Machine Learning Approaches for Sentiment Analysis: A Review

Sachin Balvir
Dept of Information Technology
DMIETR, Wardha
Maharashtra, India

Jayanat Rohankar
Dept of Information Technology
TGPCET, Nagpur
Maharashtra, India

Jagdish Yadav
Dept of Information Technology
DMIETR, Wardha
Maharashtra, India

***Abstract*— Today YouTube is very useful and it is most popular site in social media. It is also the most trendy site in social media, where users interact with each other through sharing, commenting and rating on videos Sentiment analysis is the field of study related to analyze comments & emotions of users .Due to comment analysis we can easily find best or most popular video among the various video. In YouTube there are many videos on similar topic and we cannot decide which videos are good. So by performing the sentiment analysis we can decide which is video is best .This analysis helps to find out the most relevant and popular video of YouTube according to search. This papers describes the techniques of machine learning that can be used for sentiment analysis.**

Keywords*— *Machine Learning; Sentiment Analysis; Comment Analysis

I. Introduction

The World Wide Web is a popular space where people express their individual opinions and emotions. It also includes an influencing aspect of life, with implication for marketing and business purpose. Sentiment analysis is a technology that helps to extract the information about the user's feedback. Sentiment analysis is a classification of emotions such as positive, negative and neutral within users review using text analysis techniques. Sentiment analysis model identify duality within a text e.g. a positive, negative or neutral opinion whether it is a whole document, paragraph, sentence or word. For example, in business it is essential to understanding people's emotions. Automatically analyzes customer review to listen attentively to their customers and product and services to meet their needs. One of our customers used sentiment analysis to analyze 1,495+ reviews about their product, and identify that customers were happy about their pricing but complained a lot about their customer service [1, 2].

The popularity of social media is increasing rapidly .YouTube is the most popular in social media because we can easy use and share video as well as we also share image. In YouTube there are many comments and videos are uploaded by users .Sometimes users cannot decide which video is good because of many videos are seen in YouTube. By sentiment Analysis we can decide which video is good and proper. We can use Algorithm for Analysis the video. Sentiment Analysis describes its keywords "opinion mining" in two ways, namely 1). Analysis and evaluative text, and 2). Predictive rating tracking [8] Sentiment analysis is a process of picking up and favorability of a natural language [8]. A systematic approach to Opinion Mining (OM) on YouTube comments by (i) modeling classifiers for predicting the opinion polarity and the type of comment and (ii) proposing robust shallow syntactic structures for improving model adaptability [6]. Automatic process for finding useful video by sentiment analysis of user's comments based on Natural Language Processing (NLP). The approach evaluated the quality, relevancy and popularity of YouTube videos considering the relationship of user's sentiments expressed in comments [9].

Sentimental analysis can be considered in following areas:

Area	Attributes for Sentiment Analysis
Fashion	Accessories , outlets , designers , brands
Malls and stores	brands, services , prices, products ,location, facilities
Online services	online stores, gaming services , search engine ,maps ,social networking, tools, application
Personal finance	finance institutions , loans , financial services ,branch , cards
Travel	hotels, location , airlines
Health & food	restaurants , pubs, health club , health products & accessories , hospitals , day care centers
Books	shops, title, authors, resale, libraries
Entertainment	Movies, plays, music, shows, multiplexes, channels
Electronics	computers, cell-phones, ACs, TVs, cameras

Table-1: Areas of sentiment Analysis

II. Machine Learning Techniques

There are two different types of Machine learning technique. First techniques, it trains a model on input and output data so that it can predict future output. Such type of techniques is called Supervised Machine learning. Supervised machine learning builds a model that makes predictions based in evidence in the presence of uncertainty. In Second type of techniques, it finds hidden patterns or intrinsic structure in input data. Such techniques are called Unsupervised Machine Learning.

