

Analysis of student opinions on online classes during lockdown

Dr M Sree Vani, Associate Professor, MGIT, Hyderabad, India, sreevani@mgit.ac.in

Abstract The educational system across the world has immensely been affected due to outbreak of COVID-19. This is forced to shut down of educational institutions, which adversely affected student fraternity across the globe. In the absence of traditional classroom teaching and one-to-one interaction, computer-based learning has emerged as closest substitute for off-line teaching. Although some people could adapt easily, some of them found it inefficient. This analysis reveals students' positive perception towards e-learning and thus acceptance of this new learning system. This work explore the tweets related to distance learning to understand people's opinions (a.k.a opinion mining) and to discover facts and For the present study, lexicon-based approach has been used to determine the tweets' polarities. We will also build a machine learning model to predict the positivity and the negativity of the tweets by using Naive Bayes Classifier.

Keywords — Lockdown, Data mining, Online classes, Tweets data, Student opinions, Naive Bayes Classifier

I. INTRODUCTION

Online learning is still new, and it is continuously evolving by the rapidly changing technologies. More efforts are required by all online educators to integrate pedagogy with technologies in order to enhance student learning. Factors such as age, gender, prior knowledge of computer literacy and learning styles of individual are the vital predictors of technology acceptance by students.

As a matter of fact, as this e-learning wave is a recent development, teachers as well as students are in the process of adapting this new teaching and learning methodology. At the same time, when considering the economic feasibility in different countries in the world, practically some of the distance learning approaches are not fruitful. For example, continues technical issues because of not having proper infrastructure and internet bandwidth as required, the difficulty of evaluation and assessment, the limited facility for collaborative learning and delivery of course content may not productive as productive in physical lecture hall [1]. At this prevailing situation of virtual teaching and setting of new normal of teaching-learning methodology, it becomes all more important to get to know the opinion of learners and to explore learners' inclination towards this novel teaching methodology, such as their degree of adaptation and amendment if any they would like to suggest for the same or want to reject it altogether [2]. This study mainly focuses on students' opinion about the ongoing online classes, and whether its being effectively utilised. Main conduction of online classes during this quarantine is to not waste time but also be effective in an educational aspect. Advantageous sources for the classes are technological aids like gadgets and different apps, which has been improving and emerging on new advanced platforms, working hands in concentration and paying

attention while working. The people across India may posted their opinions on demonetization in Social media. In this paper, Twitter data will be consider and build models for classifying tweets into positive, negative and neutral sentiment [3]. Then analyze the twitter data to find the effect of demonetization on the people [6][5]. One advantage of this data is that the tweets are collected in a streaming fashion and therefore represent a true sample of actual tweets in terms of language use and content. The sentiment of the Twitter tweets can be analyzed using naive Bayes algorithm and Bayes theorem [9]. Every tweet is taken and each word is matched with the database of positive and negative words [8]. Section 2 discuss about Related work and Section 3 describes sentiment classification of demonetization data and section 4 discuss about sentiment analysis of demonetization data followed by conclusion.

II. RELATED WORK

In the recent time, the research on exploring students' perception and their expectation from e-learning has been on surge [8,9]. "National Center for Education Statistics" has reported the growing requirement of e-learning due to its increasing acceptance recently. Several studies indicate that most of the students enrolled in online courses are satisfied with the mode of learning. However, studies also reveal that perceptions of learners are affected by a host of factors [10]. The paper [10] discusses the method for collecting corpus tweet data and segregates as positive and negative sentiment. This is similar to [8] by using emotions for positive and negative tweets. Classification of tweets with polarity method having the SentiStrength scales from 1 to 5 for both positive (+1 weak positive to +5 extreme positive) and negative (-1 weak negative to -5 extreme negative) sentiments.

• **Lemmatization** - grouping together the inflected forms of a word so they can be analysed as a single item.

Figure 4 Shows sample data frame after applying the above pre-processing techniques.

```
>>> from textblob import Word
>>> w = Word("octopi")
>>> w.lemmatize()
'octopus'
>>> w = Word("went")
>>> w.lemmatize("v") # Pass in WordNet part of
speech (verb)
'go'
```

