

# Analysis Through Machine Learning Techniques: A Survey

Raghavendra Reddy  
Ph.D Research Scholar  
School of C and IT  
REVA University  
Bangalore, India  
reddycs2004@gmail.com

Gopal K. Shyam  
Assoc. Prof.  
School of C and IT  
REVA University  
Bangalore, India  
gopalkrishna@revainstitution.org

**Abstract** – Analysis and search for meaningful associations in customer purchase data are considered as best applications of data mining techniques. Machine learning is the fundamental nature of imitation of intelligence. The machine learns from the past information to improve the performance of intelligent programs. We consider unsupervised machine learning techniques to analyze various sort of the data. Techniques used are clustering, feature extraction and classification. Machine learning is mainly employed to exhibit accurate estimate. The major objective of this paper is to present the outline of machine learning and discuss unsupervised machine-learning techniques for various applications. Further, this paper reviews the different machine learning techniques.

**Keywords:** Unsupervised Machine Learning Techniques, Data Analysis, Marketing Application.

## I. INTRODUCTION

Machine learning represents the multidisciplinary field in an artificial intelligence, statistics, probability, philosophy, information theory, neurobiology and psychology. Machine learning has been developed as an effort to explore computer learning mimic human brain. It requires developing newer methods to precede the knowledge of machine learning and massive quantity of effort that desires to be complete to replace active techniques as a novel method due to the enlargement in the database ensuing market desires for method that can be capable of extracting expensive data from huge data stores, and knowledge discovery in databases which have freshly appeared as a novel systematic and engineering regulation.

Figure 1 shows the operational model of machine learning. Learning procedure in machine learning techniques are separated into two stages as training and testing. In training procedure, training data are used as an input in which features learned with learning technique are taken for constructing the learning replica. In the testing procedure, knowledge replica are utilized by execution machine to construct the calculation for test or assembly information. Tagged information is the output of knowledge model which can give the prediction or classified information.

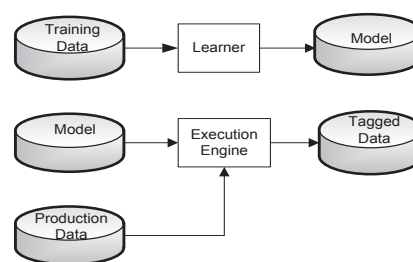


Figure 1 : Operational model of machine learning

Machine learning is consists of the following categories:

- Supervised Machine Learning
- Unsupervised Machine Learning
- Other learning techniques: Reinforcement Learning, Active Learning, Online Learning

Figure 2 shows the supervised machine learning model. Supervised learning can be trained by labeled instances, from which required production is recognized. It provides dataset which is consisting of characters and labels. A part of equipment contains training data point labeled either as failed or runs. The goal of supervised learning is to build an estimator which can be able to forecast the label of object features.

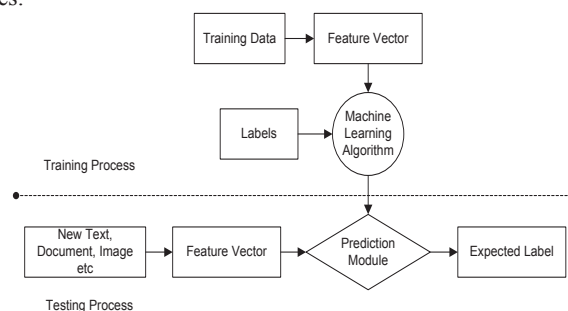


Figure 2 : Supervised learning model

Unsupervised learning utilizes data that has no historical labels and the aim is to discover the data and locate similarity among the substances as shown in figure 3. It is a method of discovering label as the information itself. Unsupervised learning works fine on transactional data that recognize segment of consumers by comparable attributes, it is also able to

discover the major attribute that divide customer segment into number of groups.

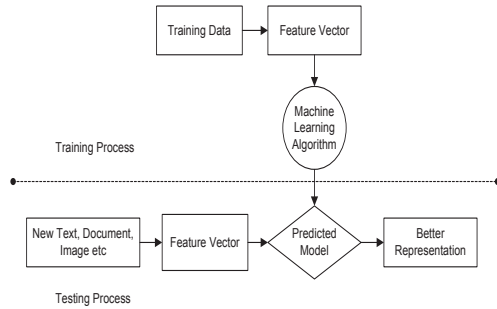


Figure 3: Unsupervised learning model

Figure 4 shows the machine learning tree which contains the supervised and unsupervised learning techniques. A huge set of machine learning algorithms can be designed to construct the machine learning applications and execute an iterative machine learning procedure. These algorithms can be classified on the basis of learning style as follows: regression, clustering, dimensionality reduction etc.