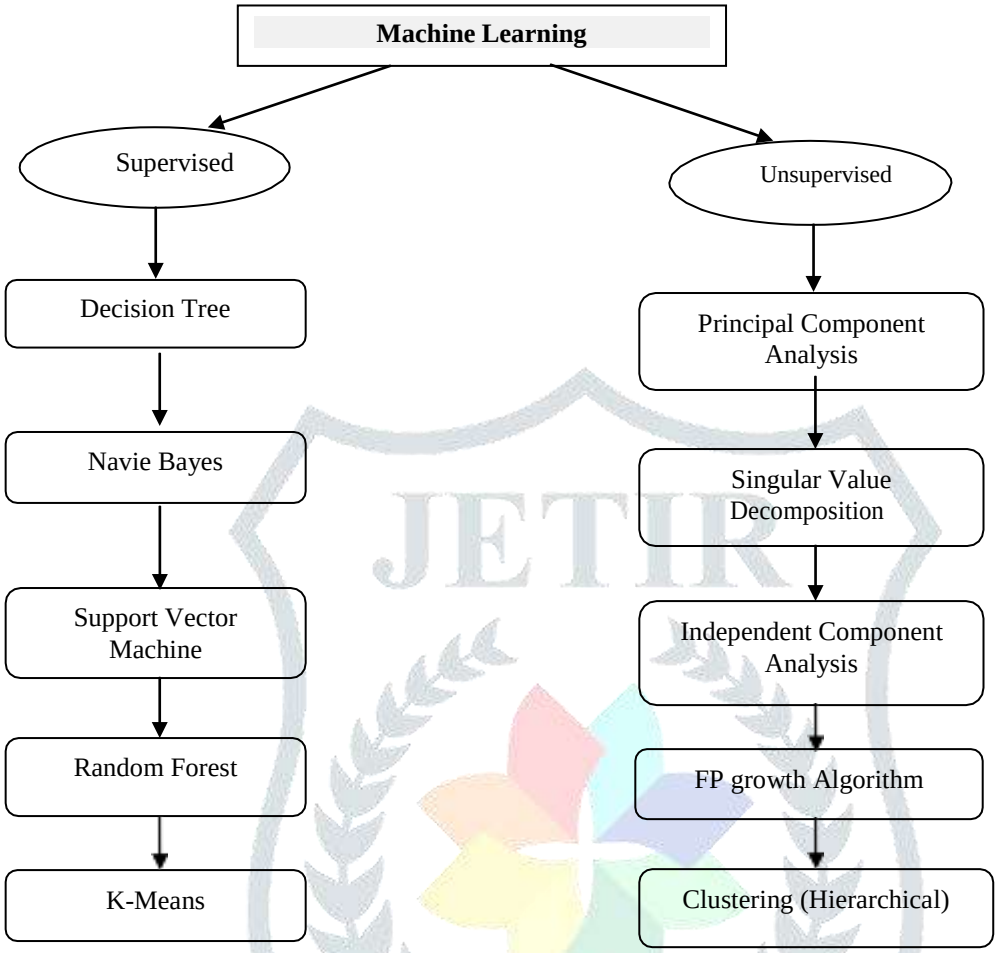

Fig-1: Types of Machine Learning

III. Sentiment Analysis Using Supervised Learning

A. *Decision Tree*

Decision tree algorithm is a two step process, learning step and prediction step in decision tree machine learning. First step is learning step, the developed model is based on given trading data. In next step is prediction step, the model is used to predict the response for given data.

Machine learning using decision tree is a most popular classification algorithm and easiest algorithm to understand. Decision tree it can be used concept of classification and concept of regression. It is nothing but a classifier.

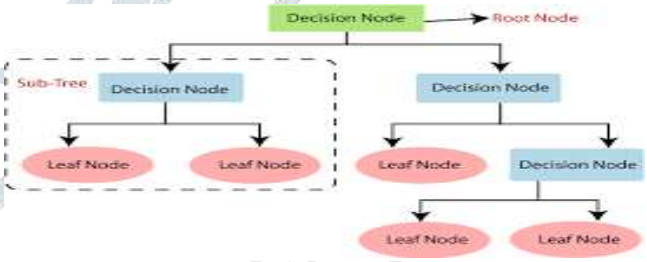

Fig-2: Decision Tree

Decision tree used in 2 node are decision node and leaf node .decision node are nothing but test. This test is performed on the feature and attribute. Decision tree is a tree structure First step is a decision tree .Then second leaf node is break down a data set into the smaller subset while at the same time an associated decision tree. Leaf node is a represent a classification or decision. There are other kinds of predicates

which depend on the similarity of documents to correlate sets of terms which may be used to further partitioning of documents [5].

