ECONSPEAK: A Journal of Advances in Management IT & Social Sciences,
Vol. 13, Issue 2, February 2023 Impact Factor 8.049 ISSN: (2231-4571)
www.skirec.org Email Id: skirec.org@gmail.com
RESEARCHERID

# A STUDY ON MULTI-OBJECTIVE FRAMEWORK FOR CLOUD COMPUTING WORKLOAD PREDICTION

**Parag Digambarrao Thakare**

**(Computer Science Engineering)**

**Dr. Rajeev Yadav**

**(Professor)**

**Glocal School of Technology and Computer Science**

## ABSTRACT

The elasticity of cloud computing and the fact that it does not require an initial expenditure of funds have both contributed to its popularity among enterprises in recent years. The benefits and risks of cloud computing, as well as the question of how to transition legacy programs to the cloud, have been debated extensively. Cloud computing has also been the subject of substantial discussion. The purpose of this research was to propose a paradigm for the optimal allocation of resources in cloud computing that is based on multi-objective workload prediction. By forecasting the amount of work that will be done in the future and allocating resources accordingly, the framework that has been developed intends to reduce the overall cost of resources, as well as the amount of time and power that will be consumed. A number of different machine learning strategies have been utilized in order to make projections regarding the future workload of cloud users. Through the utilization of real-world workload data obtained from a public cloud provider, this study compares the proposed framework with other frameworks that are already in existence. In terms of the total cost and the amount of time that is necessary, the results indicate that the suggested framework has an edge over the frameworks that are already in use. According to the findings of this study, the Decision Tree model is the most accurate model.

***KEYWORDS:*** *Cloud Computing; Cloud Environment; Workload Prediction; Multi-objective framework.*

## INTRODUCTION

Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) are all examples of scalable on-demand services that can be provided using cloud computing (Saxena et al., 2021). These services are delivered over the Internet. The Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and/or Software as a Service (SaaS) are all provided by a cloud through its own virtualized resources, which are formed over its underlying physical resources. In most cases, a virtualized resource in the cloud is referred to as a Virtual Machine (VM), which is a collection of files that contain specifications and configurations (Devi and Valli, 2021).

It is possible to obtain a significant number of the advantages that are connected with the cloud model as a consequence of the multiplexing. Studies have shown that servers in many existing data centers are frequently underutilized because they are overprovisioned for peak loads. This is the case because of the fact that they are overloaded with resources. An additional result of the frequent changes in workload demand is the placement of Virtual Machines on Physical Machines in an inefficient way, as well as an uneven distribution of resources within the data center (Mishra& Manjula, 2020). This is a consequence that affects both the physical machines and the virtual machines. The underutilization of physical machines is responsible for the waste of resources and the excessive consumption of electricity, whereas the overutilization of physical machines is responsible for the degradation of performance as well as a decline in the quality of service (Ghetas, 2021). This study addresses an issue that is completely different from the one that has been investigated in the past: what are the most effective methods for a cloud service provider to multiplex its virtual resources onto physical hardware in an efficient manner while incurring the least amount of overhead. The following section focused on the previous literatures that are associated with this topic.

## LITERATURE REVIEW

A list of previous literatures that are relevant to this research is included in table 1.

Table 1: Related works

| AUTHORS AND YEARS | METHODOLOGY | RESULTS AND FINDINGS |
|---|---|---|
| Reddy & Reddy (2023) | The framework uses Universal Unique Identification (UUID) -Blake, Manhattan Distance – Partition Around Algorithm (MD-PAM), Linear Scaling – Crow Search Optimization (LS-CSO), and Anova-Recurrent Neural Network to schedule workflows dynamically in this article. This framework was implemented in three phases. | The results showed that LS-SCO outperformed Crow Search Optimization (CSO), Particle Swarm Optimization (PSO), and Round Robin (RR). |
| Kruekaew&Kimpan (2022) | The multi-objective optimization-scheduling model using the ABC algorithm and Q-learning (MOABCQ) method, a reinforcement learning technique that speeds up the Artificial Bee Colony Algorithm (ABC) algorithm, is a multi-objective task scheduling optimization based on the ABC algorithm that is proposed as an independent task scheduling approach in cloud computing. | According to the testing results, the MOABCQ approach algorithms performed better than the other algorithms in terms of makespan reduction, cost reduction, imbalance reduction, throughput increase, and average resource usage. |
| Konjaang& Xu, (2021) | In order to simultaneously lower execution costs and execution makespan, this study expanded on the earlier work "Cost Optimised Heuristic Algorithm (COHA)" and introduced a novel workflow scheduling technique called Multi-Objective Workflow Optimization Strategy (MOWOS). | When compared to the state-of-the-art work, the suggested algorithm's performance increases greatly in large and extra-large workflow jobs compared to small and medium workflow tasks. It can enable all activities to be completed before the deadline and significantly cut costs by 8%, makespan by 10%, and resource consumption by 53%. |
| Ismayilov &Topcuoglu (2020) | This paper proposed the neural network based non-dominated sorting genetic algorithm (NN-DNSGA-II) algorithm, a prediction-based dynamic multi-objective evolutionary algorithm that combines the non-dominated sorting genetic algorithm (NSGA-II) method with an artificial neural network. | According to metrics used for dynamic multi-objective optimization problem (DMOP)s with unknown true Pareto-optimal front, such as the number of non-dominated solutions, Schott's spacing, and Hypervolume indicator, the NN-DNSGA-II algorithm significantly outperforms the other alternatives in the majority of cases, according to an empirical study based on real-world applications from the Pegasus workflow management system. |

