Impact Of The Clustering Algorithms And Its Applications: Machine Learning

Sunita[1], Vijay Rana[2]

Research Scholar, Dept. of Computer Science, Arni University, India

Assistant Professor[2], Sant Baba Bhag Singh University, India

sunitamahajan2603@gmail.com   vijay.rana93@gmail.com

Abstract

Machine learning is more demanding technology today's with their applications is subset of artificial intelligence. Behind the popularity of machine learning is depends upon its algorithms. The algorithms are based on training phase and testing phase. Machine learning is a powerful tool when it's comes to with their applications like pattern recognition, deep learning, character recognition, object recognition and any other research areas. In this paper introducing the machine learning algorithms and focus on the advancement of the research based on clustering methods. This paper looks at the probable applications as well as the existing progress of the combine of machine learning algorithms.

Keywords: *Machine learning, AI, Algorithms.*

1. Introduction

Today's machine learning is a popular research [4] in the area on how to make a machine more intelligent. In machine learning a machine to prepare and train for a specific task. Machine learning is a field of computer science and gains more popularity in AI (Artificial Intelligence)[5]. The machine is trained on some datasets and then applied to algorithms on the behalf of these process machines can make a prediction and learn on a given particular dataset. Machine learning is a specific focus on automatically identify complicated patterns and building intelligent decisions-based on the information.

Machine learning is a process to convert experience into knowledge, is the development of computer science that can change when entered into new data [6]. In this section we discuss various types of machine learning algorithms which includes

decision tree, support vector machine (SVM), liner regression, Naïve Bayes, K-nearest neighbor (k-NN) Figure 1.
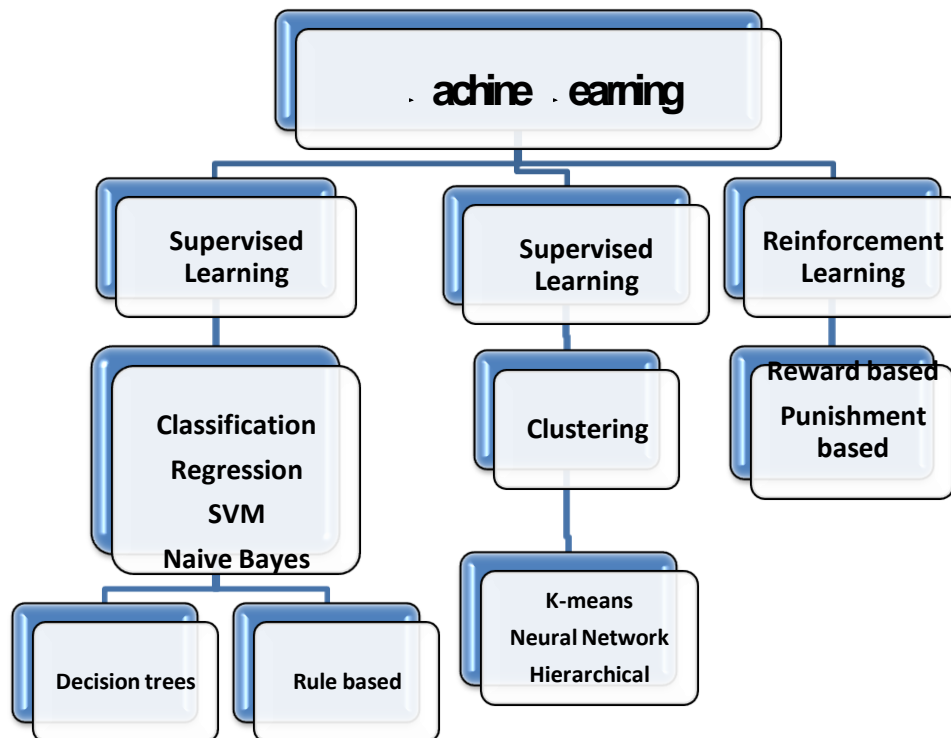

Figure 1 Flow chart of machine learning and its algorithms [2]

1.2 Comparison between Supervised and unsupervised machine learning

In the above section 2 we discuss the flow chart of machine learning process and its algorithms. In the section 2.1 we define the comparison between the supervised and unsupervised machine learning.