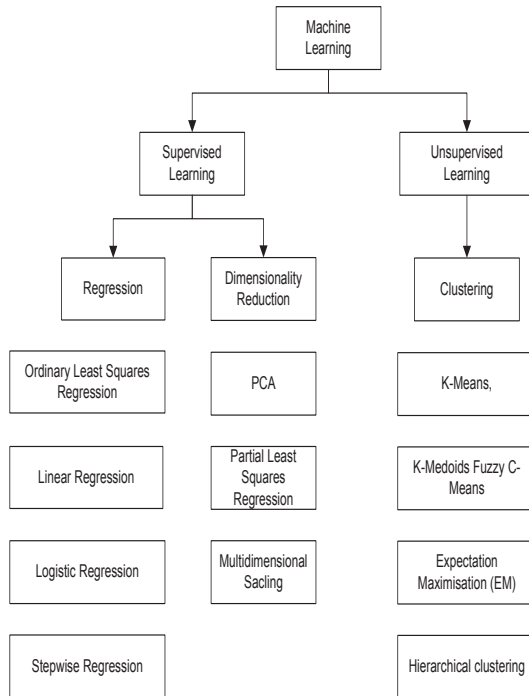


Figure 4: Machine Learning Tree

The main purpose of this paper is to review the literature and analyze the existing research gap in terms of:

- Performing systematic review and study of the supervised and unsupervised machine learning techniques.
- Discussing some problems associated with supervised and unsupervised machine learning techniques and providing instructions to the current challenges.

- Describing research gaps in case of unsupervised machine learning techniques and future researches.

The rest of this article is organized as follows. In Section II we discuss literature survey on unsupervised and supervised machine learning techniques. Section III consists of conclusion.

## II. LITERATURE SURVEY

P. Ajitha et. al[1] proposes the method of opinion mining for online marketing by sentiment thesaurus with idea of search engine. The consumer reviews the product with related service have become the de-facto for the vendor's business stuff online. Consumer re-evaluates for the product which become more significant now a days because e-commerce business grow imminently. The prominent inventions, the amount of consumer review run to the various counts. It can make the trade choice for client more difficult. Then summarization can describe information like opinion of product unlike the conservative text summarization counting the positive and negative.

Ya-Han Hu et. al[2] proposes the online travel forum with social network that become most familiar proposal for contribution of travel information, by massive statistics of the reviews which can post on daily. Mechanically produce hotel summaries assist travelers to select the hotels. Preceding study on re-evaluate summarization mostly observed contented analysis, disregards significant factor similar to author integrity and incompatible opinion. They measured the factors and urbanized new verdict significance metric. Mutually satisfied and sentiment similarity are utilized to establish the resemblance of two sentence. To recognize top- $k$  sentence,  $k$ -medoid clustering methods are utilized to divide sentence into the  $k$  groups. Then medoids from those groups are selected from the final summarization result. To estimate the recital of proposed system and two sets of the review are collected on two hotels posted on TripAdvisor.com. Then results signify that proposed method outperforms compare to other methods and projected method can present complete hotel information.

H. Suresh et. al[3] proposes an unsupervised method of fuzzy clustering method for twitter sentiment analysis. The main task is to investigate the data mining along with familiar approaches utilized in machine learning system. During contrast to supervised learning methods, cluster based algorithms generate accurate tentative result with no manual dispensation, linguistic information or guidance time. This can describes the new fuzzy clustering replica to analyze twitter feed concerning the sentiment of exacting brand by actual dataset composed in one year. Proposed method make use of K-Means and Expectation Maximization techniques by considering metrics like accuracy, recall, precision and execution time. According to analysis, the projected method can be experienced to be feasible in performing higher quality of the twitter sentiment analysis.

Gorynin et. al[4] proposes an unsupervised learning of the Markov-switching stochastic volatility for the purpose of analyzing market data. This introduces the novel approach for assess the regime-switching stochastic volatility model as of historical price. Then method is based on the new account of assumed density filter (ADF). The simulation results demonstrate the efficiency of proposed system. Thus they are analyzed various market cost histories for constancy by regime-shifting replica.

