Bonus Work - 3

Name: Yashwanth Reddy Samala

SJSU ID: 016014583

GitHub Repo: https://github.com/YashReddyS/Twitter-Sentiment-Analysis

**Notebook Link for model training:**
https://colab.research.google.com/drive/1Xs63_7PmwQOWRtv6WwVbghr52T_1mRZY?usp=sharing

Goal: Given the twitter handle of a user get the sentiment analysis of on their 20 most recent tweets. Sentiments are categorized as Positive, neutral and Negative.

Dataset Used: Used a labeled dataset that has tweets labeled as positive, negative and neutral. The distribution of these tweets are as shown below.
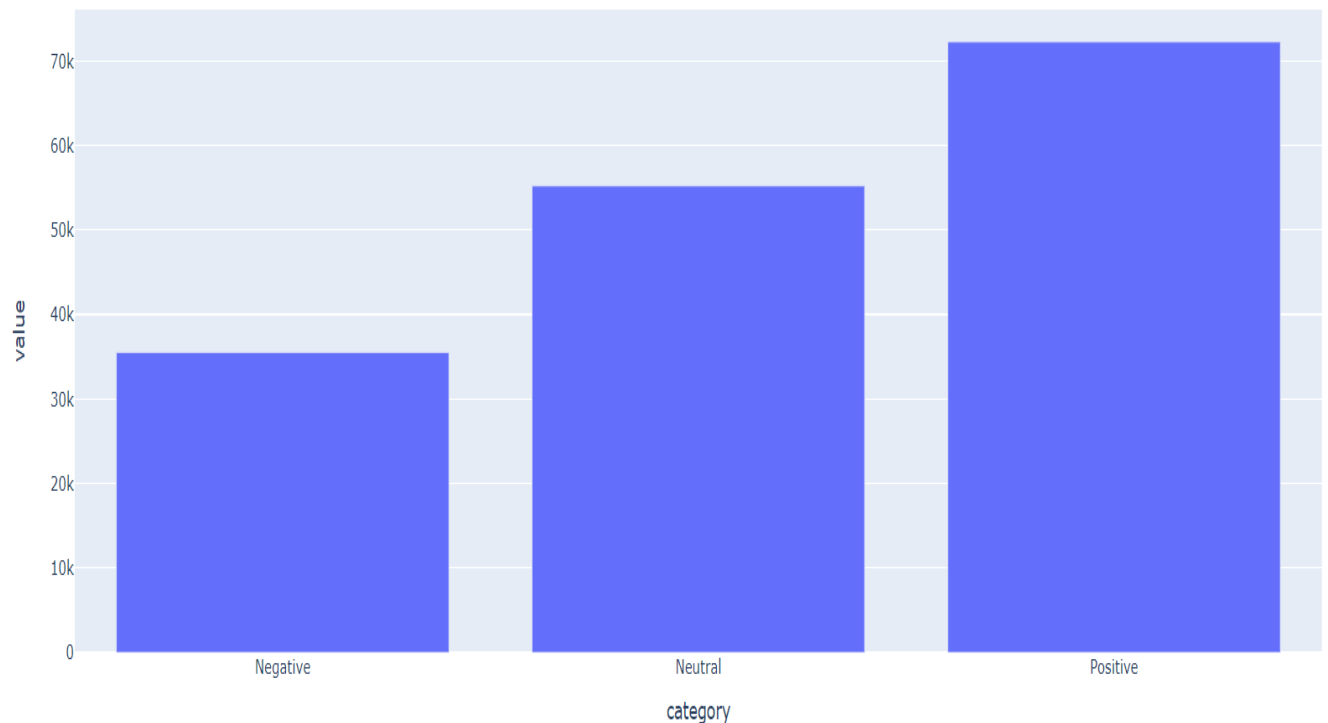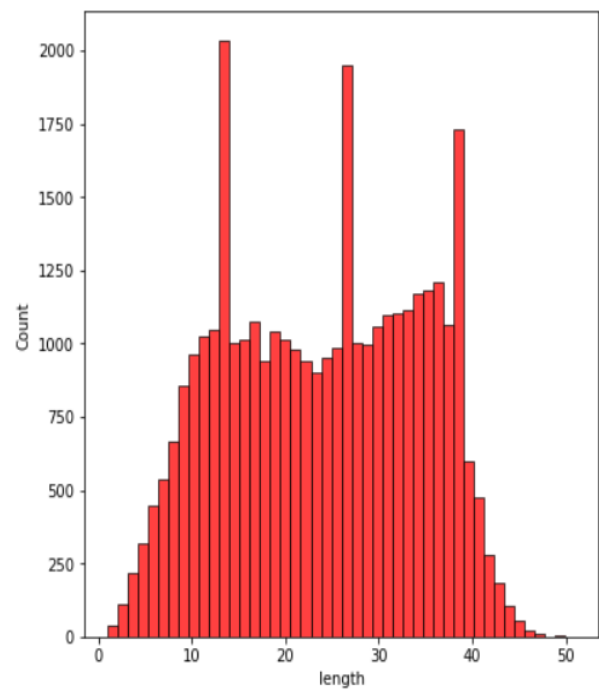
Distribution of text length for positive sentiment tweets.

	length
count	72249.0
mean	22.65
std	10.6
min	1.0
25%	14.0
50%	22.0
75%	32.0
max	50.0
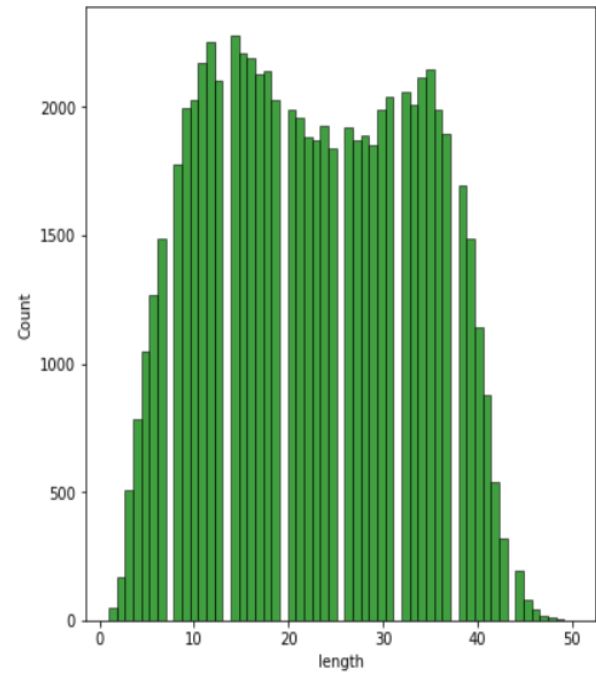
Distribution of text length for Negative sentiment tweets.

	length
count	35509.0
mean	23.88
std	10.31
min	1.0
25%	15.0
50%	24.0
75%	33.0
max	51.0

