Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Enhancing Cloud Job Failure Prediction with A Novel Multilayer Voting-Based Framework

B Charishma[1], U Siva Teja[2], P Swathi[2], Y Venkata Charu Vikram[2], M Pavaneswar[2]

[1]Assistant professor, Dept. of CSE, Sai Rajeswari Institute of Technology, Kadapa – 516360.

[2]UG Scholar, Dept. of CSE, Sai Rajeswari Institute of Technology, Kadapa – 516360.

Correspondence

B. Charishma

Assistant Professor, Dept. of CSE, Sai Rajeswari Institute of Technology, Kadapa – 516360.

- Received Date: 25 Jan 2026
- Accepted Date: 14 Mar 2026
- Publication Date: 20 Mar 2026

Keywords

Cloud Computing, Job Failure Prediction, Ensemble Learning, Weighted Voting, Random Forest, Cat Boost, Machine Learning, Google Cluster Dataset.

Abstract

Accurate prediction of job failures in modern cloud data centers is essential for ensuring system reliability, optimizing resource utilization, and maintaining uninterrupted service delivery. Cloud jobs frequently fail due to resource insufficiency, misallocation, scheduling anomalies, and system-level faults, highlighting the need for intelligent predictive models. This paper presents a multilayer voting-based predictive framework for cloud job failure detection and failure type classification using the Google Cluster 2019 trace dataset. In the first layer, an ensemble of machine learning and deep learning models Decision Tree, K-Nearest Neighbours, XG-Boost, AdaBoost, and Artificial Neural Networks is employed. These models are optimized using Grid Search-based hyperparameter tuning, and their probabilistic outputs are combined through a weighted voting mechanism to improve prediction accuracy. The second layer applies a Random Forest classifier to identify specific failure types, including Lost, Kill, Finish, Evict, and Fail. Additionally, a Cat Boost classifier is incorporated, demonstrating superior performance due to its efficient gradient boosting and resistance to overfitting. Experimental results show that Cat Boost achieves 100% accuracy, outperforming all other models and providing a reliable solution for predicting cloud job failures.

Introduction

Computational resources by delivering scalable, flexible, and on-demand services over the Internet. As enterprises increasingly migrate applications, workloads, and critical operations to cloud platforms, maintaining uninterrupted service availability has become fundamental to modern IT infrastructure. Cloud providers are expected to ensure consistent performance, high availability, and seamless execution of user jobs, as any disruption directly affects service quality and user experience. However, the rapid expansion and growing complexity of cloud environments have introduced significant challenges in resource management, workload optimization, and system reliability. Job failures in the cloud can arise from various factors, including hardware faults, software defects, scheduling issues, and resource insufficiency. These failures not only degrade service quality but also lead to substantial wastage of computing and storage resources. Real-world incidents, such as large-scale outages experienced by major service providers, demonstrate the critical importance of early failure detection to prevent service disruptions and maintain operational continuity. As the volume and diversity of cloud workloads continue to expand, understanding job execution behaviours and identifying potential failures before they occur has become a priority for ensuring dependable cloud services. This work focuses on enhancing the accuracy and effectiveness of cloud job failure prediction to support more resilient and efficient cloud operations.

Literature Survey

Recent studies have explored various machine learning and deep learning techniques for predicting job failures in cloud data centers to enhance reliability and resource utilization. In, the authors analyze cloud traces and demonstrate a strong correlation between resource allocation and job failures, achieving high prediction accuracy. Similarly, machine learning-based frameworks using datasets such as Google Cluster, Mustang, and Trinity traces have shown improved performance in terms of accuracy, recall, and F1-score. Hybrid approaches combining Long Short-Term Memory (LSTM) and K-Nearest Neighbors (KNN) have also been proposed to capture temporal patterns and improve classification of job outcomes. Furthermore,

Citation: Charishma B, Siva Teja U, Swathi P, Venkata Charu Vikram Y, Pavaneswar M. Enhancing Cloud Job Failure Prediction With A Novel Multilayer Voting-Based Framework. GJEIIR. 2026;6(3):0199.

comparative studies indicate that models such as XG-Boost, Decision Tree, and Random Forest perform effectively in failure prediction, with resource-related features playing a crucial role. In addition, workload failure prediction models in high-performance computing environments have demonstrated significant improvements in reducing resource consumption when integrated with schedulers. Despite these advancements, most existing works rely on single or limited hybrid models, highlighting the need for more robust ensemble-based frameworks for accurate failure prediction and classification.

