A Novel Hybrid Framework For Twitter Sentiment Analysis Using Feature Selection Method Isomap and Machine Learning Methods

Iqra Hussain[1], Dr. Hamid Ghous[2], Tahreem Zahra Rizvi[3]

[2]Institute of Southern Punjab, Multan, Pakistan

Abstract – Twitter tweets have become an exceedingly main stream and advantageous method of conveying and communicating feelings. Every day millions of users use twitter to express their views and feelings. Twitter is rich with data that can be utilized for sentimental analysis and nostalgic investigation. Since twitter has shown up generally as of late, a handful of an appraising works were not devoted to this point. There is a need of framework to perform sentimental analysis using feature selection methods. In this paper, a novel hybrid framework is based on feature selection method, Isomap, classification methods, SVM, decision tree and Naïve byes . The results clearly show that SVM method performed better than the others.

Keywords: Sentiment Analysis, Isometric Feature Mapping (ISOMap, Decision Tree, Naïve Bayes, Support Vector Machine (SVM).

I. INTRODUCTION

Twitter is a social media platform introduced in 2006 by Jack Dorsey. It is mostly used for sharing ideas with public where anyone can read ideas as tweets i.e. Users on this social network are increasing day by day which leads to a large number of data. All those tweets can definitely get a lot of useful information when it comes to data processing.

Data mining is used for multiple purpose which help to get patterns, information or any correlations available in that large number of raw datasets. Data mining is playing a major role in today businesses where we can predict about the business growth with the help of processing industry's data. Machine learning and data mining work together to perform this task. Machine learning algorithms access data and perform some actions on that during the process of information extraction.

Machine learning is a branch of an artificial intelligence which is playing a major role in the technology. Machine learning algorithms help systems to learn, train and become more predictable about any given data input. It is further divided into two types supervised and unsupervised machine learning. Supervised machine learning method is used here on twitter tweets. Supervised machine learning method usually works on a labelled input data. On the basis of labelled data or provided

answer key as input, it can further make predictions by comparing them with the given input. It has then further types like classification methods such as support vector machine (SVM), linear regression, decision tree and naïve Bayes. In this paper, there are some supervised methods of data mining are used.

Supervised

- Naïve Bayes
- Support Vector Machine (SVM)
- Decision Tree

Sentiment Analysis on Twitter

Sentimental Analysis might be a field that is developing rapidly. 81% of the Internet users (or then again 60% of Americans) have done online exploration on an item at least once [5], which means per annum more articles are focusing on various content spaces over years, where the surveys speak to adjust the 49.12% of the articles [6]. One wouldn't continuously need to utilize estimation to item surveys; there are an excessive number of different fields. One genuine case of this has been tested [7] is that the correlation of Tweeter estimation vs Gallup surveys of purchaser certainty. The outcomes submitted are positive in this way the relationship was 0.804, surmising that will utilize Twitter live general sentiment. This is accurately what we're getting the chance to utilize Twitter during this explore: to extricate conclusions from these and specified the tweets' extremity progressively.

II. SYSTEM MODEL

Sentimental analysis is the computationally definitive method of the authors' viewpoint or attitude as neutral, positive and negative. In order to grow their business, it is very convenient to follow the knowledge of the client. In this work, we propose a novel framework for tweets' sentiment analysis based on feature selection method Isomap and machine learning methods: SVM, decision tree and Naïve Bayes.

Twitter dataset

In this section data is collected about Twitter using, Twitter API through Develop account. Total dataset are 31962 tweets in training datasets in which the positive value is 29959 and the negative value are 2003. Data is collected with using the following steps.

1) Credential

2) Consumer Key

3) Consumer Token

4) Access Token

Table 1

ID	Label	Tweets
31963	0	#PMLN #MaryamNawaz #requires #passion # Bilawal #willpower to find #newmaterialsÃ¢Â€Â¦
31964	1	@user #Blue #Tahreem is a good girl Ã¢Â€Â˜ #birdsÃ¢Â€Â™ #movie Ã¢Â€Â” and hereÃ¢Â€Â™s why
31965	2	A better to talk with #acne!! #fitness #relieve!!
31966	3	of the hp and the doomed little one of rise for bound before? if yes,why? if no, whom? Ã°ÂŸÂ˜Â□Ã°ÂŸÂ˜Â□Ã°ÂŸÂ˜Â□ #harrypotter #pottermore #favorite

Here, two classes are selected   0 and 1, sad or hated as "0" , happy or cheery as "1". Twitter API Python covering for API solicitations to recover Twitter information from Twitter. Twitter API Includes the accompanying Steps Installation(s) of the required programming. Authentication(s) of client data. The principle establishment software(s) incorporates mass NLTK, tweepy text and so forth Verification includes the distinctive step(s) Visit the twitter website(s) and click the catch 'make new app'. It will be diverted to the application page where the purchaser keys, buyer access, access token and access token mystery that is expected to get to the twitter information will be available Implement in python. Fill the detail(s) in the structure gave and submit. There are numerous ways to save the twitter info. They are like, MongoDB, the database for open source document management, is the "No SQL" go-to.

