



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 2, February 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

Performance Prediction for Cloud Services of Different Attributes using Random Forest Classification

N. Jabeen^{1*}, M.Mubeen^{*2}, R. Hao³, A.Niaz⁴, S. Hussain⁵, M.U.Shoukat⁶

Shanxi Key Laboratory of Micro Nano Sensors & Artificial Intelligence Perception, College of Information and Computer, Taiyuan University of Technology, Taiyuan, China¹

Department of Computer Science, Government College University Faisalabad, Pakistan²

Shanxi Key Laboratory of Micro Nano Sensors & Artificial Intelligence Perception, College of Information and Computer, Taiyuan University of Technology, Taiyuan, China

Shanxi Key Laboratory of Power System Operation and Control, Taiyuan University of Technology, Taiyuan, China⁴

Department of Earth Sciences and Engineering, Taiyuan University of Technology, Taiyuan, China⁵

Hubei Key Laboratory of Advanced Technology for Automotive Components, Wuhan University of Technology, Wuhan, China⁶

*Corresponding author

ABSTRACT: Cloud computing is a type of computer technology that has seen major advancements in recent years. Cloud service providers offer users a variety of services that differ in terms of efficiency, cost, and configuration options. A large number of organizations and enterprises are planning to move their goods to the cloud network shortly. Companies focus on the selection of the most appropriate program that meets their needs. So, it makes for them a critical decision point. According to this research, the selection process is automatic, and it is based on real activity trends from cloud services as well as resource requirements identified in current operating history information. We made every effort to overcome this problem. This model, which is the default performance prediction model, employs a random forest model to anticipate the performance metrics of cloud nodes in response to the numerous options available for configuring their services. A random forest model was investigated to account for changes in task allocation as well as to enhance the accuracy of prediction. An annual database with thousands and hundreds of computer records from a functional country served as the basis for a thorough examination of our concept. A Jupyter simulation was carried out, and the results revealed that the proposed model illustrates how random forest may give accurate results in certain circumstances.

KEYWORDS: Cloud computing; random forest model; cloud services; database; performance.

I. INTRODUCTION

Cloud computing software has emerged as a new and much-recognized revolution in computer technology and the distribution of Internet resources, and it is gaining popularity. Cloud computing is regarded as a software model for establishing pools of services such as software, basic structure, and a speaker service, all of which are required to free up more time for businesses to innovate and create value for their stakeholders [1-2]. Cloud computing is a virtualized repository of services that are made accessible on-demand, ranging from raw computer power to device capabilities, and is becoming increasingly popular. When clients sign up for cloud services, the provider executes their requests utilizing robotic automation rather than hand provisioning. In particular, the ability to deploy abstracted computing, storage, and network capital to workloads on-demand, as well as to access an array of pre-built services, is a significant advantage. Due to financial and environmental factors, energy use has become a major problem for cloud service providers [3-4].

As with actual clouds, the term “cloud” refers to a group of networks in cloud computing, similar to how genuine clouds are made up of water modules. Software as a service (SaaS), platform as a service (PaaS), and infrastructure as a

service (IaaS) are the three kinds of cloud computing services. For many people, Facebook is one of the most well-known examples of cloud computing since it is used regularly. Some examples are cloud-based applications such as YouTube, Dropbox, and Gmail, to name a few. Numerous factors, including its scalability, flexibility, agility, and simplicity, all contribute to its fast development in commercial applications, which is attributed to a variety of factors [5], as shown in Fig. 1.



Fig -1: Illustration of the cloud network.

Amazon Web Services (AWS) was established in 2002 to provide storage and computing services to customers. Major corporations such as Google, Microsoft, HP, and Oracle began offering cloud computing services in 2009 and have continued to expand. Every person in today's society takes use of cloud computing services regularly. Take, for example, the services offered by Google Photos, Google Drive, and Apple iCloud. Cloud computing services will be required in the future by the information technology sector, as well as other requirements [6-7].

Therefore, cloud service providers are looking for new ways to reduce the energy used by their data centers. This research focuses on major resource allocation challenges and suggests other possible solutions to reduce energy use in cloud data centers. Special attention is paid to energy organization technology that uses virtualization technology to save energy. Some experiments based on actual track records from the Google collections are also provided to support our claims in this research.