Proposed System

Cloud computing environments are increasingly challenged by frequent job failures caused by resource insufficiency, inefficient scheduling, dynamic workload variations, and unpredictable system behaviour. These failures lead to degraded system performance, increased latency, and inefficient resource utilization, thereby affecting the reliability and scalability of cloud data centers. With the rapid growth of large-scale and continuous workloads, ensuring stable job execution and efficient resource management has become a critical concern for cloud service providers.

Existing job failure prediction approaches predominantly rely on single machine learning models, which often lack the capability to generalize across diverse and complex failure scenarios. Such limitations result in reduced prediction accuracy and inadequate identification of failure types, making it difficult to implement proactive mitigation strategies. Consequently, organizations depending on cloud infrastructure experience service interruptions, increased operational costs, delayed processing, and reduced user satisfaction. Additionally, frequent failures contribute to resource wastage, lower system throughput, and increased recovery time, further impacting overall system efficiency.

Therefore, there is a need for a robust and intelligent predictive framework that can accurately detect job failures and classify their types in advance. An effective solution should integrate multiple machine learning techniques to enhance prediction performance, improve failure diagnosis, and ensure efficient resource utilization. Addressing these challenges is essential for improving the reliability, scalability, and operational efficiency of modern cloud data centers.

System Analysis

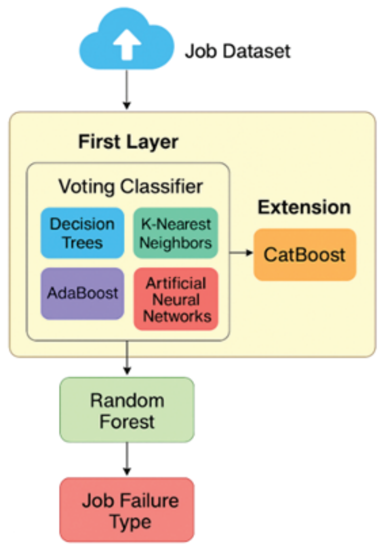
Existing System

Existing cloud job failure prediction techniques primarily utilize traditional machine learning and deep learning models that analyze workload attributes, resource usage, and execution behaviours to identify potential failures. These systems commonly depend on single-model architectures, which limits their ability to capture complex and diverse failure patterns present in large-scale cloud environments. Approaches such as clustering-based analysis with Fuzzy C-Means group similar tasks and use classifiers like Support Vector Machines, Extreme Learning Machines, and Logistic Regression to predict job outcomes. Deep learning models, including Bi-LSTM and LSTM, are applied to analyze sequential job execution patterns and system logs to estimate termination states. Although these methods help detect failures and understand job behaviour, their predictive capability is often constrained by limited feature interaction, insufficient handling of heterogeneous datasets, and an inability to classify multiple types of failures independently. Moreover, single-model dependency reduces robustness when encountering noisy or imbalanced data, making failure prediction less reliable. As cloud environments grow in scale and complexity, traditional approaches struggle to maintain accuracy and adaptability, highlighting the need for more integrated, multi-model, and multilayer prediction systems.

Proposed System

The proposed system introduces a Multilayer Multi-Prediction Voting-Based Framework (MMPF) designed to significantly improve the accuracy and reliability of cloud job failure prediction. Unlike traditional single-classifier approaches, this system incorporates multiple machine learning and deep learning models to leverage their diverse strengths and minimize individual limitations. The architecture is structured into two distinct prediction layers. In the first layer, the system focuses on classifying a job as either successful or failed by integrating several fine-tuned base algorithms, including Decision Tree, K-Nearest Neighbors (KNN), XG-Boost, AdaBoost, and Artificial Neural Networks (ANN). These classifiers are optimized through Grid Search-based parameter tuning to enhance their predictive capability. Their outputs are combined through a weighted soft voting mechanism, which aggregates probability scores from all models to generate a more stable and accurate failure prediction. Once a job is identified as failed, the second layer is activated to determine the exact type of failure. This layer employs a Random Forest model trained on the voting-based predictions to classify failure types such as Lost, Kill, Evict, Finish, and Fail. Overall, the proposed framework strengthens cloud reliability by providing a robust, multi-stage prediction system capable of detecting job failures and understanding their underlying causes.