Classification Techniques

Machine learning algorithm

Machine learning is an important domain of artificial intelligence, whereby the word implies IT systems' ability to freely find outcomes of problems through observation(s) of database design. In other words,

Machine Learning helps IT system(s), based on current algorithms and data sets, to observe designs and to organize suitable solution concepts. Therefore, may be the basis of experience(s), artificial knowledge is design in Machine Learning.

There are different Machine learning techniques but in this paper we will discuss three of them

1) Decision Tree

2) SVM

3) Naive bayes

1) Decision Tree

Some methods are selected in statistics, machine learning and and another method, data mining for simulation is decision tree learning. It also uses a decision tree (as a portending template) to figure the conclusion of the component (defined in the sections) to the choose cost of the component (defined in the petals).

In these tree systems, tree structures are called classification tree leaves, representing class labels, where a different set of values can be taken by the selected variable, and Branches reflect an association of characteristics that relate to certain class labels. Decision trees where continuous values will take up the target variable are called regression trees. Decision trees, specified their eloquence and easiness, are between the suitable machine learning algorithms. A decision tree learning method normally adopted in data mining . The goal is to make a pattern based on a few variables that forecast an objective variable's estimation.

A Decision tree is a description of models for ordering. Accept that there are small discrete areas for this segment for the entire info highlights, and there is a single objective element called the structure. As a class, each part of the order(s) region is known. A decision tree or an ordered tree is a tree in which an inclusion of information label(s) each internal (non-leaf) hub.

Each tree leaf is marked with a class or likelihood distribution over the classe(s) meaning that if the selection tree is constructed around it, the information index has been ordered into either a particular class or a particular likelihood circulation that is slanted towards specific subsets of classes.

Data appears in recorded forms:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, \ldots, x_k, Y)$$

The predicted variable we are attempting to grasp, describe or generalize is the "Y" dependent variable. The vector x consists of the characteristics of "x1, x2, x3" etc. that are used for that role.

2) Naïve bayes

The Naïve Bayes algorithm is a managed programmed algorithm based on the Bayes formula and adopted to resolve difficulties with analysis. It is mainly observed in text processing, which needs to test a data of high-training. This is a easiest and most prevailing analysis algorithms is the Naïve Bayes Classifier, that need to build models of fast machine learning that can sorts rapid prediction.

Types of Naive Bayes Algorithms

Following are the types of Naïve Bayes Algorithm

I)   Gaussian Naïve Bayes(s)

II) Bernoulli Naïve Bayes(s)

III) Multinomial Naïve Bayes(s)

I.   Gaussian Naïve Bayes

If trend values are constant in nature, then it is hypothesized that according to Gaussian, which is Natural Disposal, the values associated with each class are dissolved.

II.   Multinomial Naïve Bayes

The use of multinomial Naive Bayes on data disposed of by multinomial is preferred. It is used extensively in NLP in text allocation. In text classification, each occurrence determines the existence of a word in a document.

III.   Bernoulli Naïve Bayes

As data is scattered as confirming to multivariate Bernoulli spreading, Bernoulli Naive Bayes is utilized. This means that there are numerous attributes, but a binary value is believed to consist each one. So, to be binary-valued, it needs qualities.

3) SVM

The purpose of the support vector machine algorithm is to analyze a hyperplane that categories the selected data individually and highlights in the characteristics N-dimensional space N-number(s).

SVM model is a summary of the instance(s) as space dots, summarized so that a clear distance that is as good as possible isolates the sub-topic samples. Depending on the side of the distance they fall on, recent types are then measure and predicted in the same space to belong to a group.

In order to distinguish the selected kinds of data points, a number 0f potential hyperplanes that could be selected. The goal is to find a high-margin aircraft, is, in both categories, the highest distance between the data points. Over estimate the margin difference gives some assistance so that it is possible to classify more confidence in future data point(s).

Hyperplanes and Support Vectors

Hyperplanes are inspection procedures that help to define data points. Data points on each side of the hyperplane may be assigned to various classes.. The hyperplane(s) 's dimension also relies on the characters of functions. If 2 is the number of components reached, So, the hyperplane is almost row. When the number of input features is 3, the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine as the number of functions exceeds 3.