Paper is organized as follows. Section II describes automatic text detection using morphological operations, connected component analysis and set of selection or rejection criteria. The flow diagram represents the step of the algorithm. After detection of text, how text region is filled using an Inpainting technique that is given in Section III. Section IV presents experimental results showing results of images tested. Finally, Section V presents conclusion.

II. LITERATURE REVIEW

Due to major advances in wireless and mobile technology, applications have become more complex, and demand for more power and resources has increased. The cloud-based network paradigm (CSDWSN) has become a promising solution to these demanding tasks with the help of dynamic cloud servers. This document shows how to obtain the resource and financial allocations among CSDWSN providers for alliance development. A collaboration between CSDWSN providers as a solidarity game proposed in [8] to existing methods, the game model maximizes the revenue of CSDWSN providers with minimal power consumption by analyzing their internal user needs and dynamic pricing strategies. In addition, the proposed game achieves a stable and sustainable solidarity structure, thereby increasing the overall benefits of solidarity and improving equity among participating CSDWSN providers.

Oded Berman [9] constructed a joint decision-making model for site selection and inventory under the same inventory inspection cycle and different inventory inspection cycles. Literature [10] assumed that both distribution centers and distribution points adopt (Q, R) inventory control strategy, the address and inventory capacity of alternative distribution centers are known, the service level and inventory capacity are taken as constraints, and the minimum sum of facility cost, transportation cost and inventory cost is taken as the goal to build a joint location inventory optimization model of distribution network with random demand.

Elasticity is the vital component of distributed computing innovation, which can consequently lessen and add assets to address client's issues. Authors in [11] described cloud services to accomplish versatility, as well as they discover how and when to trigger the flexibility programmed scaling instrument. In this paper, they proposed three models to foresee the responsibility dependent on dissecting checking information. Literatures [12-14] proposed a state-of-the-art asset-based platform, as well as hardware installed with open-source software, and distributed file system to illustrate a series of applications and their performance indicators.

Literature [15] proposes a unique access control paradigm for quick and efficient data access, the cloud service provider (CSP) maintains a temporary database based on the kind of data and the chosen DO value. Cloud service providers may quickly locate data proprietors by form, decreasing data access time substantially. In terms of efficiency, the results of the assessment and analysis of the suggested system's concepts show that they are superior to the existing ones.

III. MATERIALS AND METHODS

A random forest is a machine learning technique that's used to solve regression and classification problems. It utilizes ensemble learning, which is a technique that combines many classifiers to provide solutions to complex problems. A random forest algorithm consists of many decision trees. The 'forest' generated by the random forest algorithm is trained through bagging or bootstrap aggregating. Bagging is an ensemble meta-algorithm that improves the accuracy of machine learning algorithms.

3.1 Proposed Solution

Create python module is used to compose python code. We have used Jupyter to design our research test pipeline. The model has been processed through: 1) Input Values; 2) Features Extraction; 3) Training and Testing Model; 4) Measure the accuracy; 5) Training and Testing; 6) Check Accuracy; and 7) Result/Output.

A mathematical technique of binary separation which may be accomplished in multiclass categories is a systematic system model, such as Adaline and perceptron. The layout of Sickest-learn has been heavily modified to accommodate multiclass segmentation operations. By defining selection requirements, some scholars have conducted studies on 1) picking the right cloud service, while others have concentrated on cost and protection. An overview of the proposed solution has been presented in Fig. 2.

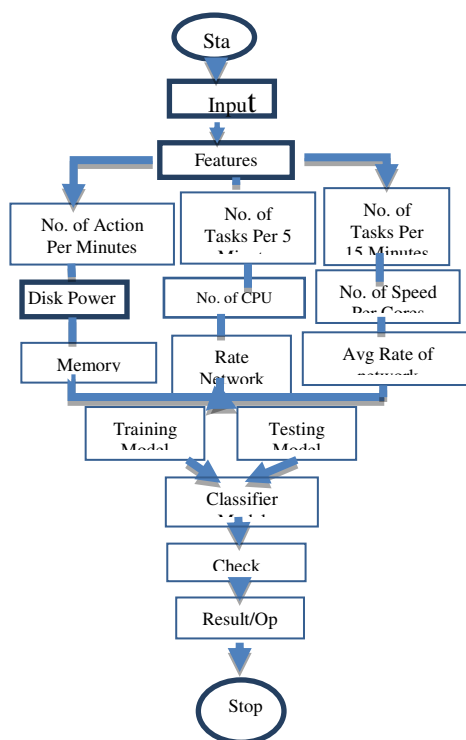


Fig -2: Proposed Solution - An overview

3.2 Random Forest Model

Random forest is a supervised learning algorithm, in which “forest” builds to ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. Put simply: random forest builds multiple decision trees and merges them to get a more accurate and stable prediction. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. Let’s look at the random forest in classification since classification is sometimes considered the building block of machine learning, as shown in Fig. 3.

