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

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 main objective of this research is to estimate the accuracy rating of predicting network operation for allocating server resources using servers data using the Novel Linear Regression algorithm versus the Support Vector Regression (SVR).

Materials and Methods: For trying to predict the accuracy rate of network performance for allocating server resources in information technology infrastructure, at varying times, the Novel Linear Regression algorithm (LR) with sample size=10 and Support Vector Regression (SVR) with sample size=10, 95 percent confidence increment, and pretest power is 80 % were restated. The Novel Linear Regression converts the original into a higher-dimensional space, which enhances accuracy.

Results: Novel linear regression appears to prove with better accuracy (88%) compared to Support Vector Regression(78%). There was an insignificant difference between LR and SVR with p=0.950 ($p<0.05$).

Conclusion: When it came to identifying server resources, the Novel Linear Regression algorithm significantly outperforms Support Vector Regression in terms of network performance.

Keywords: Novel Linear Regression, Support Vector Regression, Server resources, Standard deviation, Prediction, Infrastructure, Machine Learning.

[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

The study's goal is to forecast the percentage of network performance that is accurate when allocating server resources. Predicting Web Server Crashes: A Case Study in Algorithm Comparison (Alonso, Javier, Jordi Torres, and Ricard Gavaldà. 2009). A Bayesian Approach to Virtual Resource Prediction in a Cloud Environment(Shyam, Gopal Kirshna, and Sunilkumar S. Manvi. 2016). Self-Adaptive Resource Allocation for Energy-Aware Virtual Machine Placement in a Cloud of Dynamic Computing (Jiang, Han-Peng, and Wei-Mei Chen. 2018). 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 implemented identifying network performance for allocating server resources to raise awareness about network performance in server resources. Google scholar has around 1800 papers. Around 100 articles published in IEEE. Many algorithms used in network performance for allocating server resources in IT infrastructure. By using the Decision tree algorithm the accuracy was 76% Enhancing Web Server Performance by Suppressing Network Traffic for High Performance Clients Enhancing Suppressing network traffic to improve web server performance Clients (Jin and Tomoishi 2015). RMSE score for Support Vector Machine 0.064316 and Score Variance was 0.1294902 (Huang, Chenn-Jung, Yu-Wu Wang, Chih-Tai Guan, Heng-Ming Chen, and Jui-Jiun Jian 2020). RMSE Score Bias for Random Forest 0.036077 and Score Variance 0.644596 (Roy, Chiranjiv, Sourov Moitra, Mainak Das, Subramaniyan Srinivasan, and Rashika Malhotra 2019). "Prediction of IT Infrastructure Downtime Using Hybrid Machine Learning and Natural Language Processing" The most cited article was (Heru Widagdo et al. 2021) focused on estimating network performance of server resources with an accuracy of 80.00 percent using the Support Vector Regression technique. Because many people may not respond to field survey and may not always provide reliable information, ground-truth data collection may not provide excellent accuracy (Mok et al. 2021). According to the literature study, machine learning techniques have been utilised to forecast the accuracy of network performance of server resources. Measuring Application Performance on Android Devices: Network and Challenges (Hoque, Rao, and Tarkoma 2021). Support Vector Regression algorithm mainly used to solve the classification and regression problems. 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; Mohan 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 may be used to train servers how to anticipate network performance.. This proposed system is to improve accuracy in predicting the network performance in server resources.

2. Materials and Methods

The research study was done in the Open source lab of Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences. Two groups have been identified for the study. Using content switching to improve network speed, Firewall, and cache load balancing (Syme, Matthew, and Philip Goldie. 2004). The Novel Linear Regression technique is in Group 1 and the Support Vector Regression algorithm is in Group 2. The sample size for each group was established using prior research while keeping g power at 80%. (Jeba et al. 2016). The random sample of the Novel Linear Regression algorithm (N=10) and Support Vector Regression (N=10) were measured as an outcome (Vora, Mehul Nalin 2020). First import the dataset in colab. After importing the dataset we have to give required code for getting the accuracy for the Novel Linear Regression. Then we get the required output for the code as graphs and signified values.

