iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44672

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Machine Learning Algorithms and Its Applications: A Survey

Raja Tawseef Ahmad Mir
M.Tceh (Computer science & Engineering), Universal Group of Institutions Mohali

Abstract: ***In todays present world lots of microelectronic statistics is created in apiece and every field. The data that we obtain contains valuable info to forecast the future. Owing to the enormous in magnitude, the physical forecasting gives a intricate chore to humans. To overwhelm this problem, the data model is made in such a way so that it can predict the future by the situation with the aid of training data and test datasets. To Pullman the machine or the data model, numerous types of machine learning algorithms (ML) and tools are available. This paper will emphasis on the review of the few machine learning algorithms(ML) and methods used in numerous applications and domains in a detailed manner.***
Keywords: Artificial Intelligence (AI), Machine Learning (ML), Algorithms, Neural Networks (NN), Least Square Method.

I. INTRODUCTION

Machine Learning (ML) is a subgroup of Artificial Intelligence (AI).Using Machine Learning (ML) we can make applications acquire from experience in the same way as human do. When data is nursed into these applications, they learn grow and change giving to experience. This is done by using algorithms that cram from data in an repetitive process. Applications that use ML use pattern recognition to reply to various data that are fed as an input to the application. Machine learning is the ability of an applications to react to new data that we have fed as an input using repetitions. Machine learning algorithms helps the system to learn how to predict outputs based on previous examples that we have given to the system and the relationship among the data that we fed as input data and output data which is known as training data set . Relationship between inputs and outputs of any model is gradually improved by testing its predictions and correcting that when wrong output is obtained. Machine learning (ML) is a set of computerised methods for knowing different outlines in data. Machine Learning (ML) is a way of creating a method of something like the Line of best fit method also called as Least Square Method. It is beneficial to automate this method when the data has numerous features and is very complex.

II. DIFFERENT TYPES OF MACHINE LEARNINGS

A. Supervised Learning

Supervised learning (SL) is a type of Machine Learning in which a data that we select as a training data set with labelled data, or data with previously known output value. The glitches that we are solve through supervised learning are Classification and regression.

B. Unsupervised Learning

Unsupervised learning is a type techniques that do-not use a preparation set but finds some patterns or edifice in the already fed data by themselves or among the data. Clustering problems are resolved by an unsupervised learning approach.

C. Semi Supervised Learning

Semi-supervised learning is a type of learning approach that uses unlabelled and a minor quantity of labelled data that we have fed as an input data. Using a minor quantity of already fed labelled data can significantly surge the effectiveness of unsupervised learning errands. The model or machine essentially cram the edifice to organize the statistics is ready to make estimates.

D. Reinforcement Learning

Reinforcement learning is a type of learning approach that practices the data that we have given him as input from the surroundings as a stimulus and check how the model or the machine must respond on giving some input. Feedback is not produced over a training route like supervised learning but as rewards or forfeits in the surroundings. Reinforcement Learning is used in robot control.

©IJRASET: All Rights are Reserved

III. WORKING OF MACHINE LEARNING ALGORITHM?

Machine Learning process starts with nurturing data to an algorithm .This process of entering the data to any algorithm is called as training the algorithm. To check whether an algorithm is working, data is nurtured into the algorithm and results are tartan. If looked-for results are not achieved, then the algorithm is re-trained with the data that we have given him and tested again. This process is recurrent till we achieve the desired results. This helps the machine learning algorithm to learn and give the desired output as well as increases the accuracy of the result.

IV. HOW DOES MACHINE LEARNING WORK?

In the procedure of machine learning (ML), the worth of information is the primary factor that is provided by the external environment to the system . The information that is fed to the system from the external environment in some defined form and delivers itself in some form, and then connotes cradles of outside data; Learning can be defined as the process that routes the external facts to knowledge first it obtains the information of outside environment and then processes the information to knowledge (we get after information), and places this acquaintance into the warehouse; Warehouse stores many common principles that escort a portion of the application deed, owing to environment that delivers all types of facts for learning system, the excellence of facts impacts directly on learning understanding of the system that whether it is easy or messy.

Data Warehouse is the second factor that influences the intention of machine learning system. The mien of knowledge is diverse, such as, eigenvector, first order logic statements , rules of production, semantic networks and frameworks and many more , these customs of mien and each has its stout point. We take into consideration following four aspects when we have to elect, these four aspects are as: (a) stout in appearance (b) easy to deduce, (c) easy to alter fount, (d) the knowledge is tranquil to grow. The implementation can be defined as the process that uses the facts of fount to complete a convinced job, and to response the facts that we have obtained in the development of finalising the job to the learning, and thus will help us to guide the system for advance study.

