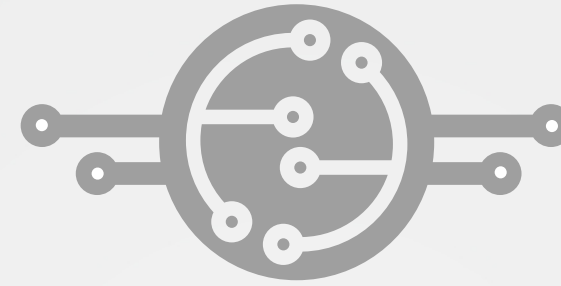


January 2023



Twitter Sentiment Analysis

Computational Intelligence Project

Project Stages

Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

Stage 1

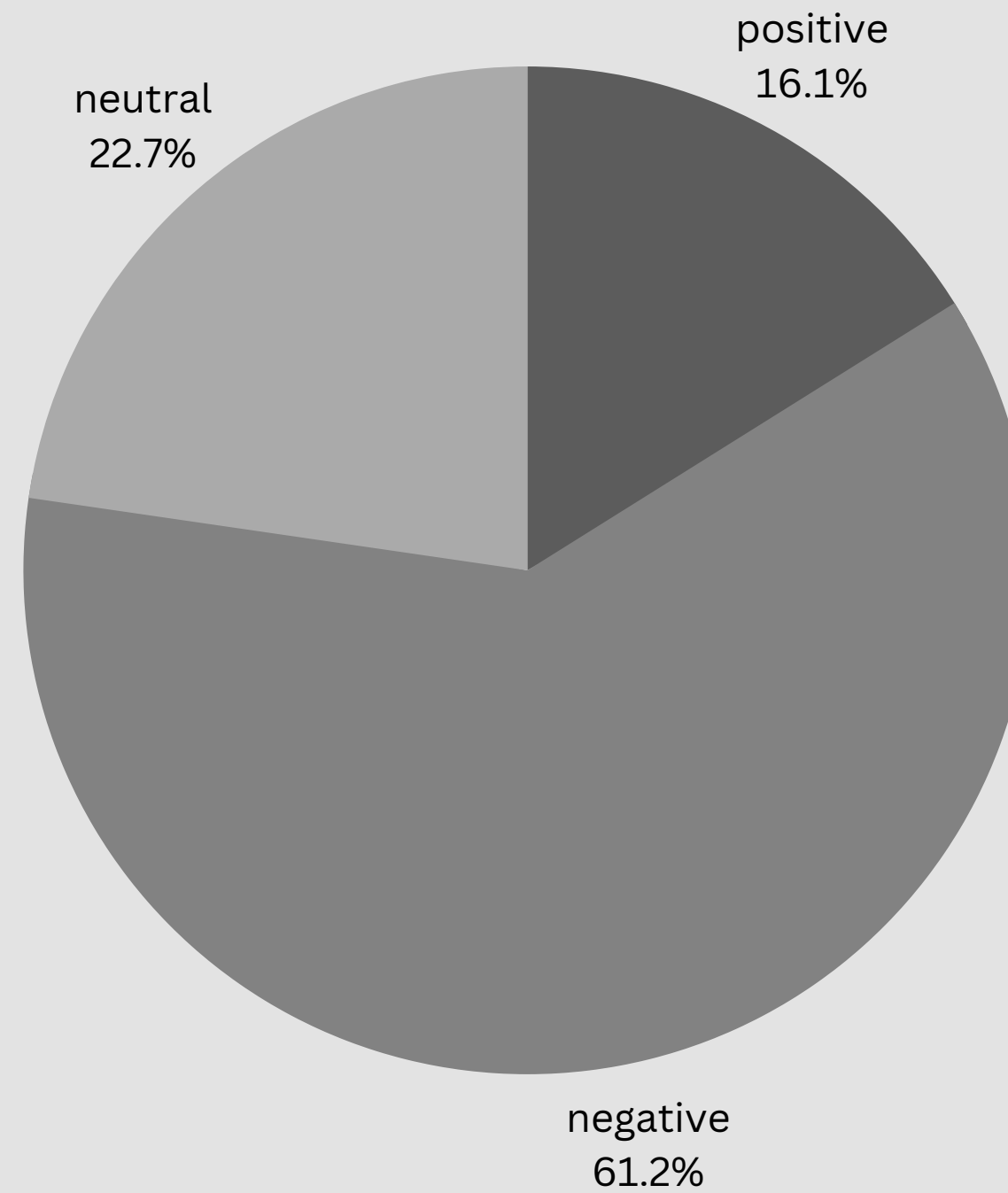
- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