Xiaodong Li et. al[5] proposes the financial news objects which can believe to contain collisions on the stock cost arrival. Earlier facility form news piece in the space of bag-of-words, are analyzed to find the hidden association among express arithmetical prototype also reserve price actions. Authors used Harvard psychological dictionary with Loughran–McDonald financial sentiment dictionary to assemble the space of sentiment. They used data set from five years past Hong Kong Stock Exchange price for evaluation. News sentiment is the significant on sequence of mapping as of the word pattern to price actions.

Y. Lan et. al[6] proposes an unsupervised representation of Learning Part-Based for store market calculation. Machine learning techniques has been widely used for forecasting financial market data. The accepted predicting model logistic regression is efficient in predicting the stock. They present the theoretical and experiential framework to be appropriate L2-regularized of logistic regression to calculate a stock market. Initially, the systems selects different factors that manipulate a stock tendency. The system exploited NMF to unsupervised learn part-based representation of information and then the classifier trained on individuals representation. Experimental results indicates that the logistic

K. Kianmehr et. al[25] proposes the efficiency of the machine learning approaches for automatic recognition of calling the community. The proposed method utilizes the data which can extract from cluster study to recognize consumer occupation pattern. Consumers calling pattern is given to the classification techniques to produce a classifier replica for forecasting the calling community of consumer. It uses support vector machine as well fuzzy-genetic classifier.

Yoshua Bengio et. al[26] proposes best training techniques for deep architecture, while those can be much more representative competent than SVMs with one hidden layer of neural networks. The three major assistances are followed. Initially they extend the RBMs as well as DBNs in novel habits to logically knob continuous-valued input, presentation instance are much more improved analytical model can be attained. Next, they execute the experiments which can supports the theory that greedy unconfirmed layer-wise training scheme assist to optimize the deep network, except propose that improved simplification can attained since the scheme initializes upper layer by improved representation of the appropriate high-level abstraction.

regression through the part-based representation of information can be a powerful analytical tool for prediction in financial market.

Table 1 and Table 2 represent the comparison of performance analysis for supervised and unsupervised learning machine techniques respectively. These tables are shows the performance for different parameters.

Ref. No	Methods	Performances Metrics			
		Accuracy	Sensitivity	Efficiency	Reliability
[5]	Bag-of-Words Space, SVM	H			
[8]	Naive Bayes and Maximum Entropy	H			
[13]	Sentiment Analysis; Bag-Of-Words; ;	H			
[14]	Principal Component Analysis (PCA), Partial Least Squares (PLS), Canonical Correlation Analysis (CCA) and Ortho-normalized PLS (OPLS)	H			
[15]	PCA, SVM	H	H		
[16]	APF-KNN Approach	H			H
[17]	Kalman Filter, ANN	H		H	
[18]	Principal Component analysis with SVM	H			
[21]	Semantic Analysis (WordNet)	H			

Table 1: Comparison Table of Performance Analysis for Supervised Learning Machine

Unnisa et. al[11] proposes the social network based behavioral analysis constraint which can enlarge the forecast accurateness and sentiment analysis. The twitter offers the huge proposal in determining public opinion beside the sensible correctness in movie review by the spectral clustering based on the machine learning techniques for the sentiment analysis. The new view removal of twitter data by unsupervised learning algorithm is anticipated with those that solves the problem of field dependence and decreases the explained training information. Unsupervised machine learning algorithms shows improved performance than the supervised learning. Thus the goal is used to defeat the difficulty of clustering multiple records by unlabeled information and execute sentiment classification. The experimental results on two thousands of tweets exhibit that method is not explicit to the movie reviews.

Eyke Hullermeier et. al[27] proposes the possibility of FST to contribute to the machine learning also data mining in different habits. According to the proposed opinion, it can be incredibly significant to focus on right concerns, to contemplate more on strength with characteristic characters of the FST.

Ref. No.	Methods	Performances Matrices				
		Accuracy	Quality	Efficiency	Precision	Speed
[7]	Extreme Learning Machine	H				H
[4]	Assumed Density Filter			H		
[10]	Fuzzy ontology-based semantic knowledge	L		H	H	
[11]	Spectral Clustering	H				
[12]	Extreme Learning Machine	H				H
[20]	Deep learning Machine			H		
[19]	ELM	H				
[22]	Affinity Propagation	H				
[23]	Learning Vector Quantization	H				
[24]	Unsupervised Semantic Orientation Approach and SentiWordNet Approach	H			L	

Table 2: Comparison Table of Performance Analysis for unsupervised Machine Learning

Jiexiong et. al[12] proposes the novel MLP training system, which is on the basis of universal estimate ability of unique ELM. The H-ELM method accomplishes the elevated level of the representation by layer wise training, which outperforms the innovative ELM in different simulation. Furthermore, on evaluation with other MLP training process, training of the H-ELM gives quicker and attains superior knowledge accuracy.