Help vector(s) are information point(s) that are near to the hyperplane and impact the hyperplane(s). The margin of the classifier is used by maximizing these assistance vectors. Deleting support vectors can shift the hyperplane(s) location. These are the points that assist us in developing our SVM

Figure 1:Framework
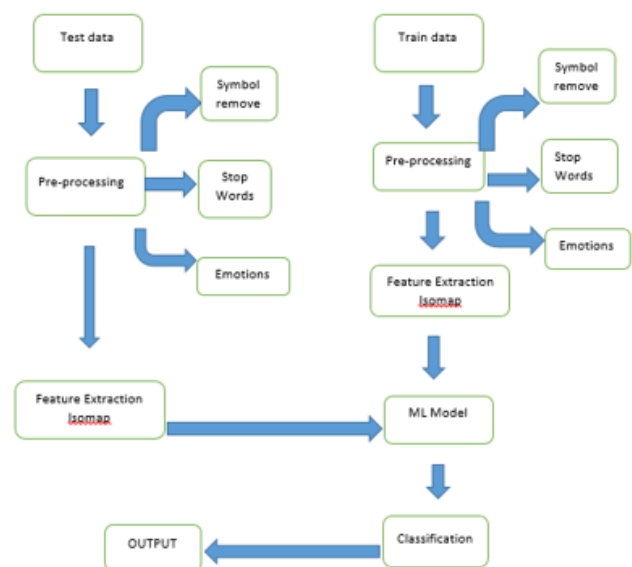
I.   DATA ANALYSIS& RESULTS

In previous section methodology was discussed which we are followed to obtain result while in this section result are described and discussed. An experiment was conducted on the classifiers described in this paper using modules from the python machine learning environment.

Implementation of Experiment

Anaconda (2017)and many tools are used here for the classification of tweets. Here for the tweeter dataset we are using machine leaning techniques (jupyter tool)are used. Here different tweets and dataset used by applying tools. The data is divided in to two classes 0 and 1. These two sets having different datasets 0 having positive tweets and 1 having negative tweets.

Experiment No 1

The experiment on the tweeter datasets was performed. In this experiment, two datasets are used.1 Training datasets

on which the machine learning model are introduced. Secondly, testing datasets on which we test the trained model. The decision tree, Naïve Bayes and SVM model are applied to this tweeter data. Here we verify the accuracy of tweeter datasets using different frameworks on these model assessments. We will review the training validation details and the testing base.

Following Model are applied on these Twitter datasets

1.Decision Tree.

2.Naive Bayes.

3.SVM

Decision Tree Model

A Decision tree is an easier presentation of models for ordering. Accept that there are small discrete areas for this segment for the entire info highlights, and there is a single objective element called the structure. As a class, each part of the order(s) region is known. A decision tree or an ordered tree is a tree in which an inclusion of information label(s) each internal (non-leaf) hub. Both of the potential target variables estimate are labelled with circular segments originating from a hub named with included information, or a subordinate choice hub is prompted by the curve on an alternate information highlight.

Each tree leaf is marked with a class or likelihood distribution over the classe(s) meaning that if the selection tree is constructed around it, the information index has been ordered into either a particular class or a particular likelihood circulation that is slanted towards specific subsets of classes.

Table 2

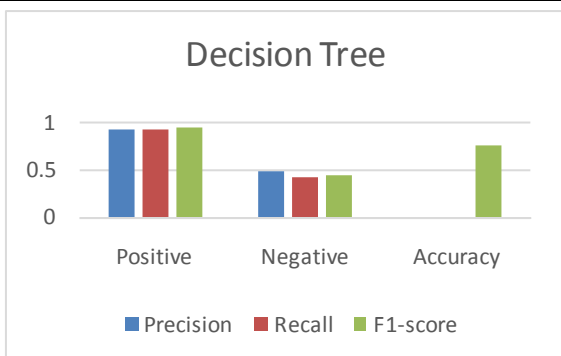|	F1-score	Precision	Recall
Positive	0.95	0.93	0.94
Negative	0.45	0.5	0.42

Figure 2: Decision Tree

Graph Representation of Decision Tree

1) After performing the Evaluation on decision tree. Following accuracy, we get the as shown in the table. We get the precision, recall, f1-score and support of positive and accuracy.

2) The result and accuracy 0.77 of decision tree.

Naive Bayes

The Naïve Bayes algorithm is a Bayes theorem-based managed learning algorithm used to solve analysis problems. It is mainly used in the processing of text, which involves high-dimensional data set training. The Naïve Bayes Classifier which helps to build quick machine learning models that can make rapid prognosis, is one of the simplest and most efficient analysis algorithms.

Following are the types of Naïve Bayes Algorithm

- Gaussian Naïve Baye(s)
- Bernoulli Naïve Baye(s)
- Multinomial Naïve Baye(s)

Graph Representation of Naïve Bayes

1) After performing the Evaluation on decision tree. Following accuracy to get as shown in the table. Then get the precision, recall, f1-score and support of positive, negative and accuracy.

