

# Performance Evaluation of Binary and Multi-Class Dataset using Ensemble Classifiers

Anupama Jha<sup>1</sup>, Dr. Meenu Dave<sup>2</sup>, Dr. Supriya Madan<sup>3</sup>

<sup>1</sup>Research Scholar, Jagannath University, Jaipur, India

<sup>2</sup>Professor, Jagannath University, Jaipur, India

<sup>3</sup>Professor, VIPS, GGSIPU, New Delhi, India

**Abstract:** Now a day's digital data is dominating the entire globe. Every day it is increasing as it is generated from various domains such as social media, healthcare, education, banking etc. as well as through smart devices, IoT Devices etc., which we call these days Big Data. Due to the availability of Big Data, ensemble machine learning methods represent an attractive approach that can be used to deal with mining large and complex datasets. Ensemble models are now standard in Big Data Mining due to the fact that combining multiple classifiers together on a large dataset can often produce a much powerful classifier. The main principle behind this approach is that when weak classifiers are correctly combined, we can get better results.

With this research paper, we have made a noble attempt to compare the performance of proposed methodology with existing research study using the various data mining classifiers. This study also compares the performance of basic data mining classifiers with the ensemble classifiers to solve the two major classes of classification problems: Binary-Class and Multi-Class in terms of accuracy. Ensemble approaches are implemented here to improve the performance of simple models and reduce overfitting of more complex models.

The experimental results show that for the Multi-class classification task, Bagging performs well in comparison of Binary-class. But for Binary Class dataset, it is found that in most of the models, basic classifiers perform better than the ensemble classifiers. Moreover, it is observed that the Bagging performs well for all types of training-testing splits of the datasets.

**Keywords:-** Big Data Mining, Classification, Ensemble method, Bagging, Boosting

## 1. INTRODUCTION

The main objective of Big Data mining is to uncover hidden insight from the large volume dataset that can be useful for many organizations to make better decision [1]. In recent year, it has attracted more and more attention due to the fact that it has been successfully applied to many domains such as Data science, Big Data Analytics, Business Intelligence, WWW, Sentiment Analysis [2] etc. In data mining, classification approach is considered to be the most important data mining approach as it becoming a fascinating topic to the researchers that precisely and effectively describes data.

In this research work, we have made an extension to improve the performance by using the ensemble learning methods. In view of that, the performance of 5 basic classifiers Decision Tree (CART and CTREE), Random Forest, Support Vector

SVM and k-NN of data mining approaches are compared with the novel ensemble learning approach Bagging and Boosting for the classification tasks. The analysis is implemented on two different datasets i.e. Binary class and Multi-class. All such classifiers have been modelled with different training-testing partitions to find out the best classifier in terms of accuracy.

## 2. RELATED WORK

It is found that there are number of techniques that are used to analyse large volume datasets are not very efficient for performing the tasks as some of them are fast but they had to compromise with the accuracy [3]. Some techniques result in good accuracy but took more execution time. In 2013, the researchers analysed 14 different classification algorithms and found no one classifiers outperformed all others in terms of the accuracy when applied to the number of datasets [4]. Researchers also highlighted that there are no classifiers available in the literature that can classify binary, multi-class and multi-label classification at the same time [5]. They proposed a novel online universal classifier based on an extreme learning machine and found that the performance was almost uniform in datasets of all classification types.

Authors Seyed Hossein Nourzad and Anu Pradhan presented 2 ensemble methods Bagging and AdaBoost for binary and multi-class classification to improve the accuracy and they were able to achieve for the binary classification, the accuracy up to 98.9% and for the multi-class classification, the accuracy is 94.6% [6]. Overall, the performance of ensemble models was found higher than that of base classifiers [7] [8].

Dewiani, Armin Lawi et al. in 2019 proposed a combined technique of Ensemble Bagging and Support Vector Machine (SVM) to improve single classification performance to detect fraud in a firm. They achieved the highest accuracy of 89.95% [9].

In 2020, Authors Isaac Kofi Nti et al. provide a wide-ranging comparison of various ensemble methods such as bagging, boosting, stacking, and blending for predicting stock-market indices. It was found that although stacking and blending achieved higher accuracy but due to their higher training and testing time, they are computationally expensive as compared to Decision Tree by boosting and bagging [10].