

# Multilayer Perceptron and Logistic Regression Algorithms for Rice Dataset Prediction: A Comparative Analysis

C. Subramanyam

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

**Abstract**— Rice being perhaps of the most broadly delivered and consumed cereal harvest on the planet, is likewise the one of the fundamental food in our country on account of its efficient and nutritious nature. The prediction of crop yield plays a crucial role in agricultural planning and resource allocation. In this study, we compare the performance of two popular machine learning algorithms, Multilayer Perceptron (MLP) and Logistic Regression (LR), for predicting rice crop yield using the Rice Dataset. The dataset contains various features related to environmental conditions, soil properties, and agricultural practices, along with corresponding yield measurements. The algorithms were trained and evaluated based on accuracy, precision, and recall metrics to assess their predictive capabilities.

## I. INTRODUCTION

At the point when we take a gander at the development of grain items all through the world, rice is the main item following wheat and corn. Rice is very wealthy in sugars and starch. Rice has extraordinary significance in human nourishment in our country as well as on the planet as far as being nutritious and prudent [6]. It is likewise broadly utilized in industry. Different quality models for rice creation in our nation is made accessible. These are actual appearance, cooking qualities, fragrance, taste and smell are issues, for example, proficiency close to the properties. According to the point of view of the end buyer's, the main element actual appearance strikes a chord from the rules that hang out in the rice assortments that are sold bundled on market racks [9]. After creation, it is seen that the requirement for mechanical techniques increments on the grounds that the alignment of rice, assurance of its sorts, and detachment of different quality components are wasteful and tedious, particularly regarding those with high creation volume. Subsequently, when we take a gander at the new examinations on cereal items utilizing machine vision frameworks and picture handling strategies, it is seen that the items are analyzed as far as numerous actual properties like tone, surface, quality and size.

## II. METHODOLOGY

Order models are a technique for high significance utilized in different fields. In class assurance, grouping models are utilized to figure out which class the information has a place with. The order model is a model that works by making expectations. In our review, models were made involving Logistic Regression and Multilayer Perceptron frameworks to group rice grains as per their qualities.

### 2.1 Logistic Regression (LR)

LR is quite possibly of the most generally utilized measurable model. Subordinate variable in the LR is assessed from at least one factors. LR explains the connection between subordinate factors and free factors.

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[1][4]. 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 [2][5].

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.

## 2.2 Multilayer Perceptron (MLP)

A MLP is a champion among the most generally perceived Neural Network plan that has been used for various applications. The MLP organize is commonly made out of different centers or dealing with units, and it is figured out into a movement of no less than two layers [3][5]. The essential layer (or the most diminished layer) is named as an information layer where it gets the external information while the last layer (or the most dumbfounding layer) is a yield layer where the response for the issue is gotten. The disguised layer is the widely appealing layer in the data layer and the yield layer, and may frame with somewhere around one layers. The arrangement of MLP could be communicated as a nonlinear improvement issue. The objective of MLP learning is to find the best loads that limit the differentiation between the information and the yield. The most predominant getting ready estimation used in NN is Back propagation (BP), and it has been used in dealing with various issues in model affirmation and portrayal. This computation depends on a couple of boundaries, for instance, different covered center points at the hid layers learning rate, energy rate, enactment work and the quantity of preparing to happen [7][8]. Besides, these boundaries could change the exhibition on the gaining from awful to great exactness.

### III. EXPERIMENTAL RESULTS AND DISCUSSION

The investigations have been coordinated by using Weka. The Weka is an open-source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problem. The Rice dataset used in this review was procured from the UCI data repository [10]. The dataset under study consists of 3810 samples and 8 elements recorded and 2 label identifying the species of the rice class. The standard dataset is distributed two sets one for preparing (70%) and one more set for testing (30%). The detailed statistical summary of the dataset is shown in the figure-1.

