National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

Machine Learning:available approaches,problems and solution

Manasi Sanjay Rane[1], Prof. Neha Lodhe[2]

[1]Department of Computer Application ,University of Mumbai
Viva School of MCA, Shirgaon, Virar(East)
Email: maurane8353@gmail.com

[2]Department of Computer Application, University of Mumbai
Viva School of MCA, Shirgaon, Virar(East)
Email: nehaachavan@gmail.com

Abstract**—***Humans learn from their past experiences , and machines follow instructions given by humans. But what if humans can train the machine to learn from past data and to what humans can do act much faster well that's called machine learning. But it is lot more than just learning, its also about understanding and reasoning. Machine learning can change your life provided you know how to implement right model with right algorithm. There are many approaches available to machine learning such as support vector machine, decision tree learning, Bayesian networks and many more. This paper focuses on decision tree learning, association rule learning, support vector machine , their problems and solution..*

Keywords—approaches,association rule, decision tree learning, machine learning, support vector machine.

I. INTRODUCTION

Over the past two decades Machine Learning has become one of the mainstays of information technology and with that, a rather central, albeit usually hidden, part of our life. With the ever increasing amounts of data becoming available there is good reason to believe that smart data analysis will become even more pervasive as a necessary ingredient for technological progress. Arthur Samuel (1959) o Machine learning: "Field of study that gives computers the ability to learn without being explicitly programmed" .Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed**.** Machine learning focuses on the development of computer programs that can access data and use it learn for themselves**..** The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. The complexity in traditional computer programming is in the code (programs that people write). In machine learning, algorithms (programs) are in principle simple and the complexity (structure) is in the data. Is there a way that we can automatically learn that structure? That is what is at the heart of machine learning. That is, machine learning is about the construction and study of systems that can learn from data. This is very different than traditional computer programming. In this paper we will see what are supervised machine learning approaches, their problems and solution.

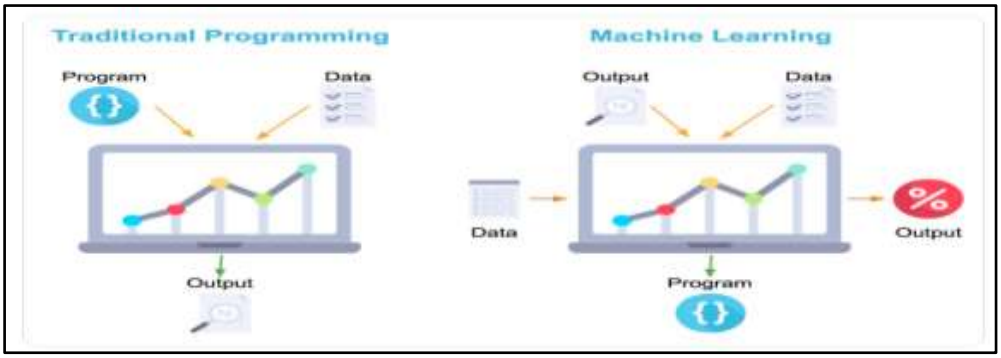


Figure. Programming vs Machine Learning

National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

.

II. TYPES OF LEARNING ALGORITHMS

Supervised learning

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. Applications in which the training data comprises examples of the input vectors along with their corresponding target vectors are known as supervised learning problems.

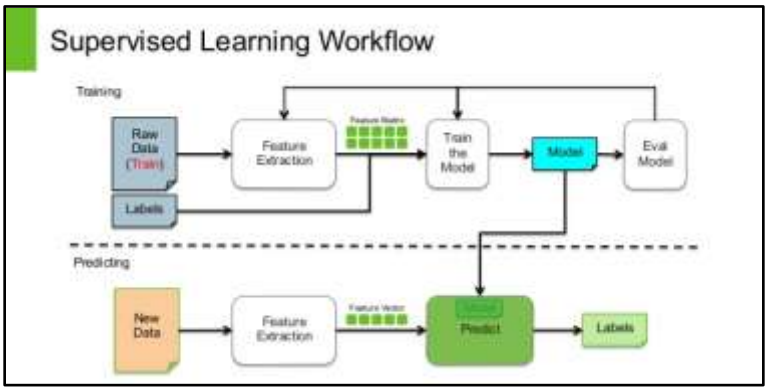
Unsupervised learning

Unsupervised learning is a type of self-organized Hebbian learning that helps find previously unknown patterns in data set without pre-existing labels**.** In unsupervised learning, there is no instructor or teacher, and the algorithm must learn to make sense of the data without this guide.