	text	sentiment
0	RT @NancyLeeGrah: How did everyone feel about...	Neutral
1	RT @ScottWalker: Didn't catch the full #GOPdeb...	Positive
2	RT @TJMShow: No mention of Tamir Rice and the ...	Neutral
3	RT @RobGeorge: That Carly Fiorina is trending ...	Positive
4	RT @DanScavino: #GOPDebate w/ @realDonaldTrump...	Positive
...
13866	RT @cappy_yarbrough: Love to see men who will ...	Negative
13867	RT @georgehenryw: Who thought Huckabee exceede...	Positive
13868	RT @Lrihendry: #TedCruz As President, I will a...	Positive
13869	RT @JRehling: #GOPDebate Donald Trump says tha...	Negative
13870	RT @Lrihendry: #TedCruz headed into the Presid...	Positive

13871 rows × 2 columns

Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation





Stage 1

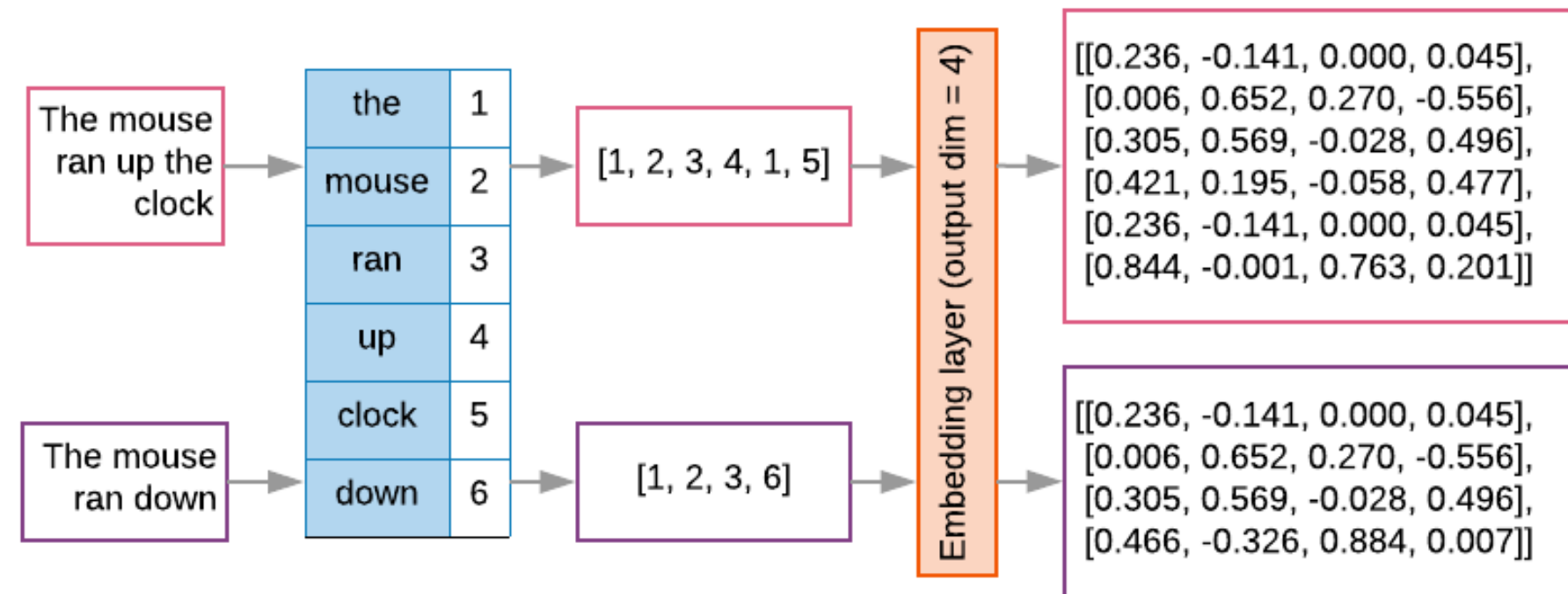
- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

