A Performance measure of KNN and MLP Supervised Study on Classification Algorithms

Sr Vinodh[1], Dr. G.V. Ramesh Babu[2]

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

Abstract— *Data mining is that the method of analyzing data from completely different views and summarizing it into useful information. Classification could be a data processing technique supported machine learning which is employed to classify each item in a set of data into a group of predefined categories or teams. Classification is method of generalizing the data consistent according to different instances. Classification algorithms as an important technology in data mining and machine learning have been widely studied and applied. Many methods can be used to build classifiers, such as the decision tree, Bayesian method, instance-based learning, artificial neural network and support vector machine. This paper focuses on the classification methods based on MLP learning and KNN, arthiyapia data set was used for the classification with 452 instances with 280 attributes as independent variable and one as dependent variable for the analysis. The results show that MLP found to be the algorithm with most precision and accuracy when compared to KNN algorithm.*

I. INTRODUCTION

Data mining is a technology that offers extracting or discovering new relations, hidden knowledge and important patterns from such data. It is also known as Knowledge Discovery in Databases (KDD). Data mining technique is important for analysis purpose. Data mining supports different techniques such as classification, clustering, association rule mining, outlier analysis etc [1][4]. Data Mining(DM) discovers hidden relationships in data, in fact it is a part of wider process called "knowledge discovery". Knowledge discovery describes the phases which must be done to ensure reaching meaningful results through research. The objective of DM process is to obtain information out of a dataset and converts it into a comprehensible outline. An understanding of algorithms is combined with detailed knowledge of the dataset an understanding of algorithms is combined with detailed knowledge of the datasets. Data mining must afford very complex and different situations to reach quality solutions. Therefore, data mining is a research field where many advances are being done to accommodate and solves emerging problems [1]. For present study purpose classification technique is investigated.

II. CLASSIFICATION

Classification plays an important role in data mining and machine learning. The purpose of classification algorithm is to construct a classifier, and then analyzes the characteristics of the unknown data to get an accurate model. The performance of the classifier is measured by its classification accuracy. Building effective classification systems is one of the central tasks of data mining. The main purpose of supervised learning is to build a simple and unambiguous model of the allocation of class labels in terms of predictor features [2][7]. The classifiers are then used to classify class labels of the testing instances where the values of the predictor features are known, to the value of the class label which is unknown [3][5].Classification of this tremendous amount of data is time consuming and utilizes excessive computational effort, which may not be appropriate for many applications.

III. METHODOLOGY

Many different types of classification techniques have been proposed in literature that includes Decision Trees, Naïve Bayesian methods, Neural Networks, Logistic Regression , SVM and KNN etc. In this paper, we evaluate the performance of the MLP algorithms on arthiyapia data set was used for the classification compared with the KNN algorithm.

3.1 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 [6]. 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. Besides, these boundaries could change the exhibition on the gaining from awful to great exactness [2].

3.2 K-Nearest-Neighbors (KNN)

The K-Nearest-Neighbors (KNN) is a non-parametric gathering technique, which is essential anyway incredible all around [1]. The essential thought for k-NN depends after determining the distances between the attempted, and the readiness data tests to recognize its nearest neighbors. The attempted model is then consigned to the class of its nearest neighbor [2].

The K-Nearest-Neighbors (KNN) is a clear anyway convincing procedure for game plan. The KNN estimation is a procedure for gathering objects reliant upon closest planning models in the part space. KNN is a kind of event based learning, or aloof acknowledging where the limit is simply approximated locally and all computation is yielded until gathering [6]

For a data record D to be requested, its K nearest neighbors is recuperated, and these constructions a neighborhood of D. Bigger part projecting a voting form among the data records in the space is by and large used to pick the request for D with or without considered distance-based weighting. Regardless, to apply KNN we need to pick a reasonable motivating force for K, and the accomplishment of collection is a great deal of wards on this value. The critical drawbacks in regards to KNN are (1) its low efficiency - being a slow learning methodology denies it in various applications, for instance, dynamic web burrowing for an enormous vault, and (2) its dependence on the decision of an "incredible worth" for K.

IV. EXPERIMENTAL RESULTS

The analyses have been directed by utilizing R programming Language. R is a sophisticated statistical software package, which provides new approaches to data mining., it is an open-source tool for analysis of data mining algorithms. The R Language is a bundle for information characterization, grouping and representation. We have considered the Pima diabetes from the UCI Machine Learning Repository datasets for assessing the productivity and adequacy of MLP calculation [8]. The characteristic data information is consolidated in Table-1. The standard dataset is parceled into two sets one for training (80%) and another set for testing (20%).

