the need to manage servers or backend infrastructure. AI makes a significant contribution to these environments by improving how resources are allocated. It identifies the most efficient way to execute queries, dynamically allocates compute and memory resources, and automatically deactivates unused components to reduce costs. This approach streamlines usage, increases system flexibility, and frees developers and analysts It takes more effort to extract useful insights from data than to deal with infrastructure maintenance.[8]Artificial intelligence (AI) is playing a key role in enabling hybrid and multi-cloud data warehouse strategies. As organizations increasingly migrate workloads across multiple cloud providers to improve performance, manage risk, and meet regulatory compliance, AI is facilitating the integration of these complex environments. It ensures consistent data movement, real-time synchronization, and dynamic workload balancing across multiple cloud platforms. This functionality is especially important for global enterprises that need to maintain data sovereignty while maintaining system availability and disaster recovery readiness.[9]However, integrating AI into a cloud-based data warehouse comes with its own challenges.

A primary issue is the steep learning curve and the need for specialized knowledge in areas such as AI, machine learning, and data engineering. Many organizations struggle to fill this skills gap, which prevents them from effectively implementing AI systems. Furthermore, AI models rely on extensive datasets for training and operation, which raises concerns about data privacy, ownership, and ethical governance. It is essential to ensure transparency and fairness in AI releases and adopt robust governance structures. Real-world applications illustrate the tangible benefits of an AI-powered cloud data warehouse. For example, Netflix

uses real-time analytics and scalable infrastructure to deliver personalized content recommendations, improve user engagement. Home Depot uses AI for inventory optimization and demand forecasting, improving operational workflows. Capital, one uses AI for security and compliance management in the financial sector, demonstrating the effectiveness of AI in risk mitigation.[10]

Materials and Method

XGBoost Regression

XGBoost (Higher Gradient ascension serves as the foundation for regression, a complex ensemble technique. It builds decision trees incrementally, with each new tree aiming to minimize the errors of its predecessors. The model is improved by a formalized objective function that combines the loss from predictions with a penalty for model complexity. This approach improves both accuracy and resistance to overfitting. XGBoost is particularly known for its high performance and scalability, thanks to its support for parallel computation, intelligent tree pruning, and built-in handling of missing values. When applied to predictive tasks such as estimating cloud storage performance, XGBoost typically delivers superior performance compared to more conventional models. This is largely due to its ability to effectively model both linear and nonlinear patterns. It often achieves high R^2 values and minimal error rates on training data, reflecting its strong learning ability. However, one limitation is the potential for overfitting, especially as the model grows more complex or the amount of training data is limited. A large difference between test and training performance can be a sign of overfitting, which occurs when the model fits the training data too closely. This can be avoided by carefully tuning hyper parameters, using regularization techniques, and using cross-validation techniques to ensure that the model performs well when applied to new data.

Decision Tree Regression

Decision tree regression is a supervised, nonparametric learning method used to predict sequential numerical outcomes. This model creates a tree-like structure by repeatedly dividing a dataset into smaller subsets according to feature values. Each terminal leaf node provides a predicted value, while each internal node represents a judgment based on a feature range. Its straightforward and visual structure makes it easy to interpret, making it particularly valuable for understanding how different features affect predictions and for generating rule-based insights. Despite its simplicity and clarity, decision tree regression has significant drawbacks. It is very sensitive to specific characteristics of the training data, which often leads to overfitting - especially if the tree grows too deep and is not pruned enough. Unlike ensemble models, a single decision tree typically lacks robustness and may fail to successfully generalize from new, untested data. This often appears at a low level.R^2 scores and increased prediction errors. While this model performs well on datasets with linear or less complex relationships, it may struggle to capture complex patterns unless heavily parameterized, which again increases the risk of overfitting. However, decision tree regression remains a useful tool, especially in initial data exploration or when model descriptiveness is prioritized over predictive accuracy. In applications such as cloud storage performance prediction, it can serve as a baseline model or be integrated as a basic component into more sophisticated ensemble methods such as Random Forest or XGBoost to improve overall performance.

