

|  |
| --- |
| DATA SCIENCE APP |
|  |
| Vidura Bandara Wijekoon  BSc. Electrical and Information Engineering  University Of Ruhuna |



# Abstract

|  |
| --- |
| In here, I am going to implement a user application based on data – science visualization and analysis. Machine learning algorithms can be used to find the patterns in data can be mainly focused and Pattern Recognition, Classification, Supervised learning, Artificial Intelligence are features of this project. This will be a user handy application in future. |
| “Big data is at the foundation of all the megatrends that are happening.” – By Chris Lynch, American Writer of Books |
|  |

Introduction:

Machine Learning is the study of a list of sub-problems, such as decision making, clustering, classification, forecasting, deep-learning, inductive logic programming, support vector machines, reinforcement learning, similarity and metric learning, genetic algorithms, sparse dictionary learning, etc. Supervised learning, or classification is the machine learning task of inferring a function from a labeled data.

In Supervised learning, we have a training set, and a test set. The training and test set consists of a set of examples consisting of input and output vectors, and the goal of the supervised learning algorithm is to infer a function that maps the input vector to the output vector with minimal error.

There is no single algorithm that works for all cases so there are many such as Neural Networks, Decision Trees, Support Vector Machines, Random Forest, Naïve Bayes Classifier, Bayes Net, Majority Classifier.

I am using random forest generator and

A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled with the max\_samples parameter if bootstrap=True (default), otherwise the whole dataset is used to build each tree.[1]

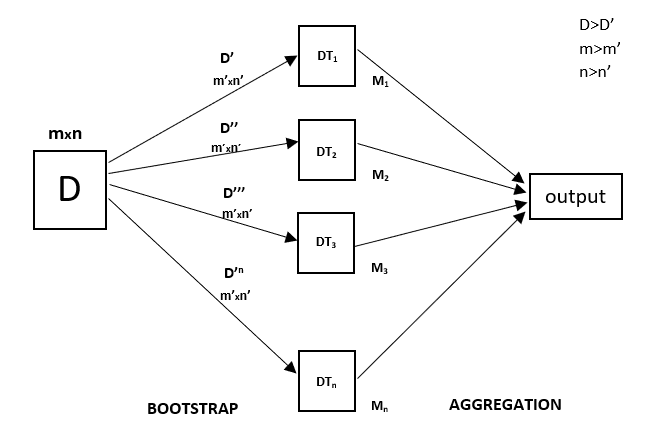


Figure 1: Random Forest Regression Neural Network

Timeline

|  |  |
| --- | --- |
| Period | Task |
| 2020-12-1 - 2020-12-14 | Gaining Resources |
| 2020-12-21 - 2020-12-28 | Building model |
| 2020-12-28 - 2021-01-14 | GUI |
| 2021-01-14 - 2021-01-31 | Further development |
| 2021-01-31 - 2021-02-14 | Further development |
| 2021-02-14 - 2021-02-28 | Testing |
| 2021-02-28 - 2021-03-15 | Re-testing and finalization |