



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sentimental Analysis on Instagram Using Machine Learning Techniques

Kinjarapu Rajasekhar

Student, Rajam, Vizianagaram, 532127, India.

ABSTRACT

Now a days, Human beings from all over the globe use social media websites to share the information. Instagram for example is a platform in which users send, read posts known as comments and the Interaction with exclusive communities. Users share each daily lives, post their opinions on whole things with brand and place. Companies can advantage from this big platform by collecting data related Opinions on them. Social media like twitter where user can post their comments in 280 characters, because Of limited number of characters in tweets it becomes easy for a sentimental analysis. On Instagram around 550 million of comments are posted daily. Sentimental analysis plays a big role here. Sentiment is procedure To identify Opinion in the text. It is also known as opinion mining. Sentimental analysis takes unstructured Textual Comments about product reviews, an event, etc., feedback posted by different customers into Distinctive Classes as either positive or negative or neutral opinion. This is also known as polar Classification. Sentimental analysis can be performed by text analysis and computational linguistics . The work aims to compare the performance of machine learning algorithms in performing sentimental Analysis on Instagram data. There are various machine learning algorithms available.. In this Paper, we will compare three kinds of machine learning algorithms, they are Naïve bayes, logistic Regression, support vector machine (SVM). we will see how sentimental analysis is done by this Classification Algorithms and what is the accuracy and precession in these cases.

Keywords: sentimental analysis , machine learning, naïve bayes, SVM, logistic regression, Instagram.

1. Introduction

Sentiment analysis is basically a process which automatically extract opinions, emotions and sentiment from text. people post their comments in social media an Event. The majority of other people had opinion on same event either it may be positive , neagative or neutral. analyzing sentiment is a process to know the users emotion of a particular item. Sentiment analysis can be done in three levels and they are sentence, aspect and document level. Instagram is a source of information to know the quality of product. Instagram uses comments which is in sentence form to denote opinions. In this paper sentiment analysis can be done Instagram data which the comment either be positive , negative or neutral. The in Instagram can be understood by users but not understood by computer.so in this paper, we use machine learning algorithms which understood by computer. They are SVM, Naïve bayes, logistic regression and this work compares the performance of SVM, Naive bayes, logistic regression in the sentence of classification.

2. Literature Survey

In paper [1] A. Poornima and K. S. Priya, In this paper, her goal is to calculate the sentiment accuracy of sentences that were extracted from the text of tweets either the tweet is positive or negative or neutral. She compared accuracy of three machine learning algorithms. the accuracy for Logistic regression is approximately 86.23%, Support vector machine is 85.69% and Multinomial naïve bayes is 83.54%. Then the Logistic regression outperforms the SVM and Multinomial naïve bayes with an accuracy of 86.23.

In paper [2] S. Zahoor and R. Rohilla, This is a case study involving a number of events that have occurred in the year 2019 like the Haryana Assembly Polls; a movie release The Sky is Pink; UNGA conference and a gathering Howdy Modi that occurred in Houston. In this paper he have used algorithms like NaïveBayes, SVM, Random Forest Classifier and Long Short Term Memory Networks (LSTMs).Naive bayes has more accuracy.

In paper [3] S. A. El Rahman, F. A. Al Otaibi and W. A. AlShehri, In this paper, he used sentiment analysis to classify specific English tweets about two restaurants, KFC and McDonald's. his research was determining which one better than other, in specific we examined weather specific tweets is positive, negative, neutral. In this paper, he extracted tweets from Twitter using R language. For statistical computing and machine learning methods, R is a programming language.

In paper [4] L. Mandloi and R. Patel, In this paper, He performed sentimental analysis on Twitter using machine learning techniques. In this paper, he use different machine learning methods to analyze the sentiments of the people. Here he used machine learning methods like Naïve Bayes Classifier, Support Vector Machine method and Maximum Entropy method. naive bayes had more accuracy.

In paper [5] M. I. Sajib, S. Mahmud Shargo and M. A. Hossain, In this paper, sentiment analysis using machine learning had been proposed and all the analysis is being carried out for English language only. In this paper clearly shows that the accuracy is better with logistic regression on Tf-Idf vectorizer with stop words on Trigrams with a 82.59 percentage.

