A Comparative Study on Logistic Regression and Neural Network Strategies

Nigella Shambhavi[1], Dr. M. Sreedevi[2]

[1]PG Student, Department of Computer Science, Sri Venkateswara University, Tirupati
[2]Assistant Professor, Department of Computer Science, Sri Venkateswara University, Tirupati

Abstract— *The motivation driving this work is to study the presentation of man-made intelligence frameworks on Vowel classifications guess utilizing Multi-facet Perceptron and K-Closest Neighbor estimations. The introduction of the assessments is studied through after execution assessments: accuracy, precision and review. The best outcome among two calculations for by and large accuracy rate was refined by Logistic Regression model with a speed of 99.2%. The proposed model is studied utilizing Vowel classifications structure UCI dataset educational documents. It is clear from the outcomes that the model has performed very well in anticipating high reality.*

I. INTRODUCTION

The target of combination learning is to help a model that separates the data into the different classes, thoroughly reason on referencing new models later on. Get-together learning frameworks rather produce different models. Given another model, the association passes it to the sum of its different base models, gets their doubts, and at some point later obliges them in some fitting manner (e.g., averaging or projecting a looking over structure). The greater part of outfit learning methodologies is customary, material across wide classes of model sorts and learning endeavors. Association learning is a sensible structure that has dynamically been embraced to join different learning evaluations to besides empower figure accuracy, by and large, [3]. Perhaps the most amazing spaces of assessment in managed simulated intelligence have been to examine techniques for making unprecedented outfits of students. The key revelation is that outfits are a critical piece of the time stunningly more distinct than the singular students [5]. While arranging an association learning procedure, correspondingly as picking the framework by which to accomplish assortment in the base models and picking the joining technique, one essential to pick the sort of base model and base model learning evaluation to use. The joining framework might restrict such base models that can be used.

Data burrowing coordinates searching for charming models or data from colossal data. It changes a huge social event of data into data. Data mining is an imperative improvement during the time spent data openness. The data mining has turned into an intriguing mechanical party regarding evaluating data as indicated by substitute perspective and changing over it into huge and fundamental information [6]. Data digging has been by and large applied in the space of clinical finding, Interruption ID structure, Training, Banking, Extortion divulgence. Get-together is an organized learning. Measure and diagram in data mining are two kinds of data evaluation task that is used to bind models portraying data classes or to expect future data plans. Portrayal measure has two phases; the first is the learning connection where the orchestrating educational records are destroyed by friendly event evaluation. The learned model or classifier is presented as plan rules or models. The accompanying stage is the use of model for social event, and test instructive assortments are used to concentrate on the exactness of portrayal rules.

II. CLASSIFICATION

Approach is the way toward finding a model or an end that portrays and sees information classes and contemplations, to utilize the model to expect the classes of things whose class mark isn't known. Information deals can be seen as a two-stage measure: learning step in which a classifier is made portraying a predestined blueprint of classes or experiences by disengaging the status set contained illuminating summary tuples and their connected names [4][5]. In the subsequent progression model is utilized for demand by first assessing the sensible exactness of classifier worked during the critical development. It is finished utilizing the test information. The precision of classifier on a given test set tuples is level of tuples that are successfully referred to by the classifier. On the off chance that the exactness is over some adequate level, the classifier can be utilized to expect future tuples whose class mark isn't known.

Depiction is a sort of information assessment that can be utilized to make models depicting immense information classes. Framework is an information mining approach used to foresee pack pay for information models. It is one of the crucial designs in information mining and is utilized in different applications, for example, plan confirmation, trouble attestation, client relationship the trailblazers, and administered appearance. The objective of the depiction examinations is to gather a model

from a gigantic heap of preparing information whose target class names are known and subsequently this model is utilized to pack covered cases [6] [8].

Plan is the most customary and most prestigious information mining strategies. Framework maps information into predefined social gatherings or classes. It is common proposed as controlled getting the hang of contemplating how the classes are settled going preceding looking at the information. Procedure is the way toward finding a model that sees information classes, to utilize the model to expect the class of things whose class name is dull. The picked model depends on the assessment of an immense heap of arranging information. Enlightening assortments are rich with disguised data that can be utilized for cautious dynamic.

Building unequivocal and significant classifiers for monstrous information bases is one of the crucial undertakings of information mining and man-made intelligence research. Building valuable mentioning frameworks is one of the focal assignments of information mining.

III. METHODOLOGY

This fragment gives the compact thought about picked managed models of Logistic Regressionand Multilayer Perceptron.

3.1 Artificial Neural Network (ANN)

ANN mirrors the limits and activities of the psyche of individual which is perceived as the centers, which is really known as or called fake neurons [11]. The neurons give and convey data and information among themselves in kind of 0 s and 1 s or blend and each neuron has a specific weight given to it, which shows its abilities and tasks to complete in the system [5]. The development of ANN is divided into layers, straightforwardly from data gathering layer, input layer, focus or mystery layer to yield layer which is called extraction or gathering layer. Each layer has a specific ability to perform and change data into the critical information to get an authoritative and optimal result. The Activation and move work expect an essential part in the activities do by neurons.

