IMPROVED ACCURACY IN PREDICTING THE NETWORK PERFORMANCE FOR ALLOCATING SERVER RESOURCES USING LINEAR REGRESSION COMPARED WITH DECISION TREE.

Gade. Mary Spandana[1] , K. Sashi Rekha[2*]

Article History: **Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The major goal of this study is to evaluate the accuracy of the Novel Linear Regression technique for forecasting network performance for allocating server resources to the Decision tree (DT) approach utilizing meteorological data.

Materials and Methods: For predicting the accuracy percentage of network performance for allocating server resources in information technology infrastructure, the Novel Linear Regression algorithm (LR) with sample size=10 and with sample size = 10, 95 percent confidence interval, and pretest power of 80 percent were iterated at various times. The Novel Linear Regression transforms the data into a higher-dimensional space, which improves accuracy.

Results: When compared to the accuracy of Decision Tree (80%), Novel Linear Regression appears to be more accurate (88%). There was a insignificant difference between LR and the Decision Tree with ($p=0.936$) ($p<0.05$) as shown in Table 3. As demonstrated in Table 3, there was a significant relationship between LR and the Decision Tree ($p=0.001$) two tailed ($p>0.05$).

Conclusion: Novel Linear Regression algorithm performed more effectively than Decision Tree for predicting network performance.

Keywords: Novel Linear Regression, Decision Tree, Machine Learning, Network, Algorithm, Dependent, Accuracy, Server Resources.

[1]Research Scholar, Department of Computer Science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105
[2*]Project Guide, Department of Computer Science Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105

1. Introduction

For predicting network performance, Novel Linear Regression looks to be more effective than Decision Tree. Improving Ensemble Decision Tree Performance Using Adaboost and Bagging Using Adaboost and Bagging to Improve Ensemble Decision Tree Performance (Hasan, Siraj, and Sainin 2015). Here in this work we will show the machine learning algorithms to compare the network performances in server resources. "Dynamic Resource Management for Longevity in Web Server Systems " (Choi, Seok-Bong, and Jong-Kook Kim. 2009). Cloud Services, Networking, and Management (da Fonseca and Boutaba 2015). A constructed database table combining performance metrics will be analyzed to find the relation and for better understanding of the application of network performance. Design and Implementation of Server Monitoring System Based on SNMP (Zeng and Wang 2009). The applications of the research are students will be able to demonstrate digital literacy through hardware, operating systems, networking etc. Machine learning applications in cloud computing resource management (Huang et al. 2013).

Many researchers have implemented identifying network performance for allocating server resources in order to draw attention about network performance in server resources. In Google Scholar, there are around 2000 papers. Toward an Architecture for Monitoring Private Clouds (Chaves et al. 2011). IEEE has published around 120 publications. Network and Server Resource Management Strategies for Data Center Infrastructures (Tso, Fung Po, Simon Jouet, and Dimitrios P. Pezaros 2020). There are numerous strategies for allocating server resources in IT infrastructure that are employed in network performance. Network and Application Performance Measurement Challenges on Android Devices (Hoque, Rao, and Tarkoma 2021). With the biggest variance score (R2) of 0.85260 and the smallest bias score (RMSE) of 0.02323, Novel linear regression modeling had a respectable score. With KNN modeling, a decent result was attained, with a variation score of 0.57422 and a bias score of 0.08049. The highest referenced study (Heru Widagdo et al. 2021) focused on estimating network performance of server resources using Decision Tree, which had an accuracy of 81%. Collecting ground-truth data may not provide exact accuracy because many people may not reply to the data and may not always provide accurate information. Web Server Performance Enhancement by Suppressing Network Traffic for High Performance Client (Jin and Tomoishi 2015). Machine learning techniques have been employed for predicting accuracy for network performances of server resources, according to the literature study (Mok et al. 2021). The Decision Tree technique is mostly used to tackle problems of classification and regression. Our team has extensive knowledge and research experience that has translated into high quality publications(Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa et al. 2022; Yaashikaa et al. 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022)

The research gap identified for the existing work is that there are some failures in the web server crashes in prediction of network performance for allocating the server resources. Machine learning can be used to teach how to predict the network performance for servers. This proposed system is to improve accuracy in predicting the network performance in server resources.