Index	Text
0	[innovate, innovative, approach, quoteoftheday, digitalmarketing, digitalearning, blogger]
1	[pandemic, raising, concerns, leens, use, technology, still, lot, know, read, continuingeducation, digitalearning, education, edutech, highereducation, learning, lifelonglearning, media, onlinelearning]
2	[sti, staying, new, normal, inquirer, technology, read, continuingeducation, digitalearning, education, edutech, highereducation, learning, lifelonglearning, media, onlinelearning]
3	[digital, learning, digital, rcrt, registration, opening, soon, website, digitalearning, easylearning, digita/rcrt, rsvk, sssy, onlineeducation, effortlesslearning, onlinecourse]
4	[upswing, classroom, virtual, school, classroom, edtechreview, education, edtech, educators, students, edchat, learning, teachers, classroom, digitalearning, teachers]
5	[digital, future, needs, workforce, conversant, digital, medium, report, digitalearning, become, need, hour, expedited, process, use, digita/transformation, talentsolutions, talent, dubai]
6	[educational, technology, share, price, shareholders, holding, read, continuingeducation, digitalearning, education, edutech, highereducation, learning, lifelonglearning, media, onlinelearning]
7	[anaheim, public, schools, launch, virtual, academies, long, haul, read, continuingeducation, digitalearning, education, edutech, highereducation, learning, lifelonglearning, media, onlinelearning]
8	[new, oriental, education, amp, technology, group, edu, earnings, call, transcript, read, continuingeducation, digitalearning, education, edutech, highereducation, learning, lifelonglearning, media, onlinelearning]
9	[instructional, considerations, school, year, blendedlearning, remotelarning, distancelearning, hybridlearning, teachingonline, elearning, remoteteaching, digitalearning, teacher/twitter, edutwitter, teachertips, teachhub, teachfromhome]
10	[uncertainty, september, bring, created, donor, choose, project, help, bring, fun, creative, digital, lessons, remote, teaching, helpateacher, respectforteachers, remoteteaching, digitalearning]
11	[check, wakelet, digital, learning, tools, via, digitalearning, wakelet]
12	[happy, friendship, day, rdnums, nagaland, kohima, friendshipday, friendshipweek, kohimaschoolsandscolleges, nagalandschoolsandscolleges, nagalandfightscovid, onlinelearning, digitalearning, ntsee]
13	[beat, summer, heat, cool, games, activities, abcya, abcya, digitalearning]
14	[startup, trilogy, mentors, raises, million, online, tutoring, platform, read, continuingeducation, digitalearning, education, edutech, highereducation, learning, lifelonglearning, media, onlinelearning]

Figure 4: Sample Tweets data set after pre-processing

B. AFFIN Dictionary

The AFINN is a dictionary which consists of 2500 words which are rated from +5 to -5 depending on their meaning. The twitter words are analyzed using AFINN dictionary and rate the twitter words as per their meaning from +5 to -5. Initially load the dictionary into pig by using the below statement:

```
dictionary=load '/AFINN.txt' using PigStorage('/t')
As(word:char array,rating:int);
```

For the second stage, sentiments were extracted from tweets using TfidfVectorizer module from Affin dictionary. After assigning individual scores to all the words, polarity and subjectivity were calculated using TextBlob. The sentiment property returns a named tuple of the form Sentiment(polarity, subjectivity). The polarity score is a float within the range [-1.0, 1.0]. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.

```
>>> testimonial = TextBlob("Textblob is amazingly simple
to use. What great fun!")
>>> testimonial.sentiment
```

```
Sentiment(polarity=0.39166666666666666,
subjectivity=0.4357142857142857)
>>> testimonial.sentiment.polarity
0.39166666666666666
```

Then classified the sentiments into three classes namely positive, negative and neutral based on their polarities as shown in Table 1. As the final stage, statistical data analysis was performed to find whether there is a relationship between user experience on distance learning in different countries and the economic stability of those countries.

Table 1:Polarity Classification

Polarity Value	Sentiment
>0	Positive
0	Neutral
<0	Negative

C. Naïve Bayes Classifier

Naive Bayes Classifier is used to predict the particular class of given words. It is used because of its easiness in both during training and classifying steps. To train the input set using Naïve Bayes classifier Pre-processed data is supplied as input data. After training phase, test data is given as input to classify sentiment words. As shown in Equation 1, the Bayes theorem shows good performance and as follows.

$$P(C|m) = P(C) \prod_{i=1}^n P(f_i|C) \tag{Equation 1}$$

Where C is the class positive, negative or objective sets, m is the twitter message and f is a feature. In our experiments the features are POS tags, unigrams or bigrams.

IV. RESULTS AND DISCUSSION

We have labeled the tweets according to their polarity scores. Let's build a Machine Learning model by using a Multinomial Naive Bayes Classifier. We will use our tf-idf vectors as the features and the labels as the target as shown in Table 2.

Table 2: Label-enc

Label-enc	
0	2
1	1
2	2
3	1
4	1

Now, we need to split our data into train and test sets. We will use the stratify parameter of train_test_split since our data is unbalanced. The word scores of the features are tested based on Chi-square method. Frequency distribution of positive sentiment and negative sentiment words is generated. Finally, the number of positive and negative

words as well as the total number of words and the dictionary of word score based on Chi-Square test is found.