System Design



Data Flow Diagram

The data flow of the proposed system represents the structured movement of job-related information from input to final prediction output. Initially, the user provides job-specific data such as resource requirements, scheduling parameters, and execution details through the system interface. This input data is -

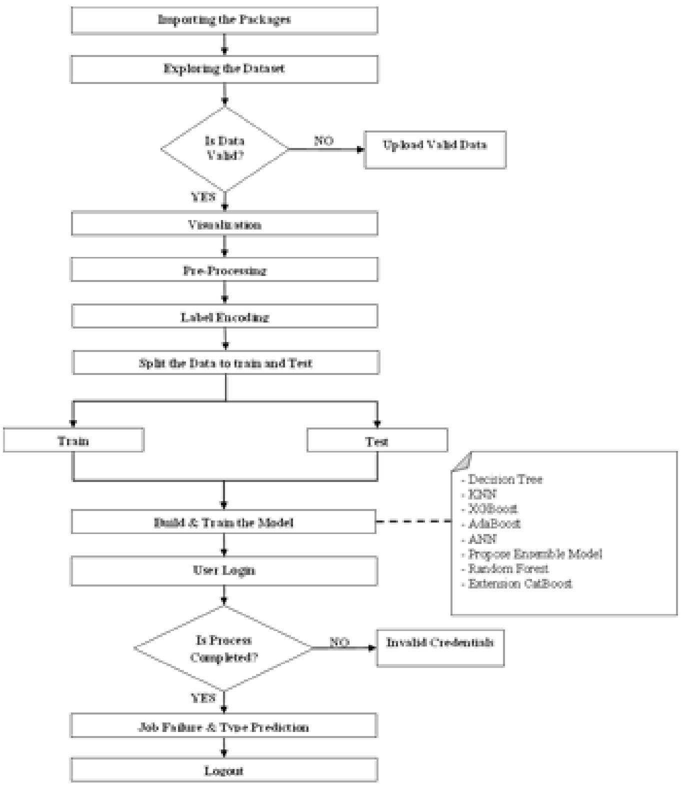
first validated and pre-processed to remove inconsistencies, handle missing values, and normalize features for effective model processing.

Once pre-processed, the data is forwarded to the first layer of the prediction framework, where multiple machine learning and deep learning models including Decision Tree, K-Nearest Neighbors, XGBoost, AdaBoost, and Artificial Neural Networks analyze the input. Each model generates prediction probabilities, which are then combined using a weighted voting mechanism to determine the likelihood of job failure.

The output from the first layer is subsequently passed to the second layer, where a Random Forest classifier identifies the specific type of job outcome, such as Lost, Kill, Finish, Evict, or Fail. In parallel, the CatBoost model processes the same input data to provide an optimized and highly accurate prediction, enhancing overall system reliability.

Finally, the system generates the prediction results, including failure status and failure type, which are presented to the user through the interface. This structured flow ensures efficient data transformation, accurate prediction, and seamless interaction between system components, thereby improving decision-making and resource management in cloud environments.

The Data Flow Diagram (DFD) illustrates the step-by-step flow of data within the proposed cloud job failure prediction system. The process begins with importing the required packages and exploring the dataset to understand its structure and features. The system then verifies whether the input data is valid; if invalid, users are prompted to upload correct data before proceeding



After model development, the system incorporates a user authentication module. Upon successful login, the system checks whether the process is completed; invalid credentials prevent further access. Finally, the trained model performs job failure and failure-type prediction, and the results are presented to the user. The process concludes with logout, ensuring secure system usage.

Results

Algorithm Name	Accuracy	Precision	Recall	F-Score
Decision Tree	99.687	99.218	99.805	99.508
K-Nearest Neighbors (KNN)	99.081	98.437	98.688	98.562
XGBoost	99.290	99.561	98.199	98.863
AdaBoost	87.965	91.723	70.091	75.040
Artificial Neural Network (ANN)	99.415	99.638	98.517	99.066
Proposed Ensemble Model	99.958	99.994	99.974	99.934
Random Forest	99.979	99.984	99.979	99.981
Extension: CatBoost	100.000	100.000	100.000	100.000

Once validated, the data undergoes visualization and preprocessing stages, including cleaning and transformation. Label encoding is applied to convert categorical values into numerical form, followed by splitting the dataset into training and testing subsets. The training phase involves building and training multiple models such as Decision Tree, KNN, XGBoost, AdaBoost, Artificial Neural Networks, and the proposed ensemble model, along with Random Forest and CatBoost for enhanced performance.