Linear Regression

Novel Linear regression was among the most well-known and well-understood algorithms in machine learning algorithms. In statistical information, the Novel linear regression methodology is used to explore the relation between a comeback and one or more explanatory variables (also known as dependent and independent variables). Simple Novel linear regression is used when there is only one regression coefficient, whereas multiple Novel linear regression could be used but there are many.

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. Split the data into two different sets one is training and other is testing.
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

Support Vector Regression

Distinct value prediction Support Vector Regression has been used as a training data approach. On the same supposition, SVMs and Support Vector Regression both accomplish well. The core premise of SVR is that the strongest line will be revealed. The best fit line in SVR is the shape parameter with the most points. Support Vector Machine may also be used as a regression method while preserving all of the algorithm's basic properties (maximal margin). With a few minor differences, the Support Vector Regression (SVR) employs the same categorization concepts as the SVM.

```
import numpy as np
import matplotlib.pyplot as plt
import pandasas pd
dataset = pd.read_csv(' mac server dataset.csv')
x = dataset.iloc[: , 1:2].values
y = dataset.iloc[: , 2].values
From sklearn.preprocessing import StandardScalar
x = sc_x.fit_transform(x)
y = sc_y.fit_transform(y)
from sklearn. svm import svr
regressor = svr(kernel = 'rbf')
regressor.fit(x,y)
y_pred = regressor.predit(6.5)
y_pred = sc_y.inverse_transform(y_pred)
x_grid = np.arrange(min(x),max(x), 0.01)
x_grid = x_grid.reshape((len(x_grid), 1))
plt.xlabel('t')
plt.ylabel('r')
plt.show
```

Fig. 2. Pseudocode of Support Vector Regression Algorithm

Testing Procedure

The collection was divided into two portions, one for training and the other for testing. The best attribute of the training set will be used to train the respective algorithm. Following training, testing will be used to compare the taught set with the

output, which will be expressed as an average accuracy. Then, using the data that we require, have used the Support Vector Regression algorithm to gather the values.
IBM SPSS (Statistical Package for Social Sciences) version 21 was employed in the analysis.It is a gathering of quantitative software for data analysis. Ten iterations with a maximum of 20 samples were done for both hypothesised and current procedures, and the anticipated accuracy for each iteration was recorded for quality analysis.

Statistical Analysis
The statistical analysis was accomplished using the Statistical Package for the Social Sciences (SPSS) (Ebbers et al. 2013) and Google collab software tools were used to analyze the data. Descriptive statistics for average, variance, and standard deviation were computed for the LR and SVR techniques. 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

For Novel Linear Regression, we used ten instances that were all different from one another. The accuracy of Novel Linear Regression looks to be 88.00 percent, whereas 10 samples in the Support Vector Regression technique yields an accuracy of 78.00 percent, as shown in Table 1 with various incidents that occurred on various days. Table 2 shows the mean, standard deviation, and standard error mean of Novel Linear Regression and Support Vector Regression techniques for network performance to server resources. For a random sample of N=10, the Novel Linear Regression algorithm has an exactness mean of 85%, standard deviation 2.95419, and the Support Vector Regression methodology has a precision mean of 77%, standard deviation 3.00730.
The correctness mean of Novel Linear Regression is extremely higher than that of Support Vector Regression, noted a large difference between the two. The mean exactness of the Novel Linear Regression methodology and the Support Vector Regression algorithm for system performance to server resources is represented by a bar graph.

