International Journal of Scientific Research & Engineering Trends
Volume 5, Issue 2, Mar-Apr-2019, ISSN (Online): 2395-566X

Analysis of Classification Algorithms using Machine Learning

M. Tech. Scholar Soniya Rathore Asst. Prof. Ankur Taneja Asst. Prof. Lakhan Singh Sisodiya
Department of Computer Science & Engineering
SAM College of Engg. & Tech.
Bhopal, M.P. , India
soniya.rathore39@gmail.com

Abstract- In this work our main focus is on regression which is one of the most important methods in machine learning algorithm. Regression is a statistical approach that is used to find the relationship between variables. It is basically used to predict the outcome from the given dataset. In this work we will discuss the regression algorithms which are available in machine learning algorithm and propose one algorithm that will have less train error and test error as compared to other existing algorithm. The accuracy measure will be in the form of train and test error.

Keywords- **Classification, Data Mining, Linear Regression, Machine Learning techniques, python.**

I. INTRODUCTION

Machine learning systems itself grasp programs or plan from data. This is generally a very impressive alternative to making or substitute constructing them and in the last some past years the utilizing of machine learning has increase rapidly in computer science. Machine learning is used in Web search i.e Query search, Network filters, recommending in many systems, for placing ad, To find-out credit scoring, fraud detection, In stock trading, drug design in medical fields, and many other applications.

A recent report from the many big and Global Institute like McKinsey asserts that machine learning (a.k.a. data mining or find-out future analysis) will be the next generation technology for society and market where we are keeping abundant amount of data [16]. So many machine learning projects extends their time to process the given data to give better results in many domains. By developing this technology knowledge is fairly easy to communicate for business requirement. In Machine Leaning major component is given below.

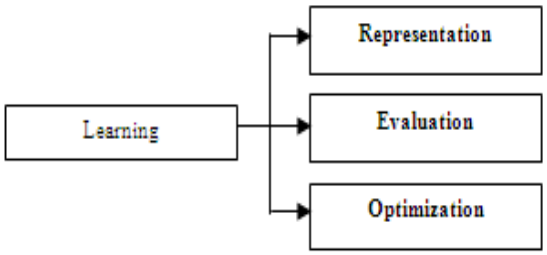

Fig.1 Evaluation of Machine Learning.

1. Representation- A classifier can represent in such manner (means a definite language) so that a computer can understand easily. If a classifier is not in the hypothesis space, it cannot be learned. Selection of classifier for any individual's problem is important.

2. Evaluation- It is like function which decides which classifier is bad and which one is good. This is also called objective function.

3. Optimization- To finding optimization technique is very important and is key to the learner, and also helps to find-out the classifier.

4. Classification of Machine Learning

There are 3 branches of machine learning we can understand this classification in details with sketch diagram.

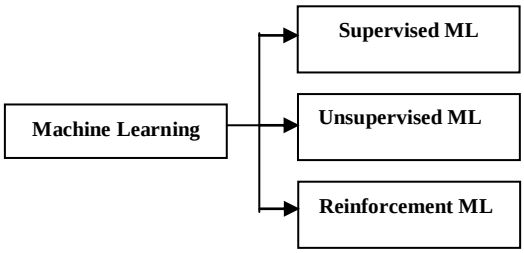

Fig.2 Classification of Machine Learning

5. Supervised Learning- In supervised machine learning; a system is trained with data that has been labeled. The labels categories each data point into one or more groups, such as 'apples' or 'oranges'. The system learns how this data known as training data is structured, and uses this to predict the categories of new or 'test' data.

© 2019 IJSRET

6. Unsupervised Learning- In this learning, learning is without labels. It aims to detect the characteristics that make data points more or less similar to each other, for example by creating clusters and assigning data to these clusters.

7. Reinforcement Learning- In this learning focuses on learning from experience, and lies between unsupervised and supervised learning. In a typical reinforcement learning setting, an agent interacts with its environment, and is given a reward function that it tries to optimize, for example the system might be rewarded for winning a game. The goal of the agent is to learn the consequences of its decisions, such as which moves were important in winning a game, and to use this learning to find strategies that maximize its rewards.