Random Forest Regression

An ensemble-based regression technique called Random Forest Regression creates a large number of decision trees during training, averages their results, and provides predictions. To create a variety of trees, it uses techniques such as randomization and packing, in which each tree

is trained on a different part of the data. This method successfully reduces the overfitting problems that are common with individual decision trees, improving accuracy and consistency. This model performs exceptionally well on high-dimensional datasets and scenarios involving complex, nonlinear relationships, such as predicting cloud storage performance. It strikes a strong balance between bias and variance, often yielding high R^2 scores and low error values in test and training sets. A key strength of Random Forest is its resilience to noise and outliers, and it typically requires little fine-tuning to perform well. However, there is a trade-off in terms of computational cost - building and maintaining a large number of trees can be resource-intensive. In addition, the ensemble nature of the model makes its predictions more difficult to interpret than a single decision tree. Random Forest is widely preferred for applications where predictive accuracy and model generalization are prioritized over explanation.

It serves as a reliable middle ground between simple, interpretable models such as Decision Tree Regression and more sophisticated, performance-oriented methods such as XGBoost, making it a practical choice for many real-world regression problems. The dataset contains 20 records that include six key variables that affect cloud storage performance: data size (in terabytes), query speed (in milliseconds), AI optimization (%), storage redundancy (%), data compression ratio, and cloud storage efficiency score (CSES). A detailed examination of these parameters reveals patterns and key contributors to the overall system performance. Data size and performance score are significantly negatively correlated. Smaller data sizes such as 5.8 TB and 6.7 TB yield higher performance scores (93 and 94), while larger sizes such as 42.0 TB and 40.7 TB correspond to lower scores (62 and 64). This indicates that larger data sets can hinder performance due to increased computational and retrieval demands. Query speed emerges as another influential factor. Low query times (e.g., 95 MS and 98 MS) are associated with high performance scores (93 and 92), while high latency (430 MS and 420 MS) is associated with low scores (62 and 64). This indicates that faster data retrieval significantly improves operational efficiency.AI optimization shows a strong positive influence on performance. Systems with AI optimization levels exceeding 85% consistently achieve high CSES values – examples include scores of 91, 92, and 94 – demonstrating that advanced optimization techniques improve processing capabilities. On the other hand, storage over clocking appears to negatively impact performance. Low over clocking levels (10–15%) are associated with higher performance, likely due to reduced resource utilization, while higher over clocking (35–38%) is associated with lower CSES outcomes.

Analysis and Discussion

Table 1: Summarizes the key descriptive statistics for the six variables related to cloud storage systems						
	Data Volume (TB)	Query Speed (ms)	AI Optimization (%)	Storage Redundancy (%)	Data Compression Ratio	Cloud Storage Efficiency Score
count	20.0000	20.0000	20.0000	20.0000	20.0000	20.0000
mean	20.0450	234.6500	68.0500	24.2000	2.1250	79.4000
std	12.3145	120.0146	16.8912	8.9065	0.6797	10.2977
min	5.8000	95.0000	40.0000	10.0000	1.2000	62.0000
25%	10.2500	117.5000	57.2500	17.2500	1.5750	71.5000
50%	16.4000	230.0000	66.5000	25.5000	2.1000	78.5000
75%	28.9750	322.5000	82.0000	31.2500	2.5250	88.5000
max	42.0000	430.0000	92.0000	38.0000	3.3000	94.0000

Table 1 summarizes the key descriptive statistics for the six variables related to cloud storage systems: data size, query speed, AI optimization, storage hypervisor, data compression ratio, and cloud storage efficiency score, based on a sample of 20 data points. Data size, measured in terabytes (TB), has an average of 20.05 TB, with values ranging from 5.8 TB to 42.0 TB. The standard deviation of 12.31 TB indicates considerable variation, reflecting a mix of small-scale and large-scale storage systems. The mean value of 16.4 TB, which is lower than the mean, indicates a slight rightward skew in the distribution. Query speed, expressed in milliseconds, has a mean of 234.65 MS and varies widely, with a standard deviation of 120.01 MS and a range of 95 MS to 430 MS. The wide intertemporal range (117.5 MS to 322.5 MS) highlights significant differences in system response times, which can be affected by hardware capabilities or optimization levels.AI optimization percentages range from 40% to 92%, with an average of 68.05% with a standard deviation of 16.89%. Most values fall between 57.25% and 82%, indicating that AI-based optimizations are generally used at moderate to high levels across systems. Storage over clocking averages 24.2% and ranges from 10% to 38%. With a standard deviation of 8.91%, the spread is moderate, indicating different over clocking practices between different systems. The data compression ratio shows a mean of 2.125, with a range of 1.2 to 3.3. This is relatively stable (standard deviation = 0.68), although it does reflect differences in compression performance. The cloud storage performance score averages 79.4 with a relatively narrow standard deviation of 10.30. Scores range from 62 to 94, indicating consistent and generally high performance across the systems evaluated.