V. DIFFERENT TYPES MACHINE LEARNING ALGORITHMS

Depending upon resemblance in rapports of their purpose how the algorithm will work algorithms are of different types . For example, tree-based methods, and neural network inspired methods. One of the most valuable method to cluster procedures and it is the technique that we are going to use here. This is a suitable grouping technique, but these are not perfect. There are also some algorithms that can fair as effortlessly fit into many categories like Learning Vector Quantization(LVQ) that is both a neural network inspired (NN) method and an instance-based method. There are also some classes that have the same name that define the problem and the class of algorithm such as Regression and Clustering.

A. Decision Tree Based Classification

Decision tree algorithm is a type of classification that is primarily used to build a model in the way of a structure that resembles a tree like having (root, branch and leaf), that is based on(inferred from) earlier information to classify/predict class or target variables of future(new data) that we can get with the help of decision rules or decision trees. The main usage of Decision Trees is in the numerical as well as categorical data. The algorithm works on greedy search approach that is it will start from top to bottom.

1) Advantages

- It Easy to implement
- It Can classify and predict categorical as well as numerical data.
- Is Less data pre-processing
- Statistical test can be done to validate the tree model. Resembles human decision-making technology.
- Tree structure is easily understandable through visualization

2) Limitations

- Low Prediction Accuracy
- Complexity in calculations if class labels are huge
- Need of redrawing for each additional data to the data set.
- Probability of over-fitting in the decision tree is high.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) Applications

- Agriculture
- Medicine
- Financial analysis

B. Support Vector Machines

The key objective of this is to discover a hyper plane that is used to divide the classes into two types. Depending upon the values obtained as hyper values the obtained data set is placed into the data set to which the similarities resembles. To draw a hyper plane we must take two rules into consideration . First is that the hyper plane needs to been chosen in such a way that it should separate the two classes and best hyper separator plane should be chosen as maximum-margin hyper.
SVM can be classified into two different types : a) Linear SVM b) Non-Linear SVM

1) Advantages:

- Robust Classifier for prediction problem
- Efficient if n(D) > n(S) where n(D) the number of dimensions is greater than the n(S) number of samples
- Suitable for high dimensional data spaces
- Memory efficient.

2) Disadvantages

- Very slow in test phase
- Not suitable for large and noisy data sets
- Classification error percentage will be increased if wrong kernel is selected
- Memory consumption is high

3) Applications

- Image classification
- Bioinformatics
- Face detection

C. K-Nearest Neighbours

This type of algorithm is used to place the data set into the predefined set by measuring the distance with the predefined data set by taking into consideration k values. The distance that it finds least puts the data set into that data set . the methods that is used to calculate the distance between two points uses k variable values among 0 an d1 normally. Most commonly used algorithm used in this are Euclidean distance , hamming distance. This method is used in such problems where we have to do classification of data set. For continuous variables we are using Euclidean distance formula and for categorical variables we are using hamming distance

1) Advantages

- Supple with traits and distance functions
- Supports multi class data

2) *Disadvantages*

- Finding appropriate K value is difficult
- Needs large sample for high accuracy
- Requires large storage space

D. Naïve Bayes Algorithm

This algorithm is used in the classification of data set if the data set is in large quantity and have much more records as the multiclass and binary class related classification problem. This can be used in the classification job in the field related to machine learning. The main objective of this is to analyse the text and natural language processing. For naïve bayes algorithm one must have the concept related to bayes theorem (based on the conditional probability) Conditional probability can be defined as an event will

happen with conditioned (based) on an event already occurred. This will help us to merge the different algorithms to form a naïve bayes by using a common principle.

Types of Naïve Bayes:

- Gaussian Naïve Bayes
- Multinomial Naïve Bayes
- Bernoulli Naïve Bayes

3) *Advantages*

- Fast
- Scalable
- Efficient for multinomial distributed and also for binary attribute values

4) Disadvantages:

- Not suitable for regression problems
- Cannot find relationships among attributes Applications:
- Text, e-mail, and symbol analysis
- Recommendation Systems 5.5

5) Linear Regression

This type of algorithm is used to find the link between an independent variable also known as (predictor (X)) and a dependent variable (criterion (Y)) variable that can be implemented to predict the future values of the dependent variable.