8. Machine Learning in Daily Life- Machine learning is using by us in day to day life in various form out of few names is given below.

- Online shopping.
- Customer support.
- Virtual Assistants.
- Email Spam and Malware Filtering.

II. REVIEW OF LITERATURE

In this papers Author's Explained How Researchers have already explored advanced regression techniques for forest variables estimation and recent literature provides examples which show their suitability in comparison with MLR, Neural networks, SVM and decision trees are the admired schemes for classification. In this paper [3] three techniques are compared by applying ML techniques on KDD CUP'99 data set. The techniques are supposed to be good for identifying the anomalies detection, but the performance may differ in terms of different algorithms.

In this paper [4] the author explain by the use of movie trailer uploaded on most popular video platform YouTube can predict the success and failure of a movie before it is released by help of these the filmmakers, In this paper they used various regression algorithms for prediction but in this paper they used only linear regression , Polynomial Regression, Gradient Boosted Tree and Simple Regression Tree they can also use SVM which can give them better performance.

After reading we realize that gradient tree boosting algorithms in this part. The Explanation follows from the same idea in existing literatures in gradient boosting. Specifically, the second order method is originated from Friedman et al. [12]. We make minor improvements in the regularized objective, which may get helpful in implementation or using.

III. DESCRIPTION OF USED ALGORITHM

1. Simple Linear Regression- It is statistical way so that it allows us to remember and study relationships between two continuous variables.

Note Linear Regression might be old but it's still useful, but there's a drawback of using linear regression because it's made on assumptions that our data have linear relationships while in many real world scenarios that not true.

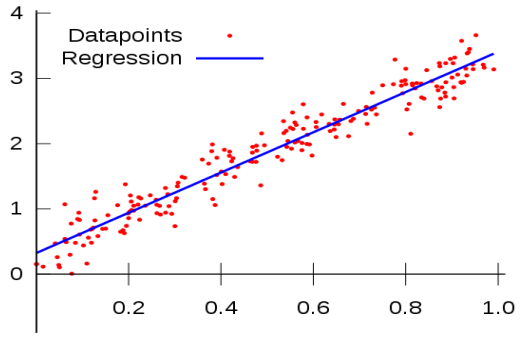

Fig. 3 Model of Regression.

2. Support Vector Machine- Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges.

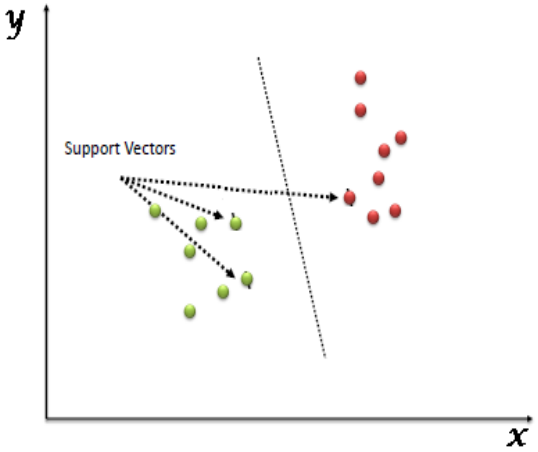

Fig. 4 Model of Support Vector Machine.

IV. REQURED FRAMEWORK

Python and its libraries are using in data science and data analysis very efficiently. They are also largely used for creating expandable machine learning algorithms. Python can apply various machine learning techniques such as Classification, Regression, and Clustering.

