International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-11, September 2019

Risk Monitoring and Quantitative Results of Various Attributes of Machine Learning Algorithms with a Time Series Data.

Rakhi Gupta, Nashrah, SD Joshi, Suhas patil

Abstract: ***The aim of this research is to do risk modelling after analysis of twitter posts based on certain sentiment analysis. In this research we analyze posts of several users or a particular user to check whether they can be cause of concern to the society or not. Every sentiment like happy, sad, anger and other emotions are going to provide scaling of severity in the conclusion of final table on which machine learning algorithm is applied. The data which is put under the machine learning algorithms are been monitored over a period of time and it is related to a particular topic in an area.***

Keywords: ***risk modelling, machine learning algorithms, emotions, sentiment analysis***

I. INTRODUCTION

Crime with time has evolved and now its presence can be seen on various digital platforms. Crime now is spreading easily and quickly through social media platforms, the most common platforms being twitter and Facebook . These platforms can be easily misused for various types of illegal type of activities as well.[1].This scaling will be of polarity negative, positive or neutral where less negative means less harmful and higher the value on scale means it's harmful. Positive words are given scaled accordingly Based on the overall sentiment value generated from final table on which machine learning algorithms applied for comparative study between classifier models, risk modelling and analytics graph is created, which would provide better visualization and help in making better judgments. The purpose of Crime Mitigation System is to find a possible suspicious activity and inform the concern authorities to help them evade a possible crime from happening. This is done by risk modelling and risk identification.[16] The system aims to analyze data from Twitter for finding criminal activity that might be going on. This would help in keeping track of people who might become threat to society or any individual and stop them before any chaos is created. [15]

Manuscript published on 30 September 2019.
***Correspondence Author(s)**
Rakhi Gupta, PHd, Department of Computer Enggineering from Bharati Vidyapeeth Deemed University Pune, Maharashtra, India.
Nashrah, Assistant Professor, Department of Information Technology from K.C. College Mumbai, Maharashtra, India.
Dr. Shashank Joshi, Professor in Computer Engineering Department Bharati Vidyapeeth Deemed University College of Engineering, Pune, Maharashtra, India.
Dr. Suhas Patil, Professor, Computer Engineering Department Bharati Vidyapeeth Deemed University College of Engineering, Pune, Maharashtra, India.

II. PROPOSED METHODOLOGY

1.1 Block Diagram :

The research work is carried to develop a way where huge amount of data is bought into desired format for later being analysed and subjected to work.

Level 1. The data is being retrieved from the twitter.

Level 2

The raw data is processed and bought into desired format.

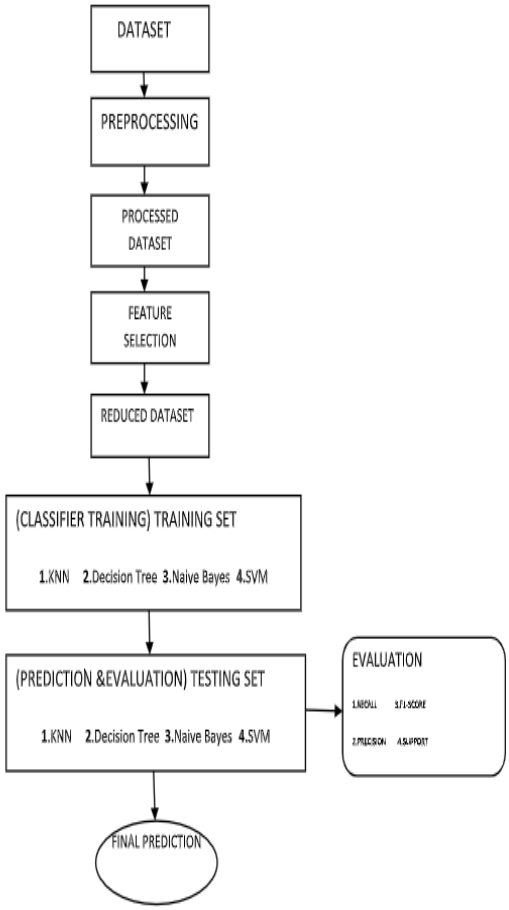


Fig 1. Block Diagram

3.2 Algorithms :

Textblob

Further the data is passed through the textblob to attain sentiment value of the tweet. Later the supervised data is been divided into test data and training data. The classifier model is trained with the help of training set which in return does the prediction. [20]

