iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** VI **Month of publication:** June 2019

DOI: http://doi.org/10.22214/ijraset.2019.6138

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: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

A Review on Machine Learning Algorithms

Anushree Ra[1], Rio D. Souza[2]
[1, 2]*Department of M.Sc. Big Data Analytics, Department of Computer Science & Engineering*

Abstract: Machine learning is a field of computer science which gives computers an ability to learn without being explicitly programmed. Machine learning is used in a variety of computational tasks where designing and programming explicit algorithms with good performance is not easy.
Machine Learning (ML) is a vast interdisciplinary field which builds upon concepts from computer science, statistics, cognitive science, engineering, optimization theory and many other disciplines of mathematics and science. Machine learning algorithms consist of identifying and validating models to optimize a performance criterion using historical, present, and future data. A large number of techniques has been developed so far to tell the diversity of machine learning. The proposed paper briefly explains the various types of machine learning algorithms.
The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically.
Keywords: Machine learning, supervised learning, unsupervised learning, reinforcement learning

I. INTRODUCTION

Machine learning is a branch of artificial intelligence that allows computer systems to learn directly from examples, data, and experience. It refers to the automated detection of meaningful patterns in data. Increasing data accessibility has endorsed machine learning systems to be trained on a bulky pool of examples, while growing computer processing power has supported the critical capabilities of these systems.
As human are often incapable of expressing what they know so for this evolution of machine learning system came into lime light. The primary goal of Modern Machine Learning is highly accurate predictions on test data. It attempts to tell how to automatically find a good predictor based on past experiences. [1]. There are several applications of Machine Learning; the main of all is Data mining. Every instance in a data-set used by these algorithms is represented using same set of features. This paper explains about machine learning in the second section and discusses different types of algorithms in the third section which follows with the conclusion.

II. MACHINE LEARNING

Machine learning is the process of teaching a computer system how to make accurate predictions when fed data. Machine learning attempts to tell how to automatically find a good predictor based on past experiences [2].
Learning (training): Learn a model using the training data.
Testing: Test the model using unrevealed test data to appraise the model accuracy by itself, continually using trial and error. This machine learns from its past experience and tries to capture the best possible knowledge to make accurate business decisions Such as Markov Decision Process. It learns to select an action to maximize payoff. Timely the algorithm changes its strategy to learn better and the best decision and accuracy. [3]

A. Machine Learning Can Be Applied To Situations Where

1) A machine is used for a specific task to be completed - obviously. The *machine* term here is an all-encompassing entity as it would involve some app/program/appliance/device/system that is being used.
2) The behavior is repeatable and predictable so that past data can be used to predict the future actions.
3) Behavior is pattern-based or rule-driven so that it can be “taught” to the machine. That will help machines to learn and match data against a pattern to take actions or decisions.
4) Large volume of data is being processed. This will typically hold true when it is humanly impossible to look at the specific data elements in the sea of data to identify a potential issue or problem.

B. Any Ml Process Would Have The Following Key Steps

1) *Data collection & Preparation:* You will need to identify the right data sources and prepare data that can be fed to the ML algorithm to be used.

2) *Learning:* This includes a key aspect of choosing the right ML algorithm to be used to get the right results. You would use the dataset as the outcome of step 1 here, to feed to the model so that it learns accordingly. The aspects considered while choosing the right algorithm could be:

a) Complexity of the data

b) Choice of data set

3) *Prediction:* In this step, you would use the model or algorithm to predict outcomes for you based on the data model supplied or provided.

III. MACHINE LEARNING ALGORITHMS

Machine learning tasks are typically classified into three broad categories: Supervise learning, Unsupervised learning, Semi Supervise learning and Reinforcement learning as shown in Fig 1.