© 2019 IJSRET

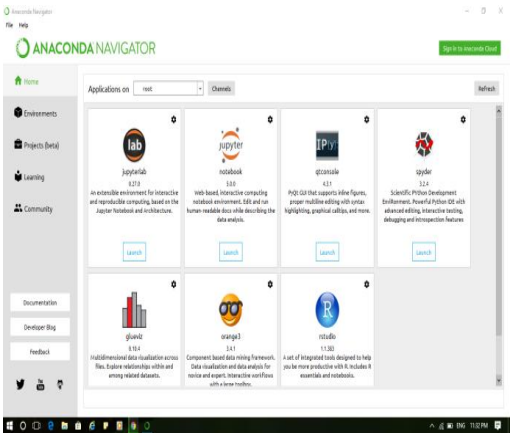

Fig.6 GUI Anaconda.

V. MACHINE LEARNING FLOW

The model employs filters for faster evaluation and lesser overall time.

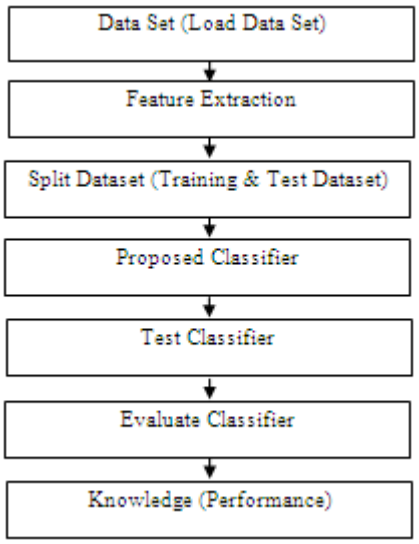

Fig.7 Proposed Data Mining Framework for Classification.

VI. CONCLUSION

After studied number of research documents or paper we can say that in machine learning algorithm like NB, SVM, LR and many more is basically depends upon data set and their features. We studied that we have number of performance measuring terms like Accuracy, Precision, AUC, ROC etc. we find that error finding and minimizing is also a very efficient part in machine learning algorithm.

REFERENCES

[1] R. Bekkerman. The present and the future of the kdd cupcompetition: an outsider's perspective.

[2] R. Bekkerman, M. Bilenko, and J. Langford. Scaling UpMachine Learning: Parallel and Distributed Approaches.Cambridge University Press, New York, NY, USA, 2011.

[3] J. Bennett and S. Lanning. The netix prize. In Proceedings of the KDD Cup Workshop 2007, pages 3{6,New York, Aug. 2007.

[4] Mining Trailers Data From YouTube for Predicting Gross Income of Movies by Md Shamsur Rahim, AZM Ehsetma chowdhury, Md.Asiful Islam, Mir Riyanul Islam in 2017 IEEE Region 10 Humanitarian TechnologyConference(R10-HTC)21 - 23 Dec 2017, Dhaka, Bangladesh

[5] C. Burges. From ranknet to lambdarank to lambdamart:An overview. Learning, 11:23{581, 2010.

[6] O. Chapelle and Y. Chang. Yahoo! Learning to RankChallenge Overview. Journal of Machine Learning

[7] T. Chen, H. Li, Q. Yang, and Y. Yu. General functionalmatrix factorization using gradient boosting. In Proceedingof 30th International Conference on Machine Learning(ICML'13), volume 1, pages 436{444, 2013.

[8] T. Chen, S. Singh, B. Taskar, and C. Guestrin. E_cientsecond-order gradient boosting for conditional random_elds. In Proceeding of 18th Arti_cial Intelligence andStatistics Conference (AISTATS'15), volume 1, 2015.

[9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, andC.-J. Lin. LIBLINEAR: A library for large linear classi_cation. Journal of Machine Learning Research,

[10] J. Friedman. Greedy function approximation: a gradientboosting machine. Annals of Statistics, 29(5):1189{1232,2001.

[11] X. Yan, X. G. Su, Regression Analysis: Theory and Computing, World Scientific Publishing Co. Pte. Ltd., 2009.

[12] M. H. Kutner, C. J. Nachtsheim, J. Neter, Applied Linear Regression Models, 4th ed., McGraw-Hill Irwin, 2004.

[13] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity. The Lasso and Generalizations, Chapman & Hall, 2015.

© 2019 IJSRET