The impact of the Covid-19 pandemic on education: Natural Language Processing

Harun YAZGAN
Computer Science Engineerring
İstanbul Kültür University
İstanbul, Turkey
1800002488@stu.iku.edu.tr

Onur ÖZBEK
Electrical and Electronics Engineering
Industrial Engineering
İstanbul Kültür University
İstanbul, Turkey
1800003894@stu.iku.edu.tr

Ahmet Can GÜNAY
Electrical and Electronics Engineering
Industrial Engineering
İstanbul Kültür University
İstanbul, Turkey
1800002666@stu.iku.edu.tr

***Abstract*—Because of the circumstances of our day, social media plays a significant part in our lives. Twitter is currently the most popular social media tool for posting relevant material. Users may post their thoughts and ideas on Twitter, which generates massive amounts of data. Examine your own thoughts, feelings, and experiences, as well as the viewpoints of others. The Coronavirus-2019 (Covid-19) is a disease carried by minute droplets transferred by close contact. Coughing, sneezing, and speaking have caused social and economic scars all over the world.More than 185 million individuals have been diagnosed with the New Coronavirus (Covid-19) as of July 7, 2021, and nearly 4 million people have perished as a result of this deadly sickness. Using the tweets that individuals communicate about the Covid-19 pandemic on the Twitter platform, this work focuses on the analysis of the feelings that Covid-19 leaves on people.**

Deep learning techniques are used in the analyses. Sentiment analysis may be quite beneficial. A network model based on Long-Short Term Memory (LSTM) was employed in this investigation. Our suggested technique may successfully execute sentiment analysis on the Twitter dataset (by using Twint), according to experimental results.

***Index Terms*—Deep Learning, Social Media, Sentiment Analysis, Covid-19, Education**

I. Introduction

The tweets on the relationship between Covid-19 and education will be included in our dataset. We used these tweets to generate our own dataset, which we used to investigate the impact of Covid-19 on schooling.

According to our definition, conditions in education, like all other fields, alter during a pandemic. Many governments have opted to continue schooling online in order to adapt to this new climate. For the past two years, and even now, certain schooling and even tests have been completed online. But what impact does this have on students? Was it better or worse for their learning? These impacts can vary from person to person, and various learning skills may react differently. The goal of this project is to use tweets to investigate this issue from many angles and dimensions.

Twitter is one of the most popular social media platforms, with a fast growing number of daily users. With 340 million active members as of 2020 and 98 thousand tweets per minute, Twitter is a social networking website where users post and exchange messages called ”tweets” and engage with one another, dubbed ”the SMS of the internet.” [3]

A field study or surveyors cannot provide such important data in such a short period of time. Big data like Twitter, where millions of individuals contribute their thoughts, is extremely tough to handle. Many academics were drawn to the analysis of these data, which allowed them to develop methodologies. Sentiment Analysis Methods are one of the most extensively utilized ways for analyzing Twitter data. Sentiment analysis is the process of automatically extracting subjective information from a text. We can determine whether a text has a favorable or negative subjective orientation using sentiment analysis. The notion of sentiment analysis has made the processing of this data easier, revealing what society says, intends, or desires.

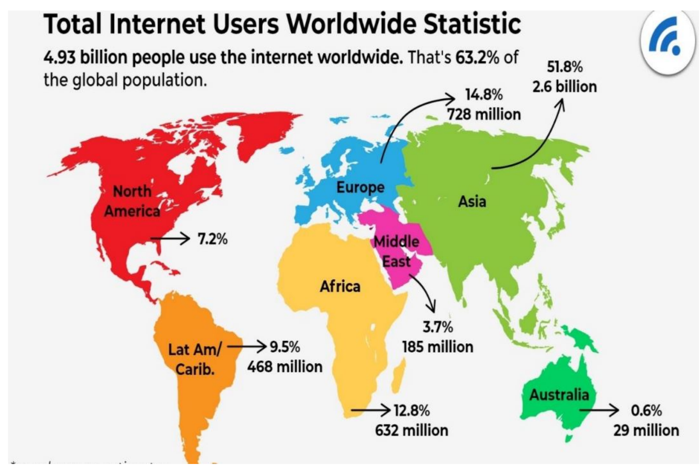


Fig. 1. Statistics on Internet Use in 2022 [1]

II. Connected Works

The fast growth of information sharing and access in social media networks has given rise to the notion of translating information into informatics in this industry. In this context, the notion of sentiment analysis, or opinion mining, is on its way to becoming the bright star of the information industry, which is growing in lockstep with technological advancements. If we look at the cause behind this, we can see that the internet is an excellent source of emotional information [4]. While emotion is an attitude, cognition is an emotional judgment.

The fast growth of information sharing and access media networks has given rise to the notion of tra information into informatics in this industry. In this the notion of sentiment analysis, or opinion mining, way to becoming the bright star of the information which is growing in lockstep with technological advan If we look at the cause behind this, we can see that the is an excellent source of emotional information [4 emotion is an attitude, cognition is an emotional judg

While technology was quickly expanding, huma countered an event that turned everything upside a time when everything they needed was just a cli on their phone in their pocket. While the coronavi first designated as a virus-related disease in the forr flu, the World Health Organization declared it a p... on March 11, 2020. More than 185 million individuals have been diagnosed with the coronavirus, which shocked the whole planet with its consequences, as of July 7, 2021.This infectious illness claimed the lives of around 4 million individuals. The coronavirus has set in motion a chain of events that induce social traumas that have a profound impact on their job and family lives, habits, holidays, and, in short, their entire existence.

Extensive research has been done employing both coronavirus and artificial intelligence technologies in this diverse process. Research concentrating on sentiment analysis of COVID-19 utilizing social media data stand out among these studies [6-10]. This is because individuals, businesses, and governments are utilizing social media to communicate with one another about the COVID-19 epidemic.

Because little is known about the concerns raised on social media platforms in relation to COVID-19, evaluating such data can assist lawmakers and healthcare organizations in assessing and responding to their stakeholders' demands [7]. During extreme occurrences such as pandemics, analyzing public mood is a must for effective crisis management [8]. Jim Samuel and colleagues [9] explain a systematic approach to evaluating Twitter data to infer public mood, monitoring the growth of dread linked with the fast spread of COVID-19 infections.Another research looks at how individuals are reacting to the epidemic, including how well they understand the disease and its symptoms, what measures they take, and if they follow political leaders' instructions [10]. Although we have discussed the positive aspects of social media shares to this point, if the shares in this formation, which we call social media, are malicious, that is, if false information is shared, while the world is already in a state of crisis due to the pandemic, sharing inaccurate information and fearing may further drag societies into psychological depression.

That is why, in circumstances of such grave concern, all governments should deploy social media fact-checkers to limit the spread of superfluous information [6].

III. Sentiment Analysis

While advances in technology and increased internet use have resulted in a rise in the amount of information communicated in this subject, problems such as what to do with it and how to handle it have established a field for researchers. Clustering, automatic text summarization, data categorization, and sentiment analysis are just a few examples of study topics produced by this surge of data. When we consider it in this perspective, we can easily observe people's remarks, what they enjoy and don't like, what they desire, their concerns, their health and economic problems, and many other characteristics when we browse different web pages or social media postings.What we saw became the basis for Idea Mining, or Sentiment Analysis [11]. Sentiment analysis is the process of automatically extracting subjective information from a text. We can determine whether a text has a favorable or negative subjective orientation using sentiment analysis.

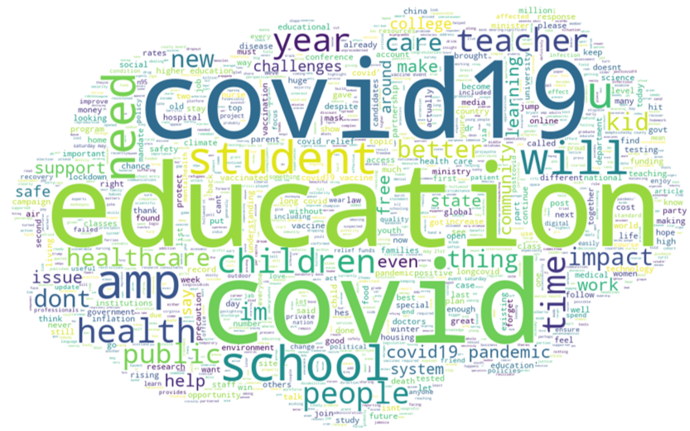


Fig. 2. Word cloud showing some words of the subject

One of the important techniques is the notion of sentiment analysis in social media posts [12]. However, with the explosion of user-generated material on the Internet, particularly in recent years, it has become an universe unto itself. The way individuals share their ideas and beliefs has also altered as a result of this atmosphere. Individuals may now submit their thoughts about a product they purchased on their website, as well as in internet forums, discussion groups, blogs, and social media about nearly anything.