Reinforcement learning

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize some notion of cumulative reward. Reinforcement learning is learning what to do — how to map situations to actions—so as to maximize a numerical reward signal. The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them.

III. SUPERVISED LEARNING –INTRODUCTION



Supervised learning is when the model is getting trained on a labelled dataset. Labelled dataset is one which have both input and output parameters. In this type of learning both training and validation datasets are labeled.

Training the system-

While training the model, data is usually split in the ratio of 80:20 i.e. 80% as training data and rest as testing data. In training data, we feed input as well as output for 80% data. The model learns from training data only. We use different machine learning algorithms(which we will discuss in detail in next articles) to build our model. By learning, it means that the model will build some logic of its own.Once the model is ready then it is good to be tested. At the time of testing, input is fed from remaining 20% data which the model has never seen before, the model will predict some value and we will compare it with actual output and calculate the accuracy.

National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

TYPES OF SUPERVISED LEARNING

User ID	Gender	Age	Salary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	1
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	1
15728773	Male	27	58000	1
15598044	Female	27	84000	0
15694829	Female	32	150000	1
15600575	Male	25	33000	1
15727311	Female	35	65000	0
15570769	Female	26	80000	1
15606274	Female	26	52000	0
15746139	Male	20	86000	1
15704987	Male	32	18000	0
15628972	Male	18	82000	0
15697686	Male	29	80000	0
15733883	Male	47	25000	1

Figure A: CLASSIFICATION

Temperature	Pressure	Relative Humidity	Wind Direction	Wind Speed
10.69261758	986.882019	54.19337313	195.7150879	3.278597116
13.59184184	987.8729248	48.0648859	189.2951202	2.909167767
17.70494885	988.1119385	39.11965597	192.9273834	2.973036289
20.95430404	987.8500366	30.66273218	202.0752869	2.965289593
22.9278274	987.2833862	26.06723423	210.6589203	2.798230886
24.04233986	986.2907104	23.46918024	221.1188507	2.627005816
24.41475295	985.2338867	22.25082295	233.7911987	2.448749781
23.93361956	984.8914795	22.35178837	244.3504333	2.454271793
22.68800023	984.8461304	23.7538641	253.0864716	2.418341875
20.56425726	984.8380737	27.07867944	264.5071106	2.318677425
17.76400389	985.4262085	33.54900114	280.7827454	2.343950987
11.25680746	988.9386597	53.74139903	68.15406036	1.650191426
14.37810685	989.6819458	40.70884681	72.62069702	1.553469896
18.45114201	990.2960205	30.85038484	71.70604706	1.005017161
22.54895853	989.9562988	22.81738811	44.66042709	0.264133632
24.23155922	988.796875	19.74790765	318.3214111	0.329656571

Figure B: REGRESSION

1.Classification : It is a Supervised Learning task where output is having defined labels(discrete value). For example in above Figure A, Output – Purchased has defined labels i.e. 0 or 1 ; 1 means the customer will purchase and 0 means that customer won't purchase.

Example: Gmail classifies mails in more than one classes like social, promotions, updates, forum.

2. Regression : It is a Supervised Learning task where output is having continuous value. Example in above Figure B, Output – Wind Speed is not having any discrete value but is continuous in the particular range. The goal here is to predict a value as much closer to actual output value as our model can and then evaluation is done by calculating error value. The smaller the error the greater the accuracy of our regression model.

