Machine Learning aspects and its applications towards different research areas

Yogesh Kumar
Dept. of CSE, Chandigarh Engineering College, Landran, Mohali.
Yogesh.arora10744@gmail.com

Komalpreet Kaur
Dept. of ECE, Punjabi University, Patiala.
chadhakomal18@gmail.com

Gurpreet Singh
Dept. of CSE, Chandigarh University, Gharuan, Mohali.
Gps_ghotra@yahoo.com

Abstract— **Machine learning is a branch of artificial intelligence that aims at enabling machines to perform their jobs skillfully by using intelligent software. The statistical learning methods constitute the backbone of intelligent software that is used to develop machine intelligence. Now a Day, a huge increase in demand for machine learning has been seen with the great number of available datasets. The knowledge acquisition mechanizing from experience improvement using computational methods comes under machine learning. There is need of knowledge specific to the domain by expert performance and number of AI expert systems has been produced by knowledge engineering. Its regular use has been seen in the industry in different domains. Due to the increase in use and applicability of machine learning, the systematic review on various aspects related to it has been presented in this paper. The paper started with giving a brief description of machine learning, and the use of different models of machine learning. The various types of machine learning algorithms that are used for various purposes like data mining, predictive analytics, image processing etc. has also presented in the comprehensive review. We have also given a review of different work done by various researchers in different application areas. It covered the use of machine learning in medical, social media, Travelling and robotics. The primary purpose behind its popularity in the different application is its ability to learn once and then it works automatically for any same type of data or input given to it.**

Keywords— *Machine learning, Data mining, Social media, Medical, Travelling, Robotics*

I. INTRODUCTION

In human learning and reasoning intelligence machines exhibited Artificial Intelligence (AI) is prove to be an effective approach [1]. How a cognitive reasoning like human can be perform by computer. In 1950, its explanation is given by "The turning Test". AI is divided into specific research sub-fields like Natural Language Processing (NLP) that helps in improving the writing experience in different applications. How to handle data efficiently is taught to machines using machine learning. In most of the cases data cannot be extracted and interpreted after reviewing it. In those cases, machine learning is applied [2]. A huge increase in demand of machine learning has been seen with the huge number of available datasets. A relevant information extraction is done by machine learning in different industries from medicine to military application

Machine translation algorithms have resulted in various applications that consider grammar structure as well as spelling mistakes. When editor or writer said to do changes in computer then there is automatically use of set of vocabulary and words. Now a days use of data mining and machine learning has become the center of attention in most research community. In this field they both collectively analyze the databases characterization into multiple possibilities [3]. From years databases have been collected with statistical purposes. Using it future behaviours can be predicted by describing the present and past using statistical curves.

The knowledge acquisition mechanizing from experience improvement using computational methods comes under machine learning. There is need of knowledge specific to domain by expert performance and number of AI expert systems has been produced by knowledge engineering. Its regular use has been seen in industry. In knowledge engineering process an increase level of automation has been provided by machine learning and replacing the time consumed in human activity with automatic technique. It helps in improving efficiency and accuracy by exploiting regularities in training data after discovering it. As compared to learning without test data use of other gives better performance.

A. Novelty in the Article

Nowadays the use of machine learning gives better results. So, use of machine learning has been found in almost every field. Due to its popularity, various researchers have worked on it but still they are unable to give a good paper on machine learning approaches and various applications in which it has been used and given good results.

In this paper, we have focused on different paradigms used in machine learning. Second section of this paper describes the model of machine learning and different algorithms that can be used for performing learning process. The third section gives review on different applications area where machine learning has been used by various researchers. Then in the last section we have given the conclusion of complete paper.