2) The result and accuracy of Naïve Bayes 0.87

Table 3

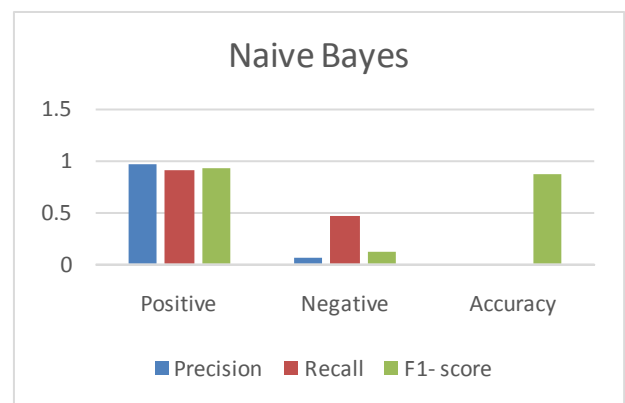|	F1-score	Precision	Recall
Positive	0.93	0.96	0.91
Negative	0.12	0.06	0.47

Graph Representation of Naïve Bayes

Figure 3: Naïve Bayes

3) **SVM**

The purpose of the support vector machine's algorithm is to find a hyperplane that separately classifies the data points in the characteristic(s) N-dimensional space N-number.

An SVM model is a description of the space points of the instance(s), mapped so that a clear distance that is as good as possible isolates the sub-topic samples. New examples

are then mapped and predicted to belong to a group in the same space according to the side of the gap they fall on.

So as to recognize the two sorts of information focuses, there are numerous potential hyperplanes that could be chosen. The purpose is to find a plane with the best edge, that is, in the two classifications, the most elevated separation between the information focuses. Boosting the edge hole offers some help with the goal that more trust in potential information focuses can be arranged.

The result and accuracy of svm is 0.91.

Table 4

	F1-score	Precision	Recall
Positive	0.97	0.98	0.96
Negative	0.57	0.48	0.69

Graph Representation of SVM
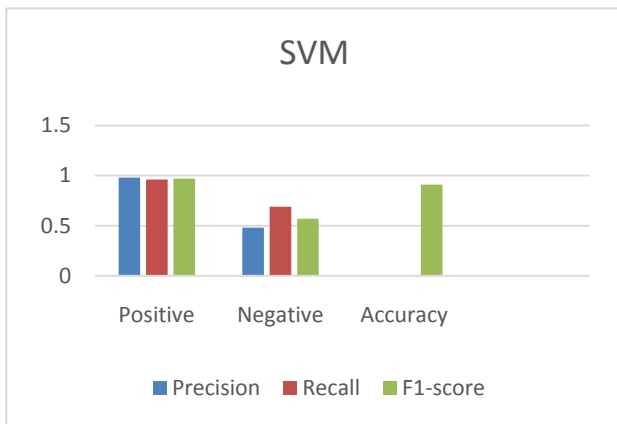
Figure 4: SVM

Table 5 Results

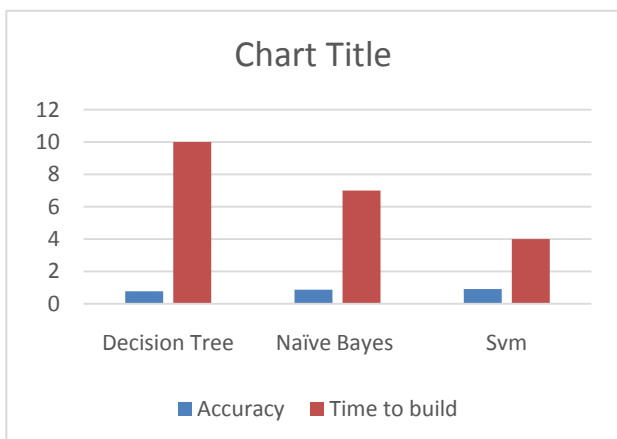| Algorithms	Accuracy	Time to build model
Decision Tree	0.77	10 sec
Naïve Bayes	0.87	7 sec
SVM	0.91	4 sec

Figure 5

Experiment No 2

In experiment 1 the accuracy is low and now in experiment 2 the Isomap is used and the accuracy is increased. The experiment is done on the tweeter datasets. Two datasets are used in this experiment. The training datasets are used to train the machine learning models. Second is the testing on which the evaluation test and trained model are required. On this tweeter data, the decision tree, Naïve Bayes and SVM model are applied. The model evaluation is performed here to check the accuracy of tweeter datasets by using different models. Here, The data is checked on training validation and testing base.

Graph Representation of Decision Tree using Isomap

1) After performing the Evaluation on decision tree. Following accuracy is appeared as shown in the table. The precision, recall, f1-score and support of positive, negative and accuracy are the results here. In the previous experimental work, the accuracy is 0.77 but in this, the Isomap and the accuracy increases.

The result and accuracy 0.92 of decision tree using the Isomap.