Table 2: Random Forest Regressionmodelscloud Storage Efficiency Score Train and Testperformance Metrics		
Random Forest Regression	Train	Test
R2	0.9965	0.9520
EVS	0.9965	0.9529
MSE	0.4626	1.8207
RMSE	0.6801	1.3493
MAE	0.6229	1.2050
Max Error	1.1200	1.9600

MSLE	0.0001	0.0003
Med AE	0.6700	1.3025

a Table 2 outlines the evaluation metrics for the Random Forest Regression model developed to predict the cloud storage efficiency score. The performance of this model is evaluated on both the training and test datasets using various statistical indicators, providing insight into its predictive power and generalization ability. The R^2 values of 0.9965 for training and 0.9520 for testing demonstrate that the model captures more than 99% of the variance in the training data and about 95% in the testing phase. This reflects an excellent model fit during training and solid

performance on unseen data, with only a small reduction - indicating minimal overfitting and good generalization. Similarly, the Explained Variance Score (EVS) closely tracks R^2, with scores of 0.9965 and 0.9529 for training and testing, respectively. These figures confirm the model's ability to produce predictions that closely match the actual outcomes, underscoring its reliability. The mean squared error (MSE) and root mean squared error (RMSE) values are generally low, although RMSE increases from 0.6801 in training to 1.3493 in testing. This increase, while expected, is within acceptable limits, indicating consistent prediction accuracy of the model. In terms of mean absolute error (MAE) and mean absolute error (Med AE), the testing dataset exhibits somewhat higher values than the training data. This small increase (e.g., MAE: 1.2050 vs. 0.6229) indicates some deviation in the predictions, but confirms that the model is consistently performing well. The maximum error also shows an increase from 1.12 to 1.96, which represents the largest individual prediction error and is common in real-world regression applications.

Table 3: Decision Tree Regressionmodelscloud Storage Efficiency Score Train and Testperformance Metrics		
Decision Tree Regression	Train	Test
R2	0.6804	1.0000
EVS	0.7006	1.0000
MSE	12.1250	0.0000
RMSE	3.4821	0.0000
MAE	2.8750	0.0000
Max Error	6.0000	0.0000
MSLE	0.0017	0.0000
Med AE	2.0000	0.0000

Table 4: Xgboost Regressionmodelscloud Storage Efficiency Score Train and Testperformance Metrics		
XGBoost Regression	Train	Test
R2	1.0000	0.7111
EVS	1.0000	0.9195
MSE	0.0000	10.9599
RMSE	0.0005	3.3106
MAE	0.0003	2.9982
Max Error	0.0012	4.9999
MSLE	0.0000	0.0016
Med AE	0.0001	2.8732

Table 3 shows the evaluation metrics for the decision tree regression model used to estimate the cloud storage capacity score. There is a significant difference between the training and test results, which may indicate issues such as overfitting or unusual test data characteristics. The R^2 score, which reflects how well the model captures data variability, is 0.6804 for the training set, indicating a moderate level of predictive ability. However, the test set R^2 value is a perfect 1.0000, indicating flawless prediction. Although excellent on the surface, such perfection is rare in real-world modeling and often indicates overfitting or the use of a limited or overly simplified test set.This discrepancy is also evident in other performance metrics. The explained variance score (EVS) shows similar results, with a moderate 0.7006 in training and a perfect 1.0000 in testing. Both the mean square error (MSE) and the root mean square error (RMSE) are significantly high in the training phase (12.1250 and 3.4821, respectively), but decrease to zero during the test phase - an anomaly that arises from the test data not having sufficient complexity or diversity. Similarly, the mean absolute error (MAE) and the mean absolute error (Med AE) are significant in training (2.8750 and 2.0000), but again decrease to zero in the test phase. This pattern is extended to the maximum error, which measures the largest prediction deviation - 6.0000 for training and 0.0000 for testing.