Orestes Appel et. al[8] proposes an alternative method to recompense for the lack of capacity of the lexicon-based system to generate the classification output. The system which is based on a mixture of classification yield in the non-lexicon based tools. Specially, the result ideals are applied to two or more non-lexicon classification technique which can be attained. Secondly, these non-lexicon results were combined by a uniform method, which proved to contain attractive return properties which are required in Sentiment Analysis (SA) context, to produce the classification yield which the lexicon based method is unable to attain. The experimental results are based on the execution of two well-known supervised machine learning techniques, such as Naive Bayes and maximum entropy, and appliance of cross-ratio uniform operator was presented.

Doaa Mohey et. al[13] proposes a method to estimate the sentiment at the word level. The contribution of the proposed work is an enhancement of Bag-of-Words model on online systematic papers reviews and

incorporates contextual polarity. The consequence of sentiment study is a challenge to recover the sentiment correctness. SAOOP aspires at estimating the reviews of systematic papers also from scientific papers are known as the CiteULike website, examines and classifies a textual content of sentiment reviews of every paper. The proposed SAOOP classifies the sentimental reviews and imagine relationships among them based on extract feature and keyword of the systematic field.

Ha-Nguyen et. al[7] says that an ensemble application of ELM with GPU for the real-time multimodal sentiment analysis. The proposed work comprises the sets of applicable aspects for content and audio-visual information, with an easy method for combining the characteristics extracted from dissimilar modalities. The proposed method utilizes a different GPU-friendly algorithm to improve the performance of feature extraction progression from various modalities. In accumulation, powerful ELM classifier is applied to construct the emotion analysis method on extracted characteristics. The visual features also participate key role to does better than previous methods.

Gorynin I et. al[9] proposes the common system of parameters assessment in switched stochastic volatility model. The explanation marks exploiting the accurate assessment of probability purpose. This process is easy, competent and permits an analysis of the stock quotes dynamics. The estimation can be intended to appreciate substantial speedups evaluated to active simulation-based algorithms. The Monte-Carlo revise proves the efficiency of the proposed methodology. This method is comprehensive to broad series of general actions like alternative price and defer curve appropriately. As the point of view, they will learn the applicability of this system to highly complex models, for example, multi-factor or multi-scale volatility model. Then they suggest a test for the method assortment.

Farman Ali et. al[10] proposes the framework by fuzzy ontology on sentiment analysis of the transportation and city characteristic assessments to make easy ITs as well as travelers. The proposed method is effectively categorized the obscure reviews, by intelligently deciding transportation and city attribute polarity. This method can mechanically gather all the traffic information, and examine that information to inform drivers before they arrive at risk regions, which can save energy. It can recover the transportation security; resolve traffic congestion troubles in huge cities, and offers travelers contented itinerant experiences in urban areas. This proposed scheme is extended through attractive the extraction process of accurate and false prejudiced phrases in the field of sentiment study of the transportation action assesses.

### III. CONCLUSION



This paper presents a review of the unsupervised machine learning techniques. Machine learning model describes the different machine learning techniques based on the types of the machine learning styles. This can deal with primarily several techniques to crisis of market basket analysis: the position of significant relations in customer purchase of the product.