3.1.1 Multilayer Perceptron (MLP)

A MLP is a saint among the most overall saw Neural Network plan that has been used for various applications. The MLP figure out is commonly made out of different concentrations or overseeing units, and it is figured out into an advancement of no under two layers [7]. The central layer (or the most reduced layer) is named as an information layer where it gets the external information while the last layer (or the most amazing layer) is a yield layer where the response for the issue is gotten. The secret layer is the comprehensively captivating layer in the data layer and the yield layer, and may diagram with some place almost one layers. The arrangement of MLP could be granted as a nonlinear improvement issue. The objective of MLP learning is to find the best loads that limit the division between the information and the yield. The most pervasive getting ready appraisal used in NN is Back inciting (BP), and it has been used in overseeing various issues in model attestation and portrayal. This computation depends a few cutoff points, for instance, remarkable covered center concentrations at the disguised layers learning rate, energy rate, approval work and the measure of wanting to happen.

3.2 Logistic Regression

Logistic Regression at times called the calculated model or logit model, dissects the connection between numerous free factors and an all out subordinate variable, and evaluations the likelihood of event of an occasion by fitting information to a strategic bend. There are two models of strategic relapse, parallel calculated relapse and multinomial strategic relapse. Twofold strategic relapse is regularly utilized when the reliant variable is dichotomous and the free factors are either constant or straight out. At the point when the reliant variable isn't dichotomous and is involved beyond what two classifications, a multinomial calculated relapse can be utilized.

In a Logistic Regression, the reliant variable just has two classifications. For the most part, the event of the occasion is coded as 1 and its nonappearance as 0. Remembering that codification changes the coefficients' sign and, in this way, their considerable understanding. To all the more likely comprehend how a calculated relapse functions, understanding the rationale of relapse investigation as a whole is important.

Let's look at the linear model's classic notation: $Y = \boldsymbol{\alpha} + \boldsymbol{\beta}X + \boldsymbol{\varepsilon}$

Y represents the dependent variable, that is, what we are trying to understand/explain/predict. X represents the independent variable. The intercept, ($\boldsymbol{\alpha}$), represents the value of Y when X equals zero. The regression coefficient, ($\boldsymbol{\beta}$), represents the

variation observed in Y associated with the increase of one unit of X. The stochastic term, ($\boldsymbol{\varepsilon}$), represents the error of the model. Technically, it is possible to estimate if there is a linear relationship between a dependent variable (Y) and different independent variables. Moreover, the model allows the observation of the effect magnitude and to test the coefficients' statistical significance (p-value and confidence intervals).

IV. EXPERIMENTAL RESULTS

In this work, a true Vowels assessment data set was taken from the UCI storehouse of AI data set [9]. It contains 990 examples, 14 attributes and grouped into 11 classes, there is no missing worth in the dataset. The class insightful circulation of names is displayed in the figure-1.

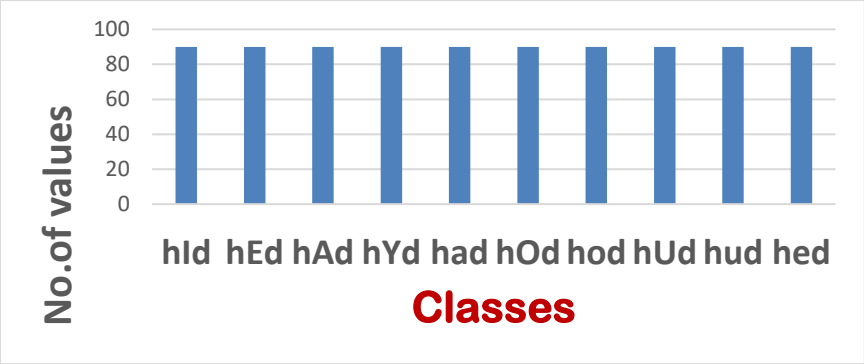

Figure-1: Class distribution

We have utilized the weka to explore our proposed calculations. Weka is a state-of-the-art office for making ML techniques and their application to genuine data mining issues. The Vowel dataset attribute wise and detailed statistical summary as shown in the figure-2.