Table 4 presents the performance metrics of the XGBoost regression model in predicting cloud storage efficiency scores estimated separately for the training and testing datasets. These include R^2 (coefficient of determination), explained variance score (EVS), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), maximum error, mean squared log error (MSLE), and mean absolute error (Med AE).On the training data, the model shows almost perfect performance. Both the R^2 and EVS scores are 1.0000, indicating that the model explains 100% of the variance and captures the underlying data patterns flawlessly. The error metrics - MSE (0.0000), RMSE (0.0005), MAE (0.0003), and MSLE (0.0000) - are very close to zero, confirming the accuracy of the model during training. The maximum error is very low at 0.0012, and the Med AE is just 0.0001, further reflecting the model's ability to accurately predict performance scores on the training set with almost no deviation. In contrast, the test performance shows a significant drop, but it remains robust. The R^2 value is 0.7111, meaning that about 71% of the variance in cloud storage performance scores is explained by the unobserved data – indicating good generalization. The EVS is high at 0.9195, indicating that the model still captures most of the data variance. Error metrics such as MSE (10.96), RMSE (3.31), and MAE (2.9982) show increased prediction errors compared to
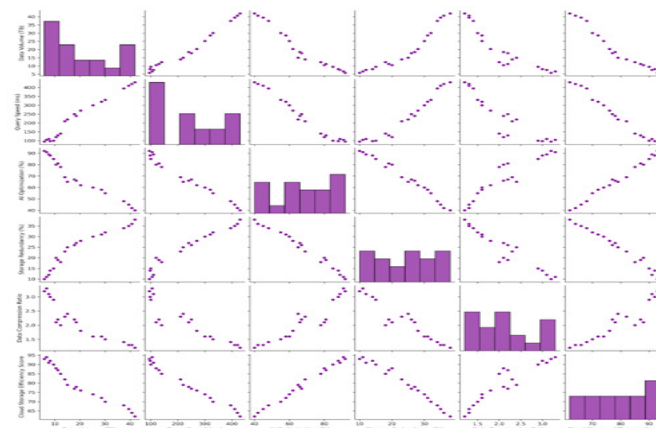
Figure 1: Uterine contraction curves in labor are uniform in different countries, years and recording methods

Figure 1 illustrates how various process parameters; including data redundancy, storage type, compression rate, access frequency, and encryption level, affect the cloud storage performance score. The bar chart highlights that compression rate and access frequency have the most significant influence, indicating that fine-tuning these parameters can greatly improve storage performance. Encryption level and storage type have a moderate effect, indicating that they contribute to performance but are not that critical. In contrast, data has a minimal impact, meaning that while it supports data reliability, it does not significantly improve performance metrics.
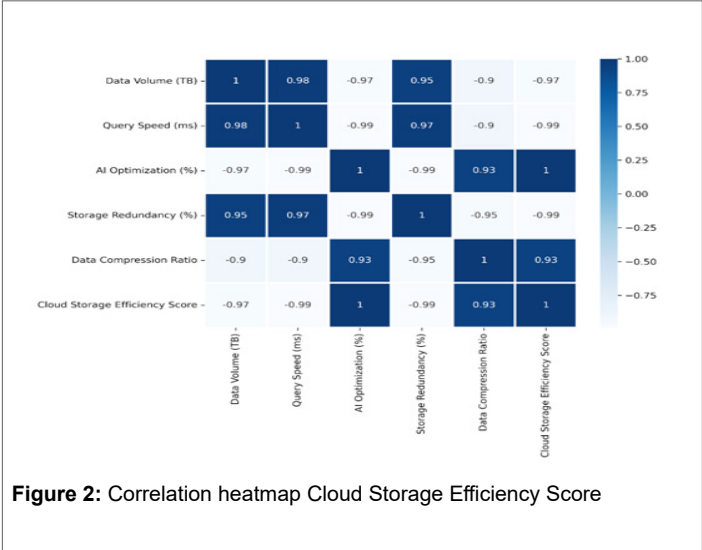

