Machine Learning: Concept, Algorithms and Applications

AJAYDILIPKUMAR MARAPATLA[1], DIVYA LANKA[2], SASUMANA VINAY KUMAR [3], AND SALADI LAKSHMI NARAYANA[4]
DEPT.OF CSE, S.R.K.R. ENGINEERING COLLEGE, BHIMAVARAM, AP, INDIA,
DEPT.OF CSE, VISHNU INSTITUTE OF TECHNOLOGY, BHIMAVARAM, AP, INDIA,
DEPT.OF CSE, S.R.K.R. ENGINEERING COLLEGE, BHIMAVARAM, AP, INDIA,
DEPT.OF CSE, S.R.K.R. ENGINEERING COLLEGE, BHIMAVARAM, AP, INDIA.

Abstract— Machine learning is one of the most exciting technologies that one would have ever come across. Machine learning is the embodiment of machine intelligence. Machine Learning (ML) has advanced from the undertaking of couple of computer lovers abusing the likelihood of computers learning to make amusements, and a piece of Mathematics that only here and there thought about computational methodologies, to a free research discipline that has not just given the fundamental base to factual computational standards of learning techniques, yet in addition has developed various calculations that are normally utilized for content translation, design acknowledgment, and a numerous other business purposes and has prompted a different research enthusiasm for data mining to recognize shrouded regularities or anomalies in social information that is developing in the current era. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. This paper compass around clarifying the idea and advancement of Machine Learning, a portion of the prominent Machine Learning calculations and attempt to analyze the most famous real time challenging problems dependent on some fundamental thoughts and also this paper audits the cutting edge in machine learning and gives a look at the pioneers of present machine-learning frameworks and systems.

KEYWORDS: Machine Learning, Algorithm, Data, Training, accuracy

I. INTRODUCTION

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that which makes it more similar to humans: ***The ability to learn***. Machine learning is actively being used today, perhaps in many more places than one would expect.

Machine learning endeavors to advise how to naturally find a decent indicator dependent on past encounters. Despite the fact that you may contend that machine learning has been around as long as insights has, it extremely just turned into a different point in the 1990's as the availability of Data has increased due to explosion in Smart Mobiles and devices. It draws its motivation from an assortment of scholarly trains, including computer science, insights, science, and brain research.

Machine Learning was the marvelous result when Computer Science and Statistics united. Computer Science centers around structure machines that take care of specific issues, and attempts to distinguish if issues are resolvable by any stretch of the imagination. The primary methodology that Statistics generally utilizes is information induction, demonstrating estimates and estimating unwavering quality of the ends.

Data mining, artificial intelligence and machine learning, these three words are so entwined and covering that it's nearly to draw a limit or hierarchy system among the three. Data mining is essentially about deciphering any sort of data, yet it establishes the framework for both artificial intelligence and machine learning. By and by, it tests data from different sources as well as its examinations and perceives example and connections that exists in those data that would have been hard to translate physically. Subsequently, data mining is anything but a negligible technique to demonstrate a speculation yet strategy for illustration pertinent theories. Artificial intelligence might be extensively characterized as machines those being able to take care of a given issue alone with no human mediation. The arrangements are not modified straightforwardly into the framework but rather the vital data and the AI deciphering that data produce an answer without anyone else. The translation that goes underneath is only a data mining calculation. Machine learning adopts elevate the strategy to a propelled level by giving the data basic to a machine to prepare and adjust appropriately when presented to new data. This is known as training. It centers around removing data from extensively huge arrangements of data, and after that recognizes and distinguishes basic examples utilizing different factual measures to improve its capacity to interpret new data and produce increasingly viable outcomes. Clearly, a few parameters ought to be "tuned" at the beginning level for better profitability.

II. FOR WHAT KIND OF APPLICATIONS, WE USE MACHINE LEARNING

We can apply a machine learning technique to a problem which can be broadly classified into one the four following categories as

- Problems where there is no deterministic algorithm (not even of evil complexity) e.g. Recognizing a 3D object from a given scene, Handwriting recognition, Speech recognition
- Problems which don't have a fix solution and goal posts keep changing. System adapts and learns from

experience e.g. SPAM emails, Financial fraud, IT Security Framework
- Problems where Solutions are Individual specific or time dependent. e.g. recommendations and targeted advertisements
- For prediction based on past and existing patterns (not defined or defined by huge number of weak rules) e.g. prediction of share prices etc.