Table 1: Comparison between Supervised and Unsupervised Learning

Sr No.	Unsupervised	Supervised
1	Unsupervised methods do not use any data to adjust internal parameters.	Supervised methods, on the other hand, exploit all given data to optimize parameters such as weights or thresholds.
2	In unsupervised learning the aim is to discover groups of similar	classification outperformed unsupervised clustering and if so which algorithms were most effective
3	Unsupervised learning has no	Supervised learning assumes that a set

	labeled data in training process. A combination of both labeled and unlabeled data	of training data has been provided
4	Unsupervised learning [38]encompasses many other problems involving summarizing and explaining various key features of data.	Supervised learning [38]applications are widely used such as decision trees, naive Bayes, and support vector machines for their easy evaluation and acceptance to human.
5	Unsupervised learning are summarized into clustering algorithms like K-Means, KNN, Decision tree, Decision list, Naïve Bayes .	Supervised learning[38] is summarized into two categories: generative algorithms and discriminative algorithms.

2. Machine Learning Algorithms

Machine learning is the subset of artificial intelligence, which increases the capability thinking and learning power of the machines being a programmer. In this section 2 we discussed machine learning algorithms. There are mainly divided into two categories.

- **Supervised Learning**
- **Unsupervised Learning**

2.1 Supervised Learning

Supervised machine learning algorithm used labelled data, but data is divided into two parts one is the training dataset and other one is testing dataset. The trained dataset needs to predict the expected outcomes as well as tested to predict real-valued outcomes. The algorithms are listed below

- *Algorithms*

The following algorithms are used in the supervised learning techniques

- **Support Vector Machine (SVM)**
- **Decision Tree**
- **Naïve Bayes**
- **Neural Network**

Support Vector Machine

SVM is classification and regression method in machine learning. SVM is dividing the input values into positive and negative groups. Find the distance to the closest positive

and negative called support vectors. In other words, support vector machines (SVMs) tend at the same time to minimize the empirical classification error and maximize the geometric margin between positive and negative examples.

- **Naive Bayes**

A Naive Bayes algorithm is a probabilistic classifier based on the purpose of Bayes' theorem. It mainly used in the text classification. It relies on the computation of the conditional probability of each sense of a word given the features is depend upon the given context.

- **Neural Networks**

A neural network is an interconnected cluster of artificial neurons that uses a computational model for processing data based on a connectionist approach. Pairs of input extract the response are input to the learning program. The utilization of input features to partition the training contexts into without overlapping sets corresponding to the preferred responses. The new groups are provided; link weights are progressively adjusted so that the output unit representing the desired response has a larger activation than any other output unit.

- **Decision Tree**

The decision tree is a one of most popular classification method. In this algorithm data depending present on ascending and descending order.

2.2 Unsupervised Learning

In this section 2.2 we discuss about the unsupervised machine learning algorithms. The unsupervised machine learning there is no need to train the data which means no labelled. The output is depending upon their inputs the algorithm analysed on the set of data. The outcomes are based on the past experience of the machine. Unsupervised learning problems grouped into clustering and association.

2.2.1 Clustering

Clustering is dividing the sampling data into similar groups. It deals with finding the structure or pattern in a collection unorganized data. There are different kinds of the clustering.

- **Partitioning**
- **Agglomerative**
- **Overlapping**

Partitioning

In partitioning clustering data are grouped in such a way that one data can belong to one cluster only. This type clustering include under the k-means clustering.

Agglomerative

In this clustering technique, every data is a cluster. The iterative unions between the two nearest clusters reduce the number of clusters. This type clustering include under the hierarchical clustering.