In Simple regression one independent variable is used and in case of multiple regressions we can use two or more independent variables depending upon the data set to predict the future. Dependent variable are those that have a continuous and independent variable are those variables that have discrete or dis-continuous values. regression models are of two kinds. One is linear and other one is non-linear. The linear regression model is one that uses straight line and non-linear regression model is a type of model that uses curved line relationships among dependent and independent variables.

6) Advantages

- Shows relationship among dependent and independent variables
- Simple and easy to understand Disadvantages:
- Not applicable for non-linear data
- Only predicts numerical output
- Data must be independent Applications:
- Observational Astronomy
- Finance

VI. COMPARISONS OF DIFFERENT ALGORITHMS

1) *Parameter Estimation Algo:* Is used to obtain the parameter calculation that are used by the system or by the model.
2) *Model Complexity Reduction:* The model used should not be complex or we should reduce the complexity of our model.
3) *Generative Or Discriminative:* The discriminative model is mainly used for machine learning particularly for supervised. The discriminative model is also known a conditional model, and this model crams the boundaries between classes or labels in a dataset. Generative models this falls in the class of statistical models that yield new data examples. This type of models are mostly used in unsupervised machine learning to do tasks related to probability and estimation, modelling data points, and distinguishing between classes using these probabilities.
4) *Loss Function:* The loss function is the function that calculates the distance between the current output of the algorithm and the estimated output. It's a method to estimate how your algorithm models the data. It can be categorized into two groups. One for classification (discrete values, 0,1,2…) and the other or regression (continuous values). Some normally used Loss functions used are Cross-entropy (CE), Log loss, Mean Square Error (MSE — L2),Mean Absolute Error (MAE — L1),Huber Loss

S.NO	LEARNING METHOD	PARAMETER ESTIMATION ALGO	MODEL COMPLEXITY REDUCTION	GENERATIVE OR DISCRIMINATIVE	LOSS FUNCTION	DECISION FUNCTION
1	GUSSAIN NAÏVE BAYES	Estimate ˆμ, ˆσ2, and P(Y) using maximum likelihood	Place previous parameters and use MAP estimator	Generative	− log P(X, Y)	Equal variance: linear boundary. Unequal variance: quadratic boundary
2	Logistic Regression	No closed form estimates. Optimize objective function using gradient descent	L2 regularization	Discriminative	− log P(Y |X)	Linear
3	Decision Trees	Many algorithms: ID3, CART, C4.5	Prune tree or limit tree depth	Discriminative	Either – log P(Y |X) or zero-one loss	Axis-aligned partition of feature space
4	K-Nearest Neighbours	Choose K using cross validation by store all training data to classify new points. .	Increase K	Discriminative	zero-one loss	Arbitrarily complicated
5	Support Vector Machines (with slack variables, no kernel)	Solve quadratic program to find boundary that maximizes margin	Reduce C	Discriminative	hinge loss: |1−y(wT x)|+	linear (depends on kernel)

VII. JEOPARDIES IN MACHINE LEARNING

A. Data Poisoning/Destroying

For the security of any ML system data plays a vital role . the reason for this is that the ML system learns to do what it does right from data that we have given to it. If a mugger can deliberately handle the data that we have used in ML system in a synchronised fashion, the whole ML system can be conceded. A special attention is needed for Data poisoning attacks . ML engineers must take those measures into consider that in what fraction of the training data an attacker can control and to what extent.

B. Data Privacy/Secrecy

In ML Data protection is problematic ample without flinging into the combination. One of the most sole challenge in ML is the caring subtle or private data that we have given to the model , through training, are built right into a model. Understated but actual extraction attacks against an ML system's data are an important category of risk.

C. Online System Manipulation

When a ML system continues to learn during operational use, modifying its behaviour over time this is said to be online system. In this situation, a cunning attacker can use system input to deliberately nudge the still-learning system in the wrong direction, gradually retraining the ML system to perform the wrong thing. It's worth noting that such an attack might be both subtle and simple to carry out. To adequately handle this risk, ML engineers must take into account data provenance, algorithm selection, and system operations.

D. Over-fitting

When a model learns the information and noise in the training data to the point where it degrades the model's performance on fresh data, this is known as overfitting. This means that the model picks up on noise or random fluctuations in the training data and learns them as ideas. The difficulty is that these notions do not apply to fresh data, causing the model's ability to generalise to be harmed. Nonparametric and nonlinear models, which have more flexibility when learning a target function, are more prone to overfitting. As a result, many nonparametric machine learning algorithms incorporate parameters or approaches that limit and constrain the amount of detail learned by the model.