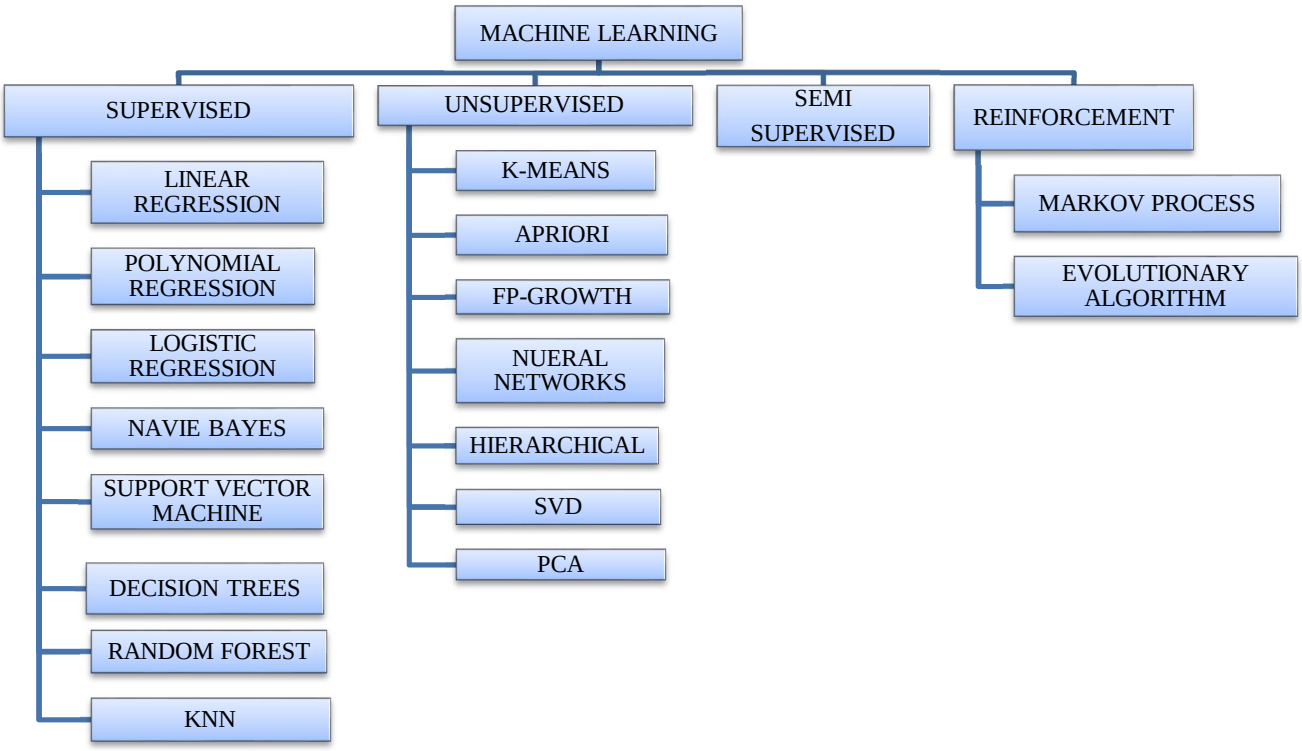


Fig. 1 Classification of Machine Learning algorithms

A. *Supervised Machine Learning Approach (SML)*

It is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. Machine learning algorithms are organized into a taxonomy based on the desired outcome of the algorithm [4]. Supervised learning generates a function that maps inputs to desired outputs. Supervised learning is fairly common in classification problems because the goal is often to get the computer to learn a classification system that we have created [5]. One standard formulation of the supervised learning task is the classification problem: The learner is required to learn a function which maps a vector into one of several classes by looking at several input output examples of the function. A supervised learning algorithm analyzes the training data and produces an inferred function that can be utilized for mapping fresh examples [6]. They can be categorized into two main types with a set of various types of algorithms under it.

1) *Linear Regression:* It's a linear model that assumes a linear relationship between the input variables (x) and the single output variable (y). It specifies set of input values (x) the solution to which is the predicted output (y) for that set of input values. Here, we establish relationship between independent and dependent variables by fitting a best line.

2) *Polynomial Regression:* It is used to estimate discrete values based on given set of independent variable(s). It's a form of linear regression in which the relationship between the independent variable x and dependent variable y is modeled as nth degree polynomial.

3) *Logistic Regression:* Logistic Regression is also called sigmoid function which forms S-Shaped curve on the graph that can take any real valued numbers and map it into a value between 0 and 1. In simple words, it predicts the probability of occurrence of an event by fitting data to a logic function.
4) *Navie Bayes:* It's a classification algorithm for binary and multi-class classification problems. It calculates the probabilities for each hypothesis rather than of each attribute values.
5) *Support Vector Machine:* SVM are simply the coordinates of individual observation. SVM is a frontier which best segregates the two classes. It finds a hyperplane in an N-dimensional space (N- number of features) that distinctly classify the data points. Hyperplanes are decision boundaries that help classify the data points. Data appoints falling on either side of the plane can be attributed to different classes.
6) *Decision Trees:* Decision Tree is a type of supervised learning algorithm that is mostly used for classification problems. It is a binary tree from algorithms and data structures, where data is continuously split according to a certain parameter, the tree can be explained by two entities namely decision nodes and leaves. The leaves are the decision of the final outcomes.
7) *Random Forest:* It creates a forest and makes it somehow random. The forest it builds is an assemble of Decision Trees, most of the time trained by 'bagging' method. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.
8) *K NN:* K nearest neighbors is a simple algorithm which stores the entire available cases and classifies new cases by a majority vote of its k neighbors. It can be used to solve both classification and regression problem. It refers on labeled input data to learn a function that produces an appropriate output when given new unlabeled data. KNN makes prediction for new instance (x) by searching through the entire training set for the K most similar instances (neighbors) and summarizes the output variable for those K instances [7].