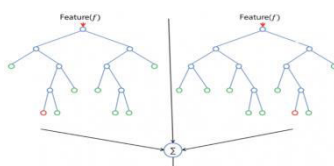


Fig -3: Random Forest Model case study

The random forest has nearly the same hyperparameters as a decision tree or a bagging classifier. Fortunately, there’s no need to combine a decision tree with a bagging classifier because you can easily use the classifier-class of random forest. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Therefore, in a random forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node.

3.3 Data Set Description

Estimates from cloud systems must be organized to analyze both workload expectations and application circumstances. Benson et al. are a group of researchers who have worked on a number. SNMP insights for 10 data centers were gathered, with a bundle resulting from multiple switches in four data centers.

Two web application datasets have been gathered and systematized. Clark Net and NASA have provided datasets. The Clark Net blog was retrieved from a Metro Baltimore, Washington, DC web server. For two weeks, the webserver's HTTP proxy logs were collected. In total, 3,328,587 requests were observed for two weeks. NASA Kennedy Space Center and its Florida-based server make up the second dataset. A total of 3,461,612 requests were seen in this dataset, which spans two months.

In previous discussions of datacenter systems, worldly and spatial inconstancy were differentiated. Although their applications may be comparable to cloud residents, Benson et al. examined link-level SNMP logs from nineteen datacenters. Benson et al. compiled SNMP insights for ten data hubs and packaged follow-up from multiple switches in four of them. Some of these are described as cloud information centers, but it's unclear whether they're IAAS networks

IV. RESULTS & DISCUSSION

Researchers are attempting to tackle these issues by testing and forecasting the required cloud resources. As a result, before moving to the cloud, they evaluate the user’s demands by examining previous service information and operations to generate the necessary resource information and anticipate performance. The advantage of existing program data is obtained in this work by acquiring a database from a real location and then running more trials using the division of Naive Bayes to forecast the system’s function.

4.1 Data Description

In this research, before migrating to the cloud, we evaluate the user’s demands by evaluating previous service information and performance to produce the necessary resource information and anticipate performance. As a result, we’ve developed a default performance model based on an asset layout divider to forecast cloud node performance metrics for different node resource configuration options (workload attributes). Asset classification is a basic

classification based on applying the theorem of random forest with independent theories and may be consumed for classification.

Include a need analysis for selecting cloud suppliers based on existing service processes. Acceptance of asset separation improves forecasting accuracy and performance on a new cloud platform with a new workload pattern. Present test findings and real-world data from the Saudi Ministry of Finance and validate the suggested solution's implementation.

4.2 Analysis of the hypothesis

The hypothesis output is the estimated frequency. When given an input X , this is used to estimate how certain the predicted value is of being the true value. Suppose we get a predicted probability of 0.8 based on the x_1 value. This result shows that there is an 80% chance that an email is spam.

4.3 Data Analysis

Data analysis is methodically describing and demonstrating, compiling and replicating, and evaluating data using mathematical and logical techniques. An accurate and precise interpretation of study results is a crucial component of maintaining data integrity. A scientist or engineer's most typical task is to evaluate data from an external source, either in text or comma format (CSV). Data analysis, such as statistics, trends, or computations, may be performed on imported data in Python to integrate data into meaningful and usable information. The lesson below demonstrates how to import data (including the internet), do basic analysis, streamline findings, and export results to another text file. NumPy and Pandas both give two examples. Below are Python source code script files with example data.

4.3.1 Types of data analysis

In any industry, there are four different forms of data analysis. Everything is related and stacked on top of each other when we split it into categories. As you progress from a simple to a more complicated form of analytics, the complexity and resources required increase. At the same time, the degree of comprehension and value grows, as shown in Fig. 4.