Overlapping

In this technique, a fuzzy set is used to cluster data. Each point may belong to two or more clusters with separate degrees of membership.

Types of clustering

- k-means clustering
- K-NN (K-Nearest Neighbors)
- Principal Component Analysis (PCA)

K-means

K means it is an iterative clustering algorithm which helps you to find the highest value for every iteration. The output of the algorithm is a group of "labels." It assigns data point to one of the k groups. In k-means clustering, each group is defined by creating a centroid for each group.

K-NN

K-NN [26]is stands for the k-Nearest Neighbor algorithm, is depend upon one of the highest-performing methods. K-NN is a simplest machine learning classification and regression algorithm but mostly used classification method classify to the target dataset.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is faster to convert two dimensional data into one dimensional. It used to transforming set of variables to new ones known principal component. This algorithm is used to data summarization.

3. Application

In section discussed the various applications of the machine learning under the supervised and unsupervised and reinforcement learning.

3.1 Supervised Learning

- **Spam Filtering**

Spam [3] mail is bulk, unwanted commercial email or junk mail. It can learn from the emails that you are mark as spam and the ones you mark as non-spam.[1]

- **Recognition**

Recognition is a process to identify something from the given data. It is most useful in the research area like face recognition, handwriting recognition and speech recognition. Face recognition is mostly used for the purpose of the security and criminal justice system. Handwriting recognition is used to match the authorized and unauthorized content of users. It also used in the area of fraud detection. [1]

3.2 Unsupervised Learning

- **Social Network**

In the social network peoples are connect with some links. The links are identifying the mutual friends on Facebook and Twitter around your friends. This is the most popular application of the supervised learning.[1]

- **Classification of DNA**

DNA classification is the approach of unsupervised learning. There is advanced information to be provided. Its results are depending upon the training of the dataset. The clustering algorithm is used to a group of individuals into a variety of people.[1]

3.3 Reinforcement Learning

- **Gaming**

AI agents are developing much success games in recent years. [1]

- **Traffic control system**

Today's traffic control system is most popular concept of to predict in area. [1]

4. Discussion

In this section performance can be evaluated on the three parameters Accuracy, Precision, Recall. These methods are evaluated the overall performance.

No. Of Experiment	Number of training dataset	Accuracy			
		Naïve Bayes' (movie reviews)	K-NN (movie reviews)	Naïve Bayes' (hotel reviews)	K-NN (hotel reviews)

1	100	56.78	47.64	43.11	45.35
2	200	64.29	55.07	41.26	40.97
3	500	70.06	58.44	42.56	41.42

Table 2 Accuracy Comparison with different Dataset [13]
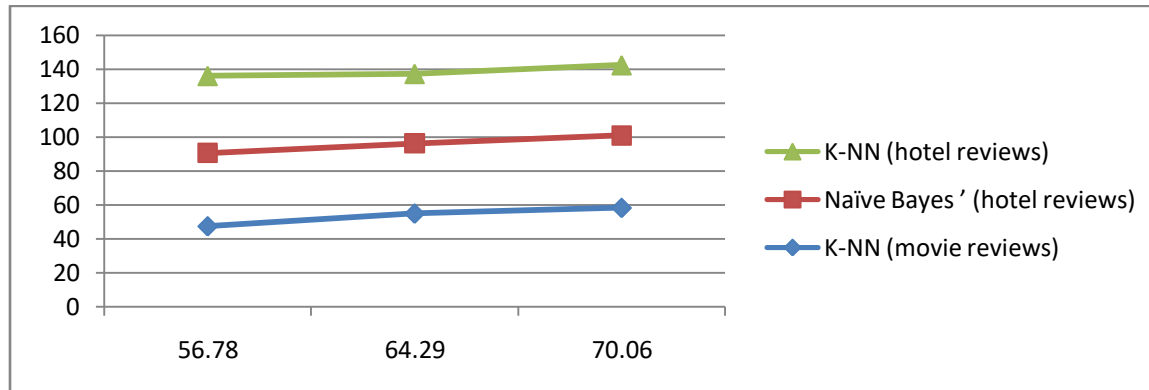
Figure 2 Representation of accuracies based different Dataset