B. Unsupervised Learning

This technique essentially addresses a specific need of machine learning where we have some understanding of output to be generated from input. However, this is not true for all the data. So, this technique is essentially a combination of supervised and unsupervised learning. It is the machine learning task of inferring a function to depict concealed structure from "unlabeled" data. The goal of unsupervised learning is to discover patterns of regularities and irregularities in a set of observations. Unsupervised machine learning used to draw conclusions from datasets consisting of input data without labeled responses [8] or we can say in unsupervised learning desired output is not given. Since the examples specified to the learner are unlabeled, there is no assessment of the accuracy of the structure that is output by the relevant algorithm. We can further classify unsupervised Learning into K Means Clustering, Apriori Algorithm, FP Growth Algorithm, Neural Networks, Hierarchical Clustering, Singular Value Decomposition and Principal Component Analysis

1) K Mean Clustering: K-mean is a partitioned - clustering algorithm. It aims to partition the given n observations into K clusters The mean of each cluster is found and the image is placed in an cluster, whose mean has the least Euclidean distance with the image feature vector. Due to the complex distribution of the image data, the k-mean clustering often cannot separate images with different concepts well enough [9]. It aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster K-means algorithm identifies K number of centroids, and then allocates every data point to the nearest cluster, which keeping the centroids as small as possible [10]. K-means algorithm in data mining starts with a first group of randomly selected centroids, and then performs iterative calculation to optimize the position of the centroids.
2) *Apriori Algorithm:* Apriori is a basic machine learning algorithm which is used t sort information into categories. It identifies a particular characteristics of a datasets and attempting to note how frequently that characteristics pops up throughout the set. A frequent data characteristics s one that occurs above the pre-arranged amount known as support. Apriori is mainly used for sorting large amounts of data. Sorting data often occurs because of associated rules [11]. Rules help show what aspect of data different sets have in common. Apriori can be used as a basis for an artificial neural network. It can help the network make sense of large recons of data and sort data into categories by frequency almost instantaneously.
3) *FP Growth Algorithm: It* is an alternative way to find frequent elements without using candidate generation, thus improving performances. Algorithm first compresses the input database creating a FP – tree instance to represent frequent items. Next, it divides the compressed data base into a set of conditional databases, each one associated with one frequent pattern [12]. Finally, each data base is mixed separately.

4) *Neural Networks:* NN are set of algorithms, modeled loosely after the human brain, that are designed to recognize patterns [13]. They interpret sensory data through a kind of machine perception labeling o clustering raw input Neural Networks. NN are a subset of algorithms built around a model of artificial neurons spread across three or more layers[14].
5) *Hierarchical Clustering:* It is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types.
a) Agglomerative (bottom up) approach, here the pairs of clusters are merged as one moves up the hierarchy.
b) Division (top-down) approach, here the split are performed recursively as one moves down the hierarchy.
6) Singular Value Decomposition: SVD states that any matrix A can be decomposed into 3 matrix A=U Sigma V^T. it s a matrix decomposition method for reducing a matrix to its constitute parts in order to make certain subsequent matrix that we wish to decompose, U is an m x n diagonal matrix and V^T is the transpose of n x n matrix where T is a superscript. The diagonal values in the Sigma matrix are known as the singular values of the original matrix A. The column of U matrix are called the Left – singular vector of A.
7) Principal Component Analysis: It reduces the dimensionality of the data set consisting of many variables correlated with each other, either heavily or lightly, while retaining the variation present in the dataset, up to the maximum extent. The dataset on which PCA technique is to be used must be scaled. The results are also sensitive to the relative scaling [15]. It's a method of summarizing data. PCA can be defined as a linear combination of optimally – weighted observed variables. The output of PCA are their principal components, the number of which is less than or equal to the number of original variables.