III. Machine Learning in Our Daily Lives

Everyday examples of Machine Learning include
- Email Spam and Malware Filtering
- Search Engine Result Refining
- Postal Mail Routing
- Online Fraud Detection
- Product Recommendations
- Vehicle Driver Assistance
- Web Advertisements
- Social Networks
- Speech Recognition
- Virtual Personal Assistants
- Traffic Predictions
- Videos Surveillance
- Social Media Services
- Online Customer Support
- Plagiarism Checkers

IV. Machine Learning Algorithms

According to a recent report, machine learning algorithms are required to replace 25% of the occupations across the world, in the following 10 years. With the rapid growth of huge information and accessibility of programming instruments like Python and R – machine learning is picking up mainstream presence for information researchers. Machine learning applications are exceptionally robotized and self-adjusting which keep on improving over time with insignificant human intervention as they learn with more information.

To address the perplexing nature of various real-world information problems, specific machine learning algorithms have been built up that take care of these problems perfectly.

Machine Learning algorithms are broadly classified as Supervised Machine Learning Algorithms, Unsupervised Machine Learning Algorithms and Reinforcement Machine Learning Algorithms.

1) Supervised Machine Learning Algorithms:

Supervised Machine Learning Algorithms will learn from examples which provide desired outputs for given inputs. In Supervised Machine Learning Algorithms, the Data points have a known outcome. Based on these basic features the Supervised Machine Learning Algorithms are divided into following types

- **Linear Regression**
- **Logistic Regression**
- **Naïve Bayes Classifier Algorithm**
- **Support Vector Machine Algorithm**
- **Decision Trees**
- **Nearest Neighbours**
- **Neural Networks**
- **Bagging and Random Forests**
- **Boosting and Stacking**

1.1) Linear Regression

In general, the Regression techniques are used for the applications whose outcome is a continuous (numerical) one. The Regression analysis is part of predictive analytics and exploits the co-relation between dependent (target) and independent variables.

Linear Regression algorithm demonstrates the connection between 2 variables and how the adjustment in one variable effect the other. The algorithm demonstrates the effect on the reliant variable on changing the autonomous variable. The independent variables are alluded as explanatory variables, as they explain the factors the impact the dependent variable. Dependent variable is often referred to as the factor of interest or predictor. Hence the Linear Regression is used for applications where the desired output will be a real value e.g. Predicting housing price, or predicting price of a share in stock market.

In most cases we have multiple dependent variables, and we call it multiple linear regression.

1.2) Logistic Regression

The name of this algorithm could be a bit of confounding as in Logistic Regression machine learning algorithm is for classification tasks and not for the regression issues. The name 'Regression' here infers that a linear model is fit into the feature space. This algorithm applies a logistic function to a linear combination of features to predict the outcome of a categorical dependent variable based on predictor variables.

Hence the Logistic Regression is used for the applications where the desired output will be a category e.g. whether this Medical Image depicts Tumour or not, whether the selected picture belongs to cat or dog class etc.,
Based on the nature of categorical response, logistic regression is classified into 3 types as
- **Binary Logistic Regression –** The most ordinarily utilized logistic regression when the Categorical response has two possible results for example either yes or no e.g. Predicting whether an understudy will pass or bomb a test, foreseeing whether an understudy will have low or hypertension, anticipating whether a tumor is dangerous or not.
- **Multi-nominal Logistic Regression**- Categorical response has at least three potential results with no requesting. Model Predicting what sort of internet searcher (Yahoo, Bing, Google, and MSN) is utilized by dominant part of US residents.
- **Ordinal Logistic Regression**- Categorical response has at least three potential results with regular requesting e.g. Classification of a demographic image into one of the targeted cancer categories, how a client rates the administration and nature of nourishment at a café dependent on a size of 1 to 10.

1.3) Naïve Bayes Classifier Algorithm

Naïve Bayes Classifier is among the most well-known learning strategy assembled by similarities, that takes a shot at the famous Bayes Theorem of Probability-to manufacture machine learning models especially for disease prediction and record grouping. It is a basic arrangement of words dependent on Bayes Probability Theorem for emotional investigation of substance.

Naïve Bayes is a classification algorithm and is extremely fast. Naïve Bayes Classifier Algorithm is mainly used to work on problems related to Sentiment Analysis, Document Categorization, Email Spam Filtering and also used for classifying news articles about Technology, Entertainment, Sports, Politics, etc.

1.4) Support Vector Machine Algorithm

Support Vector Machine (SVM) is a supervised machine learning algorithm for classification or regression issues where the dataset instructs SVM about the classes so that SVM can arrange any new data. SVM can be group the new data of its own. It works by classifying the data into various classes by finding a line (hyperplane) which isolates the preparation data set into classes. As there are numerous such linear hyperplanes, SVM algorithm attempts to maximize the separation between the different classes that are included and this is alluded as edge amplification. In the event that the line that maximizes the separation between the classes is recognized, the likelihood to sum up well to concealed data is expanded.

As such, the algorithm builds up an ideal hyperplane using input data or preparing data and this choice plane in turns classes new models. In view of the portion being used, SVM can perform both linear and nonlinear order.

SVM's are grouped into two classes:

- Linear SVM's – In linear SVM's the preparation data for example classifiers are isolated by a hyperplane.
- Non-Linear SVM's-In non-linear SVM's it is beyond the realm of imagination to expect to isolate the preparation data utilizing a hyperplane. Under such conditions, the preparation data is too intricate that it is difficult to discover a portrayal for each element vector.

In general, SVM is used for forecasting the stock market, face detection, classification of images and handwriting recognition. It is also used field of bioinformatics for protein classification and cancer classification.

1.5) Decision Trees

A classification tree, famously known as decision tree is a standout amongst the best supervised learning algorithm. It builds a diagram or tree that utilizes spreading strategy to exhibit each likely consequence of a decision. In a decision tree portrayal, each interior node tests a feature, each branch relates to result of the parent node and each leaf at last allots the class name. To characterize a case, a top-down methodology is connected beginning at the base of the tree. For a specific feature or node, the branch agreeing to the estimation of the data point for that property is considered till a leaf is come to or a label is chosen.

A decision tree builds a tree like structure including of potential answers for an issue dependent on specific limitations. It is so named for it starts with a solitary basic decision or root, which at that point forks off into various branches until a decision or expectation is made, shaping a tree.

Sorts of Decision Trees are

- Classification Trees-These are considered as the default sort of decision trees used to isolate a dataset into various classes, in light of the target variable. These are commonly utilized when the reaction variable is categorical in nature.
- Regression Trees-When the reaction or target variable is continuous or numerical, these decision trees are utilized. These are commonly utilized in prescient kind of issues when contrasted with grouping.

Decision trees can likewise be grouped into two sorts, in view of the kind of target variable-Continuous Variable Decision Trees and Binary Variable Decision Trees. It is the objective variable that chooses what sort of decision tree would be required for a specific issue.

Applications of decision trees are option pricing in the finance sector, pattern recognition, energy consumption investigation, fraud detection. It is also used for customer relation management.

1.6) Nearest Neighbours

Nearest Neighbour algorithm is a non-parametric and lazy learning algorithm. Non-parametric means there is no assumption for underlying data distribution. In other words, the model structure determined from the dataset. This will be very helpful in practice where most of the real-world datasets do not follow mathematical theoretical assumptions. Lazy algorithm means it does not need any training data points for model generation. All training data used in the testing phase. This makes training faster and testing phase slower and costlier. Costly testing phase means time and memory. In the worst case, Nearest Neighbour needs more time to scan all data points and scanning all data points will require more memory for storing training data.

In K- Nearest Neighbour, K is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2. When K=1, then the algorithm is known as the nearest neighbor algorithm. This is the simplest case. Suppose P1 is the point, for which label needs to predict. First, you find the one closest point to P1 and then the label of the nearest point assigned to P1.
K- Nearest Neighbour performs better with a lower number of features than a large number of features. we can say that when the number of features increases than it requires more data. Increase in dimension also leads to the problem of overfitting. To avoid overfitting, the needed data will need to grow exponentially as you increase the number of dimensions.
K- Nearest Neighbour stores the whole preparing dataset which it utilizes as its representation does not learn any model.

This algorithm makes expectations without a moment to spare by figuring the closeness between an input sample and each training case. That it is a smart thought to rescale your data, for example, utilizing standardization, when utilizing Nearest Neighbours.

Applications of k-nearest neighbour algorithm are visual pattern recognition which can be used to look for hidden things. In banking sector, we can use this for credit rating and loan management**.** KNN can also be used in search engines to group the webpages so to increase the relevance rate.

1.7) Neural Networks

Neural Networks are made out of different nodes, which copy organic neurons of human cerebrum. The neurons are associated by connections and they communicate with one another. The nodes can take input data and perform straightforward activities on the data. The consequence of these activities is passed to different nodes. The yield at every node is called its activation or node value.

A model dependent on the assembled and tasks of real neural systems of people or creatures. Neural Networks are viewed as non-linear models as it attempts to find complex relationship among input and output data where the mind-boggling connections among input information sources and output results are demonstrated or some specific patterns are found. However, it draws test from data instead of thinking about the whole set and accordingly diminishing expense and time.

Data that courses through the system influences the structure of the Neural Networks in light of the fact that a neural system changes - or learns, one might say - in view of that input information and output.

Basic Structure of a Neural Network will contain at least one input layer, one output layer with variable number of hidden layers. The depth of the Neural Networks will depend on the number of hidden layers in the Neural Network.

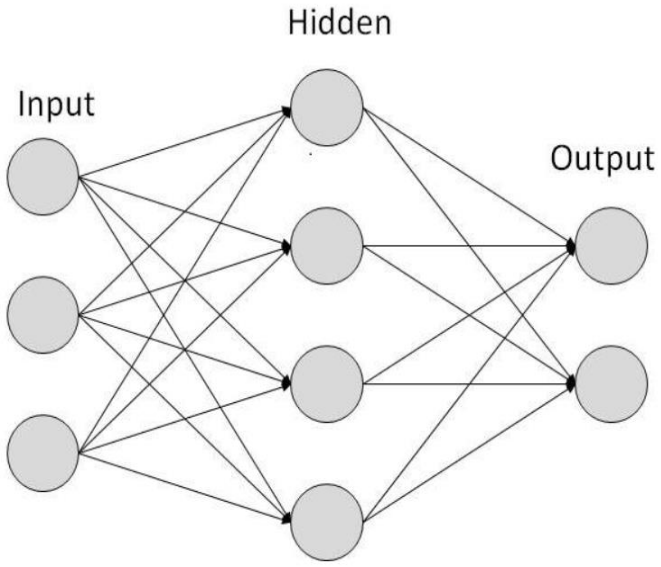

Figure 1.7.1 Basic Structure of a Neural Network

The Neural Networks are mainly categorized into four types. They are

- Feedforwarding Neural Networks
- Feedbackward Neural Networks
- Convolutional Neural Networks
- Recurrent Neural Networks

We can use neural networks in voice recognition, intelligence searching and also for credit rating.in medical field, we can use neural networks for medical diagnosis.

1.8) Bagging and Random Forests

Decision trees are very basic and require little data to develop. Likewise, decision trees are instinctive and simple to consolidate with different strategies. The essential disadvantage to a decision tree is that it will tend to overfit the preparation data, which is generally bothersome in a learning algorithm.

Random Forest is a standout amongst the most famous and most dominant machine learning calculations. Random forest is a type of ensemble machine learning algorithm called Bootstrap Aggregation or bagging. An ensemble strategy is a method that consolidates the forecasts from numerous machine learning algorithms together to make more exact expectations than any individual model.

In general Bagging ensemble algorithm and Random Forest are used for predictive modelling. When bagging with decision trees, we are less worried about individual trees overfitting the training information. Thus, and for proficiency, the individual decision trees are developed profound (for example scarcely any preparation tests at each leaf-nodes of the tree) and the trees are not pruned. These trees will have both high variance and low bias. These are significant portray of sub-models when joining predictions using bagging. The main parameters when bagging decision trees is the quantity of tests values and consequently the quantity of trees to incorporate. This can be picked by expanding the quantity of trees on pursue keep running until the precision starts to quit appearing (for example on a cross validation test saddle). Exceptionally huge quantities of models may set aside a long effort to get ready, yet won't overfit the preparation information. Much the same as the decision trees themselves, Bagging can be utilized for classification and regression issues.

The Random Forest calculation that makes a little change to Bagging and results in an exceptionally ground-breaking classifier.

The random forest is a classification calculation comprising of numerous decisions trees. It uses bagging algorithms and highlight feature randomness when building every individual tree to attempt to make an uncorrelated forest of trees whose expectation by board is more precise than that of any individual tree.

Random Forest algorithm is utilized by banks to foresee if a credit candidate is a conceivable high hazard, in the vehicle business to anticipate the disappointment or breakdown of a mechanical part.

Using random forest, we can classify images and texts, average social media shares. In automobile industry we can use for prediction of failure of automobile parts.

Application of bagging are mainly movie recommendations, intrusion detection and malware

1.9) Boosting and Stacking

Boosting techniques work in a similar soul as bagging strategies: we manufacture a group of models that are accumulated to acquire a solid learner that performs better. In any case, dissimilar to bagging that essentially goes for diminishing variance, boosting is a system that comprises in fitting consecutively various weak learners in an adaptative manner: each model in the arrangement is fitted giving more significance to data tuples in the dataset that were severely dealt with by the past models in the grouping. Naturally, each new model spotlights its endeavours on the most troublesome perceptions to fit up to now, so we get, toward the finish of the procedure, a solid learner with lower bias (regardless of whether we can see that boosting can likewise have the impact of diminishing variance). Boosting can be utilized for regression just as for classification issues.

Applications of boosting are to increase character recognition accuracy, calculating credit risk, detecting coronary diseases.

Stacking is to gain proficiency with a few distinctive weak learners and consolidate them via preparing a meta-model to yield forecasts dependent on the numerous expectations returned by these weak models. Along these lines, we have to characterize two things so as to fabricate our stacking model: The L learners we need to fit and the meta-model that consolidates them. For instance, for a classification issue, we can pick as frail students a KNN classifier, a calculated relapse and a SVM, and choose to become familiar with a neural network as meta-model. At that point, the neural network will take as sources of info the yields of our three frail students and will figure out how to return last forecasts dependent on it.

Stacking can be used for classification of cancer microarrays,

2) Unsupervised Machine Learning Algorithms:

Unsupervised learning is a sort of machine learning technique used to draw patterns from datasets comprising of unlabelled information. In Unsupervised Machine Learning Data points have unknown outcome. Here the errand of a machine is to assemble unsorted data as per likenesses, examples and contrasts with no earlier preparing of information. Based on these basic features the Unsupervised Machine Learning Algorithms are divided into following types

- **Apriori Machine Learning Algorithm**
- **K Means Clustering Algorithm**
- **Dimensionality reduction**

2.1) Apriori Machine Learning Algorithm

Apriori algorithm is an unsupervised machine learning algorithm that produces affiliation rules from a given data set. Affiliation principle infers that in the event that a thing A happens, at that point thing B additionally happens with a specific likelihood. The greater part of the affiliation standards produced are in the IF-THEN group. For instance, IF a person purchase bread THEN they likewise purchase jam Case to ensure it. For the algorithm to determine such ends, it initially watches the quantity of individuals who purchased bread case while obtaining jam.

Essential rule on which Apriori Machine Learning Algorithm works:

- In the event that a thing set happens every now and again, at that point every one of the subsets of the thing set, likewise happen much of the time.

- In the event that a thing set happens inconsistently, at that point all the supersets of the thing set have rare event.

It is anything but difficult to actualize and can be parallelized effectively. Apriori execution utilizes enormous thing set properties. In general, Apriori algorithm used for Detecting Adverse Drug Reactions, Market Basket Analysis, Auto-Complete Applications in real life.

2.2) K Means Clustering Algorithm

Clustering is a procedure that includes the gathering of data tuples. Given a lot of data values, we can utilize a clustering algorithm to characterize each point into a particular gathering. In principle, information focuses that are in a similar gathering ought to have comparative properties as well as highlights, while those in various gatherings ought to have profoundly different properties and additionally includes. The likeness between data values is normally measured by a separation metric dependent on a few kinds of highlight variable set.

K-means is a prevalently utilized unsupervised machine learning algorithm for group examination. K-Means is a non-deterministic and iterative strategy. The algorithm works on a given informational collection through pre-characterized number of bunches, k. The yield of K Means algorithm is k bunches with given data divided among the groups. K Means clustering algorithm can be connected to gather the website pages that talk about comparative ideas.

If there should arise an occurrence of globular bunches,
K-Means produces more tightly groups than various leveled clustering. Given a littler estimation of K, K-Means clustering figures quicker than various leveled clustering for enormous number of factors.

K-means clustering is an apportioning strategy that treats perceptions in your information as items having areas and separations from one another. It segments the items into K totally unrelated groups, such that items inside each bunch are as near one another as would be prudent, and as a long way from articles in other groups as could reasonably be expected. Each bunch is portrayed by its centroid, or focus point. Obviously, the separations utilized in clustering regularly don't speak to spatial separations.

K Means Clustering algorithm is utilized by the majority of the web indexes like Yahoo, Google to group pages by likeness and distinguish the 'significance rate' of query items. This helps web crawlers lessen the computational time for the clients.

2.3) Dimensionality reduction

Dimensionality Reduction alludes to the way toward changing over a lot of information having tremendous dimensions into information with lesser dimensions guaranteeing that it passes on comparative data compactly. These systems are normally utilized while tackling machine learning issues to get better highlights for a classification or relapse task.

Dimensionality Reduction helps in information packing and lessening the extra space that is required. It speeds up the time required for performing same calculations. Less dimensions prompts less registering, additionally less dimensions can permit use of algorithms unfit for countless dimensions

It takes care of multi-collinearity that improves the model execution. It expels repetitive highlights. Decreasing the dimensions of data to 2D or 3D may enable us to plot and imagine it exactly. You would then be able to watch designs all the more unmistakably. Underneath you can see that, how a 3D information is changed over into 2D. First it has distinguished the 2D plane at that point spoke to the focuses on these two new axis z1 and z2.

Dimensionality Reduction can be done by two ways. They are Linear and Non-Linear. Linear Dimensionality Reduction Methods are Principal Component Analysis, Factor analysis, Linear Discriminant Analysis. Non-linear Dimensionality Reduction Methods are Multi-dimensional scaling, Isometric Feature Mapping, Locally Linear Embedding, Hessian Eigen mapping, Spectral Embedding, t-distributed Stochastic Neighbor Embedding.

3) Reinforcement Learning Algorithms:

Reinforcement learning is a significant sort of Machine Learning where an agent figures out how to carry on in a domain by performing actions and seeing the outcomes.

Reinforcement learning known as a semi-supervised learning model in machine learning, is a computational methodology used to comprehend and computerize the objective coordinated learning and basic leadership. It is recognized from other computational methodologies by its accentuation on taking in by the person from direct association with its condition, without depending upon some predefined named dataset.

The learning model isn't advised which moves to make, as in many types of machine learning, yet rather, must find which activities yield the most reward by attempting them. In the most fascinating and testing cases, activities may influence the quick reward as well as the following circumstance and, through that, every single ensuing prize. These two attributes: trial-and-error search and delayed reward are the distinctive highlights of Reinforcement Learning. So, it is a technique to allow an agent to take actions and interact with an environment so as to maximize the total rewards. Reinforcement learning is usually modeled as a Markov Decision Process.

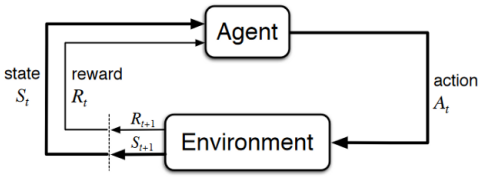

Figure 3.1: The agent–environment interaction in a Markov decision process.

Source: Reinforcement Learning: An Introduction.

In daily life we are implementing the concept of Reinforcement Learning in many fields. Some of those are listed Resources management in computer clusters, Traffic Light Control, Robotics, Web System Configuration, Chemistry, Personalized Recommendations, Bidding and Advertising, Game Design.

V. CONCLUSION

The first focus of Machine Learning analysts is to plan increasingly productive (as far as both time and space) and useful universally useful learning techniques that can perform better over an across the board area. With regards to Machine Learning, the proficiency with which a technique utilities information asset that is additionally a significant presentation worldview alongside reality multifaceted nature. Higher precision of expectation and humanly interpretable forecast principles are likewise of high significance.

Improvement of a product to explain observation errands utilizing sensors, similar to discourse acknowledgment, computer vision and so forth. It is simple for anybody to mark a picture of a letter by the letter set it signifies, however planning a calculation to play out this errand is troublesome.

Machine Learning gives a product the adaptability and versatility when vital. Notwithstanding some application (e.g., to compose framework augmentation programs) where Machine Learning may neglect to be valuable, with increment of information assets and expanding request in customized customization programming, Machine Learning will flourish in not so distant future. Other than programming advancement Machine Learning will most likely yet help change the general viewpoint of Computer Science. By changing the characterizing question from "how to program a PC" to "how to engage it to program itself," Machine Learning cloisters the advancement of gadgets that are self-checking, self-diagnosing and self-fixing, and the utilities of the information stream accessible inside the program as opposed to simply preparing it. In like manner, it will help change Statistical principles, by giving progressively computational position. Clearly, the two Statistics and Computer Science will likewise adorn Machine Learning as they create and contribute more advanced theories to adjust the method for learning.

REFERENCES

[1] T.M. Mitchel, The Discipline of Machine Learning, CMU-ML-06-108, 2006

[2] T. Mitchell, W. Cohen, E. Hruschka, P. Talukdar, J. Betteridge, A. Carlson, B. Dalvi, M. Gardner, B. Kisiel, J. Krishnamurthy, N. Lao, K. Mazaitis, T. Mohamed, N. Nakashole, E. Platanios,A. Ritter, M. Samadi, B. Settles, R. Wang, D. Wijaya, A. Gupta, X. Chen, A. Saparov,M. Greaves, J. Welling, Never-Ending Learning, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2014

[3] W. Dai, G. Xue, Q. Yang, and Y. Yu, Transferring Naive Bayes classifiers for text classification. AAAI Conference on Artificial Intelligence, 2007.

[4] P. Donmezand J. G. Carbonell, Proactive learning: cost-sensitive active learning with multiple imperfect oracles. In Proceedings of the 17th ACM conference on In- formation and knowledge management, 619–628. ACM, 2008

[5] A Survey on Machine Learning: Concept, Algorithms and Applications, Kajaree Das, Rabi Narayan Behera, February 2017.

[6] Gregory, P. A. and Gail, A. C. Self-supervised ARTMAP Neural Networks, Volume 23, 265-282, 2010

[7] https://www.dezyre.com/article/top-10-machine-learning-algorithms/202.

[8] https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12

[9] S.-B. Cho and H.-J. Park. Efficient anomaly detection by modeling privilege flows using hidden Markov model. Computers & Security, 22(1):45–55, Jan. 2003.

[10] B. Biggio, B. Nelson, and P. Laskov. Poisoning attacks against support vector machines. In Proceedings of the 29th International Conference on Machine Learning, ICML'12, pages 1467–1474, USA, 2012. Omnipress.

[11] C. Annachhatre, T. H. Austin, and M. Stamp. Hidden Markov models for malware classification. Journal of Computer Virology and Hacking Techniques, 11(2):59–73, 2015.

[12] S. Attaluri, S. McGhee, and M. Stamp. Profile hidden Markov models and metamorphic virus detection. Journal in Computer Virology, 5(2):151–169, 2009.

[13] N. Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, 2000.

[14] G. Gu, R. Perdisci, J. Zhang, and W. Lee. Botminer: Clustering analysis of network traffic for protocol- and structure-independent botnet detection. In Proceedings of the 17th ConferenceonSecuritySymposium, SS'08, pages139–154, Berkeley, CA, USA, 2008.USENIX Association.

[15] W. Hu. Robust support vector machines for anomaly detection. In Proceedings of 2003 International Conference on Machine Learning and Applications, ICMLA03, pages 23–24, 2003.

[16] E. Mooi and M. Sarstedt. Cluster analysis. In A Concise Guide to Market Research, chapter9, pages 237–284. Springer, Berlin, 2011.