2. Materials and Methods

The proposed work was done in the Open Source Lab of Saveetha School of Engineering Saveetha Institute of Medical and Technical Sciences' (SIMATS). The study will be divided into two groups. Extracting usage patterns from web server log (Jeba, J. Monisha Privthy, J. Monisha Privthy Jeba, M. S. Bhuvaneswari, and K. Muneeswaran. 2016). The Novel Linear Regression algorithm is in group one, and the Decision Tree algorithm is in group two. The sample size for each group was derived using data from a prior study with a g power of 80%. Optimizing network performance with content switching: Server , Firewall, and cache load balancing (Syme, Matthew, and Philip Goldie. 2004). The sample size of the Novel Linear Regression algorithm (N=10) and the Decision Tree (N=10) were computed based on this information. To begin, open Collab and load the dataset. Upon the import of the dataset, the appropriate code for assessing the Novel Linear Regression's accuracy was provided. (Vora, Mehul Nalin. 2020). The required output for the code is then obtained in the form of graphs and values that have been indicated.

Linear Regression

Novel Linear regression is one of the most well-known and well-understood approaches in statistics and machine learning. In statistics, Novel linear regression is a method for measuring the relationship between a scalar answer and one or more explanatory factors (also known as dependent and independent variables). Where there's just one mediating variable, simple Novel linear regression

is performed, while multiple Novel linear regression is used when there are more

Algorithm

1. Start Program
2. Input the dataset
3. Explore the data to figure out what they look like
4. Give path to LR to configure and store it in a variable
5. Start while loop to loop frame one by one
6. Detect the objects in frames using variables that we declared initially in step 2
7. find the centroid points of each side(x,y)
7.1)calculating the words from tweets
7.2) Pre-process the data
7.3)Remove nan values and unwanted values
7.4)clearing the stop words using lemitizing
7.5)Removing the unwanted text by cleaning and stemming process
8. Split the data into attributes and tables
9. Divide the data into training and testing sets
10. Train the K-Nearest neighbor algorithm
11. Destroy all windows
12. End program.

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset = pd.read_csv(' mac server dataset.csv ')
X = dataset.iloc[: , :-1].values
Y = dataset,iloc[: , 1].values
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X , y , test_size =%, random_state = 1)
from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regressor.fit(X_train , y_train)
y_pred = regressor . predict (X_test)
plt.scatter (X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict (X_train), color = 'blue')
plt.scatter (X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title(' Salary vs Experience (Test set) ' )
plt.xlabel(' Years of Experience')
plt.ylabel ('Salary')
plt.show()
```

Fig. 1. Pseudocode of Linear Regression Algorithm

Decision Tree

The main purpose of this method is to develop a model that predicts the value of a target variable, and the decision tree solves the problem by using the tree representation, where the leaf node equates to a class label and characteristics are represented on the internal node of the tree. Decision trees are a common approach in machine learning and are often used in control theory, specifically in decision - making, to help discover the ideal method for achieving a goal.

