Comparison of Binary Class and Multi-Class Classifier Using Different Data Mining Classification Techniques

Anupama Jha[a], Meenu Dave[b], Supriya Madan[c]

[a]*Research Scholar, Jagan Nath University, Jaipur, Rajasthan*
[b]*Jagan Nath University, Jaipur, Rajasthan, India,*
[c]*GGSIP University, Delhi, India,*

ARTICLE INFO

Article history:
Received 01 March 19
Received in revised form 08 March19
Accepted 08 April 19

Keywords:
Data Mining
Binary class
Multi-class
Classification techniques

ABSTRACT

Data mining (DM) is the process of retrieving information from huge datasets and transforming them into meaningful decision. Classification technique is considered to be the most important data mining techniques as it becoming an enthralling topic to the scholars that precisely and effectively describes data for the knowledge- discovery. It is used to describe and distinguish data classes or concepts. There are two major classes of classification problems: Binary-class and Multi-class. In Binary-class classifications, the given dataset is categorized into two classes whereas in Multi-class classification, the given dataset is categorized into several classes based on the classification rules.

This paper explores several DM classification approaches such as Decision tree like Classification and Regression Tree(CART) and Conditional Inference Tree(CTREE), Random Forest(RF), Support Vector Machine(SVM) and k-Nearest-Neighbour(KNN) to enhance the result of binary class and multi-class classifiers using the powerful Big data mining analytical tool R and RStudio. Various measures such as Accuracy, F-Score, Sensitivity etc. are used to evaluate the classifier's performance and also predict which classifier will perform better when the training-testing datasets are analysed with multiple partitions (%).

Peer review under responsibility of International Conference on Advancements in Computing & Management.

1. Introduction

Data mining started to be a motivating target for many organizations due to the presence of high-volume data that contain useful insight (Han &Kamber , 2000). It has attracted more and more attention in recent years, probably because it has been successfully applied to many domains such as Big Data Analytics (Jha, Dave & Madan, 2016), Business Intelligence, Web Search, Scientific Discovery, Digital Libraries, Text Mining (Jha, Dave & Madan, 2018)etc. The overall objective of the data mining process is to extract information from such dataset and transform them into an understandable structure by using some intelligent methods.

One of the fundamental and important data analysis tasks is Classification. It is a process of finding a model that describes and distinguishes data classes and concepts. The classification task is divided into two categories: Binary classification and Multi-class classification. In binary classification, the given dataset is categorized into two distinct classes (predicting which class each one belongs to) whereas in multi-class classification, the dataset is categorized into multiple classes on the basis of a classification rule.

2. Literature Review

This section of the research paper presents literature review from the various journals, researchers and other internet sources. It is to be found that after review of binary and multi-class classification task, there are models available in the literature that can classify such tasks either on Binary Class or Multi-class datasets or both but not based on multiple training-testing partition.