II. MODEL OF MACHINE LEARNING AND DIFFERENT TYPES OF LEARNING

A. Model of Machine Learning

In Machine learning, learning process is divided into training and testing. Firstly, we train the system for some of the iterations then it is tested. In training process, input is taken of training data samples in which features are learned by learning algorithm and learning model is built. Then it is

tested in which execution engine is used by learning model to make the prediction for production or test data. Learning model output is known as tagged data which gives output as classified or final prediction data.
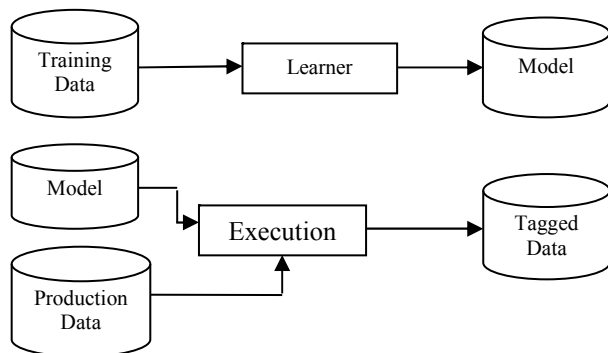

Fig. 1. Machine learning operational model

B. Different Types of Learning

It is divided into different categories. In this section we have given review of different types of learning.

1. **Supervised learning:** In this type of machine learning algorithm there is need of external assistance. Train and test dataset are two categories in which input dataset is divided. There is need to classify and predict the output variable of train dataset. From training dataset some kind of patterns is learn by these algorithms and then classification or prediction is done by applying test dataset [4]. Under this category there are three types of machine learning algorithms given below:

Decision Tree: In this type of tress where group of attributes are made by sorting them according to their values. In classification use of decision tree prove to be efficient and nodes and branches are part of tree. The value taken by node is represented by each branch and in a group attributes are represented by each node that is further classified [4]. Dt Inducer, Tree growing and Tree pruning are three main steps.
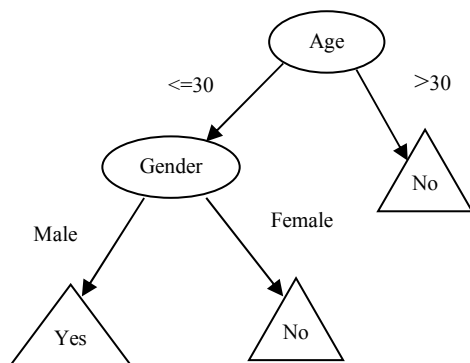

Fig. 3. Decision Tree [9]

Naive Bayes: It mainly targets the industry of test classification. Its use is mainly found in classification and clustering purpose [5]. It mainly depends on conditional probability and on the basis of probability of happening an event a trees is

created in it. Bayesian network is another name of these kinds of trees.

Support Vector Machine: This is the third and most widely used type comes under supervised learning. It is mainly used in classification and margin calculation is the principle followed in it. Margins are drawn between the classes and they are drawn in such a fashion that distance between margin and classes become maximum that results in minimization of classification error.
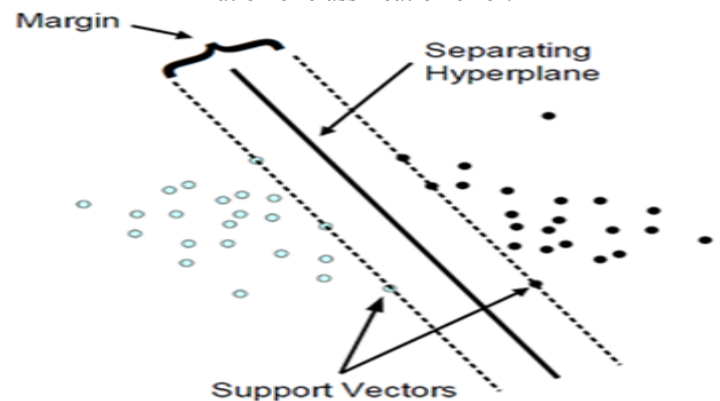

Fig. 4. Working of Support Vector Machine [6]

2. **Unsupervised Learning:** Few features from data are learnt by unsupervised learning algorithms. In case when new data in introduced the data class is recognized using previously learned features. Mostly its use has been found in feature reduction and clustering. There are two main algorithms comes under unsupervised category that is used for dimensionality reduction and clustering.

K-Means Clustering: grouping or clustering is a type of unsupervised learning technique in which groups are automatically created whenever it initiates. Similar characteristics items are put into one cluster. In this algorithm k distinct clusters are created thats why it is also known as k-means algorithms. The mean value in particular cluster is measured as center of that cluster [7].
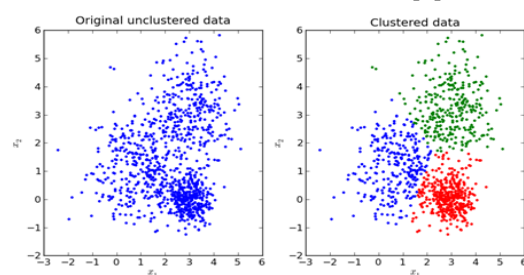

Fig. 5. K-Means Clustering [7]

Principal Component Analysis: In order to make computations easier and faster a data dimension is reduced in case of PCA. Here we have considered the example of 2D data to understand the working of PCA. Two axes are taken up when data is plotted in a graph. The data is converted into 1D after applying PCA on it as shown in fig. 12.
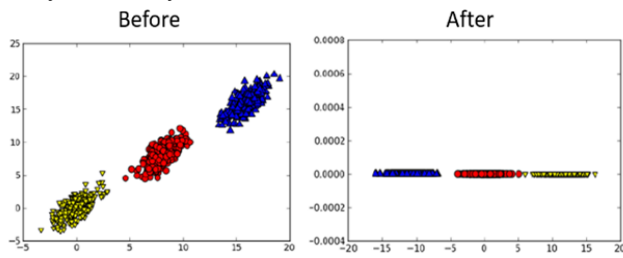
151

Fig. 6. Visualization of data before and after applying PCA [7]

3. **Semi-Supervised Learning:** Both supervised and unsupervised learning power is combined in semi-supervised learning algorithm. In cases where unlabeled data us present and getting labelled data is not an easy process in those cases use of this type of machine learning is prove to be more efficient [8]. Under Semi-supervised learning various learning algorithms are considered as given below [15]:

> **Generative Models:** This comes under the category of semi-supervised learning and structure used by it is given below:

$$p(x, y) = p(y)p(x|y) \quad \text{.. (1)}$$

In above equation mixed distribution is denoted by $p(x|y)$. e.g. Gaussian mixture models in which mixed components can be identified within the unlabeled data. For mixture distribution confirmation one labelled example per component is enough

> **Self-Training:** In this with portion of labelled data classifiers are trained. After that classifier is then fed with unlabeled data. In training set both predicted labels and unlabeled points are added and then repeat it. The name given to it is self training as in this learning of classifier is done by itself.

> **Transductive SVM:** SVM extension is known as transductive SVM. In this both unlabeled and labelled data as considered. Mainly its use is found in labelling unlabeled data that makes margin between labelled and unlabeled data as maximum. The use of TSVM in NP-hard problem for extraction of solution is most efficient way.

4. Reinforcement Learning: In this decisions are taken on the basis of actions that results in having more positive outcomes. Until a situation is not given learner can't apply any actions as they don't have any knowledge. In future actions and situations get affected by the actions taken by learner. Trial, error search and Delayed outcome are two criteria of reinforcement learning [9].

5. Multitask Learning: For other learning for performing better use of multitask learning is prove to be helpful. The procedure of solving a problem or how it reach to particular conclusion is remember by it prior of applying multitask learning algorithms. Then other similar task or problem solution is find out by algorithm using those steps. Another name of this helping of one algorithm to another is known as inductive transfer mechanism. As compared to learning individually when learners experience is shared with each other that results in more concurrent learning [10].

