Emblematic Analysis of Varied Machine Learning Algorithm

Surbhi Agarwal[1], Roli Trivedi [2], Jayati Vijayvargaya[3], Manya Srivastava[4]
[1, 2, 3, 4] Mody University of Science and Technology

Abstract- *Suitability of algorithm before its implementation is a paramount. Inventing new algorithm is not an efficient approach to imply it on an application. Existing pool of machine learning algorithm has a vast variety of implementation strategies with higher efficacy. Significance of comparative analysis of currently available algorithms is equally important. Evaluating the performance of various machine learning algorithms like K-nearest neighbor, decision tree, random forest, linear regression and others, will be the prime intend of this paper.*

I. INTRODUCTION

In the most recent years, as information that we are able to extract from the statistics has rapidly elevated. Machine Learning isn't approximately storing large quantities of facts; however, it is part of Artificial intelligence (AI).
Artificial Intelligence is the development of the computer applications to perform duties that typically require the human intervention, for instance choice making. Making the right decision for a specific hassle is the primary issue for attaining our desires. For this motive, many machine learning strategies are used for both classification and regression issues. Classification is used when the prediction aim is a discrete cost or a class label. when the prediction goal is non-stop, regression is the best approach to use.

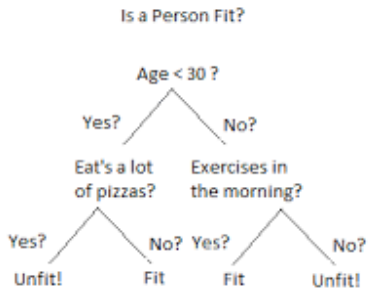
II. RESEARCH GAP

The research has been done to make a new algorithm or new application .But we can use the best from the existing algorithm if we study the comparison between different machine learning algorithm and apply different practical applications on them. Inventing new algorithm is not an efficient approach to imply it on an application. Existing pool of machine learning algorithm has a vast variety of implementation strategies with higher efficacy. Significance of comparative analysis of currently available algorithms is equally important

III. MACHINE LEARNING ALGORITHMS

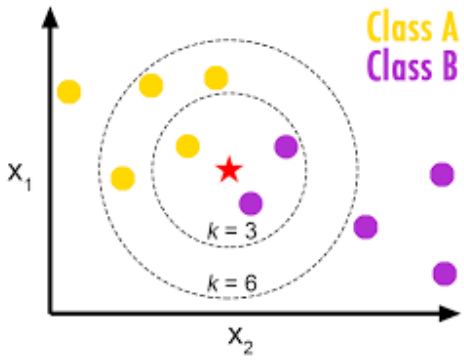
3.1 Decision Tree:

Decision tree is a standout among the most usually utilized classifiers in insights and machine learning. Decision tree is a various levelled layout that executes the divide and conquer method. It's a nonparametric approach utilized for each classification and regression. It could be especially modified over to an arrangement of fundamental if-then standards. Its clear portrayal makes the consumer geared up to decipher the final results. It's consequently worthy on the way to examine the nature of multiple decision tree, generated from a couple of decision tree gaining knowledge of algorithms. Presently, decision trees are in comparison via estimating their attainment on a few unseen records.



3.2 K-Nearest Neighbor:

K-Nearest-Neighbor is a case of incidence based on totally mastering and it's frequently utilized for type wherein the assignment is to represent the unseen illustrations in view of the database stored. The perceptions are added in a d - dimensional area, in which d is the quantity of attributes or traits which the perception has. Given any other factor, it is characterised on the premise of its similarity to whatever is left of the facts focuses saved in the version via some similitude. The belief in the back of using the nearest neighbors may be resolved by way of the subsequent saying "If it walks like a canine, barks like a canine, and looks as if a canine, then it's likely a dog."