Table 6

	F1-score	Precision	Recall
Positive	0.96	0.95	0.96
Negative	0.47	0.5	0.45
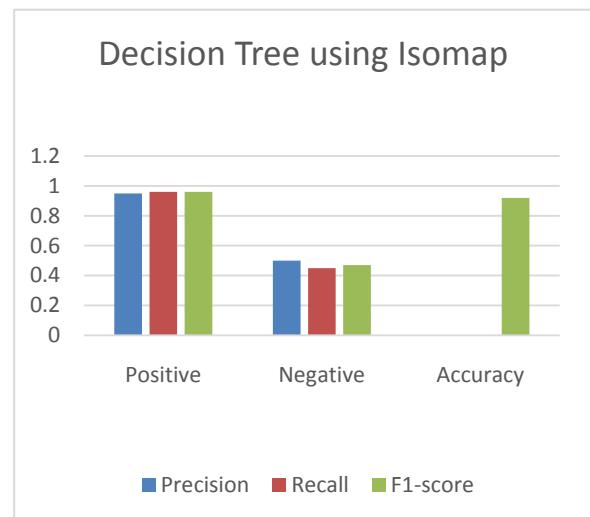
Figure 6: Decision Tree using Isomap

Table 7 Confusion Matrix of Decision Tree

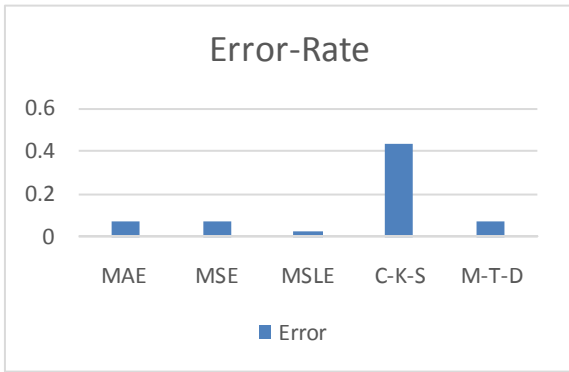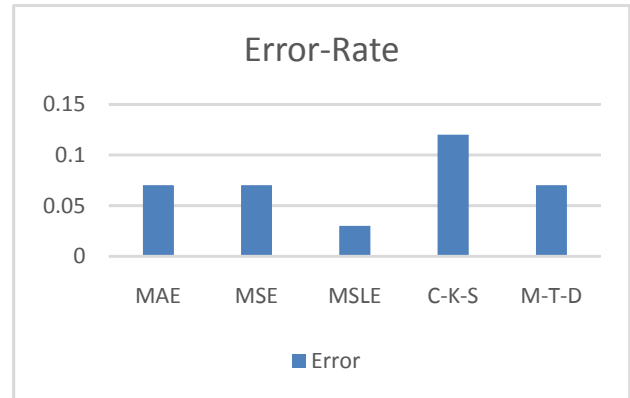|	Positive	Negative
Positive	8503	413
Negative	335	338

Figure 7

Naïve Bayes

Graph Representation of Naïve Bayes Using Isomap

1) After performing the Evaluation on decision tree. Following accuracy, we get the as shown in the table. We get the precision, recall, f1-score and support of positive, negative and accuracy. In the previous experimental work, the accuracy is 0.87 but in this we use the Isomap and the accuracy increases.

2) We get the accuracy 0.93 of Naïve Bayes using Isomap.

Table 8

	Recall	F1-score	Precision
Positive	0.93	0.96	0.99
Negative	0.50	0.14	0.08

Graph Representation of Naïve Bayes Using Isomap
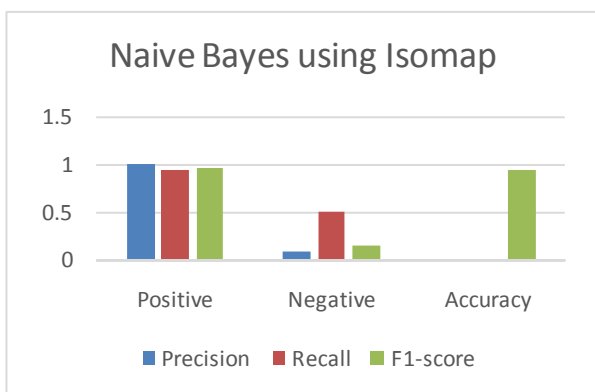
Figure 8: Naïve Bayes Using Isomap

Table 9 Confusion Matrix of Naïve Bayes

	Positive	Negative
Positive	8861	55
Negative	618	55

Figure 9 SVM

4.2.3 SVM Graph Representation of SVM Using Isomap

1. We get the accuracy of svm isomap is 0.95

Table 10

	Precision	Recall	F1-score
Positive	0.95	0.96	0.97
Negative	0.48	0.69	0.57

Graph Representation of SVM Using Isomap
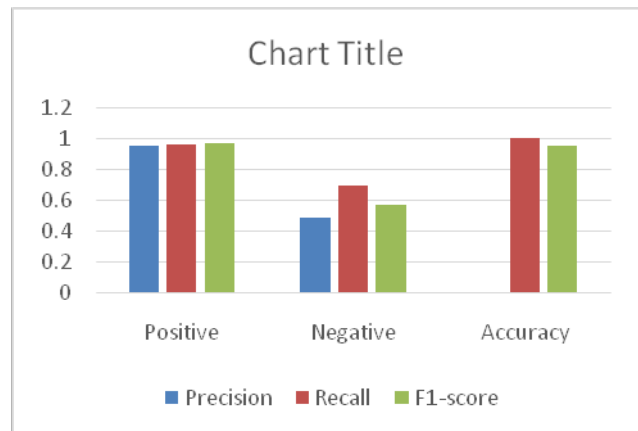
Figure 10: SVM Using Isomap