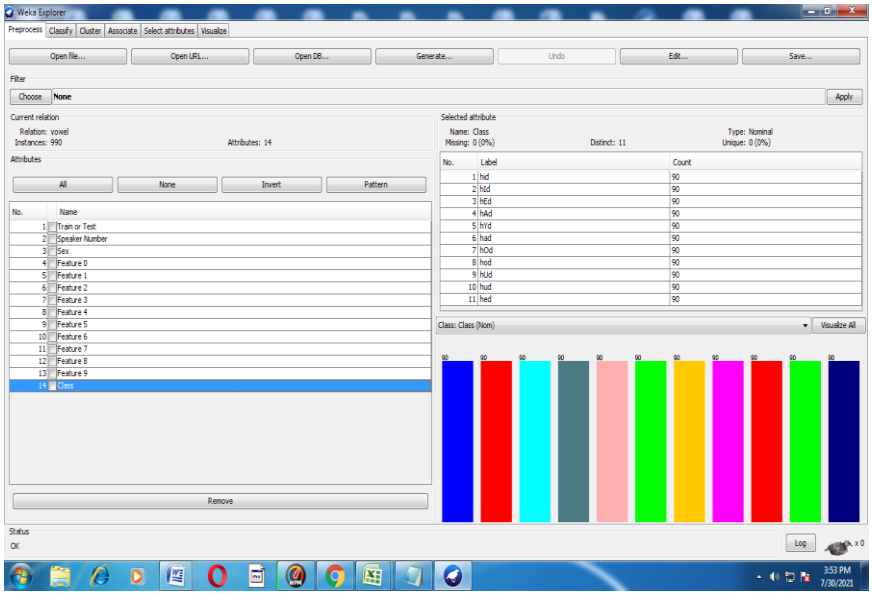

Figure-2:Vowel dataset information

4.1 Results

The dataset is separated in two sets. The planning set is 70% and the remaining 30% are used for testing. We have used the Weka simulation to investigate two supervised ML course of action estimations. We survey our two models using assorted execution estimations like Accuracy, Precision and Recall, the Experimental results are showed up in the table-1 and same showed up in the Figure-3.

TABLE 1
PERFORMANCE OF CLASSIFIERS

Algorithm	Accuracy	Precision	Recall
Multilayer Perceptron (MLP)	94.5	94	94
Logistic Regression	98.7	98	98

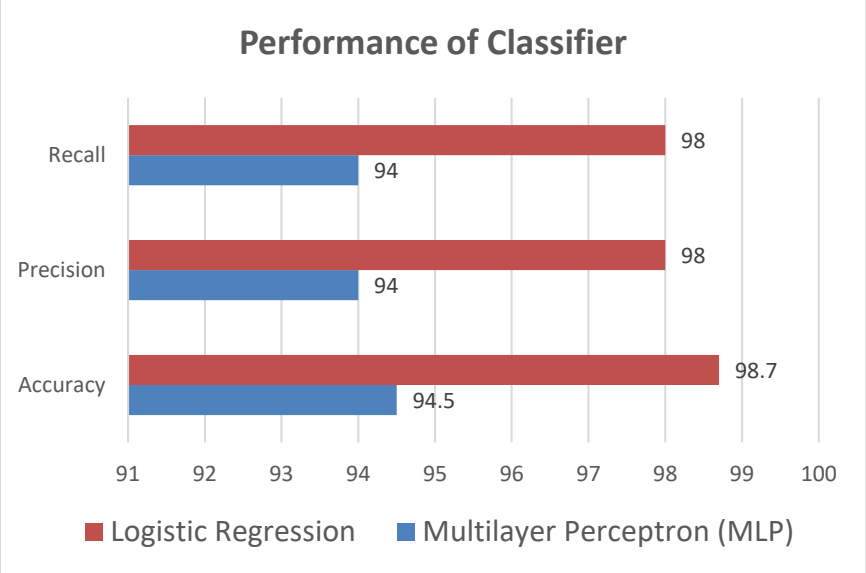

Figure-3: Experimental Results

We find in the Figure-3, the introduction of the Logistic Regression estimation has accomplished 98.7% precision and MLP has achieved 94.5%, As the result from assessment among the two computations, we find that most vital precision of Classification model is Logistic Regression (98.7%). So, the Logistic Regression algorithm have got highest accuracy, with a 4.2% difference when compared to MLP algorithm.

V. CONCLUSION

In this assessment, we will explore the presentation of Calculated Relapse for gathering Vowel class data by using MLP and Strategic Relapse in various applications like request, understanding and decisive reasoning limit learning. Our experimental outcomes show that Calculated Relapse has achieved most vital precision on Vowel dataset when appeared differently in relation to MLP. We basically tried to audit the turn out accomplished for exactness improvement and execution improvement of Calculated Relapse. This assessment which is introduced will fill in commonly representing seeking after future examination connected with self-confident backwoods classifier.

REFERENCES

[1] Baoli, L., Shiwen, Y. & Qin, L. (2003) "An Improved k-Nearest Neighbor Algorithm for Text Categorization, ArXiv Computer Science e-prints
[2] Bermejo, T. & Cabestany, J. (2000) "Adaptive soft k-Nearest Neighbor classifiers", Pattern Recognition, 33: 1999-2005
[3] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)
[4] G. Ravi Kumar, Venkata Sheshanna Kongara & Dr. G. A. Ramachandra, "An Efficient Ensemble Based Classification Techniques for Medical Diagnosis", International Journal of Latest Technology in Engineering, Management and Applied Sciences, Volume II, Issue VIII, Pages: 5-9, ISSN-2278-2540, August-2013
[5] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.
[6] 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.
[7] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems",2nd edition, Addison Wesley, 2005.
[8] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.
[9] UCI machine learning repository. http://archive.ics.uci. edu/ml/