HBRP PUBLICATION

Journal of Advances in Computational Intelligence Theory
Volume 3 Issue 2

Survey of Machine Learning Algorithms & its Applications

Akshada Sunil Shitole[1*], I Priyadarshini[2]
[1]Student, [2]Professor,
Computer Engineering Department,
K K Wagh Institute of Engineering Education & Research, Nashik, Maharashtra, India.

****Corresponding Author***
E-mail Id:-*akshadasshitole@gmail.com*

ABSTRACT
Machine Learning is a subset of Artificial Intelligence. Machine learning is one of the latest technologies which has brings new innovations in various fields. Machine learning refers to the concept of train the machine in such a way it can learns from a past experiences or it can learn from a data provided to it. The concept machine learning can be implemented in various fields using its various algorithms. The machine learning contains various algorithms like KNN, K means, decision tree, random forest, support vector machine etc. Machine Learning can be further classified into Supervised Learning, Unsupervised Learning, Reinforcement. Supervised learning performs predictions and Unsupervised learning performs clustering. Further Machine Learning also consists of Deep Learning. Deep learning consists of studies of neural networks.

Keywords:-*Machine learning, KNN, K means, decision tree, random forest, support vector machine.*

INTRODUCTION-

The main aim of this paper is to study machine learning, machine learning algorithms and its application. The machine learning can be used in various fields such as internet, education, health, entertainment, robotics, banking, industries etc. Machine learning is one of the interesting branches of computer engineering which helps innovators and researchers to make Innovations. It refers to learn from the data and take decisions based on it. The goal of machine learning is making decisions and performing extraction of patterns. The machine learning consists of supervised, unsupervised and reinforcement learnings. The predictions, Classifications, Recommendations can be carried out with the help of machine learning algorithm. The Supervised machine learning means the machine will be going to learn in the presence of supervisor. In Supervised learning the labelled data is provided. Past knowledge is present in supervised learning. Supervised learning is a predictive model. One of the applications of supervised learning is in stock market Analysis [8]. The unsupervised learning means the machine will not have any supervisor machine has to learn by its own. The machine would not have any past knowledge or experience. The Unsupervised learning works on unlabeled data. Unsupervised learning is clustering model. One of the applications of unsupervised learning is fraud detection of credit cards [3]. The Reinforcement learning is based on the penalty and rewards. If the machine takes correct decision, then it will be rewarded and if the decision is incorrect then it will have a penalty. The Reinforcement learning works on try and errors basis. The Supervised machine learning algorithms are linear regression, logistic regression,

support vector machine, decision tree, random forest etc. The unsupervised learning algorithms are knn, k means, apriori, hierarchical clustering, anomaly detection etc. Application areas of supervised machine learning consist of weather prediction, stock market prediction, biometrics, Cortana etc. Application areas of unsupervised learning are segmentation of customers and their behavior towards buying a product, Recommendations of products, movies etc.

LINEAR REGRESSION

The linear regression is a supervised machine learning algorithm which determines the relationship between two variables. Linear regression is performed on dependent and the independent variables. Graph shows the straight line in linear regression. Linear regression works on a continuous variable [7]. Continuous variables are the infinite values like weight of a person etc. The output of the linear regression is the value. Linear regression is measured by loss, R squared, Adjusted R squared etc. It solves the regression related problems. For solving weather related problems linear regression shows what will be the tomorrow's temperature in Celsius or Fahrenheit. Prediction of sales value can be performed using linear regression with approximately 84 % of accuracy [11]. Face Identification can also be performed using linear regression [18].

LOGISTIC REGRESSION

The logistic regression is uses sigmoid function. It produces the results in a binary format such as yes or no etc. In a graph it will not shows a straight line like linear regression it will show sigmoid. It works on a categorical variable. Categorical variable means finite set of values. Output of a logistic is probability. It is measured by recall, precision, accuracy, confusion matrix. The Logistic regression solves the multi class classification related problems. For solving weather related problems logistic regression shows whether there is a rain tomorrow? Yes or no. One of the applications of logistic regression is identify factors contributing to the university dropout students [10]. Also, Churn of telecom customer prediction [15], predicting football match [13] can be done with the help of logistic regression.

KNN

KNN means K nearest neighbor. It is also called as lazy learner algorithm. K is the number of nearest neighbors. KNN suggest that you are near to your neighbor means you are same or similar. KNN is used when you have to find the similar items. It is used in a recommendation system. K is measured on the basis of distance like Euclidean distance, Manhattan distance etc. music recommendation system can be implemented using KNN machine learning algorithm [9].

NAÏVE BAYES

Naïve Bayes is a classification technique works on the principle of bayes theorem. Naïve Bayes algorithm states that presence of a particular feature in a class is independent or unrelated to the presence of another feature. Bayes theorem works on the basis of probability. Types of naïve bayes are gaussian, multinomial & Bernoulli. The naïve bayes classification algorithm is used in spam filtering[14], object & face recognition, weather prediction, diabetes prediction [1] etc.

DECISION TREE

Decision tree algorithm of machine learning starts forming a tree like structure. This tree consists of root, branch nodes, leaf nodes etc. Decision tree is a graphical visualization of data and by traversing these trees decisions can be formed. The main terminology of decision trees are pruning, entropy, information gain etc. the term pruning refers to removing unwanted branches from the decision tree. Entropy is

the measure of impurities and information gain is the measure of purity i.e., decrease in entropy. Decision tree can be used to detect COVID symptoms severity [4].