Classifiers:

A. Naive Bayes Classifier

Naive Bayes has been one of the easiest ways of using machine learning technologies for classifying tweets[4].
This classifier is based on simple theorem of probability for making a probabilistic model of data. The mechanics of NB algorithm are applied to numeric data. [10] .

B. Support Vector Machine :

A Support Vector Machine (SVM) is a classifying algorithm that charts data in nonlinear pattern which in result alters the training data into a higher dimension. This new method of classification completely revolves on the Type of Attributes and dimension of examples, it attempts to find a linear optimal separating hyper plane. [4]

C. K-Nearest Neighbor

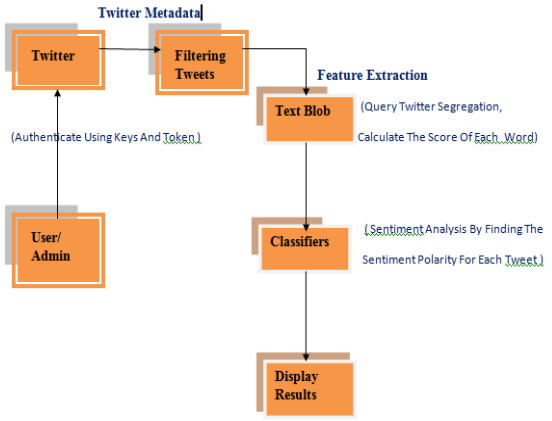
Application of KNN is for both classification and regression predictive problems.It is widely used in classification problems in the industry. Evaluation can be done through these 3 important aspects.

1. Ease to interpret output
2. Calculation time
3. Predictive Power

D. Decision Tree

A decision tree works like a decision support tool that incorporates a model of decision in a tree form and their consequences includes chance event outcomes, costs of resources and its utility.

3.3 Flow chart :



III. RESULTS ANALYSIS

TABLE 1. COMPARATIVE RESULTS

Data size (TWEETS)	K-NN	Decision Tree	Support Vector Machine	Naive Bayes
30	72.4	78.2	92.3	90.6
50	69.6	77.5	89.9	88.4
80	67.4	76.4	91.4	86.2
120	68.3	77.9	90.2	85.1
150	71.3	74.2	93.2	89.2

Description: The raw tweets are processed and classified into training data and test data .We train the classifier by providing it with the training data to prepare it for a quantitative analysis. The following points can further be elaborated based on the comparative results in Table 1.