C. *Semi-Supervised Learning*

Semi-supervised learning is used to build models from a dataset with incomplete labels. It is a class of machine learning tasks and techniques that also make use of unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data [16]. Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. The acquisition of labeled data for a learning problem often requires a skilled human agent or a physical experiment. The cost associated with the labeling process thus may render a fully labeled training set infeasible, whereas acquisition of unlabeled data is relatively inexpensive [17]. In such situations, semi-supervised learning can be of great practical value. Semi-supervised learning is also of theoretical interest in machine learning and as a model for human learning.

D. Reinforcement Learning

Reinforcement Learning is a type of *Machine Learning* which allows machines and software agents to automatically determine the ideal behavior within a specific context, in order to maximize its performance [18]. Simple reward feedback is required for the agent to learn its behavior. We can classify the reinforcement learning into two categories the Markovian and Evolutionary reinforcement learning.

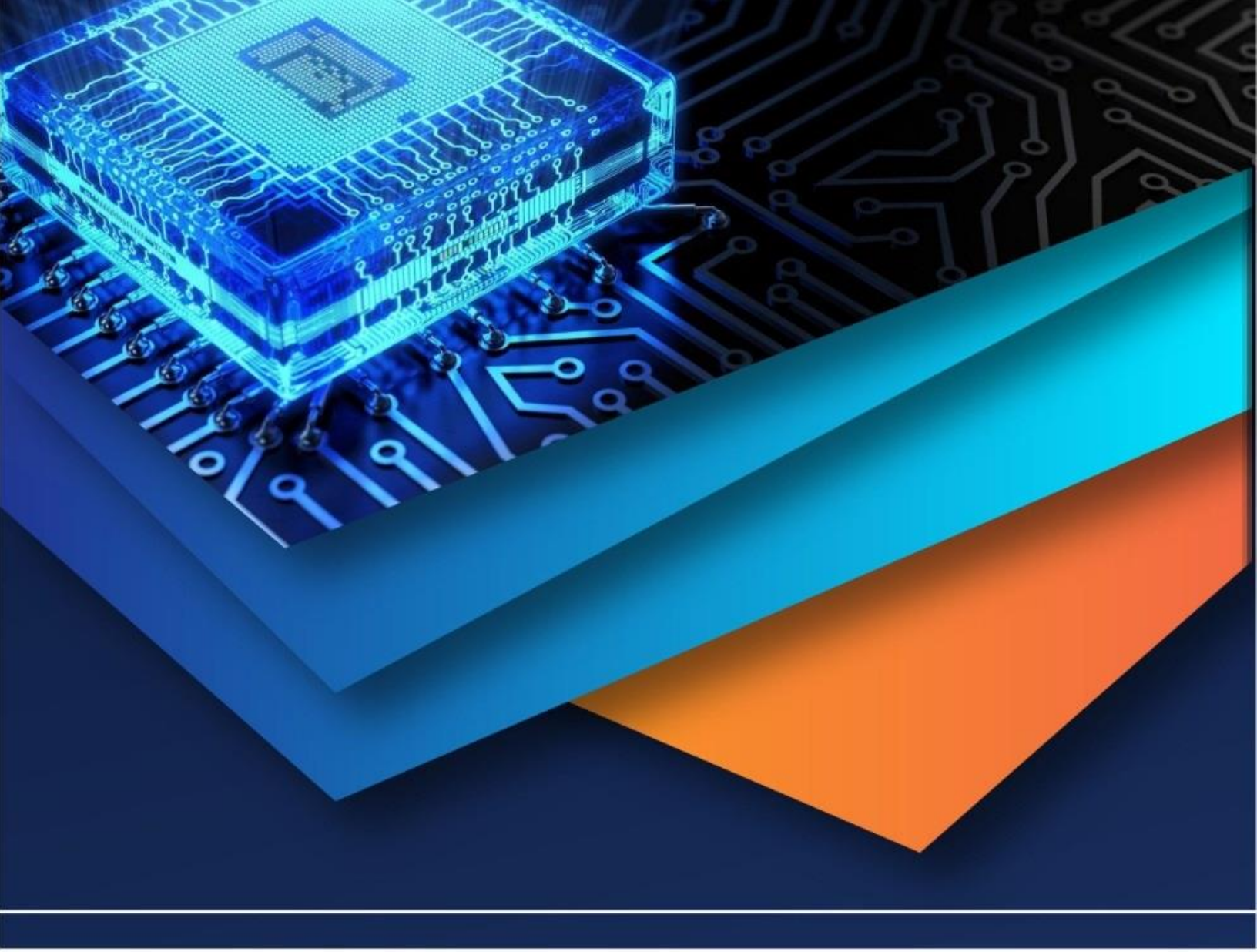
1) Markov Decision Process: In the given problem, an agent decides the best action to select based on his current state. This step is repeated many a times. These are modeling sequences with discrete states. Given a sequence, we might want to know, what most likely character to come next is, or what is the most probability of a given sequence. Markov chain only works when the states are discrete[19]. To train the transition probabilities for a markov chain where the output observation is a random variable x generated according to an output probabilistic function associated with each state.
2) Evolutionary algorithm: It works in a entirely different way than neural networks. The goal is to create computer code that solves a specific problem using an approach that is to create computer code that solves a specific problem using an approach. Evolutionary algorithm starts with code generation entirely at random. Each of these codes is tested to see whether it achieves the required goal. In this way, the code evolves. Overtime, it becomes better, and after many generations, if conditions are right, it can become better than any human coder can design. It consists of maintaining a distribution over network weight values, and having a large number of agents act in parallel using parameters sampled from this distribution. Each agent acts in its own environment, and once it finishes a set number of episodes, or steps of an episode, cumulative reward is returned to the algorithm as a fitness score. With this score, the parameter distribution can be moved toward that of the more successful agents, and away from that of the unsuccessful ones. By repeating this approach millions of times, with hundreds of agents, the weight distribution moves to a space that provides the agents with a good policy for solving the task at hand [20].