6. Neural Network Learning: Artificial neural network (ANN) is another name of neural network that is derived from neurons biological concept. A cell like structure in a brain is known as neuron and there is need to understand the working of neurons before understanding neural network. Dendrites, nucleus, soma and axon are four main parts of neuron as shown in figure given below.
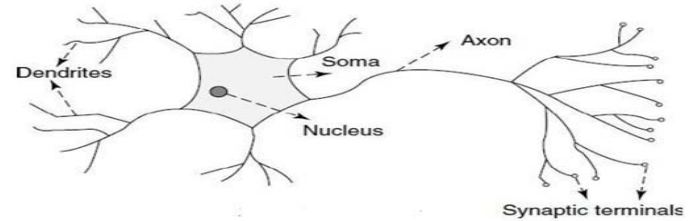

Fig. 7. A Neuron [12]

The electrical signals are received by dendrite and it is processed by soma. The process output is carried to the dendrites by axon where it is sent to another neuron. The nucleus is act as neuron heart and neural network is the inter-connection of neuron. In the same way SNN works that have three layers name as input, hidden and output layer [13]. Supervised, Unsupervised and reinforcement are three types of ANN explained below:

> **Supervised Neural Network:** In this case output is already known then the neural network predicted output is compared with actual output. The parameters are changed on the basis of error and then it is again fed into the neural network.
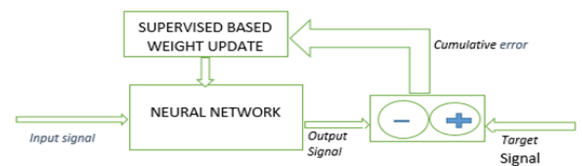

Fig. 8. Supervised Neural network

The process of supervised neural network is given in above shown fig and mainly in feed forward network its use of found.

Unsupervised Neural Network: In this case there is no prior information about the output. It mainly categorized the data on the basis of some similarities. Various inputs correlation is checked by neural network and then group it into different clusters.
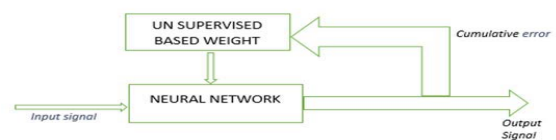

Fig. 9. Unsupervised Neural Network [11]

Reinforced Neural Network: The network will behave as human communicate in environment. Then from environment feedback is give for acknowledgement that decision taken by network is correct or not. In this case also network don't have any prior information about the output.
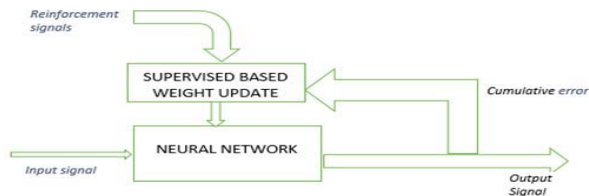
Fig. 10. Reinforcement Neural Network [10]

7. Instance Based Learning: In this case, a particular type of pattern is learnt by learner. Same pattern is applied to newly fed data and that's why its name is instance based learning. This comes under the type of lazy learner in which to arrive they wait for test data and then along with it training data is used to act on it. As data size gets increased there is corresponding increase in the learning algorithm complexity. K-nearest neighbour is an example of instance based learning [12].

      K-Nearest Neighbor: In this case well labelled training data is fed into learner. Both data is compared in it when test data is introduced to learner. In this case majority of correlated data from training data is serve as test data for new class.

III. APPLICATIONS AREA OF MACHINE LEARNING

In this section we have given the review on different application areas where machine learning has been used. We have covered only some of the application areas like Medical, Robotics, Social media and travelling.