## REFERENCES

- [1] P. Ajitha and G. Gunasekaran, "An approach of opinion mining for online marketing using sentiment thesaurus and concept search engine", IEEE, Instrumentation, Communication and Computational Technologies, pp. 208-213, 2015.
- [2] Ya-Han Hu, Yen-Liang Chen, Hui-Ling Chou, "Opinion mining from online hotel reviews –A text summarization approach", Elsevier, Information Processing and Management, Vol. 53, pp. 436–449, 2017.
- [3] H. Suresh and Dr. Gladston Raj S, "An unsupervised fuzzy clustering method for twitter sentiment analysis", International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), pp. 80-85, 2016.
- [4] Gorynin I, E. Monfrini and W. Pieczynski, "Unsupervised learning of Markov-switching stochastic volatility with an application to market data", IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-6, 2016.
- [5] Xiaodong Li, HaoranXie, Li Chen, Jianping Wang, Xiaotie Deng, "News impact on stock price return via sentiment analysis" Elsevier, Vol. 69, pp. 14–23, 2014.
- [6] M. Xu, Y. Lan and D. Jiang, "Unsupervised Learning Part-Based Representation for Stocks Market Prediction", 8th International Symposium on Computational Intelligence and Design (ISCID), pp. 63-66, 2015.
- [7] Ha-Nguyen Tran, Erik Cambria, "Ensemble Application of ELM and GPU for Real-Time Multimodal Sentiment Analysis", Springer, pp. 1-11, 2017
- [8] Orestes Appel, Francisco Chiclana, Jenny Carter, Hamido Fujita, "Cross-ratio uninorms as an effective aggregation mechanism in Sentiment Analysis", Elsevier, Vol. 124, pp. 16-22, 2017
- [9] Farman Ali, Daehan Kwak, Pervez Khan, S. M. Riazuul Islam, Kye Hyun Kim, K. S. Kwak, " Fuzzy Ontology-based Sentiment Analysis of Transportation and City Feature Reviews for Safe Traveling", Elsevier, Vol. 77, pp. 33-48, 2017.
- [10] Unnisa Muqtar, Ayesha Ameen, and Syed Raziuddin, "Opinion Mining on Twitter Data using Unsupervised Learning Technique", International Journal of Computer Applications, Vol. 148, No. 12, 2016.
- [11] Jiexiong Tang and Guang-Bin Huang, "Extreme learning machine for multilayer perceptron", IEEE transactions on neural networks and learning systems, Vol. 27, No. 4, pp. 809-821, 2016.
- [12] El-Din, Doaa Mohey, "Enhancement bag-of-words model for solving the challenges of sentiment analysis", International Journal of Advanced Computer Science and Applications, Vol. 7, pp. 244-247, 2016.
- [13] Arenas-Garcia J, Petersen K. B, Camps-Valls G, and Hansen L. K, "Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods", IEEE Signal Processing Magazine, Vol. 30, No. 4, pp. 16-29, 2013
- [14] C Salvatore, Christian, Antonio Cerasa, Isabella Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. C. Gilardi, and A. Quattrone, "Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy", Elsevier, Journal of Neuroscience Methods, pp. 230-237, 2014
- [15] Pandya, D. H, S. H. Upadhyay, and Suraj Prakash Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN", Elsevier, Expert Systems with Applications, Vol. 40, No. 10, pp. 4137-4145, 2013
- [16] Lippi, Marco, Matteo Bertini, and Paolo Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning", IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 2, pp. 871-882, 2013
- [17] Anjaria, Malhar, and Ram Mohana Reddy Guddeti, "Influence factor based opinion mining of Twitter data using supervised learning", IEEE, Communication Systems and Networks (COMSNETS), pp. 1-8, 2014.
- [18] Poria, Soujanya, Erik Cambria, Newton Howard, Guang-Bin Huang, and Amir Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content", Elsevier, Neuro computing, Vol. 174, pp. 50-59, 2016
- [19] Shang, Chao, Fan Yang, Dexian Huang and Wenxiang Lyu, "Data-driven soft sensor development based on deep learning technique", Elsevier, Journal of Process Control, Vol. 24, No. 3, pp. 223-233, 2014
- [20] Gautam, Geetika and Divakar Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis", IEEE, Contemporary computing (IC3), pp. 437-442, 2014.
- [21] Kazemian, H. B, S. Ahmed, "Comparisons of machine learning techniques for detecting malicious webpages", Elsevier, Expert Systems with Applications, Vol. 42, No.3, pp. 1166-1177, 2015
- [22] Hammer, Barbara, Daniela Hofmann, Frank-Michael Schleif, and Xibin Zhu, "Learning vector quantization for (dis-) similarities", Elsevier, Neurocomputing, Vol. 131, pp. 43-51, 2014
- [23] V. K. Singh, , Piryani, R., Uddin, A., & Waila, P, "Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches", IEEE, pp. 122-127, 2013
- [24] K Kianmehr, R. Alhajj, "Effectiveness of Machine Learning Techniques for Automated Identification of Calling Communities", 12th International Conference Information Visualisation, pp. 308-313, 2008.
- [25] Yoshua Bengio, Lamblin P, Popovici D and Larochelle H, "Greedy layer-wise training of deep networks", In Advances in neural information processing systems, pp. 153-160, 2007.
- [26] Eyke Hullermeier, "Fuzzy methods in machine learning and data mining: Status and prospects", Fuzzy sets and Systems, Elsevier, Vol. 156, No. 3, pp. 387-406, 2005.