B. Navie Bayes

It is a classification technique based on Bayes' theorem with an assumption of independence among predictors. In simple terms a Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
It is rule in probability and statistical theory that calculates an event probability based on related conditions or events. Scientists and medical professional may use these calculations to determine the like hood of various outcomes .a
Conditional probability derives Baye's theorem.
- The conditional probability of event A, given that event B occurs, is written as P (A/B).
-How can we calculate this value? Let's simplify the math world derivation a bit; you can find the full article in the resources section of this article.
- The formula for the combined probability of the events is: P (AnB) =P (A)*P (B/A).
-The probability that both event will occur equals the probability that A will occur given that B has occurred.
Since P(AnB)=P(BnA), we also have P(A n B)=P(A)*P(B/A)=P(B)*P(A/B)
-Then we arrange the equation and solve for P (A/B), to complete Bayes theorem:
P(A)*P(B/A)=P(B)=PA/B) as P(A/B)= P(A)*P(B/A)/P(B).
-There are several forms of Bayes theorem which some call Bayes rules.
-another various of Bayes rule is P (A/B) = P (A)*P (B/A)/P (B).
-This form emphasize that the term P (B/A)/P (B) adjust the "prior" probability , P(A), to given the conditional , or "Posterior "probability P(A/B).
-Note that the conditional probability relationship "i" is not commutative, P (A/B) ≠ P (B/A).
-Also, let's remember to avoid dividing by zero .these simple Bayes formula require that event B is possible
so P (B)>U
-As a reminder that the order matters , went for Bayesian probabilities are often written as the prior "Hypothesis " (H) and the "Evidence "(E): so P(H/E)=P(H)*P(E/H)/(E).
-Sometime we do not know the overall probability of the evidence rather than P(E), we have
P (~E/H): P (E) ~P (H) or P (~E/ ~H) instead.
-if we know P (E/H) and P (E/~H) we use the second form of Bayes theorem :
P(H/E)=P(H)*P(E/H)/(P(H)*P(E/H))+(P(~H)*P(E/~H))
-Again, we must not divide by zero.

C. Support Vector Machine

It lies in super wise learning used for classification & regression. It constructs a hyper lane or decision line used for classification & regression. In SVM algorithm, each data item plot in n-dimensional with the value of each feature. Then after two data item are separated by decision foundry or in generalize from it is called as hyper plane. It is a kind of boundary between the two classes which decides the new data from which class .Then after one is consider class and another class or vise verse. Here maximizing the distance between nearest data point and hyper plane will help us to decide the right hyper plane. This distance is called as margin. Margin can be called by addition of two distances.

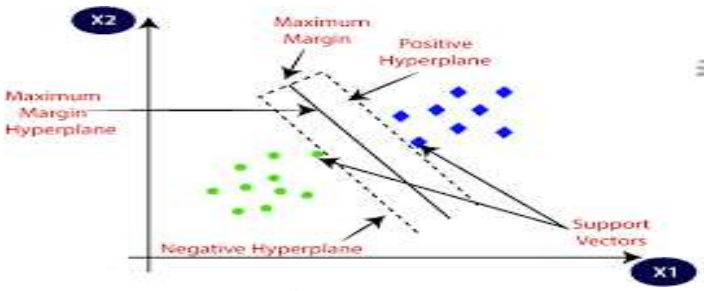

Fig-3: Example of Support Vector Machine

We also draw hyper-plane in horizontal or vertical manner. But only maximum margin hyper plane should be selected .Maximum width margin can best for future predication or less the error rate. The result of classification of weighted values according to Support Vector Machine (SVM) method has brings us to the conclusion that the value of True Positive rate is 91.1% based on the comments taken from 2015 until 2016 [4]

IN NON LINIER SVM & KERNAL FUNCTION

Non –linear SVM can be used when the data cannot be used when the data cannot be divided into two class by just simply draw a strength line .By using strength line we are not able to divide. If we used kernal function it takes the as low dimension feature space and get output high dimension feature space. When data is in high dimension the data can be easily separated.

D. Random Forest

Random forest was found by Leo Breiman .who was encouraged by earlier work by Amit and German. Even if not clear from the definition in Random Forest is an addition of Breiman sacking idea and were advance as a contestant to improve. Random Forest can be used for an unconditional response variable, mention to in as "classification", or a constant reaction, referred to as "regression". Likewise, the predicator variables can be either unconditional or constant.
Random Forest are appealing because they

- Naturally handle both regression and classification;
- Are relatively fast to train and to predict;
- Depend only on one or two tuning parameters;
- Have a built in estimate of generalization error;
- Can be used directly for high dimensional problem;
- Can easily be implemented in parallel.

Statistically, Random forest are appealing because of the additional features they provide, such as

- Measures of variable importance;
- Differential class weighting;

- Missing value imputation;
- Visualization

E. K-Means

K means algorithm is a repeated algorithm that tries to separation the data set into K pre-defined distinct non overlapping subgroups where each data point belongs to only one group. It is used to separation the data set automatically into k groups. It is start by choose k initial cluster centers and then iteratively clear them until the algorithm is connect [3, 7]. There are each data point remains in the same cluster even after the next repetition.

In calculating the distance between the data points using the formula as shown in below :

$$F = \sum^{k} j = 1 \sum^{n} i = 1 \, (Xi - Cj)^2$$

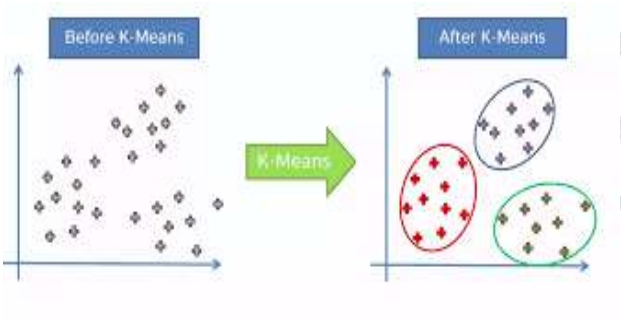

Fig-4: K-Means Sample example

The steps involved in the execution of the algorithm are as follows:

1) Let 'D' be the data set which contain of a set of data points {A1, A2, A3……An}.
2) Let C = {c1,c2,…ck} be the set of cluster centers .
3) Calculate the distance between each data point to the cluster center.
4) Set the data point to the cluster center whose distance from the cluster center is smallest when compared to other cluster centers.
5) Recalculate the new cluster center Cj by calculating the average of the data points Xi in the particular cluster.
6) Repeat the steps (4) and (5) until intersection and return the final cluster centers {c1, c2….ck} along with the data point in their respective clusters.

IV. Sentiment Analysis Using Supervised Learning

A. Principal component Analysis

In principal component analysis (PCA) is a statistical procedure. Is that uses an orthogonal transformation which convert a set of correlated variable to set of uncorrelated variables. Principle component analysis (PCA) is a most widely used tool in exploratory data analysis in machine learning. PCA is used a unsupervised, non-parametric used for a dimensionality reduction in machine learning. PCA is convert to the set of high dimensionality into the set of low dimensionality. Can be find no of PC can be less than or equal to No of attribute. Even as a data for the building of model focus on first PC and PC1 and PC2 should be independent of each other [12].

The capabilit to generalise correctly become exponentially strong as the dimensionality of the training data set grows, as the training set covers a reduce fraction of the input space. Model also become more well organized as the minimize feature set boosts learning rates and decrease calculation costs by removing unnecessary features. PCA can also be used to used to filter harsh datasets, such as image compression. The first principal component shows the most amount of difference. Each new component expresses less variance and harsher, so representing the data with a compact subset of principal components maintain the signal and reject the noise.

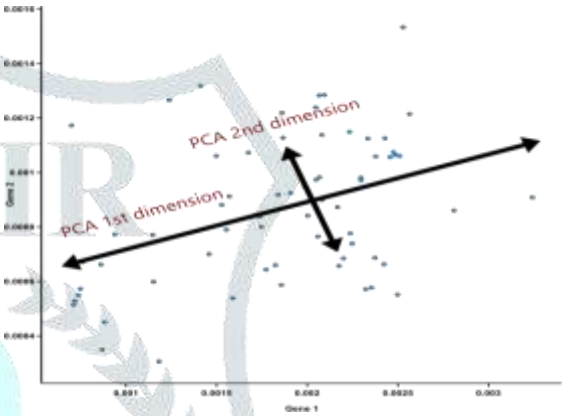

Fig-5: Principal Component Analysis example

PCA is used in a covariance formula is:

$$c = \begin{matrix} var(x,x) & cov(x,y) \\ cov(y,x) & var(y,y) \end{matrix}$$

Cov (x, y)= $\sum_{i=1}^{n} \frac{(xi-x^{-})(yi-y^{-})}{n-1}$

B. Singular Value Decomposition

The skill of singular value decomposition or SVD for short has extended and quit amazing history. It begins out in the social sciences with intelligence testing. Quick intelligence, such as verbal and spatial, were often closely match. Today singular value decomposition has increase by many branches of science, in particular psychology and sociology, climate and atmospheric science, and astronomy. It is also very useful in machine learning and in both graphic and guessing statistics. Singular value Decomposition or SVD has a vast array of application. These involve dimensionality reduction, image compression, and denoising data .In spirit, SVD states that a matrix can be represented as the product of three other matrices. [14]

The singular value decomposition or SVD for short is a matrix decomposition method for minimize a matrix to its compose

parts in order to make definite subsequent matrix calculations simpler. It is a highlight of linear algebra.
SVD is a method of decomposing a rectangular matrix into three matrixes.

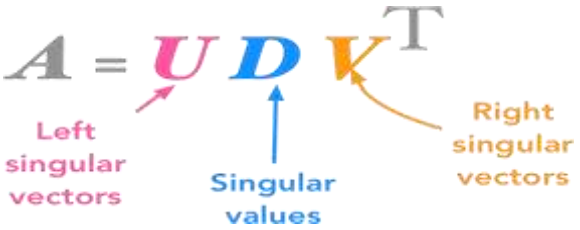

Fig-6: Decomposition of Singular matrix

Where A= input matrix which is of size m * n.
U – This matrix is called as Left singular vector
D- Singular values (r*r) diagonal matrix
r- rank of matrix A
D represents the strength of each concept
V- right singular vector ("n" term * "r" concept)
U + V are Orthogonal matrix
D is a Diagonal matrix.

C. Independent component Analysis

Independent component Analysis was first establish in the 80s by J. Herault, C. Jutten and B. Ans, and the author proposed an iterative real time algorithm. ICA is examined as an addition of the principal component analysis technique. Independent component analysis is a far used blind source separation technique. ICA has been put in many applications. ICA is usually use as a black box, without knowledge its internal details [10]. It decomposes the different signal into its independent source signals. It deals with the Independent components. It doesn't focus on the issue of conflict among the data points. It doesn't focus on the mutual orthogonality of the components. It focuses on the mutual independent of the components.

D. Clustering (Hierarchical)

Hierarchical clustering means to creating clusters that have a predetermined ordering from top to bottom. For example all school and class are organized in a hierarchy. Hierarchical clustering known as hierarchical cluster analysis, is an algorithm that groups matching objects into groups called clusters. The endpoint is a get of clusters, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly match to each other. In data mining and statistics, hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters [11].

- Two main types of hierarchical clustering
  - Agglomerative:
    - Start with the points as individual clusters
    - At each steps, merge the closest pair of clusters until only one cluster(or k clusters)left
  - Divisive:
    - Start with one, all-inclusive cluster
    - At each step, split a cluster until each cluster contains a point (or there are k cluster)
- Traditional hierarchical algorithm use a similarity or distance matrix
  - Merge or split one cluster at a time

E. FP- Growth Algorithm

FP-growth is a very fast and memory efficient algorithm. It uses a special internal structure called an FP-Tree. It is a tree based algorithm. It constructs conditional frequent pattern tree and conditional pattern base from database which satisfy minimum support. FP-growth uses a depth-first search. FP-growth utilizes a pattern-growth approach means that, it only considers patterns actually existing in the database. Runtime increases linearly, depending on the number of transaction and items. Data are very interdependent, each node needs the root. It requires less memory space due to compact structure and no candidate generation. It scan the database only twice for constructing frequent pattern tree. The FP-growth algorithm, suggest by Han in, is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth, using an extended prefix-tree structure for storing compressed and crucial information about frequent patterns named frequent-pattern tree (FP-tree) [12].
Frequent patterns are itemsets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold. For example, a set of items, such as milk and bread that appear frequently together in a transaction data set is a frequent itemset.