Table. 1: Comparison of different application area of machine learning

Author	Application area	What they did	Dataset	Machine learning approach
Ali Shoeb, et.al, (2009) [12]	Medical	They have introduced the patient-specific seizure detector based on machine learning	They have used 81 hours of intracranial EEG (IEEG) collected Data from 17 adult subjects.	Support Vector Machines (SVM)
Atif Khan, et.al, [13]	Medical	Machine learning potential is demonstrated to support task in cases where medical professionals need it by copying noisy and missing patient data	The drug prediction, drug-drug interactions and patient data rules are used by them.	Worked on decision stump, C4.5-R8, Bagging and AdaBoost
Huiying Li, et.al, (2005)	Medical	ML is used in updating the criterion of diagnosis of cancer problem. Microsoft Visual	The Ljubljana Slovenia university are used to achieve BREA and PRIM data sets.	They have used Naive Bayesian learning algorithm
		Basic 6.0 as programming tool and Bayesian formula for updating the weights of cancers diagnosis		
José Luis Seixas Jr., et.al, (2015) [14]	Medical	Comparison of different machine learning paradigms that will validate the medical image segmentation of lower members ulcers found in skin.	They have generated three datasets: Complete, border and center. For each category they have taken 263.835 examples characterized by seven attributes	They have worked on various machine learning algorithms like SVM, KNN, RF, MLP, NB, etc.
Syoji Kobashi, et.al, (2016) [15]	Medical	They have constructed a knee function prediction model of post operative using machine learning.	They have used 52 osteoarthritis (OA) knees that had been operated by using PS type knee implant (Vega, Aesculap, B/Braun, Germany) at Hyogo College of Medicine (Hyogo, JAPAN) between May 14, 2014 and April 22, 2016.	Support vector regression and Neural network
Yen-Chiao Lu, et.al, (2017) [16]	Medical	They have find important risk factors using different machine learning algorithm. Using it they have predicted the recurrence proneness for ovarian cancer.	The Chung Shan Medical University Hospital Tumor registery pathology and medical data is used.	Random Forest (RF), C5.0, Multivariate Adaptive Regression Splines (MARS), Support vector machine (SVM) and extreme learning machine (ELM)
Chen Peng, et.al, (2009) [18]	Travelling	They have proposed a Relevance vector machine (RVM) that solve the problem of existing	The Caltrans performance measurement system (PeMS) collected time data is	Deep Belief Network (DBN) is used as a machine learning approach

153

		methods.	used by them.	
Muneki Yasuda, et.al, (2008) [17]	Travelling	New restricted Boltzman machines and approximate learning algorithm is proposed by them in which they have used loopy belief propagation and EM algorithm.	They have used continuous restricted Boltzmann machine (CRBM)	-
Olga Narushynska, et.al, (2017) [19]	Travelling	In this buyer at store gets optimal route using algorithm described by author whose main motive is to reduce the problem of travelling salesman.	--	Nearest neighbour algorithm
Tasfiqul Ghani, et.al, (2018) [20]	Travelling	The first tourism smart phone application is Amar Bangladesh that introduced the program of machine learning in which Google Maps API is used and admin is able to find the all points of interest to the database in which the use ML program	SQLite	Machine learning
Aqsath Rasyid Naradhipa, et.al, (2011) [21]	Social network	They have introduced the method that classify the sentiments on social media for Indonesian language.	Social media messages	SVM (Support Vector Machine)
Morgan Madeira, et.al, (2013) [22]	Social network	Their presented machine learning approach helps in predicting a	They have taken data of 61 users who have installed facebook application	Naïve Bayes decision tree, Sequential minimual optimizatio