- 
- 
- , ? / : # @
 - links
 - RT

Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

- Text to Number
- no Stemming
- no POS tagging



Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 28, 128)	256000
spatial_dropout1d (SpatialD ropout1D)	(None, 28, 128)	0
lstm (LSTM)	(None, 196)	254800
dense (Dense)	(None, 3)	591
Total params: 511,391		
Trainable params: 511,391		
Non-trainable params: 0		

None

Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

One-hot encoding

	cat	mat	on	sat	the
the =>	0	0	0	0	1
cat =>	1	0	0	0	0
sat =>	0	0	0	1	0
...					



A 4-dimensional embedding

cat =>	1.2	-0.1	4.3	3.2
mat =>	0.4	2.5	-0.9	0.5
on =>	2.1	0.3	0.1	0.4
...				



Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

- 
- trained on "First GOP Debate Twitter Sentiment"
 - randomly masks 15% of the words
 - predict masked words
- 



Stage 1

- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

Accuracy	Train	Validation
LSTM	97.3 %	83.1 %
BERT	93.8 %	74.2 %



Project Stages

Stage 1

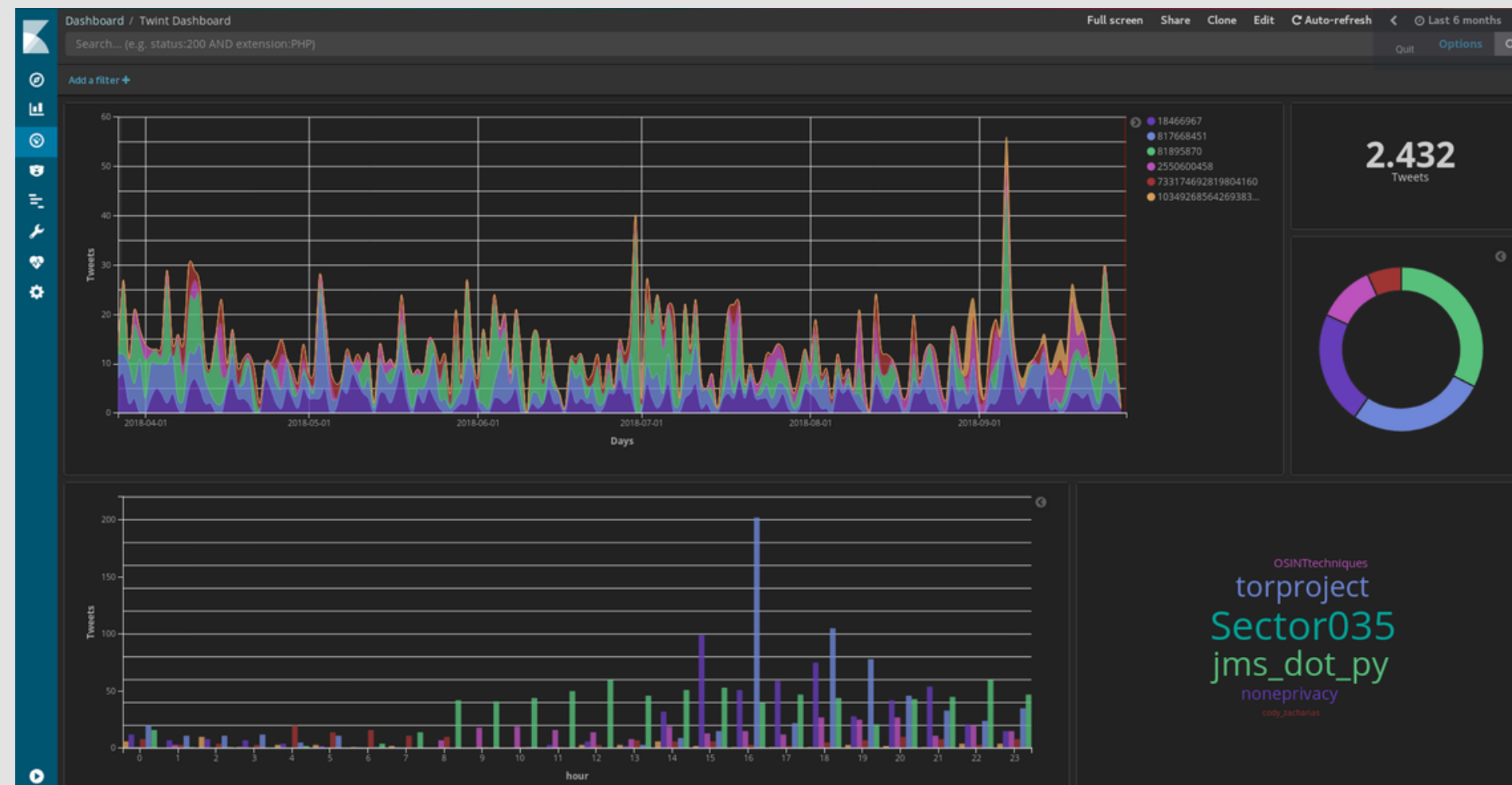
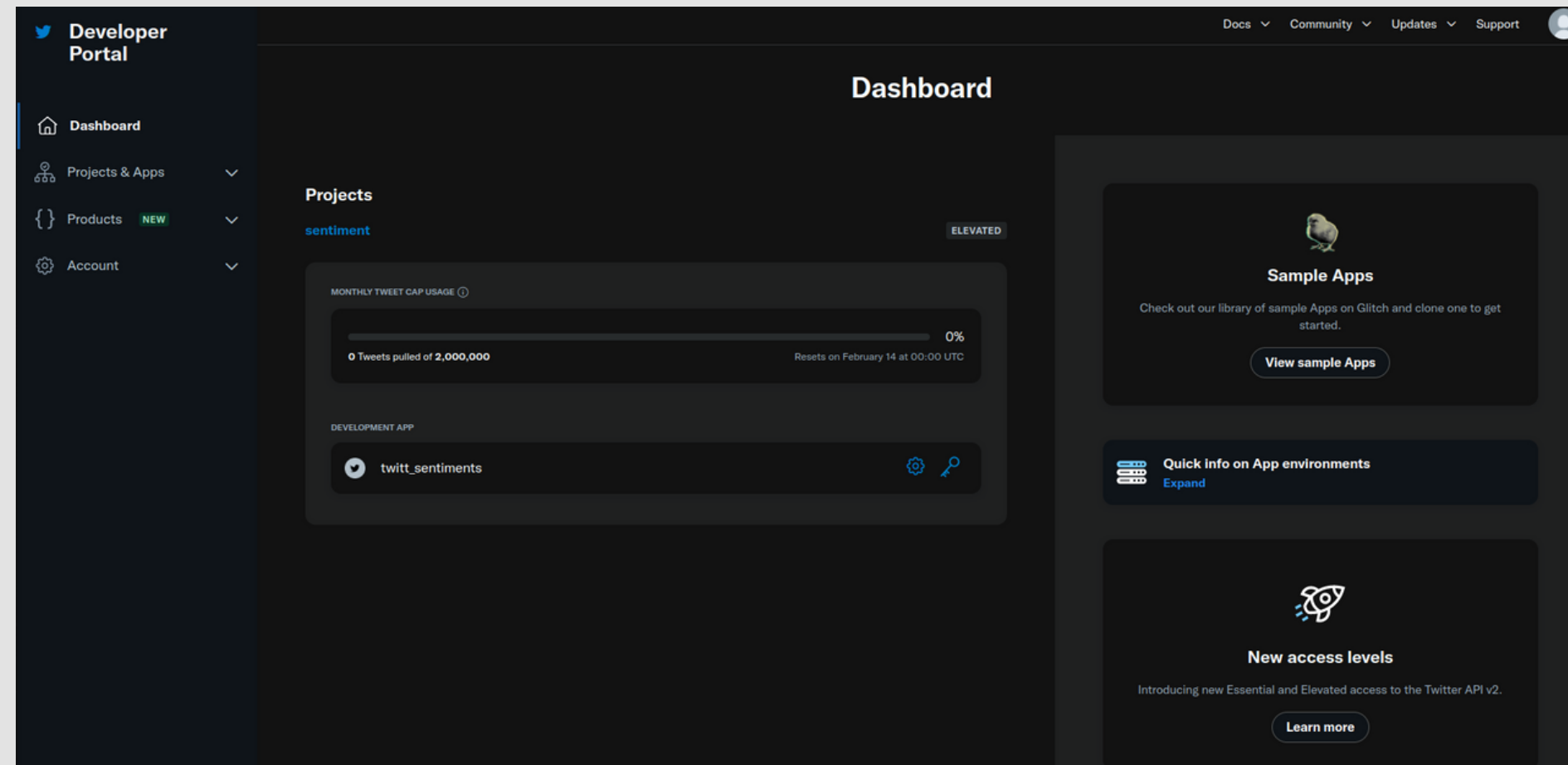
- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

