

# Comparative Analysis of Multilayer Perceptron and Logistic Regression Algorithms for Prediction Tasks: A Comprehensive Study

C. Sai Kumar

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

**Abstract**— In predictive modeling, the accurate prediction of outcomes using suitable algorithms is of utmost importance. This research paper presents a comprehensive study comparing the performance of two popular algorithms, Multilayer Perceptron and Logistic Regression, for prediction tasks. Multilayer Perceptron is a powerful neural network-based algorithm, while Logistic Regression is a widely used technique for modeling categorical outcomes. The study investigates the effectiveness of these algorithms in terms of prediction accuracy, precision, and recall. Experimental results on diverse datasets demonstrate the strengths and limitations of both algorithms, providing valuable insights for selecting the most suitable algorithm for prediction tasks in different domains. This research paper presents a comprehensive study on the application of two popular algorithms, Multilayer Perceptron and Logistic Regression, for vote dataset prediction. The objective is to compare the performance of these algorithms in terms of accuracy, precision, recall, and F1-measure. Additionally, we discuss the interpretability and computational efficiency of each algorithm to provide valuable insights for decision-making in the context of voting prediction. The experimental results demonstrate the strengths and limitations of both algorithms and provide recommendations for selecting the most suitable algorithm for vote dataset prediction..

## I. INTRODUCTION

Machine learning is a subfield of artificial intelligence that focuses on the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed. It involves the study of statistical and computational methods that allow machines to learn patterns and extract insights from data.

Accurate prediction plays a vital role in various applications, such as medical diagnosis, financial forecasting, and customer behavior analysis. Multilayer Perceptron and Logistic Regression are widely employed algorithms in the field of predictive modelling [3]. This paper aims to conduct a comprehensive study to compare their performance and analyze their applicability for different prediction tasks. The study investigates multiple evaluation metrics to assess the prediction accuracy and effectiveness of both algorithms on vote dataset.

## II. SUPERVISED LEARNING

Supervised learning is a specific type of machine learning task where the algorithm learns from labeled data. In supervised learning, the training dataset consists of input data (features) and corresponding output labels. The goal is to learn a mapping function that can predict the output labels for new, unseen input data.

The training process in supervised learning involves presenting the algorithm with a set of labeled examples and allowing it to learn the underlying patterns or relationships between the input features and output labels. The algorithm iteratively adjusts its internal parameters to minimize the difference between its predicted outputs and the true labels in the training data. This process is typically guided by an objective function, such as minimizing the mean squared error or maximizing the likelihood of the observed labels [3][4][6].

Once the supervised learning algorithm has been trained on the labeled data, it can be used to make predictions on new, unseen data by applying the learned mapping function. The algorithm takes the input features of the new data as input and produces predicted output labels as output [8][9].

Supervised learning algorithms can be further categorized into two main types: regression and classification. Regression algorithms are used when the output variable is continuous, and the goal is to predict a numerical value [1][2]. For example, predicting the price of a house based on its features. Classification algorithms, on the other hand, are used when the output variable is categorical or discrete, and the goal is to assign new data to predefined classes or categories. For example, classifying emails as spam or non-spam based on their content.

Some popular supervised learning algorithms include linear regression, logistic regression, support vector machines, decision trees, random forests, and neural networks. Each algorithm has its own strengths and weaknesses, and the choice of algorithm depends on the specific problem and the characteristics of the data. Supervised learning has a wide range of applications in various fields, including image and speech recognition, natural language processing, recommendation systems, fraud detection, and many more. It is a fundamental and powerful technique in machine learning that has revolutionized the way computers learn from and make predictions on data.

### III. METHODOLOGY

#### 3.1 Artificial Neural Networks

Artificial Neural Networks (ANN) is a computational model and is made by human cerebrum and it is made out of neurons. Brain organization's essential unit is neuron or hub. An ANN has enormous number of processors; it is worked in the equal mode and organized in layers [3]. The ANN has three layers; they are input, stowed away and yield layers. The main layer gets the natural info data. Each layer acquires the result from the layer going before, and it is covered up layer. The framework gets yield from last layer. Each cycle hub has its own subject matter, as well as what it saw and every one of the guidelines it at first modified or produced for itself [4][5].