No. Of Experiment	Number of training dataset	Precision for positive corpus:				Precision for negative corpus:			
		Naïve Bayes ' (movie reviews)	K-NN (movie reviews)	Naïve Bayes ' (hotel reviews)	K-NN (hotel reviews)	Naïve Bayes ' (movie reviews)	K-NN (movie reviews)	Naïve Bayes ' (hotel reviews)	K-NN (hotel reviews)
1	100	59.4	41.35	42.11	44.51	55.43	38.12	48.39	46.21
2	200	64.22	50.25	40.43	40.42	63.67	49.56	42.61	41.63
3	500	69.24	54.34	41.34	40.34	70.59	57.25	50.62	47.32

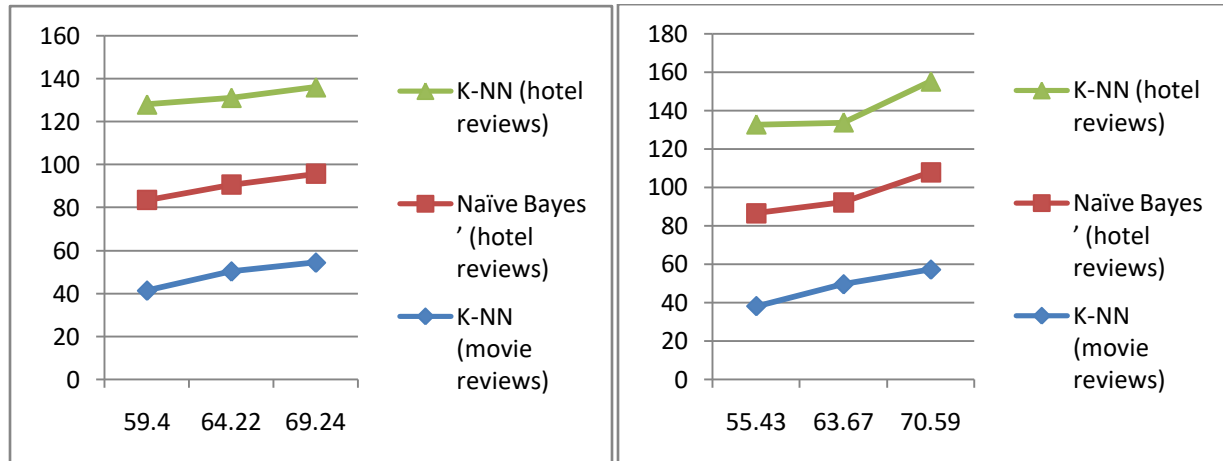**Table 3 Precision comparison based on Positive and Negative Corpus[13]**

Figure 3 Representation based on Positive Precision   Figure 4 Representation based on Negative Precision

Table 4: Recall comparison based on Positive and Negative Corpus [13]