		user's closest friends on facebook and it gives accuracy of 85% in predicting the user's closest friends.	in between feb and mar 2013.	n and Bayesian network.
Sharmishta Desai, et.al, (2015) [23]	Social network	They have given various phases of social media data mining. Their main focus is to study the algorithms in which training time don't vary with the size of data	They have used Reality Mining, Gowalla, Brightkite, Twitter	Naive Bayes, Adaboost and random decision tree
Anees Ul Hassan, et.al, (2017) [24]	Social network	They presented the way to find the depression level of a person using machine learning techniques, emotion theories.	Twitter dataset and 20newsgroups.	Maximum Entropy, Naïve Bayes and Support vector machine
Fatih Ertam, (2017) [25]	Social Network	They have used given machine learning approaches for analyzing the data taken from facebook and 16compared them in terms of performance metrics like F1 score, recall and precision.	Facebook data	Adaboost (AB), Logistic Regression (LR) and Random Forest (RF)
Nicolas Bredeche, et.al, (2006) [26]	Robotic	They have analyzed that perceptual learning mechanisms prove to be efficient in building low abstraction operators that mainly deal with real world	Real world object	Perceptual learning

154

		data. The main motive behind author is to construct a method and enable learning of robot to identify a attacks occurred in environment.		
Alessandro Giusti, et.al, (2015) [27]	Robotic	This is the first paper in which an approach to perceive forest trials is described	Real-world dataset	Deep Neural Network
Agata Nawrocka, et.al, (2018) [28]	Robotic	They have focused on content, objects or users based neighbourhood filtering algorithms. They did a comparison of these algorithms.	MovieLens portal database	neighbourhood based algorithms
George W. Clark, et. al, (2018) [29]	Robotic	Cyber attack on machine learning policy of an autonomous vehicle is evaluated by them and robotic vehicle operating in a dynamic environment is attacked.	For different attacks a Q table data has been used by them.	Reinforcement learning (RL)

Other applications fields of ML that are not covered in this paper are financial services, marketing and sales, Transportation, Virtual personal assistance, etc.

IV. CONCLUSION

Machine learning is used to teach machines how to handle the data in an efficient way. In most of the cases we are not able to interpret pattern or extract information from data after viewing it. In those cases use of machine learning has been found and rise in its demand has been seen due to increase in available datasets. In this paper, we have focused on different paradigms used in machine learning and divided the complete paper into different sections. The first section gives the brief introduction about machine learning then describes the model of machine learning that consist of two things training and testing. After that it contains different types of learning that is categorised into different categories like supervised learning, unsupervised learning, semi-supervised learning, Reinforcement learning, Multi-tasking learning, Neural network and instance based learning. Now a days machine learning has been used by every person like in updating social networking sites photos to online shopping product recommendation. So, in the third section gives a review on different applications area where machine learning has been used by various researchers. We have covered medical, travelling, social network and Robotics application area in which machine learning has been used and gives improved results. In future we can give review on use of ML in financial services, marketing and sales, Transportation, Virtual personal assistance, etc.