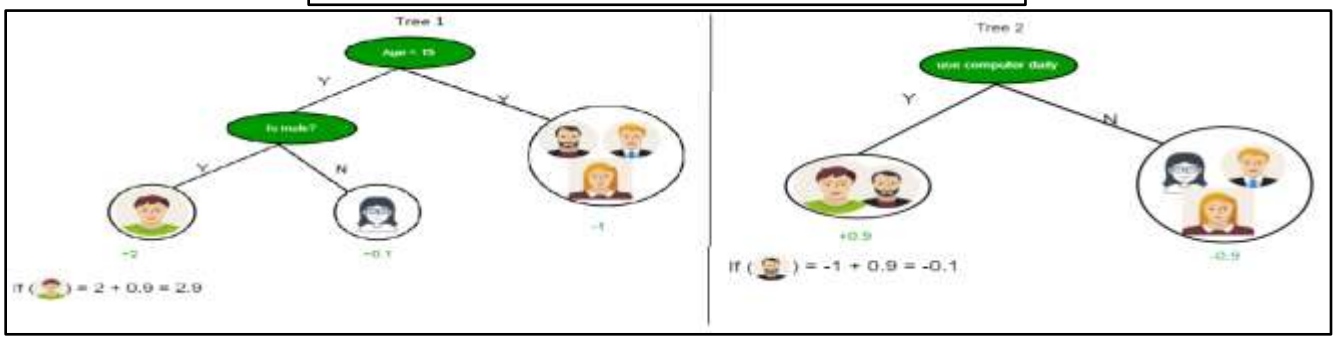
3.1 DECISION TREE LEARNING

Decision tree algorithm falls under the category of supervised learning. They can be used to solve both regression and classification problems.Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree.We can represent any boolean function on discrete attributes using the decision tree. Decision tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. A tree can be "learned" bt splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. This recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of decision tree does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision tree classifier has good accuracy.

National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

Top-Down Algorithmic Framework for Decision Trees Induction.

```
TreeGrowing (S,A,y)
Where:
S - Training Set
A - Input Feature Set
y - Target Feature
Create a new tree T with a single root node.
IF One of the Stopping Criteria is fulfilled THEN
    Mark the root node in T as a leaf with the most
    common value of y in S as a label.
ELSE
    Find a discrete function f(A) of the input
        attributes values such that splitting S
        according to f(A)'s outcomes (v1,...,vn) gains
        the best splitting metric.
    IF best splitting metric > treshold THEN
        Label t with f(A)
        FOR each outcome vi of f(A):
            Set Subtreei= TreeGrowing (σf(A)=viS,A,y).
            Connect the root node of tT to Subtreei with
                    an edge that is labelled as vi
        END FOR
    ELSE
        Mark the root node in T as a leaf with the most
            common value of y in S as a label.
    END IF
END IF
RETURN T
```



In the above image we are predicting the use of computer in the daily life of the people. In decision tree the major challenge is identification of the attribute for the root node in each level. This process is known as attribute selection. We have the following approach to solve attribute selection measure.

Entropy-

Entropy is the measure of homogeneity in the data. Its value is ranges from 0 to 1. Its value is close to 0 if all the example belongs to same class and is close to 1 if there is almost equal split of the data into different classes. Now the formula to calculate entropy

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

Information gain -
Information Gain measure the reduction in entropy by classifying the data on a particular attribute. The formula to calculate Gain by splitting the data on Dataset 'S' and on the attribute 'A' is :

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

3.1.1 PROBLEM IN DECISION TREE LEARNING: OVERFITTING

Now the main problem with decision tree is that it is prone to overfitting.We could create a tree that could classify the data perfectly or we are not left with any attribute to split. This would work well in on the training dataset but will have a bad result on the testing dataset.

Over-fitting is the phenomenon in which the learning system tightly fits the given training data so much that it would be inaccurate in predicting the outcomes of the untrained data. During the data classification process, some branches of the decision tree may contain noise or outliers in the training data and these results in a complex tree which is difficult to understand. In decision trees, over-fitting occurs when the tree is designed so as to perfectly fit all samples in the training data set. Thus it ends up with branches with strict rules of sparse data. Thus this effects the accuracy when predicting samples that are not part of the training set. One of the methods used to address over-fitting in decision tree is called pruning which is done after the initial training is complete[1].

3.1.2 SOLUTION TO THE PROBLEM : PRUNING

Pruning techniques are applied in order to remove those unwanted branches with the aim of improving the accuracy, also removing non-productive parts of the tree results in less complex tree with small size.

There are two main pruning approaches: post-pruning and pre-pruning approaches. Post-pruning is implemented after the tree is grown. In practice, post-pruning methods have better performances than pre-pruning. In pre-pruning, pruning is implemented during the tree building process and tries to stop the process when over-fitting is encountered. Hence, it prevents the generation of non-significant branches but suffers from horizon effect. Pre-pruning method navigates the tree in a top-down approach while post-pruning navigates the tree in a bottom-up approach. Nevertheless, in term of simplification and complexity post-pruning algorithm is more robust since it has access to the full tree.