Table 11 Confusion Matrix of SVM

	Positive	Negative
Positive	8773	143
Negative	349	324

Table 12 RESULTS GRAPH

Algorithms	Accuracy	Time to build model
Decision Tree	0.92	7 sec
Naïve Bayes	0.93	5 sec
SVM	0.95	2 sec

Figure 11

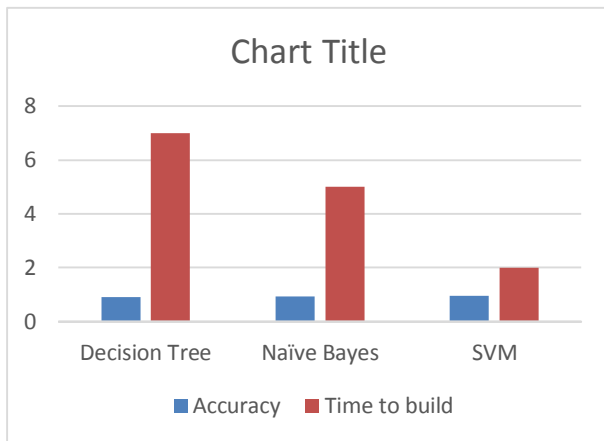### III. PREVIOUS WORK

Nikil Prakash T et al (1), worked data pre-processing tool which is significantly for Data Mining (DM) Algorithm. It's a collection of people who entered their feelings, opinions, attitude, products review, emotions, etc. The main focus is to analyse the twitter trending information and collect various information from the users. Twitter improves the accuracy. They find the people sentiment (opinion) related to election results in 2019, there is total positive sentiment is 32% and negative 18% neutral is 39%. Data Processing which is a process of DM algorithms, they work on algorithm enhances the accuracy of the Twitter data. They gathered and filtered using data cleaning algorithms to better results are given. In future improvement in the data quality, finding context words in sentiment and improvement in good accuracy of twitter data is stated.

Sanjeev Dhawan et al (2) worked on sentiment analysing the feelings of twitter info. They worked on polarity of each tweet is positive or negative. The emotions of the consumer, including rage, sorrow, happiness and joy, are a polar feeling. They used Python tool. They used twitter dataset are obtained from Twitter API to evaluate feelings of various users. There they check every tweet's feeling polarity. The feeling polarity is consumer feelings such as joy, pleasure, sorrow and indignation. The tweet is negative if the emotion polarity is equal to zero and the polarity greater than zero is positive otherwise.

Sahar A.El_Rahman et al [3] worked on sentiment analysis on twitter data. They can removed URL,stop words. They also worked to analyse tweets in English language. They used SVM, Naïve bayes and maximum Entropy in R studio. They used machine learning methods and collect the data from two different restaurants Mcdonald and KFc. They found the accuracy of positive, negative and neutral tweets and found the best result of Maximum Entropy to give better results.Each tweet has been extracted according to its positive, negative or neutral feeling. Data on two McDonald and KFC topics

were collected to Display which of the most famous restaurants. Several algorithms were used for machine learning. The outcome of these model(s) different test parameters such as cross validation and f-score have been checked. In addition, our model is high Extracted directly from Twitter on mining texts results.

Their proposed model has been improved by several algorithm(s) accuracy of the positive, negative and neutral categorization of tweets. Our approach presented combined the implementation of unattended machine learning algorithm where previous knowledge on the Lexicon is not available at first the algorithm for this. After that, several supervised models were entered with the results. It is shown that different methods for testing are used. Maximum entropy has the highest precision dependent on cross validation. McDonalds are therefore more common than KFC poor as well as optimistic evaluations. The same thing is real. Methodology for the identification of Twitter rumour's concerning disease transmission can be used in different fields. For planned events, a tweet-classifying algorithm would be an important area of research.

Shikha Tiwari et al (4) discussed that people are used to checking something before they do it like checking it films, pubs, shopping online and more. They used NLP Tool for sentiment analysis on twitter data. The decision tree and random force algorithm are also used. For present analysis, provide more precision than the SVM technique. They found that further research work and decision making may be used.

Jessica et al (5) described Twitter also introduces a large amount of data that can be related to a company's emotions. They worked stock predictions can be considered a good inventory market forecasts are not only intended to produce high results. The experimental results show that the forecast approach hits without feelings with a moveable five days method average maximum benefits. By using this with a sentiment analysis on CNBC information data, however, the average daily benefit is better.

Amar Deep Gupta et al (6) worked on the general techniques for feeling analysis. They used python and his libraries to evaluate the feeling of data obtained from Twitter Website.They worked on the GST, the Citizen Amendment Act (CAA), the Demonetization and the Bharat Money Interface (BHIM). They gathered the data and figured out that the finding(s) showed people neutral about most Internet decision(s), and will talk of the anxiety of people that they speak out openly.