Stage 2

- Twitter API/Twint
- EDA
 1. Most Common Words
 2. Word Cloud
 3. Sample Tweets
- Tweets Classification
- +/- Tweets Percentage

Stage 2

- Twitter API/Twint
- EDA
 1. Most Common Words
 2. Word Cloud
 3. Sample Tweets
- Tweets Classification
- +/- Tweets Percentage

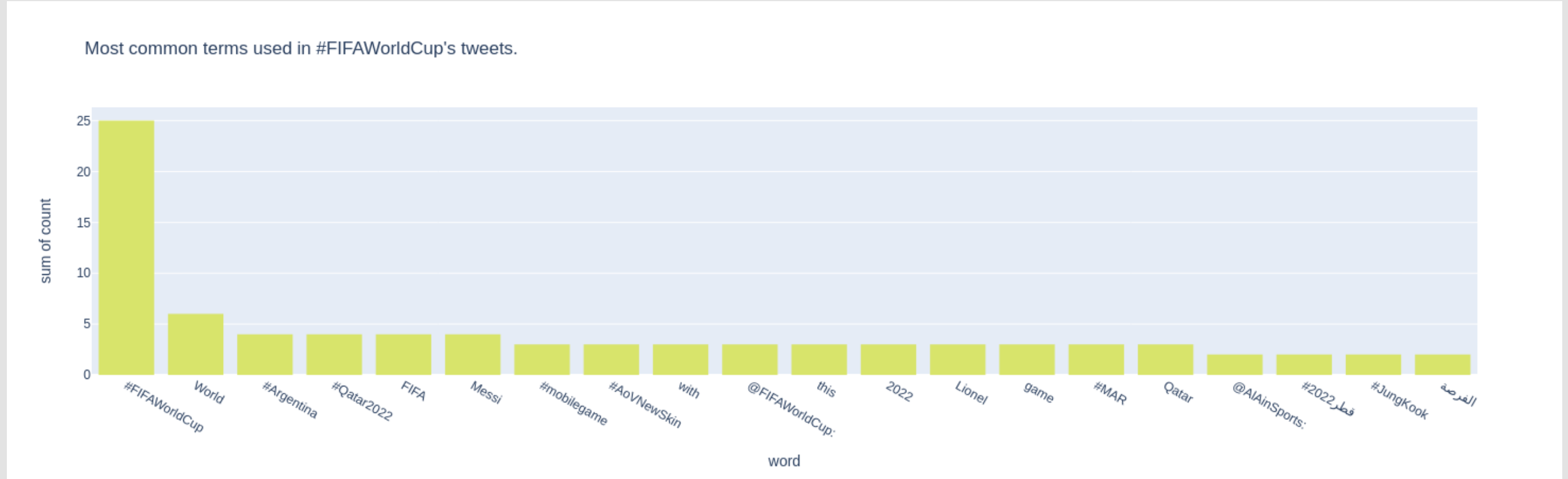


Stage 2

- Twitter API/Twint
- EDA
 1. Most Common Words
 2. Word Cloud
 3. Sample Tweets
- Tweets Classification
- +/- Tweets Percentage

```
[ 'Positive tweets:', 'RT @NICKIMINAJ: #TukohTaka 🌟🌟🌟🌟\n#fifaworldcup official song 🎧 \nFULL VIDEO TMRW @ 10AM EST 🌐 🌐 \nBARBIE X MALUMA X MYRIAM \nEnglish X Spanish...', "RT @creculture_lab: #football\n#FIFAWorldCup\u202f\u202f\u202f\u202f\u202f\u202f\n#Qatar2022\u202f\u202f\u202f\u202f\u202f\u202f\n#KOREAGHANA\nEvery Girl's Heart Stealer?!\nCheer for Cool and Gentle Striker!...", 'RT @juventusfcen: Juventus, the club with the most #FIFAWorldCup winners! 🏆\n\nFind out who all our 27 champions are 🌟👤', 'RT @Crypto_DeGenz: Find the Missing Ball is back!!!\n\nOnly interesting content this football season also an opportunity to win cool prizes...', 'RT @bts_bighit: "Look who we are, we are the dreamers\nWe make it happen, cause we believe it"\n\nHuge respect & love for all the players, sta...']