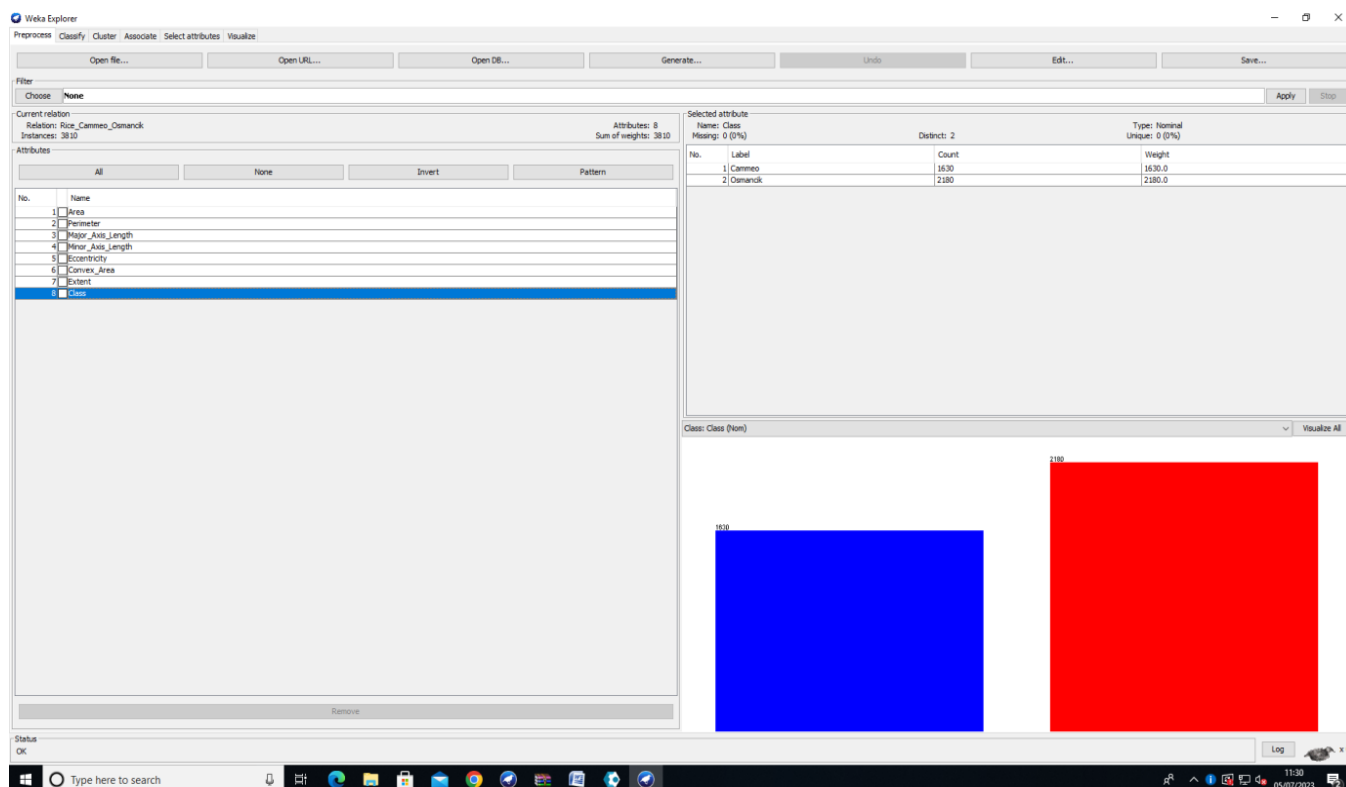
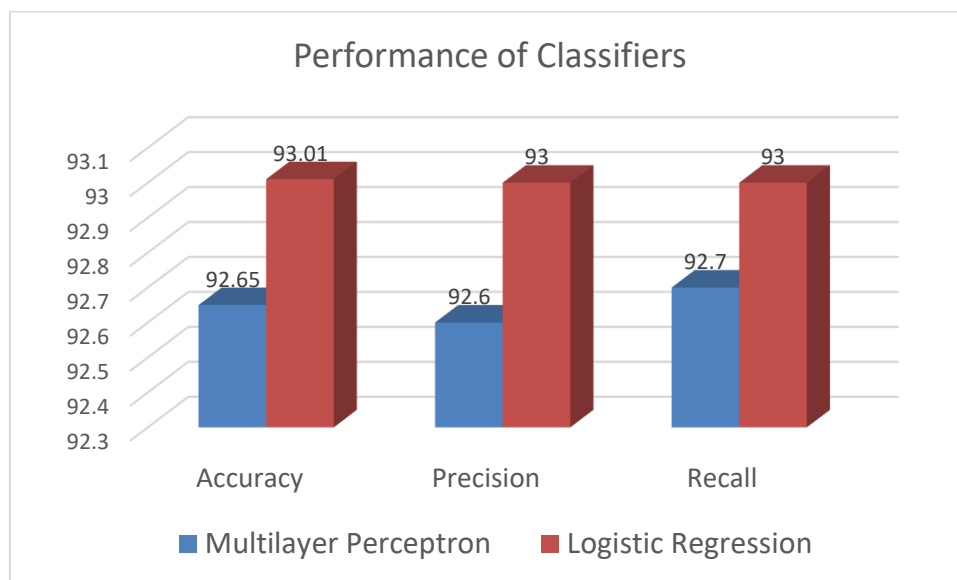