Employing tightly stopping criteria tends to create small and under–fitted decision trees. On the other hand, using loosely stopping criteria tends to generate large decision trees that are over–fitted to the training set. Pruning methods originally suggested in (Breiman et al., 1984) were developed for solving this dilemma. According to this methodology, a loosely stopping criterion is used, letting the decision tree to overfit the training set. Then the over-fitted tree is cut back into a smaller tree by removing sub–branches that are not contributing to the generalization accuracy. It has been shown in various studies that employing pruning methods can improve the generalization performance of a decision tree, especially in noisy domains. When the goal is to produce a sufficiently accurate compact concept description, pruning is highly useful.

3.2 SUPPORT VECTOR MACHINE

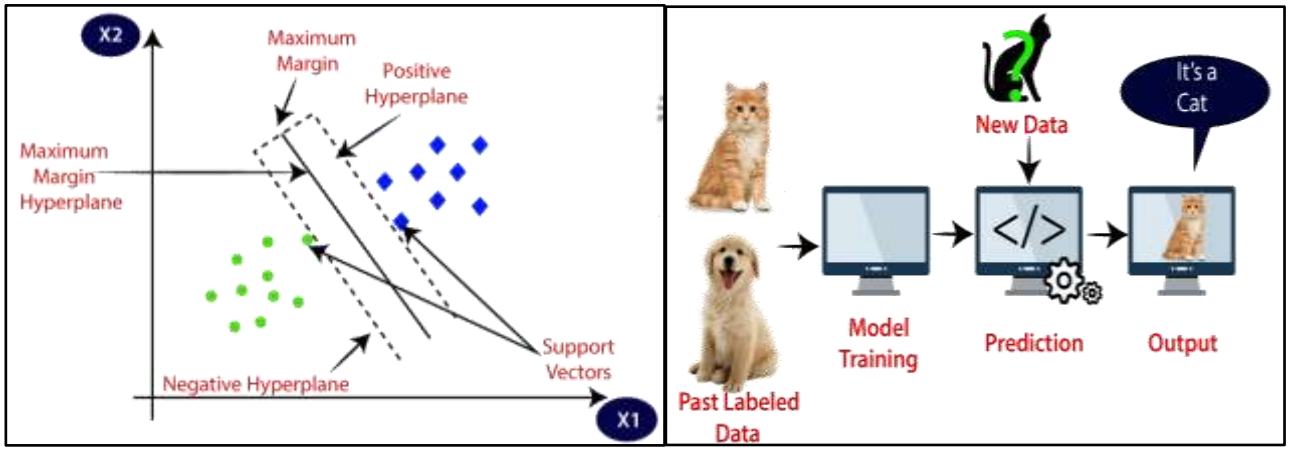
Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92.

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [1]. They belong to a family of generalized linear classifiers. In another terms, Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. Support vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

National Conference on
"Role of Engineers in Nation Building" organized by VIVA Institute of Technology, Mumbai
(6th and 7th March 2020)
International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

"A hyperplane in an n-dimensional Euclidean space is a flat, n-1 dimensional subset of that space that divides the space into two disconnected parts."

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane-



Example: SVM can be understood with the example. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat. Consider the above diagram.

3.2.1 PROBLEM IN SUPPORT VECTOR MACHINE : NON-LINEAR SEPARATION PROBLEM

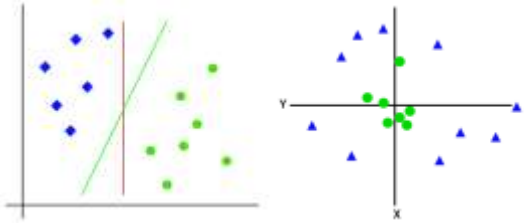


Fig. Linearly separable Fig. Non-linear

If data is linear, a separating hyper plane may be used to divide the data. However it is often the case that the data is far from linear and the datasets are inseparable. We cannot draw a straight line that can classify this non-linear data. We can't have linear hyper-plane between the two classes, so how does SVM classify the two classes? SVM can solve this problem easily! It solves this problem by introducing additional feature. Here, we will add a new feature z. For this case, if we take a new feature z as |x|. Projecting data that is not linearly separable into a higher dimensional space can make it linearly separable.
But, one of the question which arises is, should we need to add this feature manually to have a hyper-plane. No, SVM has a technique called the kernel trick[10].