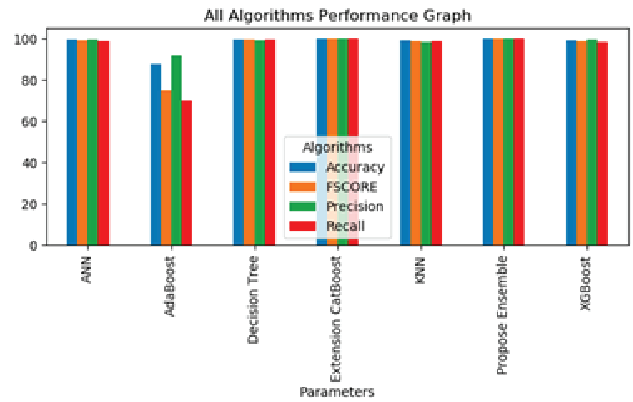
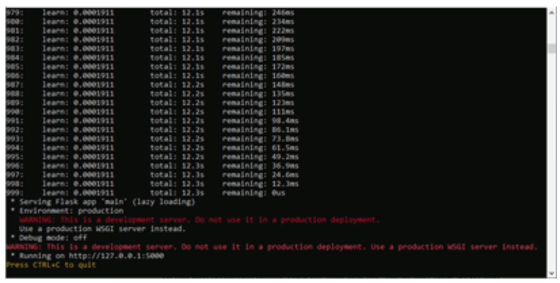


Fig – 3: Comparison Graph

To run web prediction double click on 'runFlask.bat' file to start flask server and then will get below page.

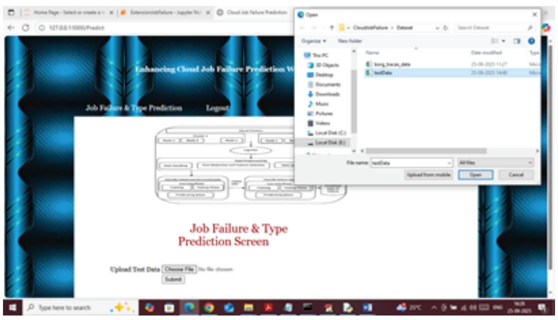


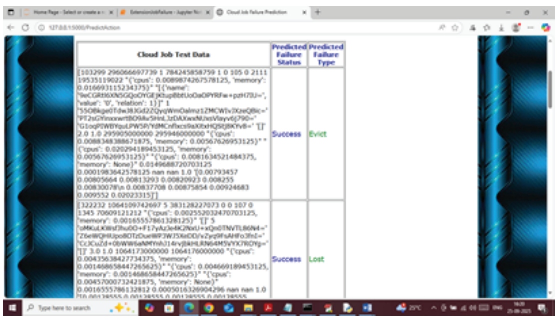
and then press enter key to get below page.

In above screen click on 'User Login' link to get below page

In above screen user is login by entering username and password as 'admin and admin'. After login will get below page

In above screen click on 'Job Failure & Type Prediction' link to get below page





In above screen in first column can see test data values and in second column can see Predicted Job status as 'Success or failed' and in second column can see predicted type as 'Evict, Kill etc.

Conclusion

This study presents a robust multilayer voting-based predictive framework for accurate cloud job failure detection and failure-type classification using the Google Cluster 2019 dataset. By integrating multiple machine learning and deep learning models, the proposed system effectively captures complex patterns in cloud workloads and improves prediction reliability. The weighted voting mechanism enhances overall classification performance by reducing the impact of noise and inconsistencies, while the second-layer Random Forest model enables precise identification of failure types, providing deeper operational insights.

Furthermore, the incorporation of CatBoost demonstrates superior predictive capability, achieving the highest accuracy among all evaluated models. This highlights the effectiveness of advanced boosting techniques in handling large-scale and complex cloud data. The proposed framework not only improves prediction accuracy but also supports better resource management, reduces system failures, and enhances service continuity.

Overall, the developed model offers a scalable, efficient, and highly reliable solution for cloud job failure prediction, making it suitable for real-time deployment in modern cloud data centers. It contributes significantly toward improving system performance, optimizing resource utilization, and strengthening the resilience of cloud computing environments.

References

1. Jassas, M. S., & Mahmoud, Q. H. (2022). Analysis of job failure and prediction model for cloud computing using machine learning. Sensors, 22(5), 2035.
2. Bommala, H., Aluvalu, R., & Mudrakola, S. (2023). Machine learning job failure analysis and prediction model for the cloud environment. High-Confidence Computing, 3(4), 100165.
3. Golani, B., Datta, J., & Singh, G. (2021). Prediction of cloud server job failures using machine learning based KNN classification and LSTM modelling methods. Int. J. Eng. Res. Technol.(IJERT), 10(05).
4. Tengku Asmawi, T. N., Ismail, A., & Shen, J. (2022). Cloud failure prediction based on traditional machine learning and deep learning. Journal of Cloud Computing, 11(1), 47.
5. Li, J., Wang, R., Ali, G., Dang, T., Sill, A., & Chen, Y. (2023, July). Workload failure prediction for data centers. In 2023 IEEE 16th International Conference on Cloud Computing

(CLOUD) (pp. 479-485). IEEE.