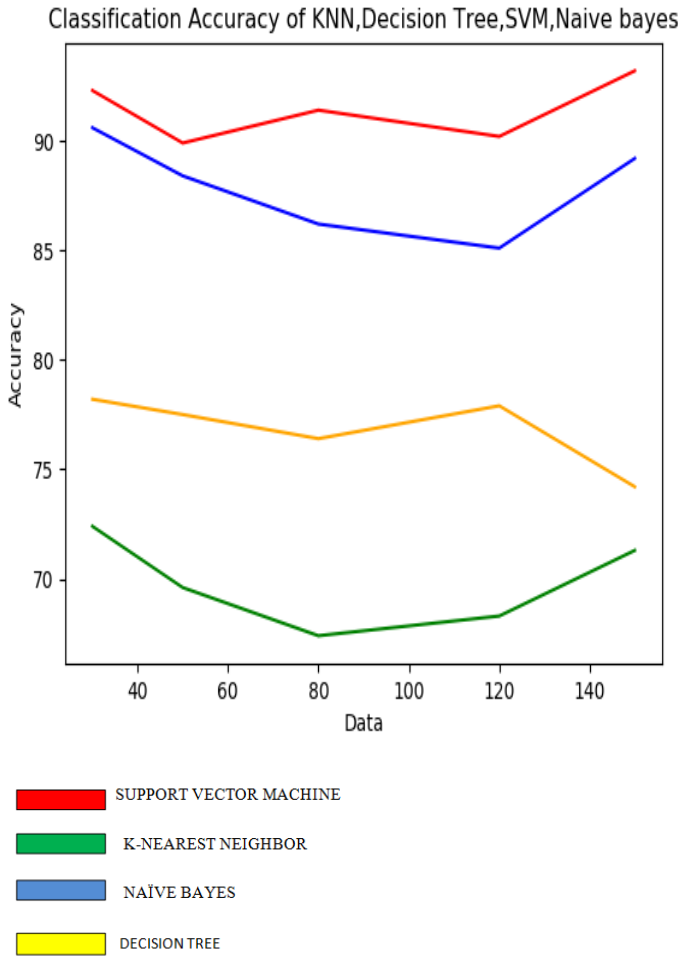


Fig 2. Statistical Comparison of Accuracy levels of various classifiers based on the data set

4.1 Tabular Representation of confusion matrix

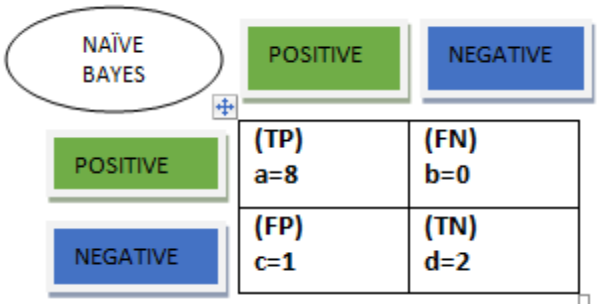


Fig 3. Confusion Matrix of Naive Bayes with Dataset Of 150 Tweets

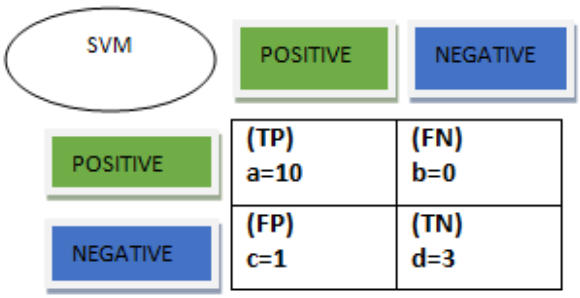


Fig 4. Confusion Matrix of Support Vector Machine with Dataset Of 150 Tweet

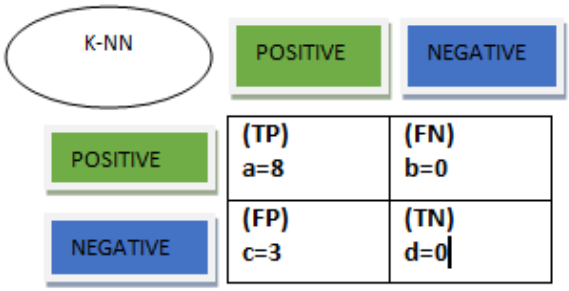


Fig 5. Confusion Matrix of K-Nearest Neighbor with Dataset of 150 Tweets

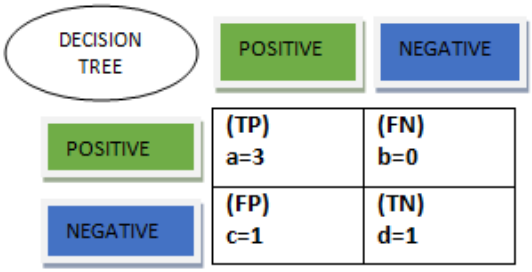


Fig 6. Confusion Matrix of Decision Tree with Dataset Of 150 Tweets.

4.2 Time Series Analysis

Time series data are a collection of ordered observations recorded at a specific time, for instance, hours, months, or years.

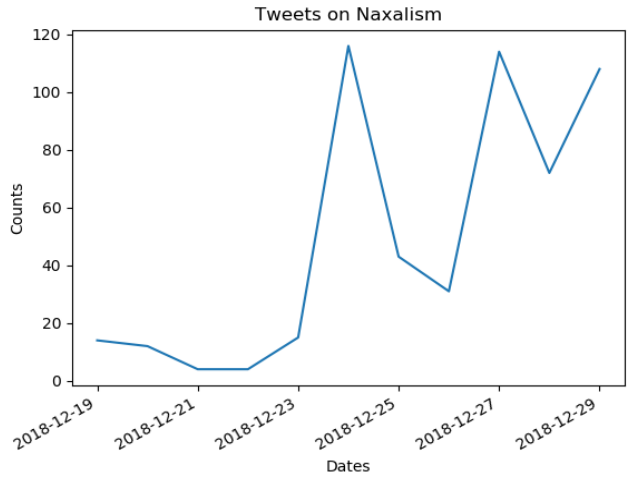


Fig 7. Time series analysis: This statistical graph is a collection data monitored hourly, daily and on a weekly basis. (District Gadchroli , Maharashtra)

4.3 Accuracy Comparison

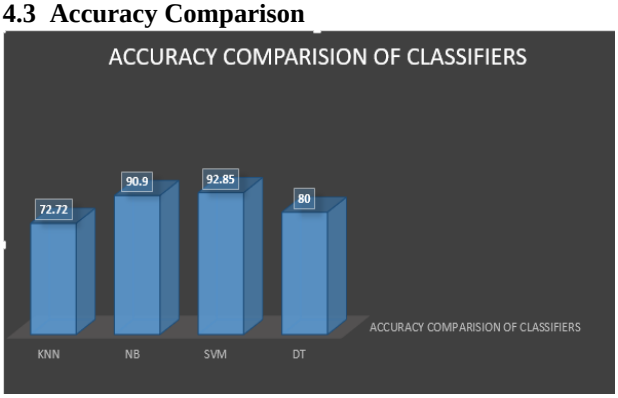


Fig 8. Accuracy Comparison of Classifiers.

Table 2. Classifiers With their Respective Accuracy Percentage

CLASSIFIERS	ACCURACY(%)
K-NEAREST NEIGHBOR	72.72
NAÏVE BAYES	90.90
SUPPORT VECTOR MACHINE	92.85
DECISION TREE	80

4.4 Quantitative Analysis of Individual Classifiers

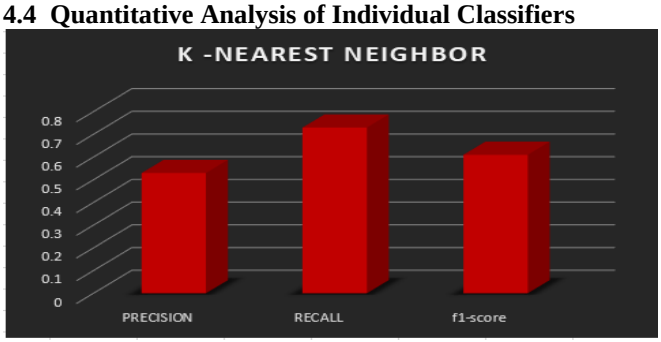


Fig. 9 : Graphical Representation of K-Nearest Neighbor: precision, recall, f1-score.

	K-NEAREST NEIGHBOR
PRECISION	0.53
RECALL	0.73
F1-SCORE	0.61
SUPPORT	11

Table 3 : K-Nearest Neighbor: precision, recall, f1-score and support values.

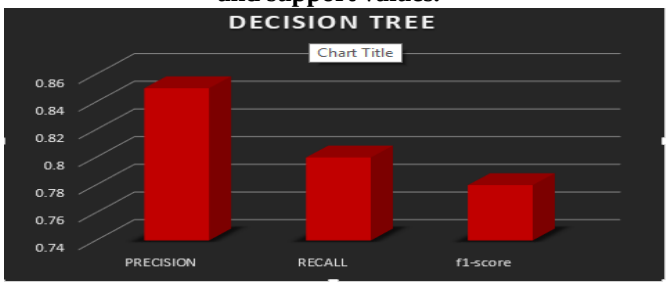


Fig. 10 : Graphical Representation of Decision Tree: precision, recall, f1-score

	DECISION TREE
PRECISION	0.85
RECALL	0.80
F1-SCORE	0.78
SUPPORT	5

Table 4 : Decision Tree: precision, recall, f1-score and support values.