3.2.2 SOLUTION TO THE PROBLEM : KERNEL FUNCTION

To resolve the above problem with non-linear data, kernels are used to non-linearly map the input data to a high-dimensional space i.e. it converts not separable problem to separable problem, these functions are called kernels. It is mostly useful in non-linear separation problem.The new mapping is then linearly separable. For this kernel functions are used. The idea of the kernel

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

function is to enable operations to be performed in the input space rather than the potentially high dimensional feature space. Hence the inner product does not need to be evaluated in the feature space. We want the function to perform mapping of the attributes of the input space to the feature space. The kernel function plays a critical role in SVM and its performance[5].

3.3 ASSOCIATION RULE LEARNING

Association Rules is one of the very important concepts of machine learning being used in market basket analysis. In a store, all vegetables are placed in the same aisle, all dairy items are placed together and cosmetics form another set of such groups. Investing time and resources on deliberate product placements like this not only reduces a customer's shopping time, but also reminds the customer of what relevant items (s)he might be interested in buying, thus helping stores cross-sell in the process[7]. Association rules help uncover all such relationships between items from huge databases. Understanding the buying patterns can help to increase sales in several ways. Association rules analysis is a technique to uncover how items are associated to each other. An association rule has 2 parts:

- an antecedent (if) and
- a consequent (then)

An antecedent is something that's found in data, and a consequent is an item that is found in combination with the antecedent. Have a look at this rule for instance:

"If a customer buys bread, he's 70% likely of buying milk."

In the above association rule, bread is the antecedent and milk is the consequent.
There are three common ways to measure association.

Measure 1: Support - This measure gives an idea of how frequent an itemset is in all the transactions. Consider itemset1 = {bread} and itemset2 = {shampoo}. There will be far more transactions containing bread than those containing shampoo. So as you rightly guessed, itemset1 will generally have a higher support than itemset2. Mathematically, support is the fraction of the total number of transactions in which the itemset occurs.

$$\text{Support}(\{X\} \Rightarrow \{Y\}) = \frac{\text{Transactions containing both X and Y}}{\text{Total number of transactions}}$$

Measure 2: Confidence - This measure defines the likeliness of occurrence of consequent on the cart given that the cart already has the antecedents. Technically, confidence is the conditional probability of occurrence of consequent given the antecedent.

$$\text{Confidence}(\{X\} \Rightarrow \{Y\}) = \frac{\text{Transactions containing both X and Y}}{\text{Transactions containing X}}$$

Measure 3: Lift - Lift controls for the support (frequency) of consequent while calculating the conditional probability of occurrence of {Y} given {X}. Lift is a very literal term given to this measure. Think of it as the *lift* that {X} provides to our confidence for having {Y} on the cart. To rephrase, lift is the rise in probability of having {Y} on the cart with the knowledge of {X} being present over the probability of having {Y} on the cart without any knowledge about presence of {X}. Mathematically,

$$\text{Lift}(\{X\} \Rightarrow \{Y\}) = \frac{\text{(Transactions containing both X and Y)/(Transactions containingX)}}{\text{Fraction of transactions containing Y}}$$

3.3.1 PROBLEM IN ASSOCIATION RULE MINING:NON INTERESTING RULES

Now that we understand how to quantify the importance of association of products within an itemset, the next step is to generate rules from the entire list of items and identify the most important ones. This is not as simple as it might sound. Supermarkets will have thousands of different products in store. After some simple calculations, it can be shown that just 10 products will lead to 57000 rules!! And this number increases exponentially with the increase in number of items. Finding lift values for each of these will get computationally very very expensive. How to deal with this problem? How to come up with a set of most important association rules to be considered? Apriori algorithm comes to our rescue for this.

3.3.2 SOLUTION TO THE PROBLEM : APRIORI ALGORITHM

Apriori algorithm is given by R. Agrawal and R. Srikant in 1994 for finding frequent itemsets in a dataset for boolean association rule. Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties. We apply an iterative approach or level-wise search where k-frequent itemsets are used to find k+1 itemsets.