6. I. K. Kim, W. Wang, Y. Qi, and M. Humphrey, ‘‘Forecasting cloud appli cation workloads with CloudInsight for predictive resource management,’’ IEEE Trans. Cloud Comput., vol. 10, no. 3, pp. 1848–1863, Jul. 2022, doi: 10.1109/TCC.2020.2998017.
7. D. Saxena, I. Gupta, A. K. Singh, and C.-N. Lee, ‘‘A fault tolerant elastic resource management framework toward high availability of cloud services,’’ IEEE Trans. Netw. Service Manage., vol. 19, no. 3, pp. 3048–3061, Sep. 2022.
8. F. H. Bappy, T. Islam, T. S. Zaman, R. Hasan, and C. Caicedo, ‘‘A deepdive into the Google cluster workload traces: Analyzing the application failure characteristics and user behaviors,’’ in Proc. 10th Int. Conf. Future Internet Things Cloud (FiCloud), Aug. 2023, pp. 103–108.
9. B. K. Ray, A. Saha, S. Khatua, and S. Roy, ‘‘Proactive fault-tolerance technique to enhance reliability of cloud service in cloud federation envi ronment,’’ IEEE Trans. Cloud Comput., vol. 10, no. 2, pp. 1–12,Apr. 2022, doi: 10.1109/TCC.2020.2968522.
10. X. Chen, L. Yang, Z. Chen, G. Min, X. Zheng, and C. Rong, ‘‘Resource allocation with workload-time windows for cloud-based software services: A deep reinforcement learning approach,’’ IEEE Trans. Cloud Comput., vol. 11, no. 2, pp. 1–15, Apr. 2023, doi: 10.1109/TCC.2022.3169157.
11. A. Priyadarshini, S. K. Pradhan, S. R. Laha, and D. S. K. Nayak, ‘‘Enhancing fault tolerance and load balancing in cloud computing for improved e-healthcare systems performance,’’ in Proc. 2nd Int. Conf. Ambient Intell. Health Care (ICAIHC), Nov. 2023, pp. 1–6, doi: 10.1109/icaihc59020.2023.10431436.
12. S. Chouliaras and S. Sotiriadis, ‘‘An adaptive auto-scaling framework for cloud resource provisioning,’’ Future Gener. Comput. Syst., vol. 148, pp. 173–183, Nov. 2023, doi: 10.1016/j.future.2023.05.017.
13. K. Vani and S. Sujatha, ‘‘A machine learning framework for job failure prediction in cloud using hyper parameter tuned MLP,’’ in Proc. 2nd Int. Conf. Adv. Technol. Intell. Control, Environ., Comput. Commun. Eng. (ICATIECE), Dec. 2022, pp. 1–6, doi: 10.1109/ICATIECE56365.2022.10047809.
14. M. M. A. Moni, M. Niloy, M. F.-U.-A. Juboraj, and A. Banik, ‘‘AI based cloud failure detection and prevention algorithm,’’ in Proc. Innov. Power Adv. Comput. Technol. (i-PACT), Kuala Lumpur, Malaysia, Dec. 2023, pp. 1–5, doi: 10.1109/i-pact58649.2023.10434295.
15. K. Vani and S. Sujatha, ‘‘Enhancing fault tolerance system in cloud environment through failure prediction using ensemble learning,’’ in Proc. 2nd Int. Conf. Intell. Data Commun. Technol. Internet Things (IDCIoT), Jan. 2024, pp. 851–858, doi: 10.1109/idciot59759.2024. 10467215.
16. V. Salanke and V. Kowshik, ‘‘Task failure prediction and migration in cloud environment,’’ in Proc. 1st Int. Conf. Commun. Comput. Sci. (InCCCS), May 2024, pp. 1–6, doi: 10.1109/InCCCS60947.2024.10593580.
17. S. Jomah and S. Aji, ‘‘Google cloud trace: Characterization of terminated jobs,’’ in Proc. 2nd Int. Conf. Advancement Comput. Comput. Technol. (InCACCT), May 2024, pp.517–522, doi: 10.1109/incacct61598.2024.10551146. 140612
18. A.-I. Tuns and A. Spataru, ‘‘Cloud service failure prediction on Google’s Borg cluster traces using traditional machine learning,’’ in Proc. 25th Int. Symp. Symbolic Numeric Algorithms Sci. Comput. (SYNASC), Sep. 2023, pp. 162–169, doi: 10.1109/synasc61333.2023.00029.