RANDOM FOREST

Random forest is one most accurate algorithm which works well on classification as well as regression. The forest is nothing but the collection of trees similarly random forest algorithm of machine learning is also collection of different decisions trees. These decision trees are combined to find final outcome results. It is called as random because the features are selected randomly. Random forest is accurate, versatile and scalable algorithm. One of the applications of random forest algorithm is to credit card risk or fraud detection, phishing detection [5], opioid use disorder prediction [6].

SUPPORT VECTOR MACHINE

Support vector machine is a supervised machine learning algorithm mainly used to classify data. SVM can be used for face recognition, colon cancer classification etc. SVM makes use of hyperplane also called as a decision boundary to divide data into various segments. The Support Vector Regressor (SVR) is used for regression related problems. SVM can be performed on non-linear dataset by using kernel function. One of the applications of SVM & KNN is in classification of grape leaves to determine whether the leaf is healthy or unhealthy [2], classify student attentiveness [16].

K-Means

K means is an unsupervised machine learning algorithm which forms clusters of data. To form the number of clusters the elbow method is used. K means is simple & very easy to understand [12]. But k means algorithm unable to handle noisy data & outliers. Fuzzy k means is extension to k means. In fuzzy clustering each data points belongs to more than one cluster. Color recognition can be done using k means [17].

CONCLUSION

In this paper, the aim was to do analysis of machine learning its types, its algorithms and their various applications in various fields. The advantage, disadvantage and the overview of each algorithm is discussed. For innovating any real-world application of machine learning one must know all these basic working of each and every algorithm of machine learning. Machine learning will definitely help innovators, researchers, students to develop modern world problems solutions.

REFERENCES

1. Priya, K. L., Kypa, M. S. C. R., Reddy, M. M. S., & Reddy, G. R. M. (2020, June). A Novel Approach to Predict Diabetes by Using Naive Bayes Classifier. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)* (pp. 603-607). IEEE.
2. Bharate, A. A., & Shirdhonkar, M. S. (2020, March). Classification of Grape Leaves using KNN and SVM Classifiers. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 745-749). IEEE..
3. Rai, A. K., & Dwivedi, R. K. (2020, July). Fraud detection in credit card data using unsupervised machine learning based scheme. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 421-426). IEEE.
4. Rochmawati, N., Hidayati, H. B., Yamasari, Y., Yustanti, W., Rakhmawati, L., Tjahyaningtijas, H. P., & Anistyasari, Y. (2020, October). Covid Symptom Severity Using Decision Tree. In *2020 Third International Conference on*

Vocational Education and Electrical Engineering (ICVEE) (pp. 1-5). IEEE.

5. Sindhu, S., Patil, S. P., Sreevalsan, A., Rahman, F., & AN, M. S. (2020, October). Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 391-394). IEEE.
6. Wadekar, A. (2019, July). Predicting Opioid Use Disorder (OUD) Using A Random Forest. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 960-961). IEEE.
7. Ray, S. (2019, February). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE..
8. Pahwa, K., & Agarwal, N. (2019, February). Stock market analysis using supervised machine learning. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 197-200). IEEE.
9. Li, G., & Zhang, J. (2018, October). Music personalized recommendation system based on improved KNN algorithm. In *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)* (pp. 777-781). IEEE.
10. Hori, G. (2018, July). Identifying Factors Contributing to University Dropout with Sparse Logistic Regression. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 430-433). IEEE.
11. Qi, J., Yu, Y., Wang, L., & Liu, J. (2016, October). K*-means: An effective and efficient k-means clustering algorithm. In *2016 IEEE international conferences on big data and cloud computing (BDCloud), social computing and networking (SocialCom), sustainable computing and communications (SustainCom)(BDCloud-SocialCom-SustainCom)* (pp. 242-249). IEEE.
12. Gopalakrishnan, T., Choudhary, R., & Prasad, S. (2018, December). Prediction of Sales Value in Online shopping using Linear Regression. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1-6). IEEE.
13. Gopalakrishnan, T., Choudhary, R., & Prasad, S. (2018, December). Prediction of Sales Value in Online shopping using Linear Regression. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1-6). IEEE.
14. Solanki, R. K., Verma, K., & Kumar, R. (2015, August). Spam filtering using hybrid local-global Naive Bayes classifier. In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 829-833). IEEE.
15. Ross, M., Graves, C. A., Campbell, J. W., & Kim, J. H. (2013, December). Using support vector machines to classify student attentiveness for the development of personalized learning systems. In *2013 12th international conference on machine learning and applications* (Vol. 1, pp. 325-328). IEEE.
16. Li, P., Li, S., Bi, T., & Liu, Y. (2014). Telecom customer churn prediction method based on cluster stratified sampling logistic regression.
17. Zheng, X., & Liu, N. (2012, July). Color recognition of clothes based on k-means and mean shift. In *2012 IEEE International Conference on Intelligent Control, Automatic*

Detection and High-End Equipment (pp. 49-53). IEEE.

18. Naseem, A. I., Togneri, B. R., & Bennamoun, C. M. (2009, November). Face identification using linear regression. In *2009 16th IEEE International Conference on Image Processing (ICIP)* (pp. 4161-4164). IEEE.