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

To improve the efficiency of level-wise generation of frequent itemsets, an important property is used called Apriori property which helps by reducing the search space[8].
Apriori Property – All non-empty subset of frequent itemset must be frequent. Apriori assumes that ,all subsets of a frequent itemset must be frequent(Apriori propertry).If an itemset is infrequent, all its supersets will be infrequent.
Rule-generation is a two step process. First is to generate an itemset like {Bread, Egg, Milk} and second is to generate a rule from each itemset like {Bread → Egg, Milk}, {Bread, Egg → Milk} etc. The challenge is the mining of important rules from a massive number of association rules that can be derived from a list of items. Frequent itemsets are the ones which occur at least a minimum number of times in the transactions. Technically, these are the itemsets for which support value (fraction of transactions containing the itemset) is above a minimum threshold — minsup.
Apriori principle allows us to prune all the supersets of an itemset which does not satisfy the minimum threshold condition for support.Once the frequent itemsets are generated, identifying rules out of them is comparatively less taxing. Rules are formed by binary partition of each itemset.From a list of all possible candidate rules, we aim to identify rules that fall above a minimum confidence level (minconf). With these two steps, we have identified a set of association rules which satisfy both the minimum support and minimum confidence condition. The number of such rules obtained will vary with the values of minsup and minconf.

IV. CONCLUSION

This paperl has introduced you to Machine Learning , their possible approaches and problem. Also the solution is mentioned. Now, you know that Machine Learning is a technique of training machines to perform the activities a human brain can do, albeit bit faster and better than an average human-being. Today we can see that the machines can beat human champions in games such as Chess, AlphaGO, which are considered very complex. You have seen that machines can be trained to perform human activities in several areas and can aid humans in living better lives.Machine Learning can be a Supervised or Unsupervised. If you have lesser amount of data and clearly labelled data for training, opt for Supervised Learning. Unsupervised Learning would generally give better performance and results for large data sets. We have seen three different supervised learning algorithms which are used for classification. Each approach has some problem associated with it which can be solved.

ACKNOWLEDGEMENTS

I am thankful to my college for giving us opportunity to make this project a success. I give my special thanks and sincere gratitude towards Prof. Neha Lodhe for encouraging me to complete this research paper, guiding me and helping me out through all the problems which I encountered while doing the research.
Without her guidance, I wouldn't have completed my research paper. Also I present my obligation towards all our past years teachers who have bestowed deep understanding and knowledge in us, over the past years.

REFERENCES

[1] Overfitting in decision tree ,Ebenezer R.H.P. Isaac[2015 Sept 9] https://www.researchgate.net/post/What_is_over_fitting_in_decision_tree
[2] Association Rule Mining A Survey ,Gurneet Kaur[2014].
[3] Association Rules: Problems, solutions and new applications, María N. Moreno, Saddys Segrera and Vivian F. López , http://www.lsi.us.es/redmidas/CEDI/papers/892.pdf
[4] A Novel Decision tree Classification Based On Post-Pruning With Bayes Minimum Risk,Ahmed AM, Rizaner A, Ulusoy AH [2018 April 4] , https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0194168
[5] An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Nello Cristianini and John Shawe-Taylor, Cambridge University Press, 2000.
[6] A tutorial on support vector machines for pattern recognition, In Data Mining and Knowledge Discovery, Burges C. Kluwer Academic Publishers, Boston.
[7] Complete guide to association rules , Anisha Garg , https://towardsdatascience.com/association-rules-2-aa9a77241654
[8] Apriori Algorithm, https://www.geeksforgeeks.org/apriori-algorithm/
[9] MLSupportVectorMachine,https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_support_vector_machine.htm
[10] How SVM performs non-linear classification,Sunil[2018 June 17], https://discuss.analyticsvidhya.com/t/how-svm-performs-non-linear-classification/5664

International Journal of Engineering Research & Science (IJOER)
ISBN:[978-93-5391-287-1] ISSN: [2395-6992] [Vol-6, Issue-3, March-2020]

[11] Decision Trees in Machine Learning,Jinde Shubham[2018 June 28], https://becominghuman.ai/decision-trees-in-machine-learning-f362b296594a

[12] Study of Various Decision Tree Pruning Methods, Nikita Patel[December 2012], https://pdfs.semanticscholar.org/025b/8c109c38dc115024e97eb0ede5ea873fffdb.pdf

[13] Machine Learning: Pruning Decision Trees,Jake Hoare, https://www.displayr.com/machine-learning-pruning-decision-trees/

[14] An Introuction to Support Vector Machine,Bruno Stecanella[2017 June 22], https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/

[15] Types of Learning in Machine Learning,Jason Brownlee[2019 November 11], https://machinelearningmastery.com/types-of-learning-in-machine-learning/