V. Conclusion

This paper illustrates a various machine learning algorithm of supervised and unsupervised type. Supervised and Unsupervised algorithms described in this paper are very much useful for sentimental analysis on data. Supervised types of algorithms are decision Tree, Navie Bayes, SVM, Random forest, k –means and Unsupervised algorithm are principal component analysis, Singular value decomposition, independent component analysis, FP growth algorithm, clustering(Hierarchical). Each algorithm specified here are suited for specific type of problems. It is difficult to know which one is the best .But based on the review of various research papers, the SVM gives the best results where you have huge amount of features.

References

[1] F. Neri, C. Aliprandi, F. Capaci, M. Cuodras, T. Bay,"Sentiment Analysis on Social Media", ACM International Conference on Advances in Social Network Analysis and Mining, 2012

[2] S. Xiong, "Improving Twitter Sentiment Classification via Multi-Level Sentimental-Enriched Word Embeddings", Association for the Advancement of Artificial Intelligence, 2017.

[3] G. Ravali, P. Swathi Moulika, A.A parna, J.Abhinaya, T.Anuradha., " Analysing YouTube Data Using K Means Clustering", International

Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 6, Issue 2,March –April 2017

[4] Fiktor Imanuel Tanesab, Irwan Sembiring and Hindriyanto Dwi Purnomo., "Sentiment Analysis Model Based On YouTube Comment Using Support Vector Machine", International Journal of Computer Science and Software Engineering (IJCSSE), Page- 180-185, Volume 6, Issue 8, August 2017 voloume 6, Issue 8, August 2017

[5] Walaa Medhat, Ahmed Hassan, Hoda Korashy., "Sentiment analysis algorithm and application: A survey", Production and hosting by Elsevier B.V. on behalf of Ain Shams University. Received 8 sept 2013; revised 8 april2014; accepted 19 April 2014

[6] Aliaksei Severyn, Alessandro moschitti, Olga Uryupina, Barbara Plank, Katja Filippova., "Opinion mining on YouTube", Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pages 1252–1261, Baltimore, Maryland, USA, June 23-25 2014. c 2014 Association for Computational Linguistics.

[7] Jain, Anil K. "Data clustering: 50 years beyond Kmeans." Pattern recognition letters 31.8 (2010): 651- 666.

[8] Bo Pang., Lillian Lee., "Opinion Mining and Sentiment Analysis" Vol 2, 2008

[9] Hanif Bhuiyan, Jinat Ara ,Rajon Bardhan, Md. Rashedul Islam., "Retrieving YouTube Video by Sentiment Analysis on User Comment" Proc. of the 2017 IEEE International conference on signal and Image processing Applications(IEE ICSIPA 2017), Malaysia, sept 12-14,2017

[10] Dominic Langlois, Sylvain Chartier, and Dominique Gosselin , "An Introduction to Independent Component Analysis: InfoMax and FastICA algorithms", Tutorials in Quantitative Methods for Psychology, 2010, Vol. 6(1), p. 31-38

[11] Pelin Yildrim , Derya Birant, "K-Linkage: A New Agglomerative Approach for Hierarchical Clustering", Advances in Electrical and Computer Engineering, Volume 17, Number 4, 2017, DOI: 10.4316/AECE.2017.04010

[12] B. Gupta, D. Garg, "FP-Tree Based Algorithms Analysis: FPGrowth, COFI-Tree and CT-PRO", International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 7 July 2011

[13] https://www.geeksforgeeks.org/ml-principal-component-analysispca/

[14] https://towardsdatascience.com/singular-value-decomposition-example-in-python-dab2507d85a0