Fig -4: Types of Data analysis

Descriptive analysis is the initial form of data examination. It is the foundation for all data understanding. The most basic and popular data utilization in today's company. The descriptive analysis summarizes previous data, generally in the form of dashboards, in response to "what happened." The most common application of graphic analysis in the workplace is tracking key performance indicators (KPIs). KPIs define how a business operates about a set of standards.

After pondering the basic issue of "what happened," the next stage is to deeper and inquire as to why it occurred. Detailed data is an important part of diagnostic analysis. You may have already gathered some essential information when new difficulties develop. Because you already have the data, the function is repeated, causing all the problems to be linked. Business applications for diagnostic analysis include: The transportation business is looking into the source of a slow shipment in a certain location. A SaaS company is debating which sales efforts contribute to the temptation.

The possible question is addressed through predictability analysis. This form of statistics makes predictions about future outcomes based on prior data. It's also crucial to remember which a forecast is simply a guess. The precision of the forecasts is determined by the quality and accuracy of the data. While descriptive and diagnostic analysis is common in business, forecasting analysis is when most businesses begin to show symptoms of trouble. Some businesses are unable to apply predictive analysis in all of their desired areas. Some departments are not ready to invest in analytical teams across the board, while others aren't ready to train existing units.

The latter type of data analysis is the most in-demand, but only a small number of firms are capable of performing it. Defining analysis is the pinnacle of data analysis, considering all previous results to determine the best course of action for a particular problem or decision. A definition study was carried out with the use of cutting-edge technology and data approaches. The commitment to the firm is considerable, and organizations must ensure that they are prepared to devote the

required time and resources to the endeavor. Artificial Intelligence is a fantastic illustration of arithmetic in action (AI). Even without the use of artificial intelligence, business processes may be performed and improved regularly.

The majority of large data-driven organizations (such as Apple, Facebook, Netflix, and others) already use fixed analytics and artificial intelligence to improve decision-making. Certain organizations may be able to overcome their aversion to acting on predictive data and projections. As technology advances and more people receive in-depth training, we will see an increase in the number of businesses that operate in a data-driven environment.

4.3.2 Provide honest and accurate analysis

At the heart of this issue is the pressing need to reduce the potential of statistical mistakes in the first place. A typical difficulty is the exclusion of outliers, the filling in of missing data, the editing or otherwise altering of data, data mining, and the generation of graphical representations of data, as shown in Fig. 5.

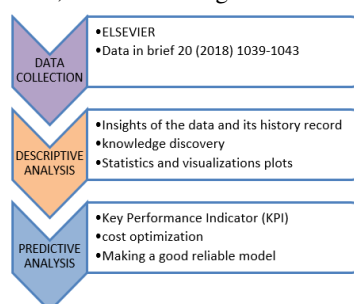


Fig -5: Qualitative Data analysis

In the field of performance measurement, the process of gathering, assessing, and reporting information on the performance of an individual(s), group(s), organization(s), system(s), or component is referred to. A common assumption underlying performance measurement definition is that the performance is being monitored for a specific reason (or reasons).

4.3.3 Performance metric Classification

Three class labels have been identified as having the potential to impact the performance of cloud services. They have been used in the configuration model to improve the performance of cloud services. To avoid issues, we have used the older model of each class in our example to avoid confusion.

4.3.4 Factor of Workload

The workload parameters (conditions) in this example represent all of the elements. That has an impact on the performance of a system. Nine items (conditions) have been discussed that can be easily removed from any software program. Table 1 is a summary of the points raised above. The average number of positions that are currently available (fi). Because a greater number of functions decreases the performance of the service provider, this feature may influence the performance of cloud services, as shown in Fig. 6.