3.3 Association Rules:

Association rule learning is an approach for locating intriguing family members between elements in tremendous databases. It is anticipated to apprehend stable regulations discovered in databases utilizing some measures of intriguing quality. Association rules are if/then statements that facilitates uncover relationships between reputedly unrelated information. A case of association rule could be "If a patron purchases twelve eggs, he's 80% at risk of likewise purchase milk." An association rule has sections, a forerunner (if) and a subsequent (then). Association rule mining is perfect for non-numeric statistics and it involves little extra than simple counting.

3.4 Random forest

Random forest algorithm is a supervised category set of rules. As the name recommends, this set of rules creates the forest with a number of trees. In standard, the greater no of trees in the forest the more robust the forest area looks. In the same way inside the random forest classifier, the better the variety of trees within the forest area offers the excessive accuracy results

In case you recognise the decision tree algorithm. You must be probably questioning are we growing more range of decision trees and how are we able to create greater variety of decision trees. As all of the calculation of nodes selection could be same for the equal dataset. Sure. You are right. To model more range of decision tree to create the forest you aren't going to apply the same apache of constructing the decision with records gain or gini index method.

3.5 Linear Regression

Linear regression is reasonably one of the most well-known and well accepted algorithms in machine learning. Regression is commonly a method to identify an equation that defines a relation you find in your data. Linear regression models this relation as a linear equation of the form y = mx + c. It is a statistical method for guessing the value of a dependent variable from an independent variable when the interconnection between the variables can be illustrated with a linear model. Therefore, it is an analytical method to achieve the specifications of the line's equation from a bunch of data.

IV. COMPARATIVE STUDY

The selection and introduction of a machine learning classifier is the following step once the record representation scheme is finalized. Numerous machine learning techniques had been hired to address various classification issues. One of the predominant variations that exist between those techniques is the philosophy at the back of the getting to know technique. We talk some of the machine learning algorithms like: decision tree, Linear Regression, random forest, association rule, k-nearest neighbour.

The main advantage of decision tree is interpretability. Decision trees are "white boxes" in the sight that the gathered knowledge can be expressed in a readable form, while KNN is generally black box, i.e. you cannot read the acquired knowledge in a comprehensible way.

Decision tree has been hired correctly in many conventional applications in one of a kind domain. In spite of the truth that it can be seemed as incredibly antique method, decision tree has stood the test of time. As an instance, decision tree has these days been employed as a machine learning technique to broaden category models that robotically classify pancreatic cancers statistics.

Decision tree uses to search for a solution in the hassle area; performance has a tendency to be a trouble, especially when coping with huge datasets. Decision tree is characterised with the aid of its relative transparent outputs, which can be easy to be read and understood via human beings. Decision tree has been proven to have superior performance over different techniques with reference to a few precise domain names with datasets that have discrete/categorical facts kind attributes.
The prediction accuracy of decision tree is >98% while k-nearest neighbour has accuracy of >99%

Your parents are visiting your town so you need to plan and decide which is the best restaurant in town where you can take your parents but you are confused which restaurant to choose. There is a friend of yours kate who suggest you to which restaurant to visit. So in order to suggest you a good restaurant first he will think if you will prefer a particular restaurant or not.So what you do is prepare a list of restaurants

that you have already visited and tell her whether you like it ornot. While you ask kate that whether you may like a particular restaurant R or not, he asks you some sort of questions like "Is "R" an open restaurant?", "Does "R" serve chinese delicacies?", "Does R have live music?", "Is R open all night ?" and so on.

Kate asks you several informative inquiries to maximize the information and offers you YES or NO answer based totally on your answers to the questionnaire. Right here kate is a decision tree for your favourite restaurant preferences.