3. Data Collection

The references used to compile the data below.

4. Methodology

In this paper, the problem mainly depends on classification modelling. So we will use machine learning predictive models to find sentimental analysis on Instagram. In this paper we need to classify either the comments are negative or positive or neutral. We employ supervised machine learning algorithms in this paper.

1. 1.Logistic regression
2. 2.Naive bayes
3. 3.Support vector machine

LOGISTIC REGRESSION:

Logistic Regression is a predictive analysis algorithm used for the classification problems. Logistic A more intricate cost function known as the "Sigmoid function" is used in logistic regression.. The cost function is usually restricted to the range of 0 and 1 according to the logistic regression hypothesis. LogisticRegression() is a function used to predict the accuracy of the sentiment classification . The Sigmoid function in a Logistic Regression Model is formulated as $1 / (1 + e^{\{-value\}})$ $1/(1+e^{-value})$. A logistic/sigmoid function in the shape of a "S" will be produced using the Logistic Regression Machine Learning. The task of forecasting values between 0 and 1 falls on this function.

Naive bayes :

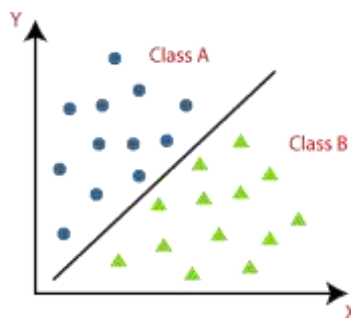
This is one of the supervised algorithm that is based on the probabilistic approach to classify the text to a particular class i.e. positive or negative . This algorithm calculates the probability of all the words in the dataset and then classifies the comments or text into particular categories. This algorithm is based on the Bayes rule. Where

$A, B = \text{events}$

$P(A/B) = \text{probability of A given B is True}$

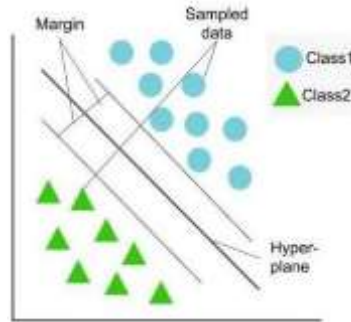
$P(B/A) = \text{probability of B given A is True}$

$P(A), p(B) = \text{Independent probabilities of A and B}$



Support vector machine :

SVM is a supervised machine learning technique that can be applied to classification and regression issues. Regression and classification are two types of supervised machine learning.. Classification is something that can be defined as predicting a label whereas Regression is about predicting quantity. So the main task of the Support Vector Machine classifier is to perform classification. That is it classifies the data in different classes by drawing a hyperplane them, which differentiate between different classes which we plot in n-dimensional space. The hyperplane that is drawn by the SVM is drawn with the help of a mathematical function called Kernels. The data point which is closest to the hyperplane is called the support vector and the method is called the support vector machine.



5. Results and Discussion

A. Formula for Calculation of Accuracy and Precision So, based on our study we have divided the comments from the users on the basis of the sentiments that is positive, negative or neutral. Here I created a contingency matrix for reference to our below tabular data

TABLE 1. CONTINGENCY MATRIX

Predicted value	Actual value			
		Positive comments	Negative comments	Neutral comments
	Positive comments	True Negative	False Positive	False positive
	Negative comments	False Negative	True negative	False negative
	Neutral comments	False neutral	False neutral	True neutral

Here, True Positive are the positive comments which are actually classified positive. Where else, False Positive are the comments which are positive but, classified as negative or neutral. Similarly True Negative are the negative comments which are classified as negative but, classified as positive or negative. where else False Negative are the Negative comments which are classified as positive or neutral comments. True Negative are the comments which are actually found neutral and classified as neutral where else False Negative are the comments which are neutral comments but, classified as positive or negative.