## RESEARCH GAP

Previous studies highlight the problem of dynamically allocating resources in cloud computing to minimize costs and maximize performance. Multiple objectives, such as maximizing throughput, lowering makespan, load balancing, energy efficiency, and cost reduction, manufacture this challenging. Cloud clients have diverse workloads that require different resource capacities, priorities, and pricing mechanisms based on their Service Level Agreements. This work focuses on multi-objective framework for cloud computing workload prediction.

## METHODOLOGY

As can be seen in the picture below, a multi-objective workload prediction framework specifically designed for dynamic resource allocation in cloud computing has been presented with the intention of assisting companies in making more effective use of resources and enhancing the quality of their services. The system takes into account a number of different performance criteria, including reaction time, throughput, and stability, and employs techniques from machine learning in order to forecast additional workloads in the future. After that, it distributes resources in a manner that is appropriate to accomplish the most critical objectives while simultaneously minimizing waste.

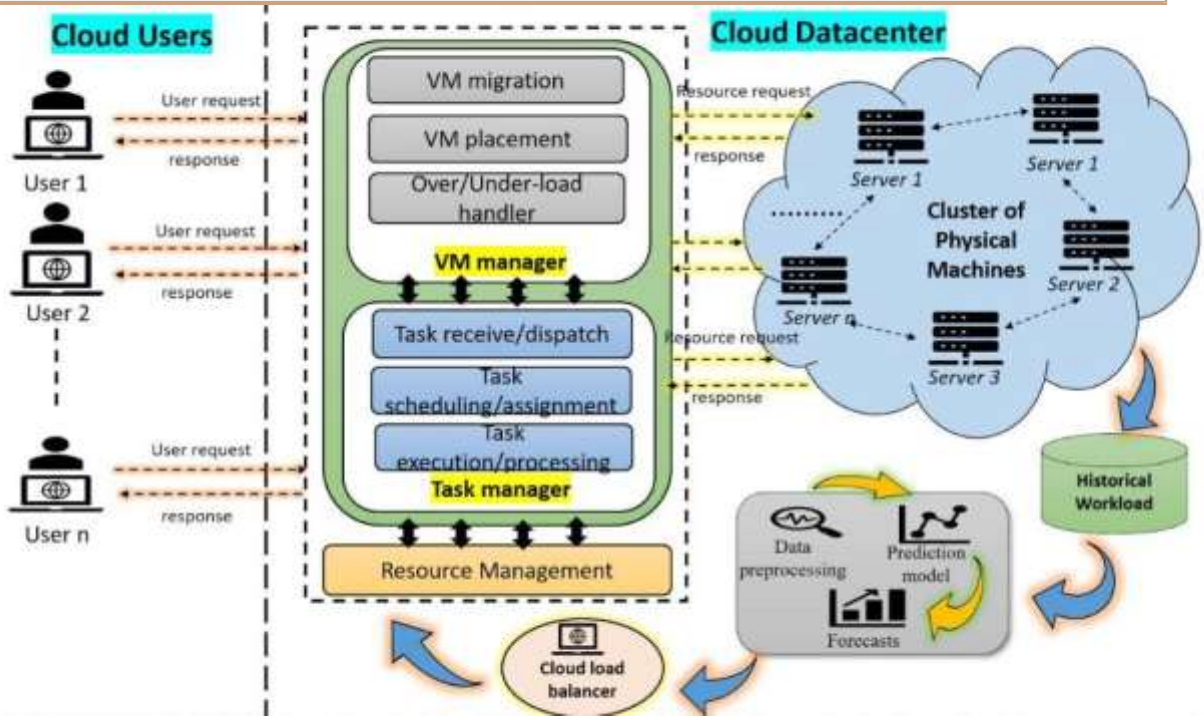


Figure 1: Proposed Multi-Objective Framework for Workload Prediction (Ali, 2022).