The interconnected layers propose that every hub  $n$  associated with different hubs in layer  $n-1$  at its passageways and in layer  $n+1$ , which gives an information record to these hubs. There may likewise be at least one hubs in the result layer, from which the arrangement it produces can be perused.

Brain networks are generally addressed regarding profundity, with the level of layers they need among info and result, or the problematic secret layers of the model. Maybe for this reason the term brain network is essentially utilized as an equivalent word for profound learning. They are even portrayed by the quantity of secret hubs in the model or by the quantity of data sources and results of every hub.

##### 3.1.1 Multilayer Perceptron (MLP)

A Multilayer Perceptron (MLP) Brain network has a place with the feedforward brain organization. It is a Counterfeit Brain Organization wherein all hubs are interconnected with hubs of various layers. The word Perceptron was first characterized by Forthcoming Rosenblatt in quite a while perceptron program. Perceptron is an essential unit of a fake brain network that characterizes the counterfeit neuron in the brain organization [4]. A managed learning calculation contains hubs' qualities, enactment capabilities, sources of info, and hub loads to compute the result.

The MLP Brain Organization works just in the forward course. All hubs are completely associated with the organization. Every hub passes its worth to the approaching hub just in the forward bearing. The MLP brain network utilizes a Backpropagation calculation to expand the precision of the preparation model.

#### 3.2 Logistic Regression

Logistic Regression is indeed a widely used technique in data analysis and machine learning for modeling categorical outcomes. It is particularly useful when the dependent variable is binary or dichotomous, meaning it has two possible outcomes. The goal of logistic regression is to estimate the probability of an event occurring based on a set of predictor variables. Unlike linear regression, which assumes a continuous dependent variable, logistic regression models the relationship between the predictors and the log-odds (logit) of the event occurring. The logit transformation allows for a linear relationship between the predictors and the log-odds, even when the relationship is nonlinear in the original scale[3][4].

Logistic regression can handle both continuous and categorical predictor variables. Continuous variables are straightforward to include in the model, as their relationship with the log-odds can be represented by a linear term. Categorical variables, on the other hand, need to be transformed into a set of binary (dummy) variables to be included in the model. Each category of the categorical variable is represented by a separate binary variable, indicating whether it is present or not.

The logistic regression model estimates the coefficients for each predictor variable, representing their contribution to the log-odds of the event occurring. These coefficients can be interpreted as the change in log-odds associated with a one-unit increase in the predictor variable, holding other variables constant. To make predictions using logistic regression, the estimated coefficients are applied to the predictor variables, and the logistic function is used to transform the resulting log-odds into probabilities. The logistic function, also known as the sigmoid function, maps any real-valued number to a value between 0 and 1, representing the probability of the event occurring.

#### IV. EXPERIMENTAL RESULTS

The investigations have been coordinated by using Python programming tongue. The Python Scikit-learn is a pack for data portrayal, gathering and portrayal. The Vote dataset used in this review was procured from the UCI ML vault data set [7]. In this vote dataset there are 435 cases and 17 elements recorded and 2 class marks. The standard dataset is distributed two sets one for preparing (70%) and one more set for testing (30%).

#### V. RESULTS AND DISCUSSION

TABLE 1

Algorithm	Accuracy	Precision	Recall	F-Measure
Multilayer Perceptron	95.52	95.5	95.53	95.5
Logistic Regression	94.42	94.5	94.3	94.4

The table presents the performance metrics of two popular algorithms, Multilayer Perceptron and Logistic Regression, for vote dataset prediction. The evaluation metrics include accuracy, precision, recall, and F1-measure.

According to the results, the Multilayer Perceptron algorithm achieved an accuracy of 95.52%. It demonstrated a precision of 95.3%, recall of 95.53%, and an F1-measure of 95.5%. On the other hand, Logistic Regression achieved an accuracy of 94.42%. It exhibited a precision of 94.5%, recall of 94.3%, and an F1-measure of 94.4%.