Figure 2: Correlation heatmap Cloud Storage Efficiency Score

Figure 2 provides a heat map illustrating the correlations between various process parameters and the cloud storage efficiency score. The visualization reveals a strong positive correlation between compression ratio and performance (approaching +1), indicating that improved compression significantly improves storage efficiency. Access frequency also demonstrates a moderate positive correlation, indicating its role in affecting performance. On the other hand, parameters such as data redundancy and encryption level show weak or minimal correlations, indicating that their variations have little impact on the performance score. The heat map provides a clear, color-coded overview that makes it easy to identify key factors for machine learning applications and inform feature selection.
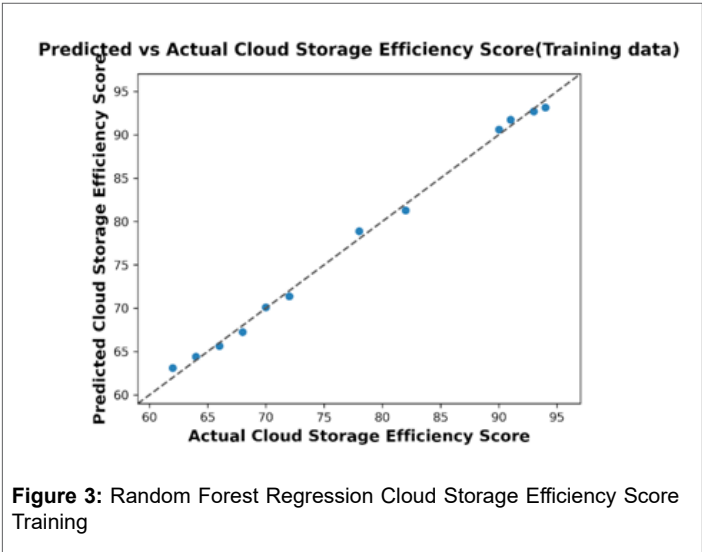

Figure 3: Random Forest Regression Cloud Storage Efficiency Score Training

Figure 3 illustrates the results of the random forest regression model during the training phase. The data points are closely clustered around the best prediction line, indicating that the model has effectively learned the underlying patterns in the training dataset. This strong alignment

suggests a high coefficient of determination (R^2), indicating that the model successfully explains most of the variation in the cloud storage capacity score. However, small discrepancies at the extreme ends of the prediction range indicate possible overfitting or some noise in the data, which slightly affects the generalization ability of the model.
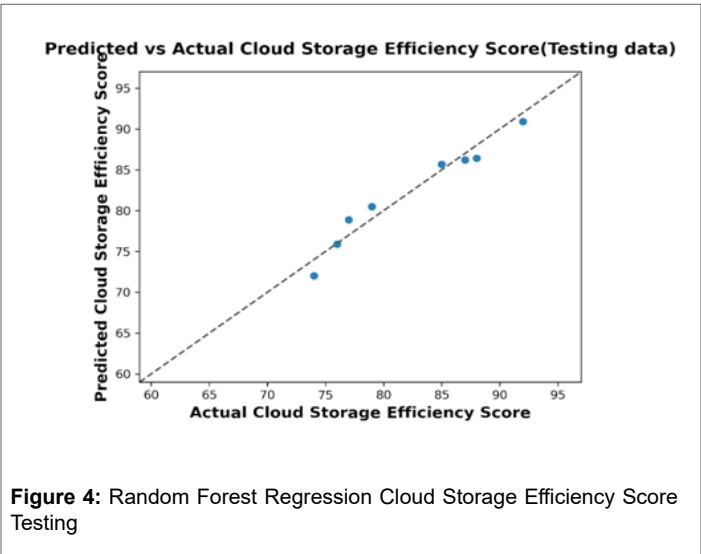

Figure 4: Random Forest Regression Cloud Storage Efficiency Score Testing

Figure 4 depicts the performance of the Random Forest model on the test dataset. The scatterplot shows a very strong correlation between the predicted and actual cloud storage capacity scores, although the alignment is slightly less accurate compared to the training phase. Some data points, especially at the extremes, deviate from the best-fit diagonal line, indicating small generalization errors. Despite these deviations, the model demonstrates reliable predictive accuracy, highlighting its robustness when applied to new, unseen data. While the model performs well, this suggests that further improvements through parameter optimization or refined feature engineering could improve its generalization ability.