Table 1: Details of the Workload parameters

<i>Feature</i>	<i>Denotation</i>
The number of jobs in minutes	f1
The number of jobs in 5 minutes	f2
The number of jobs in 15 minutes	f3
Memory capacity	f4
Disk capacity	f5
Number of CPU cores	f6
CPU speed per core	f7
Average receives for network bandwidth in kbps	f8
Average transmits for network bandwidth in kbps	f9

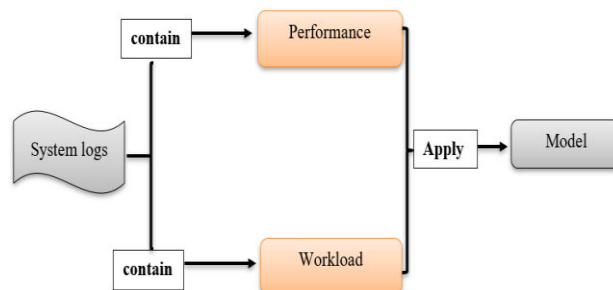


Fig -6: Proposed random forest model

Table 2, utilizing random forest to describe and estimate the relationship between a dependent variable Y, which has only two possible values coming from the existence or absence of an event, and independent factors affecting which phenomena is seen is described and calculated.

Table 2. Depiction of the dataset description

Number of instances	28,147
Number of attributes	9
Number of responses	3
Number of nodes	13
Types of nodes model	Nodes 1 and 5 are HP RP 4440, Nodes 2-4 and 6 are HP RP 7420, Nodes 7-13 are HP DL 380 G5

4.3.5 Evaluation measures

In machine learning, the confusion matrix is used to evaluate the separation model's performance. This is demonstrated in the confusion matrix for a binary section problem, which is a 2-by-2 square with columns for division predictions and lines for real-world values of the class label. If there is an unlimited amount of data, minority class recognition is considered positive, whereas class labels are considered negative.

Following are the meanings of the characters of the confusion: the TP, FN, FP, and True Negative characters mean:

True Positive stands for " True ", and is the number of positive predictions that have been recorded favorably.

False Negative = means " False ", the number of positive things that are incorrectly classified as negative, False

Positive = means " True False ", the number of negative predictions that are incorrectly classified as positive, False

Negative = means " False ", the number of positive things that are incorrectly classified as negative, False Positive =

means True False, the number of positive predictions that are incorrectly classified

True Negative is an abbreviation for non-True, and it refers to the number of negative predictions that are correctly forecasted as negative.

The following are the measures that were taken in our research:

Sensitivity (also known as the accuracy of the positive rating or memory level in other studies): Measures the real benefit rate accurately characterized as such (e.g., the percentage of patients who are correctly identified as TP / (TP + FN) in a certain study).

Direct (also known as incorrect measurement accuracy): Measuring the average of the real negatively characterized objections (for example, the proportion of instances determined to be reasonably healthy, also known as the actual negative rating), which is defined as $TN / (TN + FP)$.

Accuracy is defined as the product of the number of TP and TN divided by the product of the number of TP, FN, TN, and FP. This may be expressed as $(TP + TN) / (TP + FN + TN + FP)$, for example.

In the confusion matrix, each line represented several circumstances that were similar for the phase to which it was matching. It was hidden under the chaos, and the matrix discovered it. To describe the matrix of confusion, we make use of the python module known as learning, which was previously defined. Figure 8 depicts the results of the confusion - matrix for each of the four classes, as shown in Fig. 7.

[39	9	13	1	4	0	8]
[10	84	12	1	8	0	6]
[6	0	157	0	22	0	21]
[0	1	1	11	2	0	1]
[4	6	35	4	1275	1	79]
[0	4	1	0	2	25	5]
[0	0	10	0	19	0	116]

Fig -7: Confusion Matrix result

4.3.6 Scatter plot with histograms

Display the marginal distributions of the scatterplot as histograms on the plot's sides. Two options for achieving a nice alignment of the primary axes with the marginals are described in the next section. To determine where the axes have located, a grid specification is utilized. The Scatter Histogram (Locatable Axes) example demonstrates another method for producing a comparable image by utilizing the axes grid1 toolkit instead of the scatter histogram. Start by constructing a function that accepts x and y input as well as three axes: the primary axis of the scatter plot and two peripheral axes. This function has been used to plot the data. Following that, the scatter plots and histograms have been created within the specified axes.