[ 'Negative tweets:', 'RT @BigHayZues: Jumbotron Shit Poppin Freestyle #Drake #21savage #Herloss #viral #new #booty #FIFAWorldCup #music #rap https://t.co/prk3sTQ...', 'RT @rika_hj: 🎮 #AoVNewSkin #game #mobilegame #2022KICMENA #ClosedBeta #FIFAWorldCup', 'RT @i306_Bangtan7: Look who we are ~~\n\nI'm so addicted with #Jungkook of @BTS_twt's new single "Dreamers" for the #FIFAWorldCup\xa0\u202f\u202f\u202f\u202f\u202f\u202f\xa0\xa0 #Drea...', 'RT @SuperSportTV: Lionel Messi becomes the first player to score in every game of the #FIFAWorldCup knockout stage 🏆 https://t.co/H1j5lrCHab']
```



Stage 2

- Twitter API/Twint
- EDA
 1. Most Common Words
 2. Word Cloud
 3. Sample Tweets
- Tweets Classification
- +/- Tweets Percentage

	Postive	Negative
#TedCruz	30.4 %	13.9 %
#FIFAWorldCup	17.6 %	19.6 %

Project Stages

Stage 1

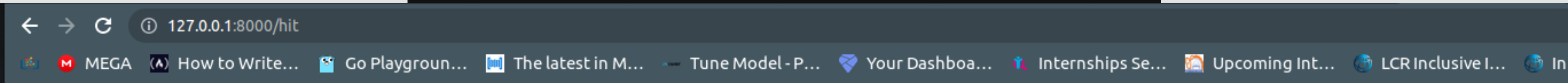
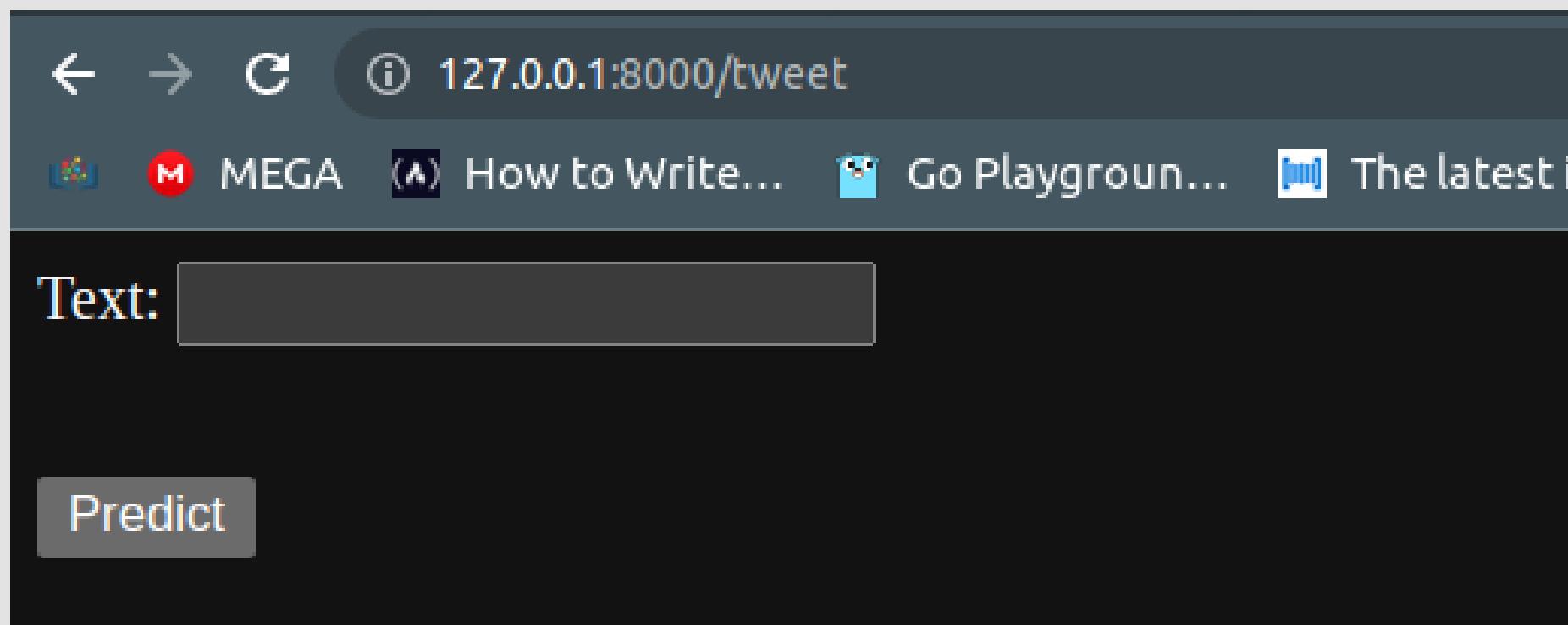
- EDA
- Cleaning Data
- Tokenizing
- Model Selection
 1. LSTM
 2. BERT
- Training Model
- Metrics Analysis
- Evaluation

Stage 2

- Twitter API/Twint
- EDA
 1. Most Common Words
 2. Word Cloud
 3. Sample Tweets
- Tweets Classification
- +/- Tweets Percentage

Stage 3

- Loading Model
- Twitter API
- Application
 1. Deephaven
 2. Heroku
 3. Django
- Future Work



Positive tweets percentage: 30.434782608695652 %

Negative tweets percentage: 15.942028985507246%

Positive tweets:

RT @solomongrundy6: Ya I see Ted Cruz looks like John Hinckley Jr. ready to strike! Impeach #JackSmithDOJ Anderson Cooper Gaetz #Boebert #M...

RT @johnmgillis: America you kindly took @tedcruz from us in 🇨🇦 . Can we interest you in a @jordanbpeterson ? Both crusty and don't play...

The Most Unpopular Senator Is Exactly Who You Would Think | Patriot Edition <https://t.co/XeCmOSH0gB...> <https://t.co/eBqCfbLdMp>

@HalSparks Look on the positive side! You can vote #tedcruz out of power next year! 👍😊 <https://t.co/DOcywTfFDW>

RT @maramcewin: The rise in Covid cases, in a country that has access to vaccines and masks, proves once again the U.S. is in the deadly gr...

Negative tweets:

@SenTedCruz #TedCruz (insert action verb here)...blah, blah, blah. Truth is he does nothing except create a false per... <https://t.co/6l6iGOZw7M>

@GalenMicheal actually not #tedcruz...,HE JUST FOLLOWS THE MORE EXTREME ELEMENTS OF #REPUBLICANS like #marjorieTRAITORgreene #FAIL

@KariLake Can you say PAPER BALLOTS? Duh! @karilake,#tedcruz,#huckabee, <https://t.co/j3pHVUW1DD>


You and all the other scumbag politicians! #TedCruz #RepublicansAreDestroyingAmerica <https://t.co/NOIGhLkKhW>

Ted Cruz Deserves ALL Of The Hate He Gets <https://t.co/FrmjhhAfc7> via @YouTube #CancunCruz #TedCruzSucks #KremlinCruz #TedCruz



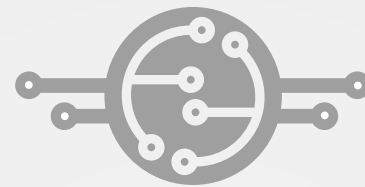
Stage 3

- Loading Model
- Twitter API
- Application
 1. Deephaven
 2. Heroku
 3. Django
- Future Work

- Testing BERT Large Model
 - Hyperparameter Tuning
 - Consider Count of Punctuation, Count all CAPS words, etc.
- 



January 2023



THANK YOU