In comparison to the solutions that are currently available, the framework that has been proposed is superior since it is able to manage many objectives simultaneously and makes greater use of past data. Furthermore, it is meant to be scalable, which enables it to be utilized in scenarios that include massive cloud infrastructures.

## RESULTS AND DISCUSSIONS

Users need a lot of data because the system allocates resources in an automated manner. But unless users use AWS or Google's data at enormous data centers, it is impossible to obtain a prepared and clean data set for accomplishing such a thing on a modest scale. For this investigation, it has made use of the GitHub data collection. This module is in charge of handling warehouse management, storing raw data, processing it, building machine learning models from the data collected, and distributing the finished models to edges. It is capable of controlling the system.

In the field of machine learning, a number of different classifier algorithms were chosen based on a cross-validation technique. This technique demonstrated that Decision Trees and other relevant models were able to classify data with an accuracy of 83%, which is a sufficient degree of accuracy for the purpose of resource allocation. The data were also evaluated using other methods, such as logistic regression, which had an accuracy of approximately 69%. Additional techniques were also utilized. Because of this, the Decision Tree (DT) classifier has been chosen as the model that maximizes efficiency the most. In order to demonstrate that it will be superior for cloud workload prediction, the following figure provides a graphical depiction of the examination and comparison of several different machine learning techniques.

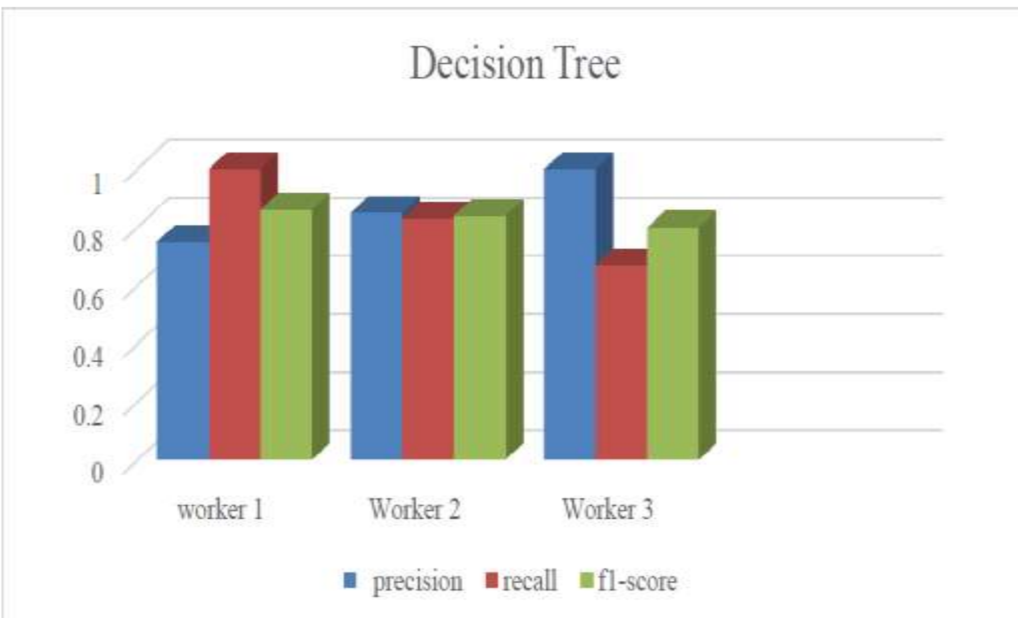


Figure 2: Machine Learning Models Comparison – Decision Tree (DT)

ECONSPEAK: A Journal of Advances in Management IT & Social Sciences,
Vol. 13, Issue 2, February 2023 Impact Factor 8.049 ISSN: (2231-4571)
www.skirec.org Email Id: skirec.org@gmail.com


Figure 3: Machine Learning Models Comparison – Logistic Regression (LR)