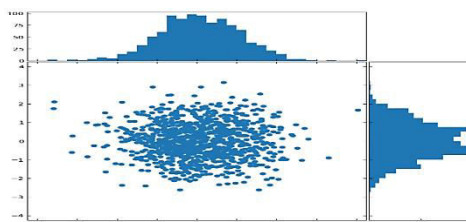


Fig -8: Scatter plot with histogram

4.3.7 Correlation checking: Heatmap

When a correlation heatmap displays a two-dimensional correlation matrix between two discrete dimensions, colored pixels indicate data on a monochromatic scale, the heatmap is called a correlation heatmap. The values of the first dimension are represented by the rows of the table, while the values of the second dimension are represented by the columns of the table. The color of the cell is determined by the number of measurements that are identical to the dimensional value. As a result of these measurements, correlation heatmaps are beneficial for data analysis since they reveal variances and variations in the same data while also making patterns readable and visible. Similar to how it works in a conventional heatmap, the use of color bars in a correlation heatmap improves the reading and comprehension of data. Methods for creating a correlation heatmap are demonstrated in the following sections: 1) To begin, import all of the necessary modules; 2) Open the file that contains your data and import it; 3) Create a heatmap of your website; and 4) Visualize it with the help of Matplotlib.

The default colormap does not work well with correlation heatmaps since it is too bright. A diverging color palette with dramatically distinct colors at the two extremities of the value range and a pale, practically colorless center performs significantly better than a constant color palette. Consider the following additional element to be included in this statement: Find out how to save a heatmap as an image file that has all of the x- and y-coordinates. readily discernible; readily discernible; readily discernible (stick labels).

Table 3. A correlation matrix with nine variables

Jobs_per_Minute	Jobs_per_5Minutes	Jobs_per_15Minutes	Mem_capacity	Disk_capacity_GB	Num_of_CPU_Cores	CPU_speed_per_Core	Avg_Receive_Kbps	Avg_Transmit_Kbps	Class_Name
17.000	16.636	16.455	4.0	133.089	16	1	61.002	22.436	High
16.585	16.589	16.502	4.0	133.089	16	1	33.959	221.007	High
11.463	11.541	11.604	4.0	133.089	16	1	20.169	22.410	High
16.273	17.091	17.162	4.0	133.089	16	1	61.002	22.436	High
6.162	5.500	3.577	4.0	133.089	16	1	61.002	22.436	High
...
6.162	5.500	3.577	4.0	133.089	16	1	10.708	40.209	High
3.091	3.162	3.162	4.0	133.089	16	1	10.708	40.209	High
20.500	20.500	21.000	4.0	133.089	16	1	61.002	22.436	High
6.162	5.500	3.577	4.0	133.089	16	1	61.002	22.436	High
17.810	16.091	17.909	4.0	133.089	16	1	10.708	40.209	High

rus = 10 columns

The default colormap does not work well with correlation heatmaps since it is too bright. A diverging color palette with dramatically distinct colors at the two extremities of the value range and a pale, practically colorless center performs significantly better than a constant color palette. Consider the following additional element to be included in this statement: Find out how to save a heatmap as an image file that has all of the x- and y-coordinates. readily discernible; readily discernible; readily discernible; readily discernible (stick labels).

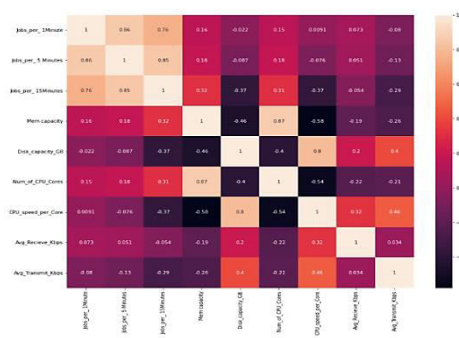


Fig -9: Basic seaborn Heatmap

4.3.8 Count plot in Python

The count plot depicts the number of times an observation appears in the categorical variable (counts). The visual representation is based on the notion of a bar chart.

4.3.8.1 Parameters

The following parameters are specified when we create a count plot, let us get a brief idea of them:

Table 4. Illustration of the parameters regarding count plots

Count plots	Description	Count plots	Description
X and Y	This parameter specifies the data we refer to for representation and then observes the highlighted patterns.	Data	This parameter specifies the data frame we would like to take for the representation. For instance, data can be an array.
Color	This parameter specifies use color which can give a good appearances to our plot	Dodge	This parameter is an optional one, and it accepts a Boolean value as input
Palette	It takes the value of the palette. It is mostly used to show the hue variable.	Saturation	This parameter accepts a float value. A variation in the intensity of colors can be observed when we specify this.
Hue	This parameter specifies the column name.	Hue order	The parameter hue order takes strings as inputs
kwargs	The parameter kwargs specifies the key and mapping value	AX	The parameter AX is optional and used to take axes on which plot is created

Now, let us see what the different ways of representing our attributes are. In the first example, we have created a count plot for different variables.