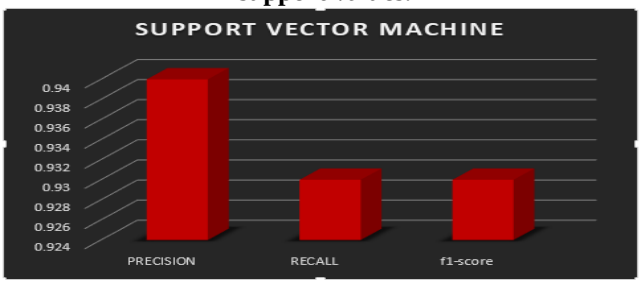


Fig. 11 : Graphical Representation of Support vector machine: precision, recall, f1-score.

	SUPPORT VECTOR MACHINE
PRECISION	0.94
RECALL	0.93
F1-SCORE	0.93
SUPPORT	14

Table 5 : Support Vector Machine: precision, recall, f1-score and support values.

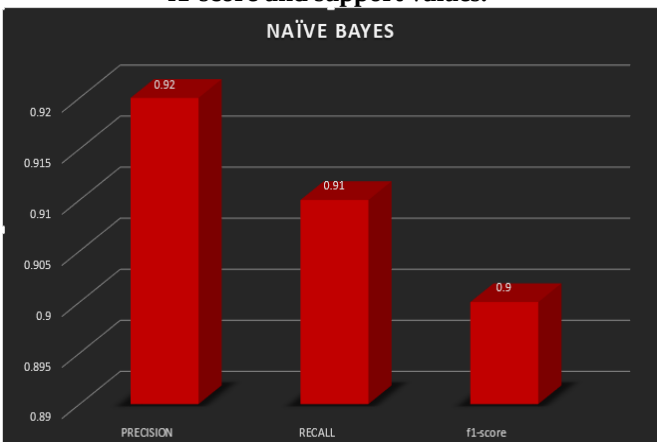


Fig. 12 : Graphical Representation of Naive Bayes: precision, recall, f1-score

	NAIVE BAYES
PRECISION	0.92
RECALL	0.91
F1-SCORE	0.90
SUPPORT	11

Table 6 : Naive Bayes : precision, recall, f1-score and support values.

4.5 Comparison Of Precision, Recall And F1-Score Of Classifiers

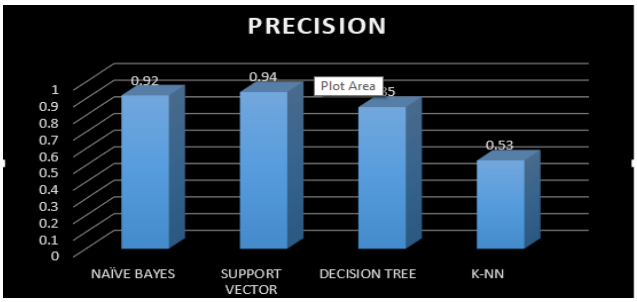


Fig.13 : Comparison of Precision of classifiers Naive Bayes, Support vector machine, Decision tree, and K-NN out of which Support vector machine gives the best results.

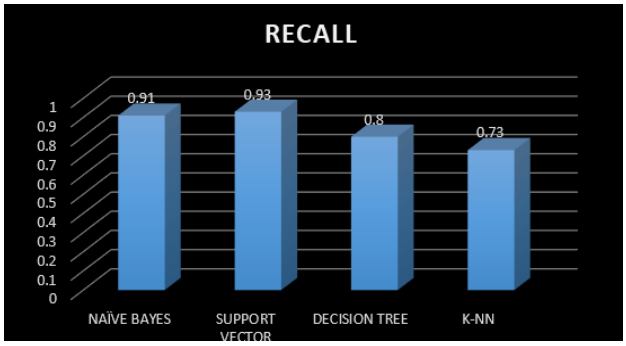


Fig. 14 : Comparison of Recall values of classifiers Naive Bayes, Support vector machine, Decision tree, and K-NN out of which Support vector machine gives the best results.