REFERENCES

[1] Abdel, O., "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. Acoustics", Speech and Signal Processing, vol. 7, pp. 4277-4280, 2012.
1. A. Mosavi, A. R. Varkonyi-Koczy, "Integration of Machine Learning and Optimization for Robot Learning", Advances in Intelligent Systems and Computing, vol. 519, pp. 349-355, 2017.
2. Y. Bengio, "Learning deep architectures for AI", Foundations and trends in Machine Learning, vol. 2, pp. 121-127, 2009.
3. S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques", Informatica, vol. 31, pp. 249-268, 2007.
4. A. B. Hillel, R. Lerner, D. Levi, G. Raz, "Recent progress in road and lane detection: a survey", Machine Vision and Applications, pp. 1–19, 2012.
5. P. Harrington, "Machine Learning in action", Manning Publications Co., Shelter Island, New York, 2012.
6. X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 3, pp. 1-130, 2009
7. V. Sharma, S. Rai, A. Dev, "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, 2012.
8. S. B. Hiregoudar, K. Manjunath, K. S. Patil, "A Survey: Research Summary on Neural Networks", International Journal of Research in Engineering and Technology, vol. 3, pp. 385-389, 2014.
9. D. Rebernak, M. Leni, P. Kokol, V. Zumer, "Finding Boundary Subjects for Medical Decision Support with Support Vector Machines", Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems (CBMS'03), pp. 1-6, 2003.
10. H. Li, D. Li, C. Zhang, "An Application of Machine Learning in the Criterion Updating of Diagnosis Cancer", IEEE, pp. 187-190, 2005.
11. A. Shoeb, D. Carlson, E. Panken, T. Denison, "A Micropower Support Vector Machine Based Seizure Detection Architecture for Embedded Medical Devices", 31st Annual International Conference of the IEEE EMBS, pp. 4202-4206, 2009.
12. A. Khan, J. A. Doucette, R. Cohen, D. J. Lizotte, "Integrating Machine Learning into a Medical Decision Support System to Address the Problem of Missing Patient Data", 2012 11th International Conference on Machine Learning and Applications, pp. 454-457, 2012.
13. J. L. Seixas, S. Barbon, R. G. Mantovani, "Pattern recognition of lower member skin ulcers in medical images with Machine Learning Algorithms", 2015 IEEE 28th International Symposium on Jos'e Computer-Based Medical Systems, pp. 50-53, 2015.
14. S. Kobashi, B. Hossain, M. Nii, "prediction of post-operative implanted knee function using machine learning in clinical big data", Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, pp. 195-200, 2016.
15. Y. C. Lu, C. J. Lu, C. C. Chang, "A hybrid of data mining and ensemble learning forecasting for recurrent ovarian cancer", ICIIBMS 2017, Track2: Artificial Intelligence, Robotics and Human-Computer Interaction, Okinawa, Japan, pp. 216-220, 2017.
16. M. Yasuda, K. Tanaka, "Approximate Learning Algorithm for Restricted Boltzmann Machines", IEEE Computer society, pp. 692-697, 2008.

17. C. Peng, Y. Xin-ping, L. Xu-hong, "Bus travel time prediction based on relevance vector Machine", IEEE, pp. 1-4, 2009.

18. O. Narushynska, V. Teslyuk, B. D. Vovchuk, "Search model of customer's optimal route in the store based on algorithm of machine learning A", CSIT 2017, pp. 284-287, 2017.

19. T. Ghani, N. Jahan, S. H. Ridoy, A. T. Khan, S. Khan, M., Khan, M. Amar "Bangladesh - A Machine Learning Based Smart Tourist Guidance System", IEEE, pp. 1-5, (2018).

20. A. R. Naradhipa, A. Purwarianti, "Sentiment Classification for Indonesian Message in Social Media", 2011 International Conference on Electrical Engineering and Informatics, pp. 1-4, 2011.

21. M. Madeira, A. Joshi, "Analyzing Close Friend Interactions in Social Media", SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013, pp. 932-935, 2013.

22. D. Sharmishta, Dr. S. T. Patil, "Efficient Regression Algorithms for Classification of Social Media Data", International Conference on Pervasive Computing (ICPC), pp. 1-5, 2015.

23. A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, S. Lee, "Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression", ICTC 2017, pp. 138-140, 2017.

24. F. Ertam, "Analysis of Data Using Machine Learning Approaches in Social Networks", IEEE, pp. 812-815, 2017.

25. N. Bredeche, Z. Shi, J. D. Zucker, "Perceptual Learning and Abstraction in Machine Learning: An Application to Autonomous Robotics", IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, pp. 172-181, 2006.

26. A. Giusti, J. Guzzi, D.C. Cire, F. L. He, J. P. Rodríguez, "A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots", IEEE Robotics and Automation letters. PREPRINT VERSION, pp. 1-7, 2015.

27. A. Nawrocka, A. Kot, M. Nawrocki, "Application of machine learning in recommendation Systems", IEEE, pp. 328-331, 2018.

28. G. W. Clark, M. V. Doran, W. B. Glisson, "A Malicious Attack on the Machine Learning Policy of a Robotic System", 2018 17th IEEE International Conference on Trust, Security And Privacy In Computing And Communication, pp. 516-521, 2018.

156