The performance of Naïve Bayes classifier will be evaluated using Accuracy. Number of correctly selected positive and negative words divide by total number of words will produce accuracy as given in the Equation 2.

$$Accuracy = \frac{\sum True\ Positive + \sum True\ Negative}{\sum Total\ number\ of\ words} \quad \text{Equation 2}$$

Table 3: Experimental results

	TP	TN
Positive	53365	11964
Negative	17394	1737
Neutral	22194	1314

Table 3 shows experimental results. Where TP is number of tweets identified as positive and TN is number of tweets identified as negative respectively with Accuracy of Naïve Bayes classifier accuracy is 0.83 as shown in Table 4. Figure 5 describes analysis of tweets data.

Table 4: Performance of Model

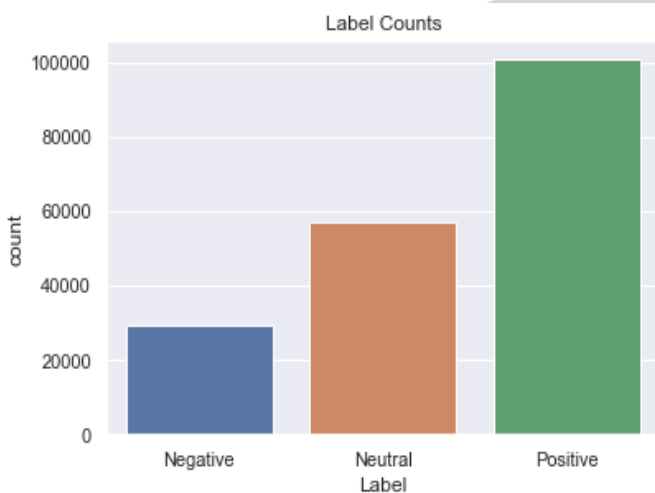


Figure 5: Analysis of Tweets Data

V. CONCLUSION

The present study analyses the perception of university students towards e-learning during the lockdown. we have collected the Tweets about distance learning by using Twitter API and the *tweepy* library. After that we applied common preprocessing steps on them such as tokenization, lemmatization, removing stopwords, and so forth. We explored the data by using summary statistics and visualization tools. After all, we used TextBlob to get polarity scores of the tweets and interpreted our findings. Consequently, we found that in our dataset most of the tweets have positive opinions about distance learning. The study reveals the preferences of students for e-learning as it provides them much freedom to connect with their teachers, fellow students and engage with their study materials at the comfort and flexibility of space and time. The easy access of study resources is found to be one of the major reasons

for the students to opt for e-learning. The study indicates that e-learning technology enables easy information access leading to positive attitude formation of students towards it. As in future work, Sentiment Analysis need to incorporate negation handling and emphasis handling to improve its classification accuracy.

	Precision	Recall	F1-Score	support
0	0.63	0.66	0.65	5840
1	0.89	0.81	0.85	1141
2	0.86	0.89	0.88	2015
Accuracy			0.83	3741

REFERENCES

- [1] J. Jaskolski, F. Sieberg, T. Tibroni, P. Cimiano, and R. Klinger, "Opinion mining in online reviews about distance education programs," 07 2016.
- [2] P. Fidalgo, J. Thormann, O. Kulyk, and J. Lencastre, "Students' perceptions on distance education: A multinational study," *International Journal of Educational Technology in Higher Education*, vol. 17, pp. 1–18, 2020.
- [3] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, "Elearning: Challenges and research opportunities using machine learning data analytics," *IEEE Access*, vol. 6, pp. 39 117–39 138, 2018
- [4] (Tutorial) *simplifying sentiment analysis in Python*. (n.d.). DataCamp Community. <https://www.datacamp.com/community/tutorials/simplifying-sentiment-analysis-python>
- [5] Lee, J. (2020, May 19). *Twitter sentiment analysis | NLP / Text analytics*. Medium. <https://towardsdatascience.com/twitter-sentiment-analysis-nlp-text-analytics-b7b296d71fce>
- [6] Li, C. (2019, September 20). *Real-time Twitter sentiment analysis for brand improvement and topic tracking (Chapter 1/3)*. Medium. <https://towardsdatascience.com/real-time-twitter-sentiment-analysis-for-brand-improvement-and-topic-tracking-chapter-1-3-e02f7652d8ff>
- [7] Randerson112358. (2020, July 18). *How to do sentiment analysis on a Twitter account in Python*. Medium. <https://medium.com/better-programming/twitter-sentiment-analysis-15d8892c0082>
- [8] *Stemming and Lemmatization in Python*. (n.d.). DataCamp Community. <https://www.datacamp.com/community/tutorials/stemming-lemmatization-python>
- [9] J. Williams, "Narrow-band analyzer (Thesis or Dissertation style)," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993.
- [10] T. H. Foundation, 2020 Index of Economic Freedom, 2020 (Accessed on: October 11, 2020). [Online]. Available: <https://www.heritage.org/index/country/southafrica?version=986>