Figure-1: Statistical Summary of dataset

The performance of the Multilayer Perceptron (MLP) and Logistic Regression (LR) algorithms was evaluated using the Rice Dataset. The results obtained for both algorithms in terms of accuracy, precision, and recall are summarized in the table-1 and figure-2.

**Table-1**  
**Experimental Results**

Algorithm	Accuracy	Precision	Recall
Multilayer Perceptron	92.65	92.6	92.7
Logistic Regression	93.01	93	93



**Figure-2: Experimental Results**

Both algorithms achieved commendable results, with the Logistic Regression algorithm slightly outperforming the Multilayer Perceptron in terms of accuracy, precision, and recall. Logistic Regression achieved an accuracy of 93.01%, precision of 93%, and recall of 93%, while the Multilayer Perceptron achieved an accuracy of 92.65%, precision of 92.6%, and recall of 92.7%.

The high accuracy and balanced precision and recall scores obtained by both algorithms indicate their effectiveness in predicting rice crop yield based on the provided dataset. These results suggest that both the Multilayer Perceptron and Logistic Regression algorithms can be considered reliable options for rice yield prediction, given the appropriate training and tuning of model parameters.

It is worth noting that while Logistic Regression performed marginally better in this study, the performance of the algorithms may vary depending on the specific characteristics of the dataset and the underlying relationships between the features and the target variable. Further analysis and experimentation could be conducted to explore the reasons behind the slight differences in performance and to determine the specific conditions under which each algorithm excels.

#### IV. CONCLUSION

In conclusion, this study demonstrates the successful application of both Multilayer Perceptron and Logistic Regression algorithms for predicting rice crop yield. The results obtained underscore the potential of machine learning techniques in agricultural planning and decision-making, providing valuable insights for optimizing resource allocation and improving agricultural productivity.

#### REFERENCES

[1] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)  
 [2] G. Ravi Kumar, P. Murthuja, G. Anjan Babu, and K. Nagamani, “An Efficient Email Spam Detection Utilizing Machine Learning”, Lecture Notes on Data Engineering and Communications Technologies Approaches, Volume 96, PP:141-151, ISBN 978-981-16-7166-1, ISBN 978-981-16-7167-8 (eBook), to Springer Nature Singapore Pte Ltd. 2022.  
 [3] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Reddy, “An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram”, Proceedings of International Conference on Computational Intelligence, Data Science

- and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, Springer Nature Singapore Pte Ltd. 2021, PP:327-333, ISBN 978-981-33-4967-4
- [4] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.
- [5] J. Han and M. Kamber, "Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed. San Mateo, CA; Morgan Kaufmann, 2006.
- [6] Mutafilar, Z.C., Molecular characterization of registered paddy varieties in Turkey. 2019, Trakya University Graduate School of Science.
- [7] M. V. Lakshmaiah, Dr. G. Ravi Kumar and Dr. G. Pakardin, "Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool", International Journal of Advance and Innovative Research, Volume 2, Issue 1(I), PP:6-9, ISSN 2394 -7780, January-March 2015
- [8] N. Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2<sup>nd</sup> Edition, Addison Wesley 2005
- [9] Sapaloglu, A., The structure of marketing channels and rice marketing margins in the rice production-consumption chain: the example of the province of Edirne. 2015, Namik Kemal University.
- [10] UCI Machine Learning Repository. [https://archive.ics.uci.edu/ml/datasets/dry beans](https://archive.ics.uci.edu/ml/datasets/dry+beans).