```
import pandas as pd
import numpy as np
import seaborn a sns
import matplotlib.pyplot as plt
from sklearn.model_selection import
train_test_split
from sklearn. tree import DecisionTreeClassifer
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn import tree
df = pd.read_csv(' mac server datset.csv')
df.describe()
dtc = DecisionTreeClassifier()
dtc.fit(x_test, y_test)
clf_gini = DecisionTreeClassifer(criterion = "gini" ,random_state = 100,max_depth = 2, min_samples
_x_2 =3)
clf_gini.fit(x_train, y_train)
y_pred = clf_gini.predict(x_test)
y_pred
print("Accuracy is :" ,accuracy_score(y_test, y_pred)*100)
```

Fig. 2. Pseudocode of Decision Tree Algorithm

Testing Procedure

The dataset was composed of alternating, one for training and the other for testing. The best attribute in the training phase was already used to train each algorithm. Following training, testing will be used to compare the taught set with the output, which will then be presented as an accuracy rate. After that, the code for the Decision Tree algorithm was added for some data we needed to get the values. The Novel Linear Regression Algorithm and

Decision Tree were evaluated using Google Collab and the Python programming language. The system configuration included an Intel Core i5 processor and 8GB of RAM. The machine had a 64-bit operating system, an X64-based processor, and a 917-GB hard drive. The operating system Windows 10 is included in the software configuration. IBM SPSS (Statistical Package for Social Sciences) version 21 was used for the analysis. It's a type of statistical software that's used to analyze data. For both proposed and current algorithms, 10 iterations with a maximum of 20 samples were performed, and the projected accuracy for each iteration was recorded for accuracy analysis.

Statistical Analysis

The Statistical Package for the Social Sciences(SPSS) was used to conduct the statistical analysis (Ebbers et al. 2013) and Google collab software tools were used to analyze the data. For the LR and Decision Tree algorithms, descriptive statistics for mean, standard deviation, and standard error were calculated. Anonymized logging parameters X 1, X 2, and so on are independent variables. Output variables were the dependent variable (Accuracy). To compare the performance of algorithms, an independent sample t-test is used.

3. Results

In this study, ten samples were examined for Novel Linear Regression and Decision Tree, and the results were 88% and 80% accurate, respectively, as shown in Table 1, with various occurrences occurring on different days.

The mean, standard deviation, and standard error mean of Novel Linear Regression and Decision Tree algorithms for network performance to server resources were also listed in Table 1, which proves that Novel Linear Regression algorithm has an accuracy mean of 85 %, standard deviation 2.95419 for sample size of N=10, and Decision Tree algorithm after predicting has an accuracy mean of 79 %, standard deviation 3.00523 for sample size of N=10. While comparing Novel Linear Regression to Decision Tree, the mean, standard deviation, and significant difference reveal that there is a meaningful difference between the two groups. The mean accuracy of the Novel Linear Regression algorithm and the Decision Tree technique for network performance of server resources is evaluated in the bar graph Fig. 1.

4. Discussion

The results reveal that the Novel Linear Regression method outperforms the Decision Tree approach in terms of accuracy. Novel Linear Regression had a standard deviation of 2.95419, while the Decision Tree approach had a standard deviation of 3.00523. For sample size (N=10), the standard error mean for Novel Linear Regression was 0.93420, whereas the standard error mean for the Decision Tree technique was 0.95034, as shown in Table 2. As a result, it is evident that the Novel Linear Regression method achieves the Decision Tree approach when it comes to network performance when distributing server resources. Pretest analysis was carried out with a gpower of 80%.

There are numerous techniques used in network performance for allocating server resources in IT infrastructure. The accuracy of the Decision tree algorithm was 81% (Heru Widagdo et al. 2021). In network performance, there are a variety of ways for allocating server resources in information technology (IT) architecture. Novel Linear regression modeling produced a decent performance, with the highest variance score (R2) of 0.85260 and the smallest bias score (RMSE) of 0.02323. With a variation score of 0.57422 and a bias score of 0.08049, KNN modeling produced an excellent outcome. (Yazdi, M. Amin, M. Amin Yazdi, Pejman Farhadi Ghalatia, and Benedikt Heinrichs. 2021. "Event Log Abstraction in Client-Server Applications."). Training and test material are created using data that has been translated into a tabular format. A graph is drawn based on the acquired findings to highlight the contrast between the Novel Linear Regression algorithm and the Decision Tree algorithm. Identifying utilisation trends in web server logs (Jeba, J. Monisha Privthy, J. Monisha Privthy Jeba, M. S. Bhuvaneswari, and K. Muneeswaran. 2016). We were able to achieve good results by utilizing the algorithms' accuracy. The data for this study was obtained and analysed through log files collected from a server device, and then transformed into a model.. This log is the most useful source of information for determining how a machine works. Data analysis employing machine learning is expected to result in a more accurate estimation of server requirements. According to the following, the following are the challenges in accessing and keeping these logs: The extensive product selection (servers, storage, networking, backup power, and so on) "Resource Monitoring on an Application Layer Active Network Server."(Liabotis, I., D. Garton, O. Prnjat, and L. Sacks). The number of goods manufacturers (multi manufacturer), as well as related items (Servers, Storage, and Networking) from separate sellers, that can be used in a single network. Each gadget (machine) communicates data in its unique way (Types of Machine Data). Some are in text format, while others are encrypted, binary files, or data from different file systems. The future work can be implemented in companies to improve the network performance.

5. Conclusion

In this proposed work, prediction of network performance for allocating server resources is performed using Linear Regression algorithm with better accuracy 88.00% when compared to the Decision Tree (DT) algorithm.

Declaration

Conflict of interests

No conflict of interest in this manuscript.

Authors Contributions

Author GMS was involved in data collection, data analysis, manuscript writing. Author KSR was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Qbec Infosol
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

6. References

Hasan, Md Rajib, Fadzilah Siraj, and Mohd Shamrie Sainin. 2015. "Improving Ensemble Decision Tree Performance Using Adaboost and Bagging." https://doi.org/10.1063/1.4937027.

Fonseca, Nelson L. S. da, and Raouf Boutaba. 2015. Cloud Services, Networking, and Management. John Wiley & Sons.

Chaves, Shirlei De, Shirlei De Chaves, Rafael Uriarte, and Carlos Westphall. 2011. "Toward an Architecture for Monitoring Private Clouds." IEEE Communications Magazine. https://doi.org/10.1109/mcom.2011.6094017.

Zeng, Wenxian, and Yue Wang. 2009. "Design and Implementation of Server Monitoring System Based on SNMP." 2009 International Joint Conference on Artificial Intelligence. https://doi.org/10.1109/jcai.2009.34.

Hoque, Mohammad A., Ashwin Rao, and Sasu Tarkoma. 2021. "Network and Application Performance Measurement Challenges on Android Devices." ACM SIGMETRICS Performance Evaluation Review. https://doi.org/10.1145/3453953.3453955.

Tso, Fung Po, Simon Jouet, and Dimitrios P. Pezaros. 2016. "Network and Server Resource Management Strategies for Data Center Infrastructures: A Survey." Computer Networks. https://doi.org/10.7763/ijmo.2013.v3.256.

Ravindran, Kaliappa, Mohammad Rabby, and Shereef El Metwally. 2010. "Management Intelligence for Optimal Resource Allocations in Network Server Systems." 2010 IEEE Network Operations and Management Symposium - NOMS 2010. https://doi.org/10.1109/noms.2010.5488497.