Shamantha Rai B et al (7) worked on tweets or reviews and compare positive and negative polarity of tweets. The tweets' feelings which are analysed on a twitter depending

on each score word attribute rangeNaive Bayes Classifier (NBC) is used to train and test word features and determine the polarity of each tweet's feelings to pick the best features. They used Random Tree, Naive Bayes and Support Vector Machines (SVM) classification(s) are three machine-learning classifier(s). They found some new functionality introduced as part of future work, which will in turn boost prediction precision. Positive, negative and neutral polarity may also be implemented to mark the tweets,

Malika Acharya et al (8) described Twitter is the famous microblog sites which has made its place in a social networking. They evaluate view(s), hence the opinion mining efficiently. They used the Hadoop which analyses the vast high volume of data present. They work on the Hadoop cluster in an effective and timely manner. The main aim was to establish the actual line paper that automate(s) the cluster setting and then filter(s) the allocated file(s) on the basis of the name-node. The Hadoop structure along with the Map Reduce helps a lot. The use of the Hive ODBC driver allows us to produce the excel graph and to display the real orientation of the present data as well.

Prakruthi V et al (9) worked on Sentiment analysis affective states and subjective knowledge for processing natural language, text analysis. The Twitter API is used for direct tweet access and classification of sentiments Tweets. The work show results are seen for positive, negative and neutral feedback on their view(s). They analysed the tweets to obtain results. They would like to develop in future by making it possible for the system to properly classify user opinion displayed in tweets enabling the study of emoticons and even adjust the tweet perspective based on the given question.

Sanjeev Dhawan et al (10) worked on sentiment analysing the feelings of twitter info. They worked on polarity of each tweet is positive or negative. The emotions of the consumer, including rage, sorrow, happiness and joy, are a polar feeling. They used Python tool. They used twitter dataset are obtained from Twitter API to evaluate feelings of various users. There they check every tweet's feeling polarity. The feeling polarity is consumer feelings such as joy, pleasure, sorrow and indignation. If the polarity of the emotion is equal to zero and the polarity greater than zero is otherwise positive, the tweet is negative.

Hee Yong Youn et al (11)worked with unique customized functionality clustering and weighting scheme based on Chi Square is implemented for Twitter message(s) opinion analysis. Extensive research was carried out on Sentiment 140, and 4 representative feature weighting systems were also tested to show results. The test result(s) shows that the suggested system significantly outperformed others in aspects of clarity, correctness, recollect, and F1-

measure(s). They presented a method for automatic corpus selection that can be used to train a sentiment classifier. For POS-tagging, they used Tree Tagger and noticed the differences in distribution(s) between good, bad and neutral set(s). To train a sentiment classifier, they used the collected corpus. There classifier is capable of evaluating good, bad and neutral document sentiment(s). The classifier is based on the multinomial classifier Naive Bayes that uses as characteristics N-gram and POS-tags.

Dae Ryong Seo et al (12) discussed the current subjects in the domain of information processing is sentiment analysis or opinion mining. They work on polarity of text data (document, sentence, paragraph) will result in a positive, negative or neutral characteristic. They use naive bayes and ensemble feature. They find investigation utilized archive text that was gotten to from Twitter about the Indonesian film survey. The after effects of the investigation indicated that the f-measure estimation of the gadget utilizing Ensemble Features is 0.88. Meanwhile,, Bag of Words highlights with a 0.94 f-measure esteem have better efficiency. There was found that the Bag of Word feature(s) with 0.96 accuracy, 0.92 recall, and 0.94 f-measure value have best performance among these individual feature(s).This combination of features has slightly with 0.91 accuracy, 0.86 recall, and 0.88 f-measure value, the lower precision.

Ching-yu Huang et al (13) worked on Twitter data referring to tweets relating to gifts, fundraising or charities. They perform sentiment analysis strategies and perform methods in exploratory data analysis to capture the polarity of people's feelings about donating for any cause. They find a tweet has a neutral, positive or negative polarity by using the Natural Language Processing Toolkit (NLTK). They collect training data and use it to prospective customers as a potential target is tweets linked to donations. We applied sentiment analysis techniques. In the type of polarity, they discovered the emotions of people. The main aim to find out more about the users. As a potential guide, corporations should research prospective donors for non-profit organizations.

Dr. K. Maheswari et al (14) worked lots of data is generated by social networks. sent value is appreciable than sentiment. The experimental results show that the classes with numerical classification produce more accurate results than text classes. According to the experiments, the accuracy of the sentiment is 36% because this attribute is text-based. Whereas, the KNN algorithm classifies into 13 classes. The feature selection algorithm ranked sent value as the first attribute for classification. The sentiment is the second-ranked attribute, therefore it produces 36%. The tuning is again needed for such work.