The Multilayer Perceptron algorithm performed exceptionally well, achieving high accuracy, precision, recall, and F1-measure scores. With an accuracy of 95.52%, the Multilayer Perceptron model accurately predicted the vote outcomes for the given dataset. The precision score of 95.5% suggests that the algorithm correctly classified positive instances, minimizing false positive predictions. Additionally, the recall score of 95.53% indicates that the model effectively identified a large proportion of actual positive instances. The F1-measure score of 95.5% highlights the balance between precision and recall achieved by the Multilayer Perceptron algorithm.

While the Logistic Regression algorithm also performed reasonably well, it obtained slightly lower accuracy, precision, recall, and F1-measure scores compared to the Multilayer Perceptron algorithm. The accuracy of 94.42% indicates that the Logistic Regression model accurately predicted the vote outcomes in the dataset, but with a slightly lower overall accuracy than the Multilayer Perceptron algorithm. The precision score of 94.5% suggests that the Logistic Regression algorithm correctly classified positive instances at a high rate, reducing false positives. The recall score of 94.3% indicates the model's ability to identify a significant proportion of actual positive instances. The F1-measure score of 94.4% reflects the harmonic mean between precision and recall, representing a balanced performance by the Logistic Regression algorithm.

Based on these results, it can be concluded that the Multilayer Perceptron algorithm outperformed Logistic Regression in terms of accuracy, precision, recall, and F1-measure for the vote dataset prediction task. The Multilayer Perceptron algorithm's ability to capture complex decision boundaries and its interpretability make it a suitable choice for this particular prediction task.

It is worth mentioning that the choice between algorithms should consider other factors such as computational efficiency, scalability, and interpretability. Multilayer Perceptron are known for their interpretability, as they provide clear decision rules and feature importance. On the other hand, Logistic Regression may have challenges with computational efficiency for large datasets, as it requires calculating distances between instances.

#### VI. CONCLUSION

In conclusion, the Multilayer Perceptron algorithm demonstrated superior performance for the vote dataset prediction, achieving higher accuracy, precision, recall, and F1-measure compared to the Logistic Regression algorithm. However, the final choice of algorithm should take into account the specific requirements of the prediction task, considering factors such as interpretability, computational efficiency, and the importance of precision versus recall in the context of the decision-making process.

The study concludes by summarizing the findings and insights gained from the comprehensive analysis of Multilayer Perceptron and Logistic Regression algorithms. It emphasizes the importance of selecting the appropriate algorithm based on

the specific requirements of the prediction task and dataset characteristics. The comparative analysis provides valuable guidance for researchers and practitioners in choosing the most suitable algorithm for accurate predictions in various domains.

### REFERENCES

- [1] G. Ravi Kumar, S. Rahamat Basha, Surya Bhupal Rao, “A Summarization on Text Mining Techniques for Information Extracting from Applications and Issues”, Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 324-332, ISSN:0973-8975.
- [2] G. Ravi Kumar, K. Tirupathaiah and B. Krishna Reddy, “Client Churn Prediction of Banking and fund industry utilizing Machine Learning Techniques”, International Journal of Computer Sciences and Engineering, Volume-7, Issue-6, e-ISSN: 2347 — 2693, PP: 871-875, June 2019,
- [3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.
- [4] J. Han and M. Kamber,” Data Mining concepts and Techniques”, the Morgan Kaufmann series in Data Management Systems, 2 nd ed. San Mateo, CA; Morgan Kaufmann, 2006.
- [5] Lam, Yee Hong Brian, Proteomic Classification of Liver Cancer using Artificial Neural Network”, 2005
- [6] Surya Bhupal Rao, S.Rahamat Basha, G. Ravi Kumar, “A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques”, Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 120-131, ISSN: 0973-8975.
- [7] UCI machine learning repository. <http://archive.ics.uci.edu/ml/>.
- [8] Yaacob, W.F.W.; Nasir, S.A.M.; Yaacob, W.F.W.; Sobri, N.M. Supervised data mining approach for predicting student performance. Indones. J. Electr. Eng. Comput. Sci. 2019, 16, 1584–1592.
- [9] Zohair, L.M.A. Prediction of Student’s performance by modelling small dataset size. Int. J. Educ. Technol. High. Educ. 2019.