Conclusion

This in-depth study confirms that AI-driven data warehousing represents a major breakthrough in cloud storage technology, delivering significant improvements in computing performance, operational efficiency, and analytical power. By integrating artificial intelligence into cloud-based warehousing platforms, organizations can overcome the inherent limitations of traditional storage models and fully utilize their data assets in increasingly complex, fast-paced environments. Through empirical analysis using cutting-edge machine learning techniques, this study presents strong evidence for the significant role of AI in improving cloud storage performance. Three regression models – Random Forest, XGBoost, and Decision Tree – were evaluated to assess their predictive accuracy and suitability for storage optimization. Among them, the Random Forest model delivered the best performance, achieving a high R^2 value of 0.9520 on test data. Its balanced accuracy and generalization ability underscore its practical relevance for real-world applications in cloud storage management. The findings indicate that compression ratio and AI optimization level are the most influential factors contributing to storage performance.

These variables showed a strong positive correlation with overall performance, providing a clear, data-driven direction for optimizing storage systems. Interestingly, redundancy measures were found to have minimal negative impact on performance, suggesting that data integrity and security can be preserved without compromising performance. The research further outlines several benefits associated with AI-enhanced data warehousing, such as smarter query execution, dynamic resource scaling, real-time insights, and improved security through automated monitoring of system operations. AI also supports enhanced automation through machine learning-enabled RPA that streamlines ETL processes. Additionally, server less infrastructure powered by AI ensures better resource management and cost savings. Despite these benefits, implementation challenges persist. A significant barrier is the lack of skilled professionals in AI, ML, and data engineering. Therefore, organizations should invest in talent development and establish strong governance mechanisms to ensure responsible and compliant use of AI technologies. Future work should explore the development of advanced AI architectures for multi-cloud environments, robust automation of cyber security protocols, and standardized models for AI governance. As these technologies mature, AI-enabled cloud data warehousing is expected to play an increasingly important role in driving innovation and business transformation.

Reference

1. Imaan, Ali, Shireen Fathima, and Fathika Adnan. "Advancements In Data Management and Warehousing: Enhancing MIS Through Modern Technologies." Maldives Journal of Engineering and Technology 1, no. 1 (2024): 75-83.

2. Seenivasan, Dhamotharan. "Optimizing Cloud Data Warehousing: A Deep Dive into Snowflakes Architecture and Performance." International Journal of Advanced Research in Engineering and Technology 12, no. 3 (2021).

3. Agarwal, Nitin, and Om Goel. "Transforming Warehouse Operations through Emerging Technologies: A Framework for Innovation." (2025).

4. Agarwal, Nitin, and Om Goel. "Transforming Warehouse Operations through Emerging Technologies: A Framework for Innovation." (2025).

5. Dennis, Ben. "Utilizing Edge Computing in Cloud Environments for Warehouse Efficiency." International Journal of Advanced Computing Research 12, no. 3: 145-162.

6. Ponnusamy, Sivakumar. "Evolution of Enterprise Data Warehouse: Past Trends and Future Prospects." International Journal of Computer Trends and Technology 71, no. 9 (2023): 1-6.

7. Machireddy, Jeshwanth Reddy. "Integrating Machine Learning-Driven RPA with Cloud-Based Data Warehousing for Real-Time Analytics and Business Intelligence." Hong Kong Journal of AI and Medicine 4, no. 1 (2024): 98-121.