Pie chart of different sentiments of tweets

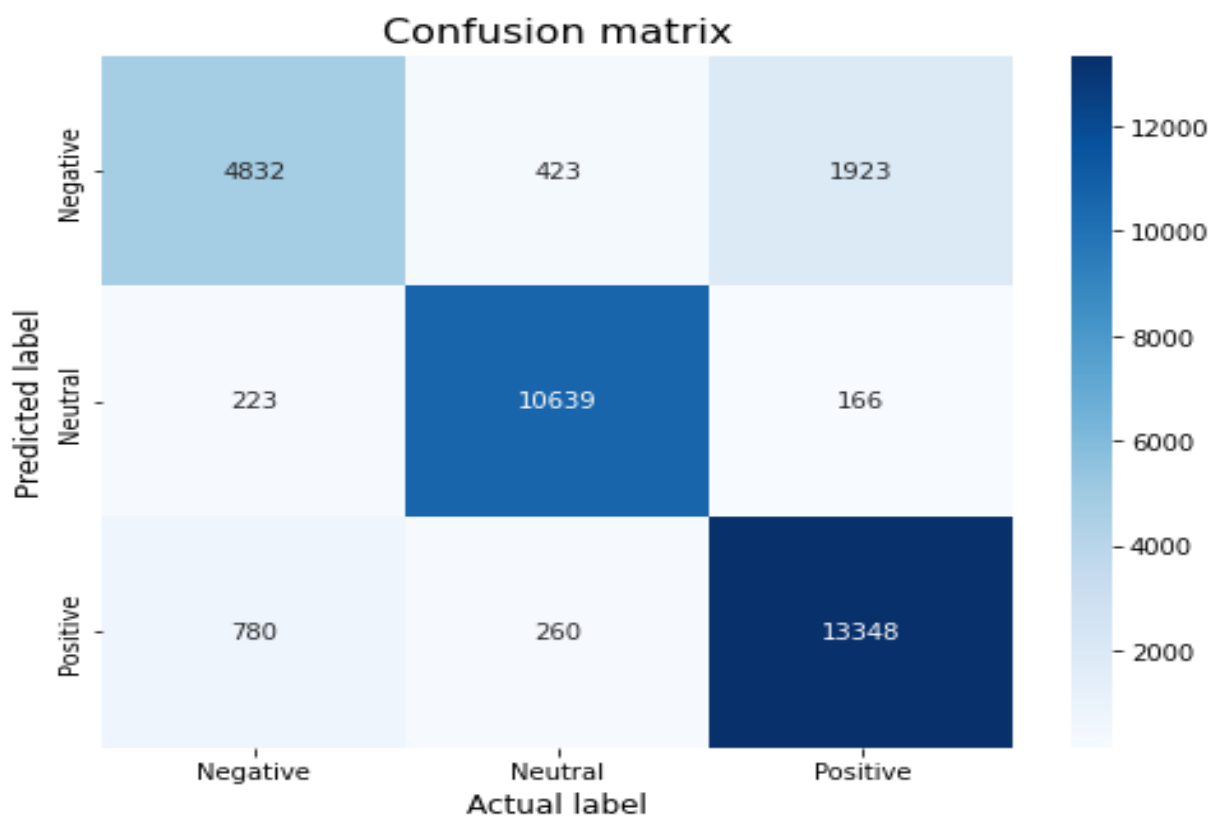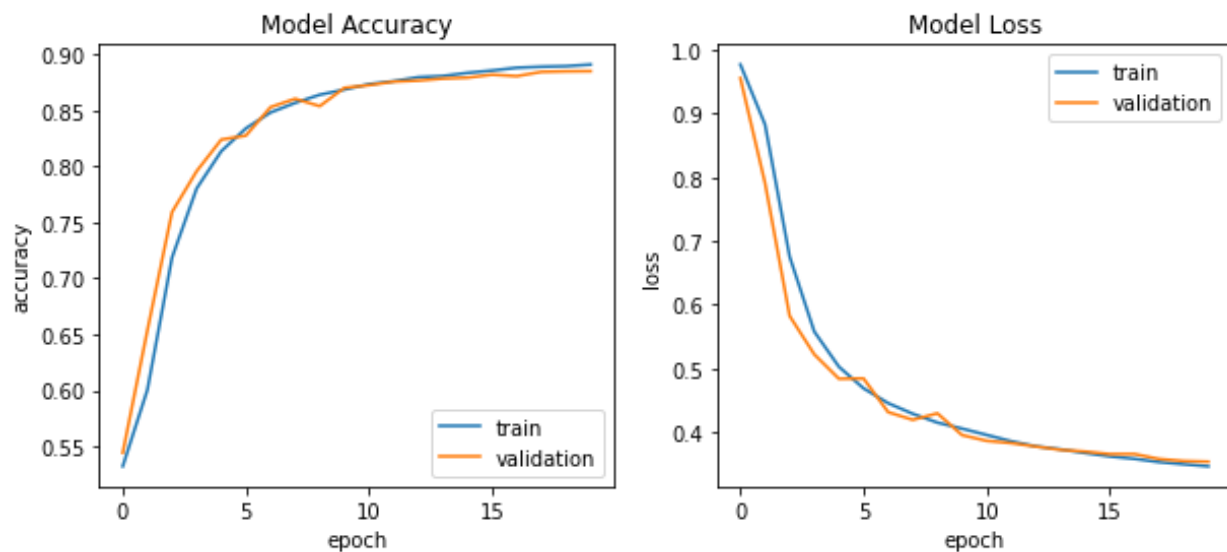**Positive**
Neutral
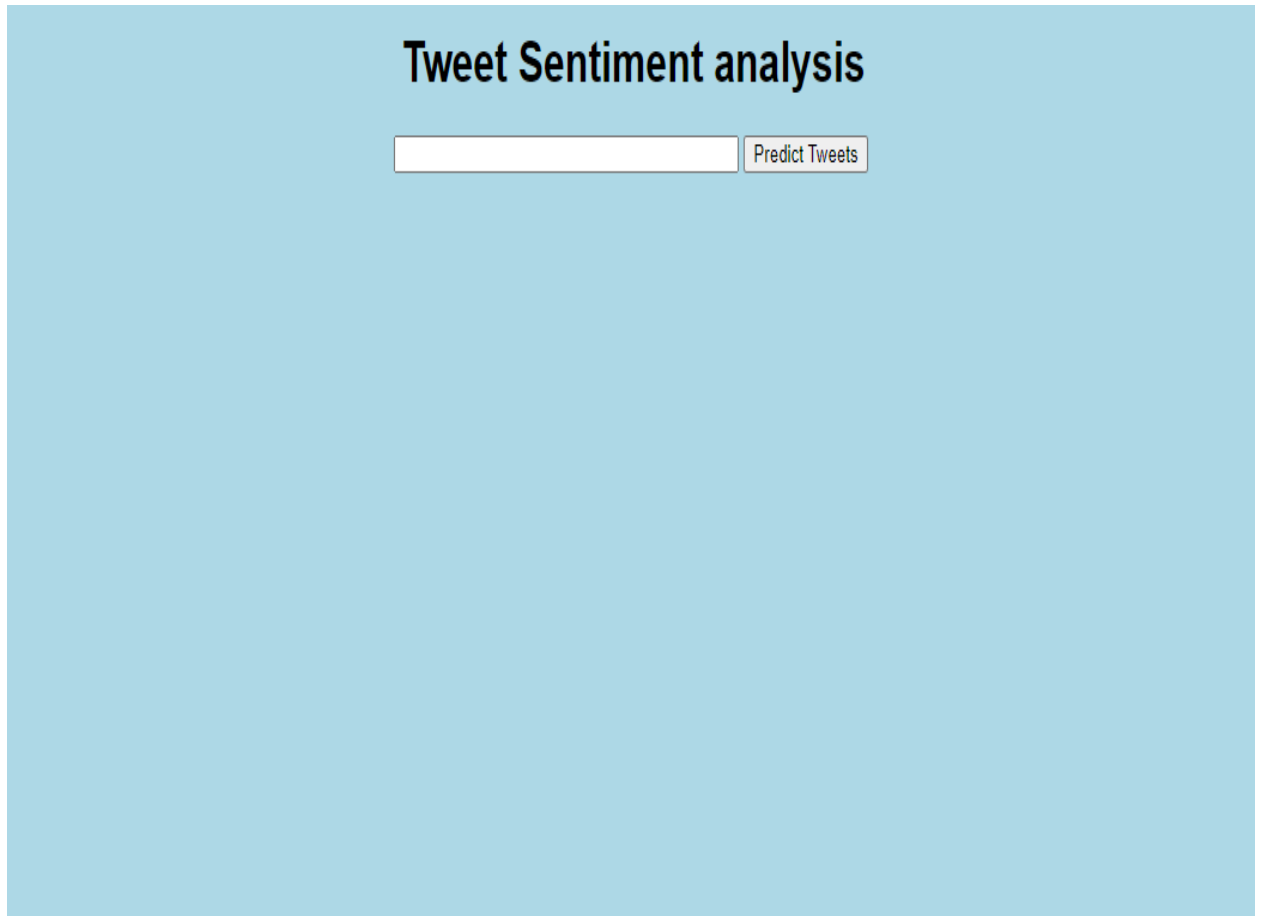**Negative**

33.9%
44.3%
21.8%
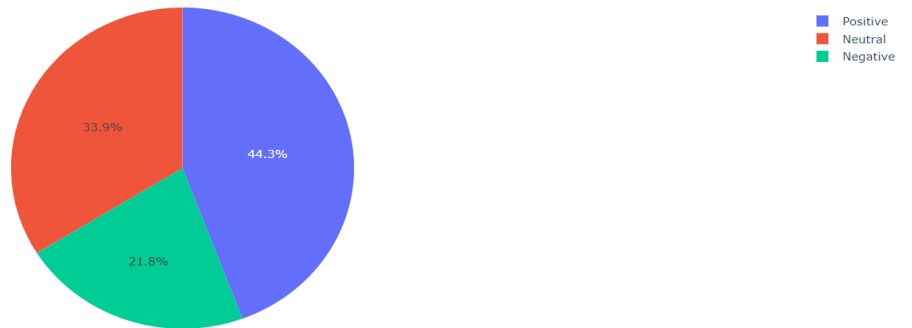
Steps for classification:

1. Initially we prepare a dataset containing positive, negative and neutral tweets and train a neural net model to get the classifications. We implement LSTM model as it is very is efficient with NLP tasks. The model architecture is as shown below

embedding_input	input:	[(None, 50)]	[(None, 50)]
InputLayer	output:		

embedding	input:	(None, 50)	(None, 50, 32)
Embedding	output:		

conv1d	input:	(None, 50, 32)	(None, 50, 32)
Conv1D	output:		

max_pooling1d	input:	(None, 50, 32)	(None, 25, 32)
MaxPooling1D	output:		

bidirectional(lstm)	input:	(None, 25, 32)	(None, 64)
Bidirectional(LSTM)	output:		

dropout	input:	(None, 64)	(None, 64)
Dropout	output:		

dense	input:	(None, 64)	(None, 3)
Dense	output:		

2. The neural net model with best accuracy from all the epochs is saved and later we load the model weights and run inference.

Model Accuracy

Model Loss

Confusion matrix

	Negative	Neutral	Positive
Negative	4832	423	1923
Neutral	223	10639	166
Positive	780	260	13348

3. After training process is done, we collet most recent tweets given the Twitter user's handle. To implement this, we use Flask python framework to create a web server and get inputs form the web app.

4. After the handle is obtained, we have to retrieve the users' tweets. There are multiple approaches to get these tweets
   - Selenium, Beautiful Soup python library
   - Twitter API
   - Other twitter libraries such as tweepy etc

   In this implementation we use a library known as snscrape. Snscrape doesn't require user credentials to access public tweets and there is no limitation to the no of tweets the script can retrieve unlike Twitter API. We save the tweet data such as date, username, tweet content into a pandas dataframe.

5. After getting the most recent tweets from the user we need to perform the sentiment analysis on the tweets. We run each tweet through the previously trained neural net model to get it classified as either positive, negative or neutral tweet and we append the sentiment as column to the previously saved tweets data frame. After this we display the tweets datatframe to HTML table using flask to get results as shown below (for user handle "elonmusk").

Tweet Sentiment analysis

Username [   ] [Predict Tweets]

	Date	Username	Tweet	Sentiment
0	2022-05-11 02:39:49+00:00	elonmusk	@WholeMarsBlog @akidesir Release notes on this version will be long	Negative
1	2022-05-11 02:18:05+00:00	elonmusk	@akidesir Probably this weekend. Lot of code updates means much is discovered during testing.	Positive
2	2022-05-11 01:00:15+00:00	elonmusk	@BillyM2k Exactly	Neutral
3	2022-05-11 00:56:54+00:00	elonmusk	Starlink has resisted Russian cyberwar jamming & hacking attempts so far, but they're ramping up their efforts https://t.co/w62yCsDA5w	Positive
4	2022-05-10 19:01:50+00:00	elonmusk	@bennyjohnson Important to listen to my full explanation	Positive
5	2022-05-10 18:43:16+00:00	elonmusk	@BillyM2k @BNONews Haha 💯	Neutral
6	2022-05-10 16:19:46+00:00	elonmusk	@BillyM2k 🤣	Neutral
7	2022-05-10 16:06:33+00:00	elonmusk	@Tesmanian_com The Tesla China team is amazing	Positive
8	2022-05-09 22:55:39+00:00	elonmusk	@ThierryBreton Great meeting! We are very much on the same page.	Positive
9	2022-05-09 20:28:58+00:00	elonmusk	@RationalEtienne @HindenburgRes You might be right	Positive

Future Scope: Other models such as BERT and roBERTa can be implemented to achieve better accuracies and results and UI could be improved.