Long forum messages and blogs are frequently a source of inspiration. As a result, finding relevant sources, extracting phrases about concepts, reading, summarizing, and organizing them into useable formats is tough [12]. As a result, systems that automatically find and summarize opinions are required. While the fundamental foundations of emotion analysis are created by processing this information, removing subjectivity, and classifying emotion, it also involves problems owing to its structure.

IV. Pre-Processing Steps

In this section, we cleaned and reorganized our raw data in preparation for usage in a deep learning model. Data preprocessing is an important stage in developing a Deep Learning model, and the outcomes are dependent on how effectively the data has been preprocessed. Text preprocessing is the initial stage in the NLP model-building process.Preprocessing processes were applied according to the table order below:

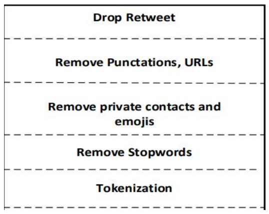


Fig. 3. Pre-processing Steps of the subject

A. Drop the Unnecessary Columns

To begin, we remove any extraneous columns from the data collection. When we acquire data from somewhere, in this example tweets from Twitter, it is often enormous. There might be a trove of data that adds no value to our model. Such information should be eliminated since it wastes memory and processing time.

```
#df.drop(columns=['id', 'conversation_id','created_at','date','time','timezone',
     #   'user_id','username','name','place','language','mentions','urls',
      #  'photos','replies_count','retweets_count','likes_count','hashtags',
      #  'cashtags','link','retweet','quote_url','video','thumbnail',
      #  'near','geo','source','user_rt_id','user_rt','retweet_id',
      #  'reply_to','retweet_date','translate','trans_src','trans_dest'], axis=1, inplace=True)
```

Fig. 4. Drop the Unnecessary Columns

B. Punctuation Marks Removal

The removal of punctuation marks, which are used to break text into sentences, paragraphs, and phrases, impacts the findings of any text processing strategy, especially those that rely on word and phrase occurrence frequencies, because punctuation marks are used often in text. Stop-words are often used in language and are deleted before any NLP procedure. Stop-words are a collection of words that are commonly used without any further information, such as articles, determiners, and prepositions. We may focus on the key terms instead by deleting these frequently used words from the text.

```
for line in lines:
    tokens=word_tokenize(line)
    tokens=[w.lower() for w in tokens]
    table=str.maketrans('','',string.punctuation)
    stripped = [w.translate(table) for w in tokens]
    words=[word for word in stripped if word.isalpha()]
    stop_words=set(stopwords.words('english'))
    words=[w for w in words if not w in stop_words]
    train_all_data.append(words)
```

Fig. 5. Punctuation Marks Removal

C. Replace Upper-case Characters to Lower-case

Words like book and book have the same meaning, yet in the vector space model, they are represented as two separate words when not transformed to lower case (It results in more dimensions). As a result, it has an impact on our processing speed and deep learning model correctness.

D. Tokenization

Identifying the words that make up a string of characters before processing a natural language. Tokenization is therefore a fundamental step in Natural Language Processing. This step is crucial because the meaning of the text may be deduced from an examination of the words in the text. Tokenization is the process of separating source text into distinct tokens for examination. Tokens are fragments of the original text that have not been reduced to their simplest form.

```
from keras.preprocessing.text import Tokenizer

train_data=train_data.astype(str)
test_data=test_data.astype(str)
tokenizer = Tokenizer()
tokenizer.fit_on_texts(train_data)

x_train = tokenizer.texts_to_sequences(train_data)
x_test =  tokenizer.texts_to_sequences(test_data)

word_index = tokenizer.word_index
print('Unique tokens:%d' %len(word_index))
```

Fig. 6. Tokenization

E. Stop Words Removal

Stop words are often used terms (a, an, the, etc.) in texts that should be removed. These terms have no actual meaning because they don't assist distinguish between two publications.

```
import string
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

train_all_data=list()
lines = counts

for line in lines:
    tokens=word_tokenize(line)
    tokens=[w.lower() for w in tokens]
    table=str.maketrans('','',string.punctuation)
    stripped = [w.translate(table) for w in tokens]
    words=[word for word in stripped if word.isalpha()]
    stop_words=set(stopwords.words('english'))
    words=[w for w in words if not w in stop_words]
    train_all_data.append(words)
```