| No. Of Experiment	Number of training dataset	Recall for positive corpus:				Recall for negative corpus:			
		Naïve Bayes ' (movie reviews)	K-NN (movie reviews)	Naïve Bayes ' (hotel reviews)	K-NN (hotel reviews)	Naïve Bayes ' (movie reviews)	K-NN (movie reviews)	Naïve Bayes ' (hotel reviews)	K-NN (hotel reviews)
1	100	43.42	30.35	31.11	29.51	68.42	40.35	63.11	59.51
2	200	61.22	44.44	42.43	40.42	65.22	54.44	52.43	50.42
3	500	70.24	51.33	40.34	39.34	67.24	52.33	51.34	51.34
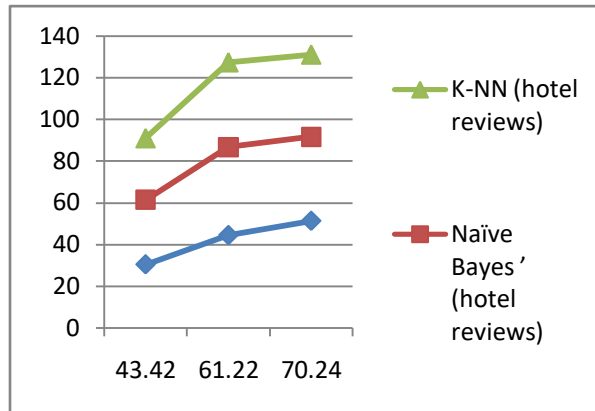
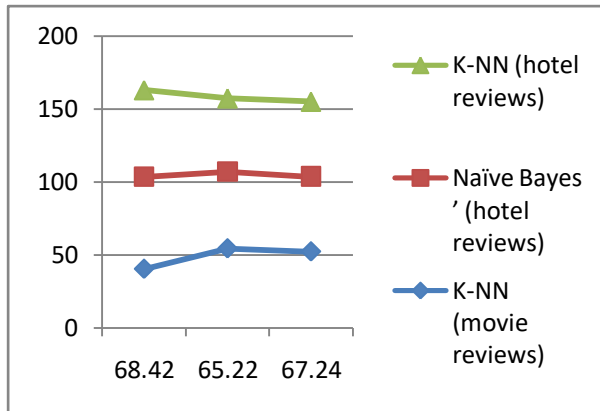Figure 5 Representation based on Positive Recall   Figure 6 Representation based on Negative Recall

5. Conclusion

In this paper give details of machine learning algorithms and listing all the modern approaches and applications under the research area AI and machine learning. Today's in the world, every user's used of machine learning techniques. The paper demonstrates the supervised and unsupervised learning techniques in detail and explains where machine learning is application areas. The aim behind this to demonstrate the accuracy, recall, precision based on clustering methods.

Reference

[1] Das, S., Dey, A., Pal, A., & Roy, N. (2015). Applications of artificial intelligence in machine learning: review and prospect. International Journal of Computer Applications, 115(9).

[2] Dey, A. (2016). Machine learning algorithms: a review. International Journal of Computer Science and Information Technologies, 7(3), 1174-1179.

[3] Mahajan, S., & Rana, V. (2017). Spam detection on social network through sentiment analysis. Advances in Computational Sciences and Technology, 10(8), 2225-2231.

[4] Rana, V. (2018, July). Optimizing performance of user web browsing search. In International Conference on Advanced Informatics for Computing Research (pp. 230-239). Springer, Singapore.

[5] Dhall, D., Kaur, R., & Juneja, M. (2020). Machine Learning: A Review of the Algorithms and Its Applications. In Proceedings of ICRIC 2019 (pp. 47-63). Springer, Cham.

[6] Rana, V. (2020). A Web Extraction Browsing Scheme for Time-Critical Specific URLs Fetching. In Proceedings of ICRIC 2019 (pp. 617-626). Springer, Cham.

[7] Deng, L.: Three classes of deep learning architectures and their applications: a tutorial survey. APSIPA Trans. Signal Inf. Process. (2012).

[8] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436 (2015)

[9] Deng, L., Yu, D.: Deep learning: methods and applications. Found. Trends® Signal Process.7 (3–4):197–387 (2014)

[10] Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. 61, 85–117 (2015)

[11] Kaur, R., Juneja, M.A.: Survey of different imaging modalities for renal cancer. Indian J. Sci. Technol. 9, 44 (2016)

[12] Valenti, R., Sebe, N., Gevers, T., Cohen, I.: Machine learning techniques for face analysis. In: Machine Learning Techniques for Multimedia, pp. 159–187. Springer, Berlin, Heidelberg (2008).

[13] Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. arXiv preprint arXiv:1610.09982.