TABLE 1
DATASET INFORMATION

S. No	Name of the Dataset	No. of Attributes	No. of Instances	No. of Classes
1	arthiyapia	280	452	16

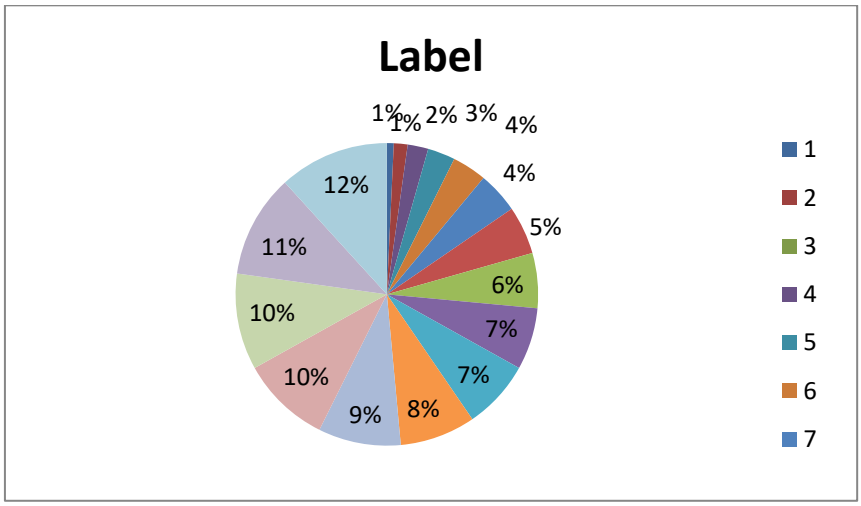

Figure-1: Class distribution

We survey our two models using assorted execution estimations like Accuracy, Precision and Recall, the Experimental results are showed up in the table-2 and same showed up in the Figure-2

TABLE 2
PERFORMANCE OF CLASSIFIERS

Algorithm	Accuracy	Precision	Recall
MLP	90.7	90	89
KNN	89.2	88	89

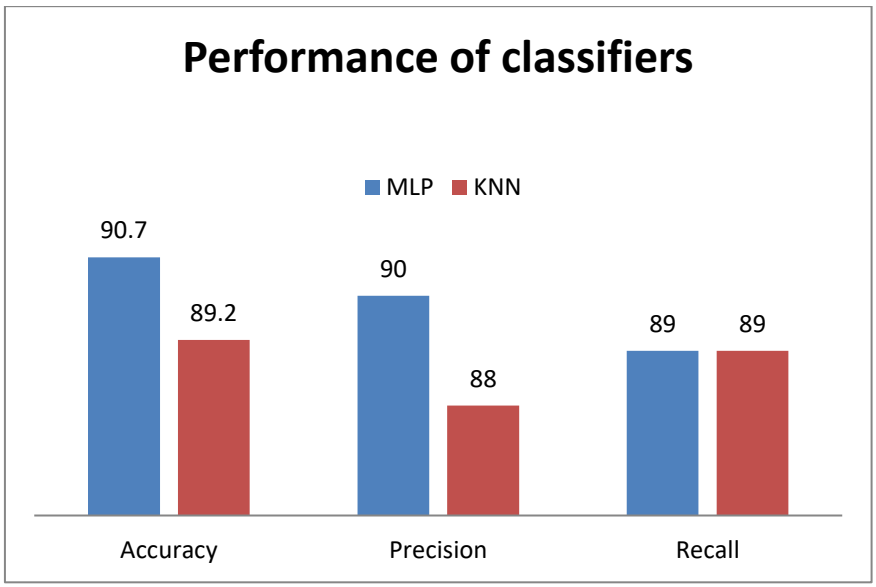

Figure-2: Experimental Results

We find in the Figure-2, the introduction of the MLP estimation has accomplished 90.7% precision and KNN has achieved 89.2%, As the result from assessment among the two computations, we find that most vital precision of Classification model is MLP (90.7%). So, the MLP algorithm have got highest accuracy, with a 1.5% difference when compared to KNN algorithm.

V. CONCLUSION

In this paper, the accuracy of two classification techniques is evaluated based on the selected classifier algorithm. An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for arthiyapia dataset. The performance of MLP shows the high level compare with KNN classifiers. Hence MLP shows the concrete results with arthiyapia data better results with accuracy, low error rate and performance.

REFERENCES

[1] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)
[2] G. Ravi Kumar and K. Nagamani, "Banknote Authentication System utilizing Deep Neural Network with PCA and LDA Machine Learning Techniques", International Journal of Recent Scientific Research, ISSN: 0976-3031, Volume 9, Issue 12(D), PP:30036-30038, 2018
[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, 2nd ed. San Mateo, CA; Morgan Kaufmann, 2006.
[5] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems",2nd edition, Addison Wesley, 2005.
[6] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.
[7] S.Rahamat Basha and G.Ravi Kumar Surya Bhupal RaoG,"A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques", Journal of Mechanics of Continua and Mathematical Sciences, ISSN: 2454 -7190, Special Issue, No.-5,PP:120-131, 2020
[8] UCI machine learning repository. http://archive.ics.uci. edu/ml/