4. Discussion
Figure 1 contrasts the mean precision of the Novel Linear Regression method and the Support Vector Regression methodology for data system performance. Standard deviation for Novel Linear Regression was 2.95419 and Standard deviation for Support Vector Regression algorithm was 2.82290. Standard error mean was 0.93420 for Novel Linear Regression and Standard error mean for Support Vector Regression obtained as 0.89268 for sample size (N=10) as tabulated in Table 2. As a result, it was completely obvious that the Novel Linear Regression algorithm outscored the Support Vector Regression methodology in addition to network performance for distributing server resources. Pretest analysis has been done with gpower 80%. Many algorithms used in network performance for allocating server resources in information technology (IT) infrastructure. By using the Decision tree algorithm the accuracy was 81% . RMSE score for Support Vector Machine 0.064316 and Score Variance was 0.1294902 (Huang, Chenn-Jung, Yu-Wu Wang, Chih-Tai Guan, Heng-Ming Chen, and Jui-Jiun Jian. 2013). RMSE Score Bias for Random Forest 0.036077 and Score Variance 0.644596 (Roy, Chiranjiv, Sourov Moitra, Mainak Das, Subramaniyan Srinivasan, and Rashika Malhotra. 2015). Information that has been listed is divided into training and test material. Based on the outcome, a graph was created to compare the Novel Linear Regression algorithm and the Support Vector Regression computation. "Multiple Evidence Combination for Web Site Search Using Server Log Analysis." (Zhou, Jin.). Here we secured good output by using the algorithms by using their accuracy. The sample for this research was collected and processed from log files received from a server device before being interpreted into a framework (Liao, Shih-Wei, Tzu-Han Hung, Donald Nguyen, Chinyen Chou, Chi Cheng Tu, and Hucheng Zhou. 2009). "Prefetch Optimization for Data Center Applications Using Machine Learning". The information in this log is the most important source for figuring out how a machine functions. It is envisaged that data analyzed using machine learning will result in a more accurate estimate of server requirements. The preceding are the obstacles in retrieving and preserving these logs, according to: The wide range of products (servers, storage, networking, backup power, and so on) The quantity of goods makers (multi manufacturer), as well as similar products (Servers, Storage, and Networking) from various sellers, enabling it to be utilized in a single network. Each device (machine) has its own way of giving data (Types of Machine Data). Some are in text, while others are encrypted, binary files, or other file systems. The future work can be implemented in companies to improve the network performance.

5. Conclusion

In this suggested work, the Linear Regression technique used to predict network performance for the purpose of allocating server resources performed better accuracy 88.00% when compared to the Support Vector Regression 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

Alonso, Javier, Jordi Torres, and Ricard Gavaldà. 2009. "Predicting Web Server Crashes: A Case Study in Comparing Prediction Algorithms." 2009 Fifth International Conference on Autonomic and Autonomous Systems. https://doi.org/10.1109/icas.2009.56.

Barrel, Josep Ll, Iñigo Goiri, Ramon Nou, Ferran Julià, Josep O. Fitó, Jordi Guitart, Ricard Gavaldá, and Jordi Torres. 2012. "Toward Energy-Aware Scheduling Using Machine Learning." Energy-Efficient Distributed Computing Systems. https://doi.org/10.1002/9781118342015.ch8.

Bowles, Michael. 2015. Machine Learning in Python: Essential Techniques for Predictive Analysis. John Wiley & Sons.

Guda, C. 2006. "pTARGET: A Web Server for Predicting Protein Subcellular Localization." Nucleic Acids Research. https://doi.org/10.1093/nar/gkl093.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2013. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.

Huang, Chenn-Jung, Yu-Wu Wang, Chih-Tai Guan, Heng-Ming Chen, and Jui-Jiun Jian. 2013. "Applications of Machine Learning to Resource Management in Cloud Computing." International Journal of Modeling and Optimization. https://doi.org/10.7763/ijmo.2013.v3.256.

James, Gareth Michael, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2021. An Introduction to Statistical Learning: With Applications in R. Springer Nature.

Jiang, Han-Peng, and Wei-Mei Chen. 2018. "Self-Adaptive Resource Allocation for Energy-Aware Virtual Machine Placement in Dynamic Computing Cloud." Journal of Network and Computer Applications. https://doi.org/10.1016/j.jnca.2018.07.011.

Liao, Shih-Wei, Tzu-Han Hung, Donald Nguyen, Chinyen Chou, Chi Cheng Tu, and Hucheng Zhou. 2009. "Machine Learning-Based Prefetch Optimization for Data Center Applications." Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09. https://doi.org/10.1145/1654059.1654116.

Shyam, Gopal Kirshna, and Sunilkumar S. Manvi. 2016. "Virtual Resource Prediction in Cloud Environment: A Bayesian Approach." Journal of Network and Computer Applications. https://doi.org/10.1016/j.jnca.2016.03.002.

Roy, Chiranjiv, Sourov Moitra, Mainak Das, Subramaniyan Srinivasan, and Rashika Malhotra. 2015. "IT Infrastructure Downtime Preemption Using Hybrid Machine Learning and NLP." Position Papers of the 2015 Federated Conference on Computer Science and Information Systems. https://doi.org/10.15439/2015f400.