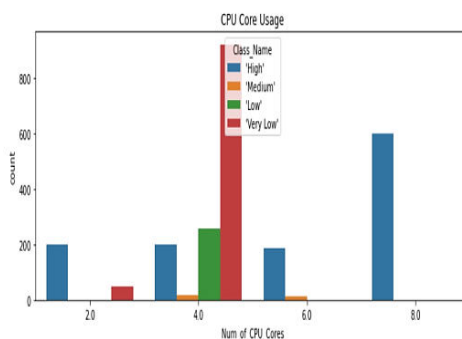


Fig -10: Depiction of the CPU core usage for four categorical variables and using hue parameter

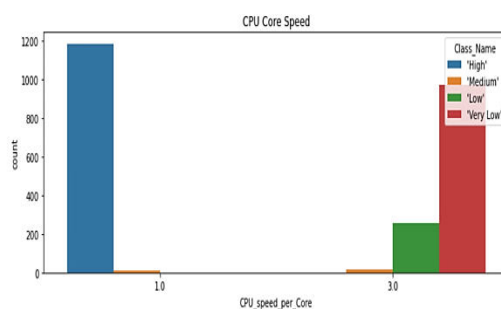


Fig -11: Illustration of the Disk capacity for four categorical variables and using hue parameter

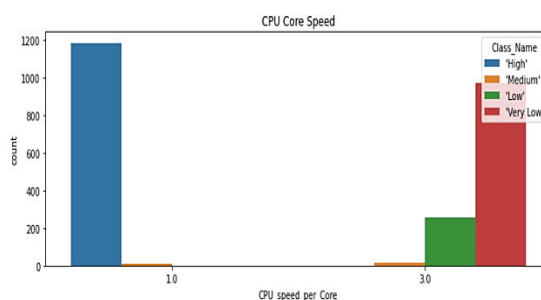


Fig -12: Illustration of the CPU core speed for four categorical variables and using hue parameter

4.3.9Confusion Matrix

Essentially, a confusion matrix is a table that represents the performance of your prediction model. Each item in a confusion matrix represents the number of predictions made by the model that properly classified the classes or incorrectly classified the classes respectively. In the case of a binary classification problem, we had to use 2×2 matrices with four values, as seen in the example below:

		Predicted Values	
		P (1)	N (0)
Actual Values	P (1)	Ture Positive	False Positive
	N (0)	False Negative	True Negative

Fig -13: 2 x 2 matrix for a binary classification

- There are two possible values for the target variable: positive or negative.
- The values of the target variable are represented by the columns of data; the projected values of the target variable are represented by the rows of data.

Table 5. Number of instances in the training and testing dataset of the model

No. of Cases in the Training Dataset	No. of Cases in the testing Dataset	Total	No. of Cases in the Training Dataset	No. of Cases in the testing Dataset	Total
25,687 Cases	2450 cases	28,147 cases	2427 cases	7420 cases	9847 cases

After the standard LR categories have been taught in model data sets, test data sets are utilized to evaluate the performance of trained classifiers. To find alternate test procedures, we must first compute the matrices of confusion for all possible responses to the question. To create confusion matrices, the number of outcomes in each division is tallied. As previously stated in the prior paragraph, these findings represent the quantities of TP, FP, TN, and FN contained in the sample. Memory usage models for the LR categories that include confusion matrices. Observe that 1247 instances in the Lower category are well separated from 2284 (54.60 percent) cases in the LR category and that 2257 cases in the same category are well divided (87.2 percent) of the LR class, based on this confusion matrix of cases. Because of LR's ability to deal with difficult problems such as injustice and a significant infringement of one's right to autonomous thought.