VIII. BENEFITS OF MACHINE LEARNING

1) Trends and patterns are easily discernible. Machine Learning can analyse enormous amounts of data and identify specific trends and patterns that people might miss. For example, an ecommerce website like Amazon uses it to better understand its users' browsing habits and purchase histories in order to provide them with the most relevant products, promotions, and reminders. It uses the information to show them ads that are relevant to them.

2) Improvements are ongoing. Machine Learning algorithms improve in accuracy and efficiency as they gather experience. This enables people to make more informed selections. Assume you're working on a weather forecasting model. Your algorithms learn to generate more accurate predictions faster as the amount of data you have grows.

3) No human intervention needed (automation) With ML, you don`t want to make the adjustments manually, because it manner giving machines the cap potential to learn, it allows them to make predictions and additionally enhance the algorithms on their own. A not unusual place instance of that is anti-virus software`s; they discover ways to clear out new threats as they may be recognized. ML is likewise exact at spotting spam.

IX. APPLICATIONS OF MACHINE LEARNING

1) Autonomous Vehicles Machine learning algorithms are an essential component of self-driving automobiles, and they will play an increasingly crucial role in their capacity to operate. These learning systems are widely employed for tasks such as image identification and scheduling, but they are challenging to train in chaotic real–world contexts. It's difficult to collect, analyse, and combine a lot of different sorts of data. So far, the solution has been to do extensive road testing in order to catch as many of these unusual instances as possible. Algorithms are employed to create replies based on this data, which are subsequently tested in simulations.

2) Medicine, Healthcare When it comes to diagnosis or decision making, machine learning algorithms are not a good replacement for clinicians - at least in most situations. A good diagnosis must take into account structured (e.g. diagnosis codes, medications), unstructured data (clinical notes), image data (X-rays) and even subtle visual cues from the patient (do they look ill, how did they answer family history questions) in a very short time frame.

3) Governmental machine learning Machine learning is now being used by a small number of government agencies, such as the Government Digital Services (GDS), which is using it to anticipate page views in order to find nomalies11, and the HMRC, which is using clustering algorithms to segment VAT clients. The adoption rate is still low, and there is a lot of untapped potential. GDS has thus far concentrated on proving machine learning algorithms' capabilities on a variety of goods and prototype services. The development of a 'data first' mindset at a much earlier level in the policy process is one of the first steps toward increased exploitation of these learning systems.

X. CONCLUSION

This paper introduces machine learning, its basic model, and its applications in numerous fields, as well as its benefits and drawbacks. It also examines numerous machine learning approaches and tools, such as classification and prediction techniques, as well as their objectives, working procedures, benefits, drawbacks, real-time applications, and implementation tools.