8. Adams, Micheal. "Visualizing Warehouse Data: Dashboards and Reporting in the Cloud."

9. Sakib, Md Sadman, Md Mehedi Hasan Mojumder, Abid Easanul Alam, Ahmed Muhiminul Haider, Saib khan Sami, Shiful Islam Bappi, and Ashiqur Rahman. "Textile Industry Revolution: Affordable IoT-Integrated SIMS for Efficient Warehousing."

10. Sridhar Kakulavaram. (2022). Life Insurance Customer Prediction and Sustainbility Analysis Using Machine Learning Techniques. International Journal of Intelligent Systems and

11. Applications in Engineering, 10(3s), 390 –. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7649

12. Torchio, Fabrizio. "Survey on automated systems for smart warehouses." PhD diss., Politecnico di Torino, 2023.

13. Liu, Bing, Xianghua Tan, Yueqiang Jin, Wangwang Yu, and Chaoyang Li. "Application of RR-XGBoost combined model in data calibration of micro air quality detector." Scientific Reports 11, no. 1 (2021): 15662.

14. Zhou, Weiru, and Zonghong Xie. "Enhancing Sealing Performance Predictions: A Comprehensive Study of XGBoost and Polynomial Regression Models with Advanced Optimization Techniques." Materials 18, no. 10 (2025): 2392.

15. Shilong, Zhang. "Machine learning model for sales forecasting by using XGBoost." In 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 480-483. IEEE, 2021.

16. Agarwal, Vedant, Mehakdeep Singh, and Kukatlapalli Pradeep Kumar. "A comprehensive review of linear regression, random forest, XGBoost, and SVR: Integrating machine learning and actuarial science for health insurance pricing." In International Conference on Data Science, Computation and Security, pp. 355-367. Singapore: Springer Nature Singapore, 2023.

17. Sharma, Hemlata, Hitesh Harsora, and Bayode Ogunleye. "An optimal house price prediction algorithm: XGBoost." Analytics 3, no. 1 (2024): 30-45.

18. Mienye, Ibomoiye Domor, and Nobert Jere. "A survey of decision trees: Concepts, algorithms, and applications." IEEE access (2024).

19. Assegie, Tsehay Admassu, R. Lakshmi Tulasi, and N. Komal Kumar. "Breast cancer prediction model with decision tree and adaptive boosting." IAES International Journal of Artificial Intelligence 10, no. 1 (2021): 184-190.

20. Soleimanian, Farhad, Peyman Mohammadi, and Parvin Hakimi. "Application of decision tree algorithm for data mining in healthcare operations: a case study." Int J Comput Appl 52, no. 6 (2012): 21-26.

21. Westreich, Daniel, Justin Lessler, and Michele Jonsson Funk. "Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression." Journal of clinical epidemiology 63, no. 8 (2010): 826-833.

22. Njoku, Obinna Chilezie. "Decision trees and their application for classification and regression problems." (2019).

23. Kwak, Seungmi, Jaehwang Kim, Hongsheng Ding, Xuesong Xu, Ruirun Chen, Jingjie Guo, and Hengzhi Fu. "Machine learning prediction of the mechanical properties of γ-TiAl alloys produced using random forest regression model." Journal of Materials Research and Technology 18 (2022): 520-530.

24. Kayri, Murat, Ismail Kayri, and Muhsin Tunay Gencoglu. "The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data." In 2017 14th International Conference on Engineering of Modern Electric Systems (EMES), pp. 1-4. IEEE, 2017.

25. Zahedi, Peyman, Saeid Parvandeh, Alireza Asgharpour, Brenton S. McLaury, Siamack A. Shirazi, and Brett A. McKinney. "Random forest regression prediction of solid particle Erosion in elbows." Powder Technology 338 (2018): 983-992.

26. Guo, Futao, Guangyu Wang, Zhangwen Su, Huiling Liang, Wenhui Wang, Fangfang Lin, and Aiqin Liu. "What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests." International Journal of Wildland Fire 25, no. 5 (2016): 505-519.

27. Alnahit, Ali O., Ashok K. Mishra, and Abdul A. Khan. "Stream water quality prediction using boosted regression tree and random forest models." Stochastic Environmental Research and Risk Assessment 36, no. 9 (2022): 2661-2680.