IV. CONCLUSION

Machine learning techniques are being widely used to solve real-world problems by storing, manipulating, extracting and retrieving data from large sources. Machine Learning is an incredibly powerful tool. In this paper we give a brief overview on what is machine learning and the different types of algorithms under various categories of Machine learning. The discussed techniques can be implemented on different type of data set i.e. health, financial etc. It is difficult to find out which technique is superior to other because each technique has its own merits, demerits and implementation issues. Besides software development, Machine learning will probably but help reform the general outlook of Computer Science. By changing the defining question from "how to program a computer" to "how to empower it to program itself," Machine learning priories the development of devices that are self- monitoring, self-diagnosing and self-repairing, and the utilization of the data flow available within the program rather than just processing it. Likewise, it will help reform Statistical rules, by providing more computational stance.

REFERENCES

[1] M. Welling, "A First Encounter with Machine Learning"
[2] P. Harrington, "Machine Learning in action", Manning Publications Co., Shelter Island, New York, 2012
[3] [Online]. Available: www.analyticsvidhya.com
[4] S.B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica 31 (2007) 249-268
[5] Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, InTech, University of Portsmouth United Kingdom. Pp 3 – 31. Available at InTech open website: http://www.intechopen.com/books/new-advances-inmachine-learning/types-of-machine-learning-algorithms
[6] http://www.simplilearn.com/what-is-machine-learning-and-why-itmatters-article
[7] T. Mitchell. Machine Learning. Boston: McGraw-Hill, 1997.
[8] Zhang D, Nunamaker JF. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information Systems Frontiers 2003; 5: 207-218. https://doi.org/10.1023/A:1022609809036
[9] http://pypr.sourceforge.net/kmeans.html
[10] K. Alsabati, S. Ranaka, V. Singh, "An efficient k-means clustering algorithm", Electrical Engineering and Computer Science, 1997 [14] M. Andrecut, "Parallel GPU Implementation of Iterative P
[11] R. Caruana, "Multitask Learning", Machine Learning, 28, 41-75, Kluwer Academic Publishers, 1997
[12] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of Artificial Intelligence Research, 11, Pages 169- 198, 1999
[13] V. Sharma, S. Rai, A. Dev, "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN 2277128X, Volume 2, Issue 10, October 2012
[14] S. B. Hiregoudar, K. Manjunath, K. S. Patil, "A Survey: Research Summary on Neural Networks", International Journal of Research in Engineering and Technology, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389, May, 2014
[15] https://en.wikipedia.org/wiki/Instance-based_learning
[16] X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, Vol. 3, No. 1, Pages 1-130
[17] X. Zhu, "Semi-Supervised Learning Literature Survey", Computer Sciences, University of Wisconsin-Madison, No. 1530, 2005
[18] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992
[19] L. P. Kaelbing, M. L. Littman, A. W. Moore, "Reinforcement Learning: A Survey", Journal of Artificial Intelligence Research, 4, Page 237-285, 1996
[20] Z. H. Zhou, "Ensemble Learning", National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY