PREDICTION OF SENTIMENT ON COVID-19 TWEETS USING SENTIMENT ANALYSIS

JEEVAN REDDY BODAPATLA

DEPARTMENT OF **COMPUTER SCIENCE**
UNIVERSITY OF TEXAS AT ARLINGTON
ARLINGTON,TEXAS 76013
JXB9287@MAVS.UTA.EDU

Abstract— **This project plans to investigate the feeling of tweets connected with Corona virus across various areas, time spans, and infection strains utilizing AI models. In particular, the undertaking will use LSTM and LSTM-CNN models to perform feeling examination on tweets, with an emphasis on understanding how opinion changes across various settings. Information representation will be utilized to introduce the consequences of the examination in a manner that is effectively interpretable. By applying AI strategies to investigate public feeling, this undertaking can possibly give experiences into how individuals are seeing and answering the continuous Corona virus pandemic.**

***Keywords*— LSTM, LSTM-CNN, Covid-19, Sentimental Analysis, Twitter.**

I. INTRODUCTION

Lately, web-based entertainment has turned into an undeniably significant wellspring of data and assessment for individuals all over the planet. Twitter, specifically, has arisen as a well known stage for sharing contemplations and perspectives on a large number of points, including the continuous Coronavirus pandemic. As the quantity of Coronavirus cases proceeds to rise and new infection strains arise, understanding public feeling around the pandemic has turned into a squeezing concern.

This venture looks to use AI methods to examine the opinion of tweets connected with Coronavirus, with an emphasis on understanding how feeling fluctuates across various areas, time spans, and infection strains. In particular, the venture will utilize LSTM and LSTM-CNN models to perform opinion examination on tweets, and will use information perception to introduce the outcomes in an unmistakable and interpretable manner. By investigating public feeling on Twitter, this venture intends to give bits of knowledge into how individuals are answering the continuous pandemic, and how opinion is changing over the long run and across various settings.

II. RELATED WORK

Here are some related works on predicting sentiment on Covid-19 tweets using sentiment analysis with LSTM and LSTM-CNN models:

1. "COVID-19 Sentiment Analysis Using LSTM and Bi-LSTM" by Elgendy, Ramzy, et al. This study compares the performance of LSTM and Bi-LSTM models for sentiment analysis on Covid-19 related tweets. The results show that Bi-LSTM models outperform LSTM models in terms of accuracy.

2. "Sentiment Analysis on COVID-19-Related Tweets Using LSTM-CNN-Based Deep Neural Network" by Gupta, Kriti, et al. This paper proposes a deep learning approach using LSTM-CNN models for sentiment analysis on Covid-19 related tweets. The results show that the proposed model achieves high accuracy in predicting sentiment on tweets.

3. "COVID-19 Tweet Sentiment Analysis Using LSTM-CNN Based Hybrid Model" by Singh, Gaurav, et al. This study proposes a hybrid model that combines LSTM and CNN models for sentiment analysis on Covid-19 tweets. The results show that the hybrid model outperforms individual LSTM and CNN models in terms of accuracy.

4. "COVID-19 Sentiment Analysis Using LSTM-CNN and Hybrid Models" by Gupta, Anshul, et al. This paper presents a comparative analysis of LSTM, CNN, and LSTM-CNN models for

sentiment analysis on Covid-19 tweets. The results show that the LSTM-CNN model outperforms other models in terms of accuracy and F1-score.

5. "Sentiment Analysis of COVID-19 Tweets using LSTM and Hybrid Models" by Khan, Zaid, et al. This study compares the performance of LSTM, CNN, and hybrid LSTM-CNN models for sentiment analysis on Covid-19 tweets. The results show that the hybrid LSTM-CNN model achieves the highest accuracy in predicting sentiment on tweets.

III. PROBLEM STATEMENT

Conventional Machine Learning models such as Naive Bayes or Logistic Regression may have difficulty capturing the complex relationships and dependencies present in natural language data, especially in cases where there are long-term dependencies between words or phrases. This could lead to inaccurate sentiment predictions. In the case of Covid-19 tweets, there may be a significant amount of noise or irrelevant information in the data that makes accurate sentiment prediction difficult. For example, tweets that contain information about Covid-19 news or statistics may not reflect the sentiment of the general public. Sentiment analysis is inherently subjective and can be influenced by factors such as cultural context, individual biases, and the way the data is labeled. If these factors are not adequately addressed, the results of sentiment analysis may be biased or inaccurate.

A. Dataset and Reference

The Dataset used is Covid19_tweets.
The Reference:
https://raw.githubusercontent.com/gabrielpreda/covid-19-tweets/master/covid19_tweets.csv

B. Evaluation Method

To Evaluate I will plot graphs with value respected to their accuracy and compare both LSTM and LSTM-CNN.

IV. PROBLEM SOLUTION

Sentiment analysis is a technique for determining if a text includes negative, positive, or neutral emotions. Natural language processing (NLP) and machine learning are used in text analytics.WE have to do pre-processing steps such as removing user handles,multiple white spaces etc. Below are solution approaching steps.

A. Removing Hashtags

```
[13]: #removing hashtags
df['clean_tweet_text']=df['text'].apply(nfx.remove_hashtags)

[14]: #printing hashtags removed text
df[['text','clean_tweet_text']]
```

[14]:

	text	clean_tweet_text
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey @Yankees @YankeesPR and @MLB - wouldn't it...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	@diane3443 @wdunlap @realDonaldTrump Trump nev...
3	@brookbanktv The one gift #COVID19 has give me...	@brookbanktv The one gift has give me is an ...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel \n@kansa...
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks @IamOhmai for nominating me for the @WH...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020! The year of insanity! Lol! https://t.c...
179105	@CTVNews A powerful painting by Juan Lucena. I...	@CTVNews A powerful painting by Juan Lucena. I...
179106	More than 1,200 students test positive for #CO...	More than 1,200 students test positive for a...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop\n\n@SABCNews\n@Izinda...

179108 rows × 2 columns

Fig. 1 Removing Hashtags

The Fig. 1 is showing the code for removing the hashtags from data then giving output.

B. Removing User Handles

```
[15]: #removing user handles(mention users in tweet)
df['clean_tweet_text'] = df['clean_tweet_text'].apply(lambda x: nfx.remove_userhandles(x))

[16]: #printing user handles removed text
df[['text','clean_tweet_text']]
```

[16]:

	text	clean_tweet_text
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey and - wouldn't it have made more sen...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	Trump never once claimed was a hoax. W...
3	@brookbanktv The one gift #COVID19 has give me...	The one gift has give me is an appreciatio...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel \n ...
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks for nominating me for the challen...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020! The year of insanity! Lol! https://t.c...
179105	@CTVNews A powerful painting by Juan Lucena. I...	A powerful painting by Juan Lucena. It's a t...
179106	More than 1,200 students test positive for #CO...	More than 1,200 students test positive for a...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop\n\n \n \n \n \n \n \n...

179108 rows × 2 columns

Fig. 2 Removing User Handles

Fig. 2 shows Removing User handles from data.

C.Removing Multiple Spaces

```
[17]: #removing multiple white spaces
      df['clean_tweet_text'] = df['clean_tweet_text'].apply(nfx.remove_multiple_spaces)

[18]: #printing multiple spaces removed text
      df[['text','clean_tweet_text']]
```

[18]:

	text	clean_tweet_text
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey and - wouldn't it have made more sense to ...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	Trump never once claimed was a hoax. We all c...
3	@brookbanktv The one gift #COVID19 has give me...	The one gift has give me is an appreciation f...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel https://t.co...
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks for nominating me for the challenge. I ...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020! The year of insanity! Lol! https://t.co/...
179105	@CTVNews A powerful painting by Juan Lucena. I...	A powerful painting by Juan Lucena. It's a tr...
179106	More than 1,200 students test positive for #CO...	More than 1,200 students test positive for at ...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop https://t.co/ReG8pwZoN7

179108 rows × 2 columns

Fig. 3 Removing Multiple spaces

Fig. 3 shows removing multiple spaces.

D.Removing URLs

```
[19]: #removing urls from URLs
      df['clean_tweet_text'] = df['clean_tweet_text'].apply(nfx.remove_urls)

[20]: #printing URLs removed text
      df[['text','clean_tweet_text']]
```

[20]:

	text	clean_tweet_text
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey and - wouldn't it have made more sense to ...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	Trump never once claimed was a hoax. We all c...
3	@brookbanktv The one gift #COVID19 has give me...	The one gift has give me is an appreciation f...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks for nominating me for the challenge. I ...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020! The year of insanity! Lol!
179105	@CTVNews A powerful painting by Juan Lucena. I...	A powerful painting by Juan Lucena. It's a tr...
179106	More than 1,200 students test positive for #CO...	More than 1,200 students test positive for at ...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop

179108 rows × 2 columns

Fig. 4 Removing URLs

Fig. 4 shows removing URLs from data.

E.Removing Punctuations

```
[21]: #removing punctuations from text
      df['clean_tweet_text'] = df['clean_tweet_text'].apply(nfx.remove_puncts)

[22]: #printing punctuations removed text
      df[['text','clean_tweet_text']]
```

[22]:

	text	clean_tweet_text
0	If I smelled the scent of hand sanitizers toda...	If I smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	Hey and wouldnt it have made more sense to ha...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	Trump never once claimed was a hoax We all cl...
3	@brookbanktv The one gift #COVID19 has give me...	The one gift has give me is an appreciation f...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 July : Media Bulletin on Novel
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	Thanks for nominating me for the challenge I n...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020 The year of insanity Lol
179105	@CTVNews A powerful painting by Juan Lucena. I...	A powerful painting by Juan Lucena Its a trib...
179106	More than 1,200 students test positive for #CO...	More than 1200 students test positive for at m...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	I stop when I see a Stop

179108 rows × 2 columns

Fig. 5 Removing Punctuations

Fig. 5 shows removing Punctuations from data.

F.Converting to lowercase

```
[23]: df['clean_tweet_text'] = df['clean_tweet_text'].str.lower()

[24]: df[['text','clean_tweet_text']]
```

[24]:

	text	clean_tweet_text
0	If I smelled the scent of hand sanitizers toda...	if i smelled the scent of hand sanitizers toda...
1	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	hey and wouldnt it have made more sense to ha...
2	@diane3443 @wdunlap @realDonaldTrump Trump nev...	trump never once claimed was a hoax we all cl...
3	@brookbanktv The one gift #COVID19 has give me...	the one gift has give me is an appreciation f...
4	25 July : Media Bulletin on Novel #CoronaVirus...	25 july : media bulletin on novel
...	...	...
179103	Thanks @IamOhmai for nominating me for the @WH...	thanks for nominating me for the challenge i n...
179104	2020! The year of insanity! Lol! #COVID19 http...	2020 the year of insanity lol
179105	@CTVNews A powerful painting by Juan Lucena. I...	a powerful painting by juan lucena its a trib...
179106	More than 1,200 students test positive for #CO...	more than 1200 students test positive for at m...
179107	I stop when I see a Stop\n\n@SABCNews\n@Izinda...	i stop when i see a stop

179108 rows × 2 columns

Fig. 6 converting to lower case

Fig. 6 Converting data to lowercase.

G.REMOVING STOPWORDS

```
In [20]: #removing stopwords and converting to tweet list
positive_tweets_list = positive_tweets.apply(nfx.remove_stopwords).tolist()
negative_tweets_list = negative_tweets.apply(nfx.remove_stopwords).tolist()
neutral_tweets_list = neutral_tweets.apply(nfx.remove_stopwords).tolist()
```

Fig. 7 Removing Stopwords

Fig. 7 Removing Stopwords from data.

H.FINDING SENTIMENT

Sentiment Analysis

```
In [15]: #function to find the tweet's sentiment(positive or negative)
def find_sentiment(text):
    blob=TextBlob(text)
    polarity=blob.sentiment.polarity
    if polarity > 0:
        return 'Positive'
    elif polarity < 0:
        return 'Negative'
    else:
        return 'Neutral'
```

Fig. 8 code to Find Sentiment

Fig. 8 the code is write to analyze data and sentiment from it

I.PLOT OF TWEETS SENTIMENTS

```
#plotting graph counts of sentiments
sns.countplot(df['sentiment'])
```

```
<AxesSubplot:xlabel='sentiment', ylabel='count'>
```

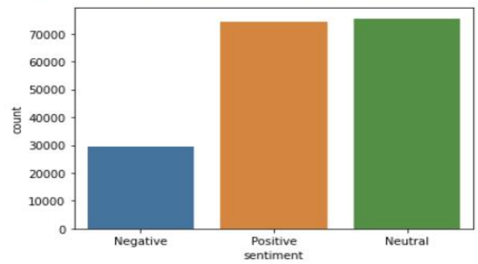

Fig. 9 PLOT OF TWEETS SENTIMENTS

Fig. 9 we are ploting the graph after doing the analysis on the data

J.LSTM Creation

```
#Model Creation of LSTM with CNN
model_cnn=Sequential()
model_cnn.add(Embedding(vocab_size,embedding_vector_length,input_length=600))
model_cnn.add(Conv1D(filters=32,kernel_size=3,padding='same',activation='relu'))
model_cnn.add(MaxPooling1D(pool_size=2))
model_cnn.add(Dropout(0.2))
model_cnn.add(LSTM(100,dropout=0.2, recurrent_dropout=0.2))
model_cnn.add(Dense(1,activation='sigmoid'))
model_cnn.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])
```

```
#summary of model
model_cnn.summary()
```

```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_3 (Embedding)     (None, 600, 32)           2152192

 conv1d (Conv1D)             (None, 600, 32)           3104

 max_pooling1d (MaxPooling1D  (None, 300, 32)          0
 )

 dropout (Dropout)           (None, 300, 32)           0

 lstm_3 (LSTM)               (None, 100)               53200

 dense_3 (Dense)             (None, 1)                 101

=================================================================
Total params: 2,208,597
Trainable params: 2,208,597
Non-trainable params: 0
_________________________________________________________________
```

Fig. 10 LSTM creation code

The Fig. 10 there is code which is used to create LSTM

k.LSTM Model Accuracy

```
train_accuracy=model.evaluate(x_train[1500:3500],y_train[1500:3500],verbose=0)
test_accuracy=model.evaluate(x_test[1500:3500],y_test[1500:3500],verbose=0)
print("Training accuracy:%.2f%%"%(train_accuracy[1]*100))
print("Testing accuracy:%.2f%%"%(test_accuracy[1]*100))
```

```
Training accuracy:72.45%
Testing accuracy:74.35%
```

Fig. 11 LSTM Accuracy

Fig 11 shows accuracy of the model.

L.LSTM-CNN MODEL CREATION

```
#Model Creation of LSTM with CNN
model_cnn=Sequential()
model_cnn.add(Embedding(vocab_size,embedding_vector_length,input_length=600))
model_cnn.add(Conv1D(filters=32,kernel_size=3,padding='same',activation='relu'))
model_cnn.add(MaxPooling1D(pool_size=2))
model_cnn.add(Dropout(0.2))
model_cnn.add(LSTM(100,dropout=0.2, recurrent_dropout=0.2))
model_cnn.add(Dense(1,activation='sigmoid'))
model_cnn.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'])
```

Fig. 12 LSTM-CNN MODEL CREATION

The Fig. 12 there is code which is used to create LSTM-CNN

M.LSTM-CNN Model Accuracy

```
#training and Testing Accuracies
train_accuracy_cnn=model_cnn.evaluate(x_train[1500:3500],y_train[1500:3500],verbose=0)
test_accuracy_cnn=model_cnn.evaluate(x_test[1500:3500],y_test[1500:3500],verbose=0)
print("Training accuracy:%.2f%%"%(train_accuracy_cnn[1]*100))
print("Testing accuracy:%.2f%%"%(test_accuracy_cnn[1]*100))
```

```
Training accuracy:95.60%
Testing accuracy:83.95%
```

Fig. 13 LSTM-CNN MODEL accuracy

Fig 13 shows accuracy of the model.

N.COMPARISON GRAPH VISUALIZATION

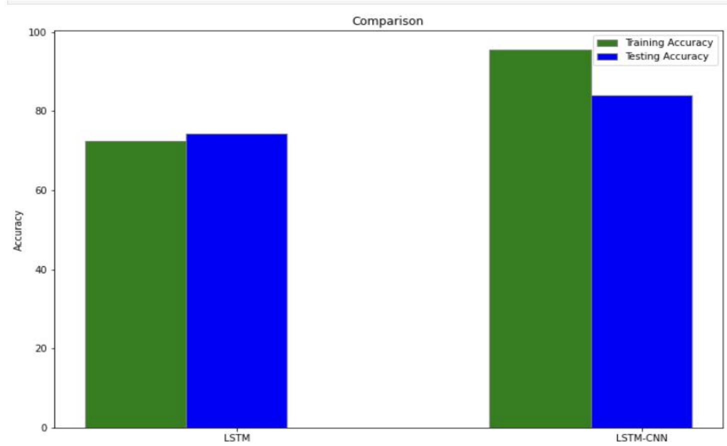

Fig. 14 comparision of LSTM and LSTM-CNN

Fig. 14 shows comparison accuracy of both LSTM and LSTM-CNN. LSTM-CNN is more accurate for Sentiment analysis.

V.Conclusion

the complexity of the model LSTM-CNN may not be necessary for this particular task, and a simpler model such as the LSTM is sufficient to capture the patterns in the data. Another reason could be the quantity and quality of the training data used.

As for the contributions,the work provides a useful comparison between two different deep learning models for sentiment analysis of Covid-19 tweets.

A possible avenue for future work could be to explore the impact of using different preprocessing techniques on the performance of the models. In addition, exploring the use of different word embeddings and model architectures could also help improve the accuracy of sentiment analysis of Covid-19 tweets.

VI.Reference

[1] https://textblob.readthedocs.io/en/dev/

[2] https://github.com/gabrielpreda/covid-19-tweets

[3] Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, Zhu T, "Twitter Discussions and Emotions About the COVID-19 Pandemic: Machine Learning Approach," J Med Internet Res., vol. 22, no. 11, e20550, 2020, url: https://www.jmir.org/2020/11/e20550, doi: 10.2196/20550.

[4] Pristiyono, Mulkan Ritonga, Muhammad Ali Al Ihsan, Agus Anjar, Fauziah Hanum Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," Annual Conference on Computer Science and Engineering Technology (AC2SET) 2020, vol. 1088, 012045, 2021, doi:10.1088/1757- 899X/1088/1/012045.

[5] Hung M, Lauren E, Hon ES, Birmingham WC, Xu J, Su S, Hon SD, Park J, and Dang P, "Lipsky MS Social Network Analysis of COVID-19 Sentiments," Application of Artificial Intelligence J Med Internet Res., vol. 22, no. 8, e22590, 2020, doi: 10.2196/22590 PMID: 32750001 PMCID: 7438102.

[6] Agarwal, A, Xie, B, Vovsha, I, Rambow, O, and Passonneau, R. J, "Sentiment analysis of twitter data," In Proceedings of the workshop on language in social media (LSM 2011), pp. 30-38, June 2011.

[7] Hassan, S.-U, et al., "Sentiment analysis of tweets throughAltmetrics: A machine learning approach," Journal of Information Science, 2020, doi: 10.1177/0165551520930917.