A main downside of decision tree machine learning algorithms, is that the results can be primarily based on expectancies. While choices are made in real-time, the payoffs and resulting results may not be the same as predicted or deliberate. There are probabilities that this will result in unrealistic decision trees leading to horrific decision making.
There are many advantages of random forest algorithm when in comparison with different kind of machine learning algorithms. The overfitting problem will in no way come while we use the random forest algorithm in any class hassle. The similar random forest algorithm can be used for both class and regression task. The random forest algorithm can be used for feature engineering. Due to this figuring out the most important capabilities out of the available capabilities from the training dataset.

kate is a decision tree on your eating place preferences. However, kate being a person does no longer usually generalize your eating place preferences with accuracy. To get greater accurate restaurant advice, you ask many of your friends and decide to visit the restaurant R, if most of them say that you'll find it irresistible. In place of simply asking Kate, you would really like to ask Jon Snow, Sandor, Bronn and Bran who vote on whether or not you will just like the restaurant R or not. This means you have built a random forest which is an ensemble classifier of decision tree. By presenting your buddies with slightly one of a kind statistic in your eating place possibilities, you make your buddies ask you exclusive questions at specific instances. In this situation simply by barely altering your restaurant choices, you're injecting randomness at model stage (unlike randomness at data degree in case of decision tree). Your bunche of friends now shape a random forest of your restaurant preferences. Random forest is the go to machine learning algorithm that makes use of a bagging method to create a bunch of decision trees with random subset of the statistics.

Linear regression is one of the most interpretable machine learning algorithms, making it easy to explain to other and is ease of use as it requires minimal tuning. It is one of the fast running widely used machine learning algorithm.

Linear regression is one of the maximum interpretable machine learning algorithms, making it clean to explain to other and is ease of use because itrequires minimum tuning. It's the most commonly used machine learning method that runs rapidly.

K Nearest Neighbour is one of the handiest method for classification as well as regression problem. That is the reason it's far extensively followed. KNN is a supervised approach that uses estimation based totally on values of neighbour's

V. CONCLUSION

Machine learning is an extraordinary fulfilment in the area of artificial Intelligence. This paper proposes a framework for comparison study of machine learning algorithm. We evaluated and compared well-known machine learning algorithms such as decision tree, k-nearest neighbor, random forest, linear regression and association rules. Furthermore, it is inconclusive whether a set of proposed features was the best.

REFERENCES

[1] Comparative Study of Four Supervised Machine Learning Techniques for Classification Amr E. Mohamed Department of Mathematical Sciences University of Essex, UK.
[2] A Comparative Study of Machine Learning Methods for Verbal Autopsy Text Classification Samuel Danso1, Eric Atwell2 and Owen Johnson 1Language Research Group, School of Computing, University of Leeds LS2 9JT, U.K.scsod@leeds.ac.uk
[3] A Comparative Study of Machine Learning Techniques for Automatic Product Categorisation Chanawee Chavaltada1, Kitsuchart Pasupa1(B), and David R. Hardoon21 Faculty of Information Technology, King Mongkut's Institute ofTechnology Ladkrabang, Bangkok 10520, Thailand.
[4] A Comparative Study on Machine Learning Classification Models for Activity Recognition Mohsen Nabian* Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, United States
[5] Comparative Study of Machine Learning Algo-rithms for Heart Disease Prediction Helsinki Metropolia University of Applied Sciences Bachelor of Engineering Information Technology

[6] Enesemble methods in machine learning, Thomas G.Dietrich, Oregon State
University, Corvallis, Oregon, USA.

[7] A Preliminary Performance Comparison of Five Machine Learning Algorithms for
Practical IP Traffic Flow Classification Nigel Williams, Sebastian Zander, GrenvilleArmitage Centre for Advanced Internet Architectures (CAIA) Swinburne University of Technology Melbourne, Australia

[8] Approximate Statistical Tests for Comparing Supervised Classification Learning
Algorithms Thomas G. Dietterich Posted Online March 13, 2006

[9] Classification and Regression by randomForest Andy Liaw and Matthew Wiener

[10] RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY Authors D. Richard Cutler, Thomas C. Edwards Jr., Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, Joshua J. Lawler First published: 1 November 2007