1) Precision calculation

$$\text{Precision} = \text{True positive} / \text{True positive} + \text{False Positive}$$

2) Accuracy calculation

$$\text{Accuracy} = \text{Number of current predicted data} * 100 / \text{Total number of data}$$

$$\text{Accuracy} = (\text{True positive} + \text{True negative} + \text{True Neutral}) * 100 / \text{Total}$$

B. Accuracy and Precision Calculation in Naïve Bayes

Based on our study we found following data According to the contingency matrix table form above Table 1.

TABLE 2. PREDICTED VALUE FOR NAÏVE BAYES

Predicted value	Actual Value				
		Positive	Negative	Neutral	Total
	Positive	612	32	46	690
	Negative	13	123	17	153
	Neutral	14	18	125	157
	Total	639	173	188	1000

$$\text{Accuracy} = (612+123+125)*100 / 1000$$

$$\text{Accuracy} = 86\%$$

$$\text{Precision} = 612 * 100 / 690$$

$$\text{Precision} = 88.69\%$$

C. Accuracy and Precision is Support Vector Machine Based on our study we found following data According to the contingency matrix table form above Table 1.

TABLE 3. PREDICTED VALUE FOR SUPPORT VECTOR MACHINE

Predicted value	Actual value				
		Positive	Negative	Neutral	Total
	Positive	516	41	123	680
	Negative	18	114	29	161
	Neutral	21	22	116	159
	Total	555	177	268	1000

Bases on the Above Table 3 we calculate Accuracy and Precision for Support Vector Machine

$$\text{Accuracy} = (516+114+116) * 100 / 1000$$

$$\text{Accuracy}=74.6\%$$

$$\text{Precision}=516*100/680$$

$$\text{Precision}=75.882\%$$

D. Accuracy and Precision in Logistic Regression Method Based on our study we found following data According to the contingency matrix table form above Table 1.

TABLE 4. PREDICTED VALUE FOR LOGISTIC REGRESSION METHOD

Predicted value	Actual value				
		Positive	Negative	Neutral	Total
	Positive	564	37	70	671
	Negative	17	108	16	141
	Neutral	18	16	154	188
	Total	599	161	240	1000

Bases on the Above Table 4 we calculate Accuracy and Precision for Logistic Regression.

$$\text{Accuracy} = (564+108+154) * 100 / 1000$$

$$\text{Accuracy}=82.6\%$$

$$\text{Precision}=564*100/671$$

$$\text{Precision}=84.05\%$$

E. Comparison of Accuracy and Precision in Machine Learning Methods

TABLE 5. COMPARISON OF ACCURACY AND PRECISION OF DIFFERENT MACHINELEARNING METHODS

Name of the method	Accuracy	Precision
Naïve Bayes	86%	88.69%
Support vector machine	74.6%	75.882%
Logistic Regression	82.6%	84.05%

6. Conclusion

The different machine learning technique of data analysis of Instagram are discussed like Naïve Bayes, SVM and Logistic Regression Method. The analysis of Instagram data is being done in various aspects to mine the sentiments. This study defines the concept of opinion in sentiment analysis of Instagram. Sentiment analysis deals with opinion classified into positive, negative and neutral. The study shows that the machine learning method such as Naïve Bayes has the highest accuracy and can be consider as the baseline learning methods as well as in some cases Logistic Regression method is very effective.

REFERENCES

1. A. Poornima and K. S. Priya, "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 493-496, DOI: 10.1109/ICACCS48705.2020.9074312.
2. S. Zahoor and R. Rohilla, "Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study," 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM), 2020, pp. 194-199, DOI: 10.1109/ICACCM50413.2020.9213011
3. S. A. El Rahman, F. A. Al Otaibi and W. A. AlShehri, "Sentiment Analysis of Twitter Data," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp. 1-4, DOI:10.1109/ICCISci.2019.8716464

-
4. L. Mandloi and R. Patel, "Twitter Sentiments Analysis Using Machine Learning Methods," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154183.
 5. M. I. Sajib, S. Mahmud Shargo and M. A. Hossain, "Comparison of the efficiency of Machine Learning algorithms on Twitter Sentiment Analysis of Pathao," 2019 22nd International Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6, DOI: 10.1109/ICCIT48885.2019.9038208