ABSTRACT:

Machine Learning is a branch of Artificial Intelligence, which enables computers to analyse the data and learn without any explicit programming. Every machine learning system require some training data, which is used by the model to learn and develop relationship between input and output variables. Accuracy of the model largely depends upon the type, quality and quantity of data used for training the model. In this paper we have studied the amount of training data required to train the models effectively. We have used two multivariate datasets having continuous and binary dependent output variables respectively. On these datasets, we implemented algorithms like *Random Forest, Decision Tree, KNN, Logistic Regression and Linear Regression* and tested accuracy by varying the sizes of training data. We then compared the accuracies obtained from different amount of training data to support our research.