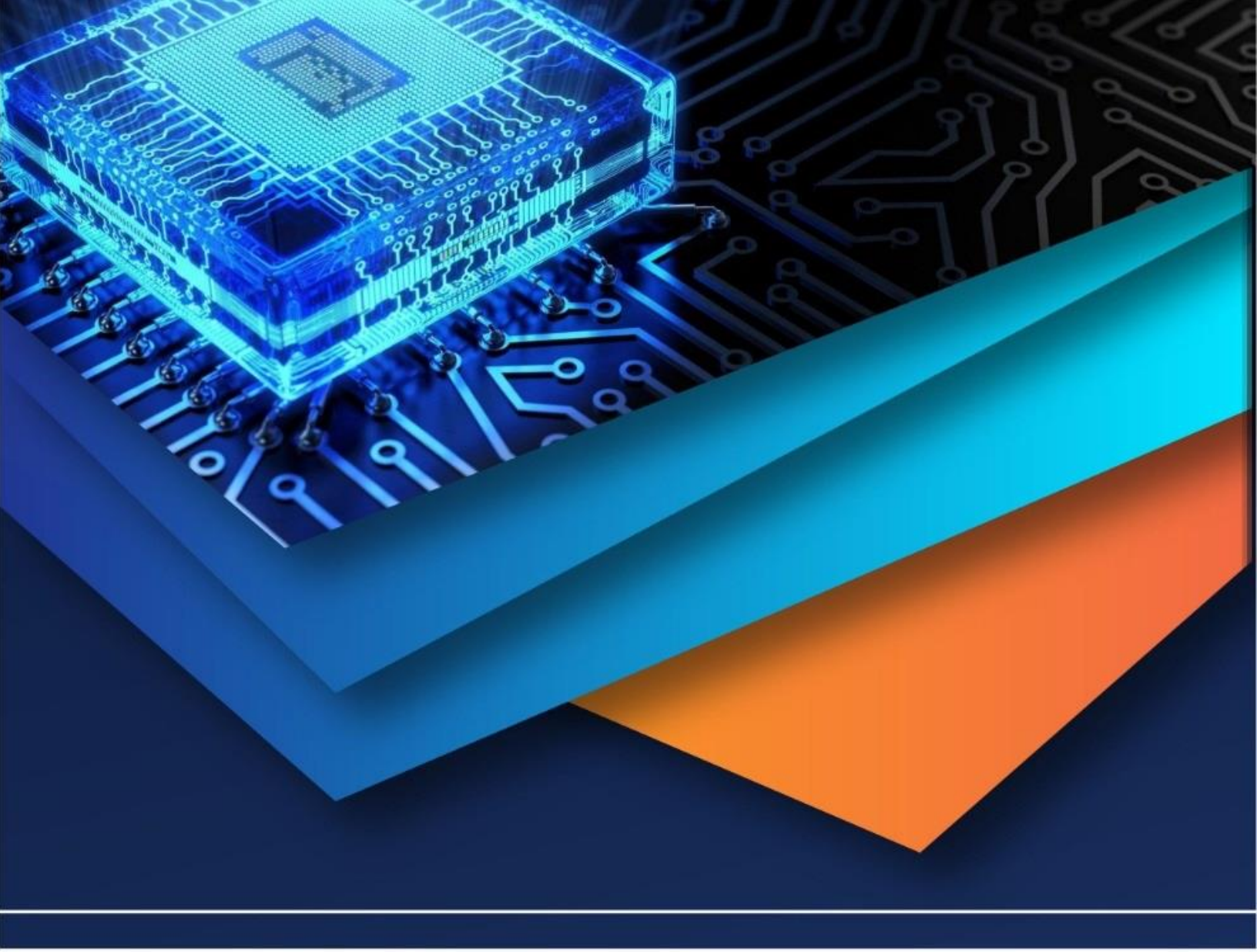
The solid underpinnings of the above-mentioned methodologies are required by emerging developments in Artificial Intelligence and machine learning, and they will be valuable in transdisciplinary domains as well.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Jafar Tanha, "Semi-supervised self-training for decision tree classifiers", International Journal of Machine Learning and Cybernetics, Volume 8, Issue 1, Page No's: 355-370, January, 2015.
[2] Khadim D, Fleur M and Gayo D, "Large scale biomedical texts classification: a k-NN and an ESA based approaches", Journal of Biomedical Semantics, 7:40, June, 2016
[3] Hong R, H. M. Wang and Jian L, "Privacy Preserving nearest Neighbour Computation in Multiple Cloud Environments, IEEE Access, ISSN: 2169-3536, Volume 4, Page No's: 9589-9603, December, 2016
[4] L. Jiang, C. Li, "Deep feature weighting for naive Bayes and its application to text classification", Journal of Engineering Applications of Artificial Intelligence, Volume 52, Page No's: 26-39, June, 2016
[5] Ahmed M, Alison H, "Modeling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm",Volume 67, PageNo's: 147- 156, January, 2018
[6] T.Razzaghi,OlegR,"Multilevel Weighted Support Vector Machine for Classification on Healthcare Data with Missing Values", PLUS ONE, Page No's:1- 18, May 2016
[7] Hui L, D. Pi, "Integrative Method Based on Linear Regression for the Prediction of Zinc binding Sites in Proteins", IEEE Access, Volume 5, Page No's:14647-14656, August, 2017.
[8] L. Wang, D. Wang, "Intelligent CFAR Detector Based on Support Vector Machine", IEEE Access, Volume 5, Page No's: 26965-26972, December, 2017.
[9] Enrico R, Michel L, "The Counter, a Frequency Counter Based on the Linear Regression", IEEE Transactions on Ultrasonics, Ferroelectrics, Volume 63, Issue 7, Page No's: 961-969, July, 2016.
[10] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques", 3rd Edition, MK Series, 2012.
[11] http://www.maxlittle.net/publications/index.php 10.
[12] https://gdsdata.blog.gov.uk/2014/08/15/anomalydete c tion-a-machine-learning-approach/ http://www.ons.gov.uk/ons/.../mwp1-onsinnovationlaboratories.pdf
[13] https://berryvilleiml.com/ results/ara.pdf
[14] X. Yu, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 9, pp. 2805–2824, 2019. doi: 10.1109/TNNLS.2018.2886017
[15] https://data-flair.training/blogs/advantagesanddisadvantages-of-machine-learning/
[16] https://www.burtchworks.com/2018/06/12/2018m achine-learning-flash-survey-results/

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)