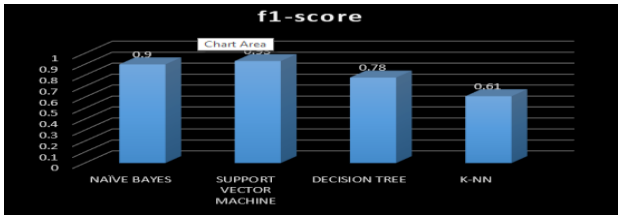


Fig. 15 : Comparison of f1-score of classifiers Naive Bayes, Support vector machine, Decision tree, and K-NN out of which Support vector machine gives the best results.

CLASSIFIERS	NAIVE BAYES	SVM	K-NN	DECISION TREE
PRECISION	0.92	0.94	0.53	0.85
RECALL	0.91	0.93	0.73	0.80
F1-SCORE	0.90	0.93	0.61	0.78
SUPPORT	11	14	11	5

Table 7: Comparison of Precision, Recall, F1-score and Support values of Naive Bayes, K-NN, Support Vector Machine and Decision Tree.

4.6 Comparison of Precision , Recall and f1-score of classifiers

CLASSIFIERS	NAIVE BAYES	SVM	K-NN	DECISION TREE
PRECISION	0.92	0.94	0.53	0.85
RECALL	0.91	0.93	0.73	0.80
F1-SCORE	0.90	0.93	0.61	0.78
SUPPORT	11	14	11	5

Table 8 .Comparison Of Precision,Recall,F1-score and support values of Naive Bayes, K-NN, Support Vector Machine And Decision Tree.

IV. CONCLUSION

This research does a comparative performance analysis on the data fetched from the social platform twitter with the main focus towards crime mitigation and risk analysis.

Performing a comparative study which is done between Naive Bayes, Decision Tree, K-Nearest Neighbor and Support Vector Machine on basis of quantitative parameters for crime mitigation ,it is realized that the Support Vector Machine gives better performance than all the other classifiers in terms of consistency. However this is applicable to the domain of data which had been monitored hourly, daily and weekly with respect to Naxalisim (District Gadchiroli, Maharashtra)

REFERENCES

1. Chirag Kansara , Rakhi Gupta , Dr. S.D. Joshi , Dr. S.H. Patil, “Crime Mitigation At Twitter Using Big Data Analytics and Risk Modelling “ , IEEE International conference on Recent Advances and innovations in Engineering (ICRAIE -2016) December 23-25 , 2016 , Jaipur , India.
2. Anuja P Jain , Padma Dandannavar , “Application of Machine Learning Techniques to Sentiment Analysis “ for 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT) , 978-1-5090-2399-8/16/$31.00_c 2016 IEEE

3. Geetika Gautam , Divakar yadav , "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis " for 978-1-4799-5173-4/14/$31.00 ©2014 IEEE
4. Mr. S. M. Vohra, 2 Prof. J. B. Teraiya," A Comparative Study Of Sentiment Analysis Techniques", Journal Of Information, Knowledge And Research In Computer Engineering Issn: 0975 – 6760| Nov 12 To Oct 13 | Volume – 02, Issue – 02 Pg 313-317
5. DATA 1 , S Shayaa, NI Jaafar ,S Bahri , AS, PSW, YWC, AZP and MAA , " Sentiment Analysis of Big Data : Methods , Applications , and Open challenges,Vol 6 2018, DOI : 10.1109/ACCESS.2018.2851311, 2169-3536 , 2018 IEEE
6. Journal of Convergence Information Technology ,Volume 4, Number 3, September 2009.