Verma, Niharika, and Anju Sharma. 2017. "Workload Prediction Model Based on Supervised Learning for Energy Efficiency in Cloud." 2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA). https://doi.org/10.1109/cscita.2017.8066526.

Kapadia, N. H., C. E. Brodley, J. A. B. Fortes, and M. S. Lundstrom. "Resource-Usage Prediction for Demand-Based Network-Computing." Proceedings Seventeenth IEEE Symposium on Reliable Distributed Systems (Cat. No.98CB36281). https://doi.org/10.1109/reldis.1998.740526.

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.

Mok, Ricky K. P., Hongyu Zou, Rui Yang, Tom Koch, Ethan Katz-Bassett, and K. C. Claffy. 2021. "Measuring the Network Performance of Google Cloud Platform." In Proceedings of the 21st ACM Internet Measurement Conference. New York, NY, USA: ACM. https://doi.org/10.1145/3487552.3487862.

Tables and Figures

Table 1. Accuracy values of Novel Linear Regression and Support Vector Regression.

S no	Linear Regression (Accuracy)	Support Vector Regression (Accuracy)
1	88.00	78.03
2	87.01	77.04
3	86.02	76.31
4	85.03	75.01
5	84.00	74.12
6	83.23	73.31
7	82.41	72.42
8	81.21	71.21
9	80.00	70.12
10	79.21	69.01

Table 2. The Novel Linear Regression technique appears to have a mean and standard deviation comparison accuracy of 88%, while the Support Vector Regression methodology appears to be 78 % to almost.

S no	N	Mean	Standard Deviation	Std.Mean Error
Linear Regression	10	85.6120	2.95419	.93420
Support Vector Regression	10	77.6580	3.00730	.95099

Table 3. An independent T is calculated with a 95 percent confidence interval and a level of significance of 0.05 (the Linear Regression technique outperforms Support Vector Regression with a value of p=0.950).

	Levene's Test for Equality of Variances		T-test for equality of Means					
	F	Sig	t	df	Sig (2-tai-	Mean Difference	Std.Error Difference	95% Confidence interval of the

						led)			Difference	
									Lower	Upper
Accu-racy	Equal Variances assumed	.004	.950	5.967	18	.001	7.95400	1.33308	5.15330	10.7547
	Equal variances not assumed			5.967	17.994	.001	7.95400	1.33308	5.15323	10.7547

Fig. 2.

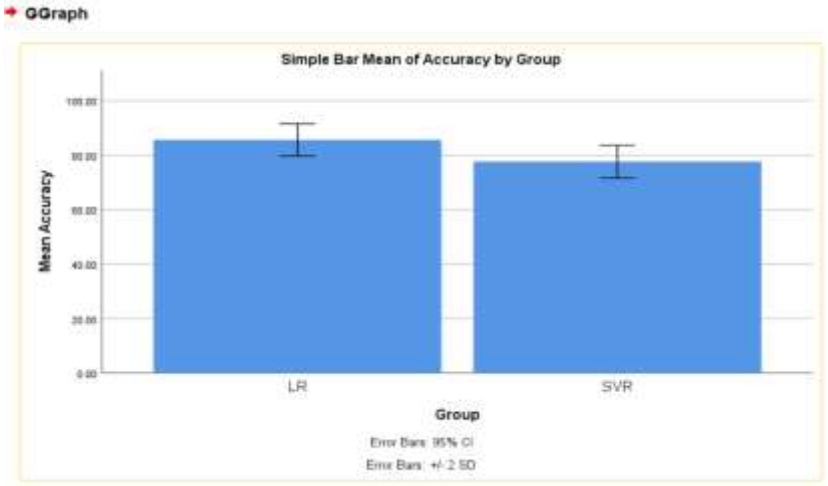

Fig. 3. In terms of the average accuracy, there is a comparison between the Novel Linear Regression and Support Vector Regression algorithms. Novel Linear regression has a higher average accuracy (85%) than Support Vector Regression (77 percent). Support Vector Regression looks to have a lesser standard deviation than Novel Linear Regression. LR vs. SVR on the X Axis Axis Y: Mean detection accuracy within 2 standard deviations.