Jin, Yong, and Masahiko Tomoishi. 2015. "Web Server Performance Enhancement by Suppressing Network Traffic for High Performance Clients." 2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS). https://doi.org/10.1109/apnoms.2015.7275385.

Khanna, G., K. Beaty, G. Kar, and A. Kochut. 2006. "Application Performance Management in Virtualized Server Environments." 2006 IEEE/IFIP Network Operations and Management Symposium NOMS 2006. https://doi.org/10.1109/noms.2006.1687567.

Terrell, Jeff, Kevin Jeffay, F. Donelson Smith, Jim Gogan, and Joni Keller. 2008. "Exposing Server Performance to Network Managers through Passive Network Measurements." 2008 IEEE Internet Network Management Workshop (INM).

Han, Jeong Soo, Seong Jin Ahn, and Jin Wook Chung. "Web-Based Performance Manager System for a Web Server." NOMS 98 1998 IEEE Network Operations and Management Symposium. https://doi.org/10.1109/noms.1998.654836.

Liabotis, I., D. Garton, O. Prnjat, and L. Sacks. "Resource Monitoring on an Application Layer Active Network Server." NOMS 2002. IEEE/IFIP Network Operations and Management Symposium." Management Solutions for the New Communications World"(CatNo.02CH37327). https://doi.org/10.1109/noms.2002.1015641.

Choi, Seok-Bong, and Jong-Kook Kim. 2009. "Dynamic Resource Management for Longevity in Web Server Systems." 2009 11th IEEE International Conference on High Performance Computing and Communications. https://doi.org/10.1109/hpcc.2009.8.

Zhou, Jin. "Multiple Evidence Combination for Web Site Search Using Server Log Analysis." https://doi.org/10.32920/ryerson.14658057.v1.

Yazdi, M. Amin, M. Amin Yazdi, Pejman Farhadi Ghalatia, and Benedikt Heinrichs. 2021."Event Log Abstraction in Client-Server Applications." Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. https://doi.org/10.5220/001065200000306.

Chapman, Chris. 2016. Network Performance and Security: Testing and Analyzing Using Open Source and Low-Cost Tools. Syngress.

Jeba, J. Monisha Privthy, J. Monisha Privthy Jeba, M. S. Bhuvaneswari, and K. Muneeswaran. 2016. "Extracting Usage Patterns from Web Server Log." 2016 2nd International Conference on Green High Performance Computing (ICGHPC).

Tso, Fung Po, Simon Jouet, and Dimitrios P. Pezaros. 2016. "Network and Server Resource Management Strategies for Data Center Infrastructures: A Survey." Computer Networks 106 (September): 209–25.

Jin, Yong, and Masahiko Tomoishi. 2015. "Web Server Performance Enhancement by Suppressing Network Traffic for High Performance Clients." 2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS). https://doi.org/10.1109/apnoms.2015.7275385.

Tables and Figures

Table 1: Accuracy values for Novel Linear Regression and Decision Tree.

S No.	Linear Regression (Accuracy)	Decision Tree (Accuracy)
1	88.00	80.31
2	87.01	79.02
3	86.02	78.11
4	85.03	77.00
5	84.00	76.09
6	83.23	75.19
7	82.41	74.11
8	81.21	73.31
9	80.00	72.21
10	79.21	71.15

Table 2:- The Novel Linear Regression algorithm has a mean and standard deviation comparison accuracy of 88%, whereas the Decision Tree technique has an accuracy of 80%.

S No	N	Mean	Standard Deviation	Std.Mean Error
Linear Regression	10	85.6120	2.95419	.93420
Decision Tree	10	79.6500	3.00523	.95034

Table 3: The result of an Independent T - test is calculated with a 95% confidence interval and a significance level of 0.05 (the Linear Regression algorithm performs considerably better than the Decision Tree method with a value of p =0.007).

	Levene's Test for	T-test for equality of Means

		Equality of Variances								
		F	Sig	t	df	Sig (2-tai-led)	Mean Difference	Std.Error Difference	95% Confidence interval of the Difference	
									Lower	Upper
Accu-racy	Equal Variances assumed	.007	.936	4.474	18	.001	5.96200	1.33262	3.16228	8.76172
	Equal variances not assumed			4.474	17.995	.001	5.96200	1.33262	3.16222	8.76178

Fig. 3.

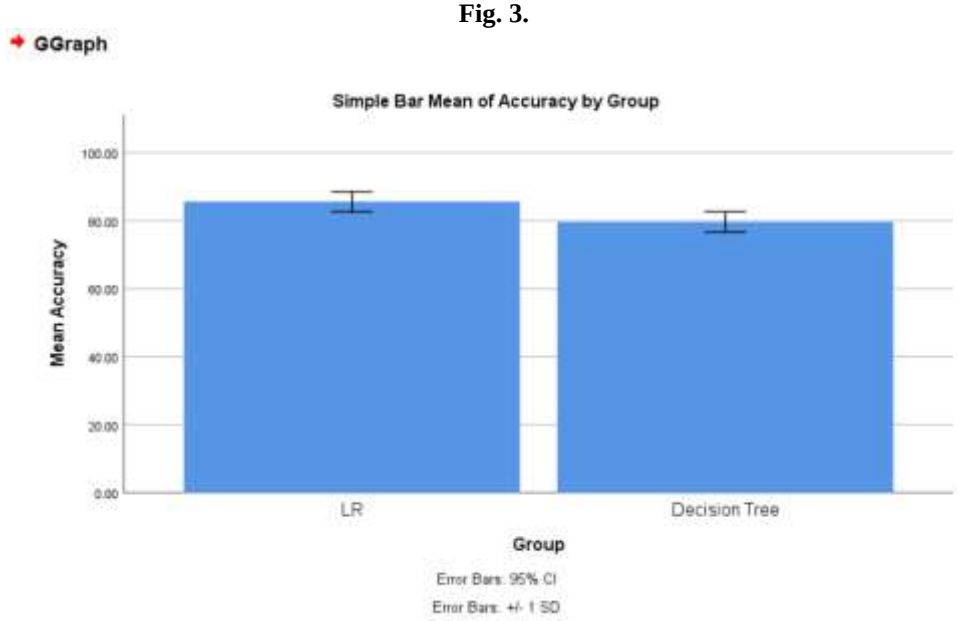

Fig. 3. In terms of mean accuracy, the Novel Linear Regression algorithm and the Decision Tree algorithm are compared. Novel Linear Regression has a higher mean accuracy (85%) than Decision Tree (79%). The Novel Linear Regression algorithm has a somewhat lower standard deviation than Decision Tree. LR versus Decision Tree on the X Axis. Y Axis: Mean detection accuracy 1 SD.