Fig. 7. Stop Words Removal

The total results shown in these figures:

```
"@AaronKaviiri Well, it's not like we've a choice when they still contribute a huge chunk to our countries ' budgets, fund our
essential services including education, military and health (donated Covid-19 vaccines), huge loans and grants! Worst of it, ou
r leaders making home hell for us to stay"
```

Fig. 8. Stop Words Removal

```
['aaronkaviiri',
 'well',
 'like',
 'choice',
 'still',
 'contribute',
 'huge',
 'chunk',
 'countries',
 'budgets',
 'fund',
 'essential',
 'services',
 'including',
 'education',
 'military',
 'health',
 'donated',
 'vaccines'.
```

Fig. 9. Stop Words Removal

We then created and trained our algorithm to categorize the emotions of tweets. We built the model using a Sequential network for this purpose. The Sequential network model is explored in depth.

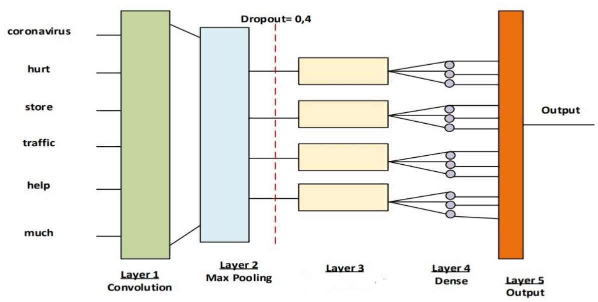
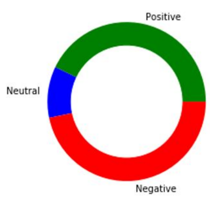


Fig. 10. Sequence Network Model's Architecture

V. Investigational Outcomes



	Total	Percentage
negative	278	46.72
positive	255	42.86
neutral	62	10.42

Fig. 11. Sequence Network Model's Architecture

As we can see from the Sentiment analysis results, we can conclude that the rates of positive and negative tweets are very close to each other. The number of people who are neutral at this point is comparatively less.

VI. Conclusion

Because of the circumstances of our day, social media plays a significant part in our lives. Twitter is the most popular social networking site, with millions of users. It provides researchers with a vast data set for data mining. Users on Twitter share 140-character tweets expressing their emotions.

We want to look at how people feel about Covid 19 pandemics in this study. A deep neural network was used to create the suggested model.

References

[1] Internet Users Worldwide Statistic, Available at: https://www.broadbandsearch.net/blog/internet-statistics, Anonymous, retrieved 28th July, 2021.

[2] He, W., Wu, H., Yan, G., Akula, V., and Shen, J. "A novel social media competitive analytics framework with sentiment." Elsevier, 1-12, 2015.

[3] Twitter. (2021, 07 13). wikipedia:https://tr.wikipedia.org/wiki/Twitter

[4] Fang, X., and Justin, Z., "Sentiment analysis using product review data." Journal of Big Data, 1-14, 2015.

[5] Christi, J., and Jain, G., "Sentiment Categorization through Natural Language Processing :A Survey.", 104- 107, (2019, 11 15).

[6] Alrazaq, A. a., Alhuwail, D., Househ, M., Hamdi, M., and Shah, Z., "Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study." JOURNAL OF MEDICAL INTERNET RESEARCH, 1-10, 2020.

[7] Gencoglu, O., "Large-Scale, Language-Agnostic Discourse Classification of Tweets During COVID-19." Machine Learning and Knowledge Extraction, 603–616, 2020.

[8] Samuel, J., Ali, G. G., Rahman, M., Esawi, E., Samuel, Y., "Covid-19 public sentiment insights and machine learning for tweets classification." Information, 11(6), 314, 2020.

[9] Iyer, P., Kumaresh, S., "Twitter Sentiment Analysis On Coronavirus Outbreak Using Machine Learning Algorithms." European Journal of Molecular Clinical Medicine, 2663-2676, 2020.

[10] BİLEN, B., HORASAN, F., "LSTM network based sentiment analysis for customer reviews." JOURNAL of POLYTECHNIC, 2021.

[11] Liu, B., "Sentiment analysis and subjectivity." Handbook of natural language processing, 2, 627-666, 2010.

[12] Xiao, Y., Yin, Y., "Hybrid LSTM neural network for short-term traffic flow prediction." Information, 10(3), 105, 2019.