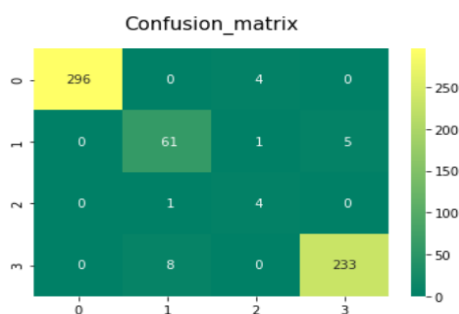


Fig -14: Confusion matrix of memory utilization

When compared to previous research, the proposed random forest identification method showed the best accuracy 96% percent. As seen by this matrix, the LR can independently balance the distribution of each type of class. The average level of sensitivity and specificity for each classification is derived by multiplying the degree of sensitivity and specificity of all categories by the number of cases that fall within each category's range. Furthermore, we noticed that while estimating the load of cloud services, the accuracy, sensitivity, and clarity of the results were more consistent.

V. CONCLUSION

The random forest model is applied for achieving the best accuracy. We find the accuracy of the model for performance prediction. We also placed our model through its steps with a one-year database from the MOF data center, which had thousands of entries and hundreds of different devices. The Random Forest model is used to determine the correctness of the results. Predicting the performance metrics of cloud nodes concerning alternative configuration options for their resources is modeled using the Random Forest classifier in the suggested technique, which is based on the classifier. Using an actual one-year dataset from the KSA finance department, which had thousands of records and hundreds of computers, we were able to measure the effectiveness of our model. This model demonstrates how a random forest classifier may deliver accurate and efficient results. The accuracy of training and testing was segregated from one another throughout model running. These parameters and attributes show the effectiveness of the model. Like this, we find the greatest level of accuracy that we were able to attain with our model was 96%. In the future, we hope to improve our model by incorporating additional characteristics and workload metrics into it. Furthermore, we will use SVM to the findings collected to execute a scientific and accurate cloud service selection process.

REFERENCES

- [1] Eisa, M., Younas, M., Basu, K., & Awan, I. (2020). Modelling and simulation of qos-aware service selection in cloud computing. *Simulation Modelling Practice and Theory*, 103, 102108.
- [2] Buyya, R., Broberg, J., & Goscinski, A. M. (Eds.). (2010). *Cloud computing: Principles and paradigms* (Vol. 87). John Wiley & Sons.
- [3] Demirkan, H., & Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, 55(1), 412-421.
- [4] Padhy, R. P., & Patra, M. R. (2012). Evolution of cloud computing and enabling technologies. *International Journal of Cloud Computing and Services Science*, 1(4), 182.
- [5] Mahmood, Z. (2011). Cloud computing for enterprise architectures: concepts, principles and approaches. In *Cloud computing for Enterprise architectures* (pp. 3-19). Springer, London.
- [6] Rajan, A. P. (2013). Evolution of cloud storage as cloud computing infrastructure service. *arXiv preprint arXiv:1308.1303*.
- [7] Keyes, J. (2016). *Enterprise 2.0: Social networking tools to transform your organization*. CRC Press.
- [8] Dabbagh, M., Hamdaoui, B., Guizani, M., & Rayes, A. (2015). Toward energy-efficient cloud computing: Prediction, consolidation, and overcommitment. *IEEE network*, 29(2), 56-61.
- [9] Berman, Oded, Dmitry Krass, and M. Mahdi Tajbakhsh. "A coordinated location-inventory model." *European Journal of Operational Research* 217, no. 3 (2012): 500-508.
- [10] Ji, Jindi, Yuxin Wang, Haitao Jia, and Mingjun Shi. "Distribution network planning and inventory management in a multi-retailing supply chain." In *Journal of Physics: Conference Series*, vol. 1903, no. 1, p. 012009. IOP Publishing, 2021.
- [11] Hassan, M. M., & Alsanad, A. (2016). Resource provisioning for cloud-assisted software defined wireless sensor network. *IEEE Sensors Journal*, 16(20), 7401-7408.
- [12] Di, S., Kondo, D., & Cirne, W. (2014). Google hostload prediction based on Bayesian model with optimized feature combination. *Journal of Parallel and Distributed Computing*, 74(1), 1820-1832.
- [13] Pramanik, M. I., Lau, R. Y., Yue, W. T., Ye, Y., & Li, C. (2017). Big data analytics for security and criminal investigations. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 7(4), e1208.
- [14] Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable cities and society*, 38, 230-253.
- [15] Siriya, P., Chadsuthi, S., Jampachaisri, K., & Kesorn, K. (2018). Dengue epidemics prediction: a survey of the state-of-the-art based on data science processes. *IEEE Access*, 6, 53757-53795.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details