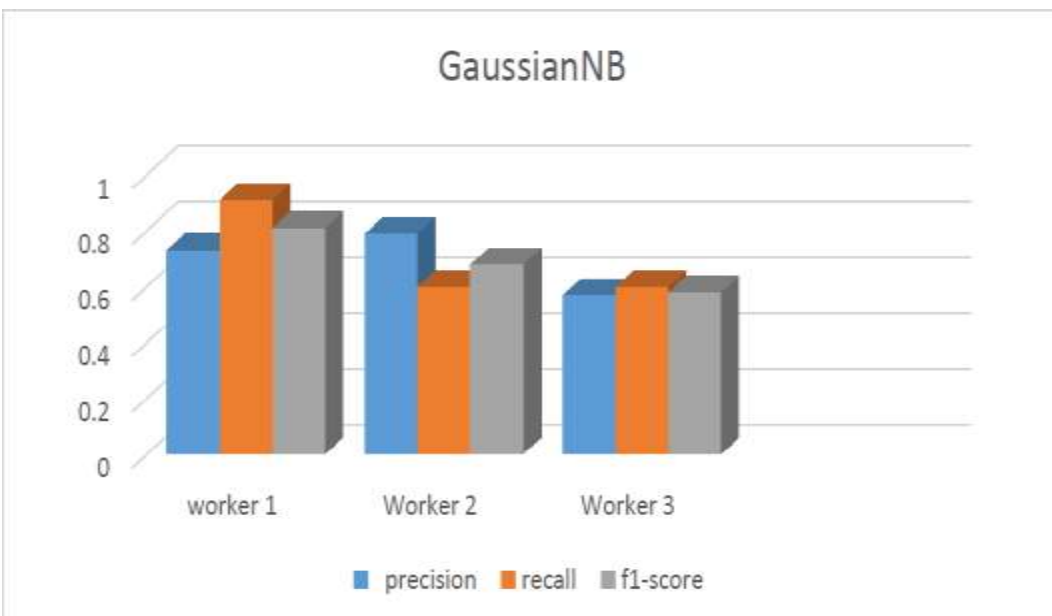


Figure 4: Machine Learning Models Comparison – Gaussian ModelNaive Bayes (Gaussian Model NB)

## CONCLUSION:

In summary, a viable approach to the issue of effective resource utilization in cloud environments is the multi-objective workload prediction framework that has been suggested for dynamic resource allocation in cloud computing. The system employs a range of prediction models to precisely estimate future workloads while taking into account a number of objectives, including cost and reaction time. Because of this, the system may allocate resources dynamically to maximize efficiency and reduce expenses.

Furthermore, various machine learning classification methods such as DT, LR, and GaussianNB are tried on the data and compared based on their training data performance measures. DT was utilized to predict the dataset because it had the highest accuracy of all the classifiers.

## DECLARATION:

I as an author of this paper / article, hereby declare that paper submitted by mefor publication in the journal is completely my own genuine paper. if any issue regarding copyright/ patent/ other real author arises. the publisher will not be legally responsible.ifany of such matters occur publisher may remove my content from the journal website/ updates. I have resubmitted this paper for the publication, for any publication matters or any information intentionally hidden by me or otherwise, I shall be legally responsible (complete declaration of the author at the last page of this paper/article).

## REFERENCES:

1. Devi, K. L., & Valli, S. (2021). Multi-objective heuristics algorithm for dynamic resource scheduling in the cloud computing environment. *The Journal of Supercomputing*, *77*, 8252-8280.
2. Ghetas, M. (2021). A multi-objective Monarch Butterfly Algorithm for virtual machine placement in cloud computing. *Neural Computing and Applications*, *33*, 11011-11025.
3. Ismayilov, G., &Topcuoglu, H. R. (2020). Neural network based multi-objective evolutionary algorithm for dynamic workflow scheduling in cloud computing. *Future Generation computer systems*, *102*, 307-322.
4. Konjaang, J. K., & Xu, L. (2021). Multi-objective workflow optimization strategy (MOWOS) for cloud computing. *Journal of Cloud Computing*, *10*(1), 1-19.

5. Kruekaew, B., &Kimpan, W. (2022). Multi-objective task scheduling optimization for load balancing in cloud computing environment using hybrid artificial bee colony algorithm with reinforcement learning. *IEEE Access*, *10*, 17803-17818.
6. Mishra, S. K., & Manjula, R. (2020). A meta-heuristic based multi objective optimization for load distribution in cloud data center under varying workloads. *Cluster Computing*, *23*, 3079-3093.
7. Reddy, P. V., & Reddy, K. G. (2023). A Multi-objective based Scheduling Framework for Effective Resource Utilization in Cloud Computing. *IEEE Access*.
8. Saxena, D., Singh, A. K., &Buyya, R. (2021). OP-MLB: an online VM prediction-based multi-objective load balancing framework for resource management at cloud data center. *IEEE Transactions on Cloud Computing*, *10*(4), 2804-2816.