IV. CONCLUSION

Twitter is one of the most used social media networks in which the officials use their account to make people aware of their deeds. According to the research there are 321 million active users in the world but twitter is mainly used by the officials. We are presenting a novel hybrid framework for the classification of twitter dataset which we got from the twitter API it was approximately 18968 tweets using Isomap and machine learning techniques. The objective of thi research is to choose the algorithms and metrics for comparing the overall performance of Machine Learning Classifiers and to examine the metrics acquired from special system getting to know algorithms relying on the dimensions of datasets. First Experiment was done using the Decision Tree, Naïve Bayes and SVM in which we got the low accuracy but after applying the Isomap technique we acquired much more accuracy then before. Our data was in binary form e.g happy or cheery as "1", sad or hated as "0". The tools we used were ANACODA using python language. In future we want to try more machine learning techniques on large dataset. We also want to work on multi classes on different dataset of twitter using different machine learning techniques.

REFERENCES

[1] Nikil Prakash T, Aloysius Amalanathan. P., 2019. DATA PREPROCESSING IN SENTIMENT ANALYSIS USING TWITTER DATA. INTERNATIONAL EDUCATIONAL APPLIED RESEARCH JOURNAL (IEARJ) Volume 03, Issue 07, July 2019 E-ISSN: 2456-6713.

[2] Sanjeev Dhawan, Kulvinder Singh, Priyanka Chauhan. P., 2019. Sentiment Analysis of Twitter Data in Online Social Network. 5th IEEE International Conference on Signal Processing, Computing and Control (ISPCC 2k19), Oct 10-12, 2019, JUIT, Solan, India

[3] Sahar A. El_Rahman, Feddah Alhumaidi AlOtaibi, Wejdan Abdullah AlShehri. P.,2019. Sentiment Analysis of Twitter Data.

[4] Shikha Tiwari, Anshika Verma, Peeyush Garg, Deepika Bansal. P., 2020. Social Media Sentiment Analysis On Twitter Datasets.2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS).

[5] Tarun Anand, Vikrant Singh, Bharat Bali, Biswa Mohan Sahoo, Basu Dev Shivhare, Amar Deep Gupta. P., 2020. Survey Paper: Sentiment Analysis for Major Government Decisions. 2020 International Conference on Intelligent Engineering and Management (ICIEM)

[6] Jessica, Raymond Sunardi Oetama. P., 2019. Sentiment Analysis on Official News Accounts of Twitter Media in Predicting Facebook Stock. 2019 5th International Conference on New Media Studies Bali, Indonesia | October 09-11, 2019

[7] Shamantha Rai B, Sweekriti M Shetty. P., 2019. Sentiment Analysis Using Machine Learning Classifiers: Evaluation of Performance. 2019 IEEE 4th International Conference on Computer and Communication Systems

[8] Malika Acharya, Shilpi Sharma. P., 2018. Semantic Analysis of Twitter Posts. 978-1-5386-7709-4/18/$31.00 c 2018 IEEE

[9] Prakruthi V , Sindhu D, Dr S Anupama Kumar. P., 2018. Real Time Sentiment Analysis Of Twitter Posts. 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018.

[10] Sanjeev Dhawan, Kulvinder Singh, Priyanka Chauhan. P., 2019. Sentiment Analysis of Twitter Data in Online Social Network. 5th IEEE International Conference on Signal Processing, Computing and Control (ISPCC 2k19), Oct 10-12, 2019, JUIT, Solan, India.

[11] Yili Wang, KyungTae Kim, ByungJun Lee and Hee Yong Youn. P., 2018. Word clustering based on POS feature for efficient twitter sentiment analysis. Wang *et al. Hum. Cent. Comput. Inf. Sci. (2018) 8:17* https://doi.org/10.1186/s13673-018-0140-y.

[12] Park, C. W., & Seo, D. R. (2018). *Sentiment analysis of Twitter corpus related to artificial intelligence assistants. 2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*

[13] Amrita Shelar and Ching-yu Huang. P., 2018. Sentiment Analysis of Twitter Data. 2018 International Conference on Computational Science and Computational Intelligence (CSCI).

[14] Dr. K. Maheswari. P., 2018. Improving Accuracy of Sentiment Classification Analysis in twitter Data Set Using knn.E ISSN 2348 –1269, PRINT ISSN 2349-5138

[15] Amit Agarwal, Durga Toshniwal. P., 2018. Application of Lexicon Based Approach in Sentiment Analysis for short Tweets. 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE-2018) Paris, France 22-23 June 2018

[16] .Davy, M., Desobry, F., Gretton, A., & Doncarli, C. (2006). *An online support vector machine for abnormal events detection. Signal Processing, 86(8),20092025.* doi:10.1016/j.sigpro.2005.09.027

[17] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). *An introduction to decision tree modeling. Journal of Chemometrics,18(6),275285.* doi:10.1002/cem.873

[18] L. Jiang, D. Wang, H. Zhang, Z. Cai and B. Huang, "Using Instance Cloning to Improve Naive Bayes for Ranking", *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 22, no. 6, pp. 1121-1140, 2008.s