

NMIMS

DISASTER TWEETS REAL-OR-NOT

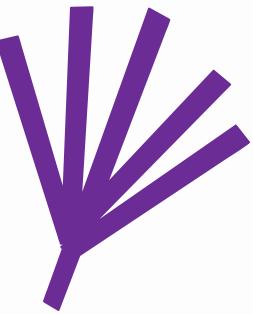
I032- Brshank Singh Negi

I033- Vishan Oberoi



TOPIC= DISASTER FROM TWITTER TWEETS

CLASSIFICATION



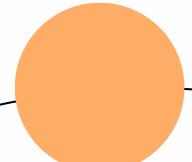
Twitter is a great source for timely information during a disaster

BUT

It can also be a source of misinformation

Target:1== Disaster

Target:0== NOT disaster



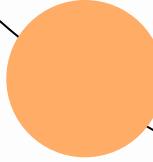


INTRODUCTION

One of the most crucial ideas in data analysis has always been text analysis. In addition to its numerous business-to-business applications, it is widely employed by several Fortune 500 businesses to comprehend consumer opinions.

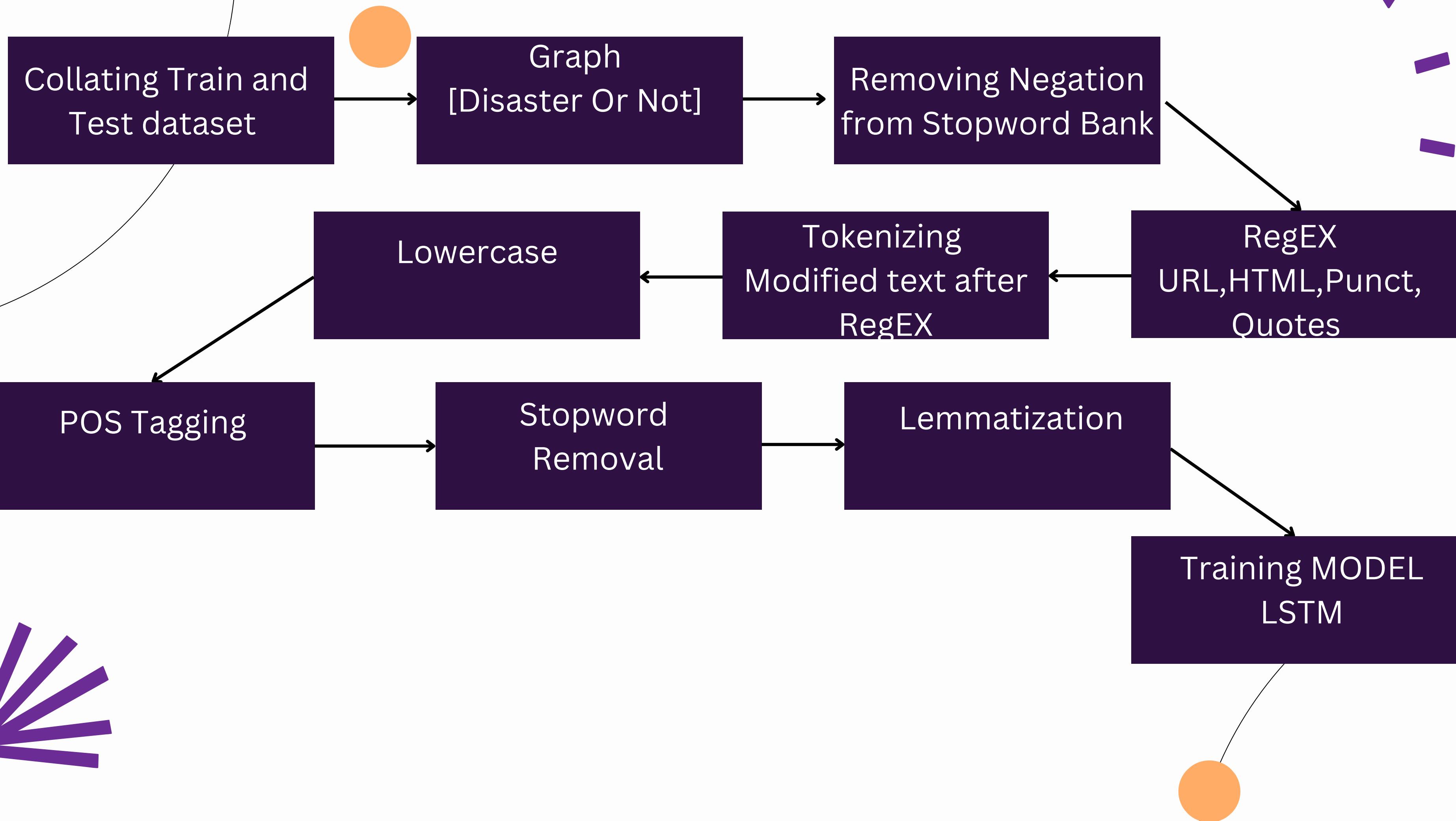
We have adopted a strategy to address the Kaggle-hosted Disaster Tweet problem statement.





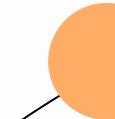
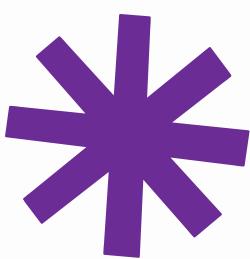
PIPELINE FLOW

- Step -1: Importing Necessary Libraries and Datasets**
 - Step -2: Data Pre-processing**
 - Step -3: Binary Classification using LSTM(DL)**
- 

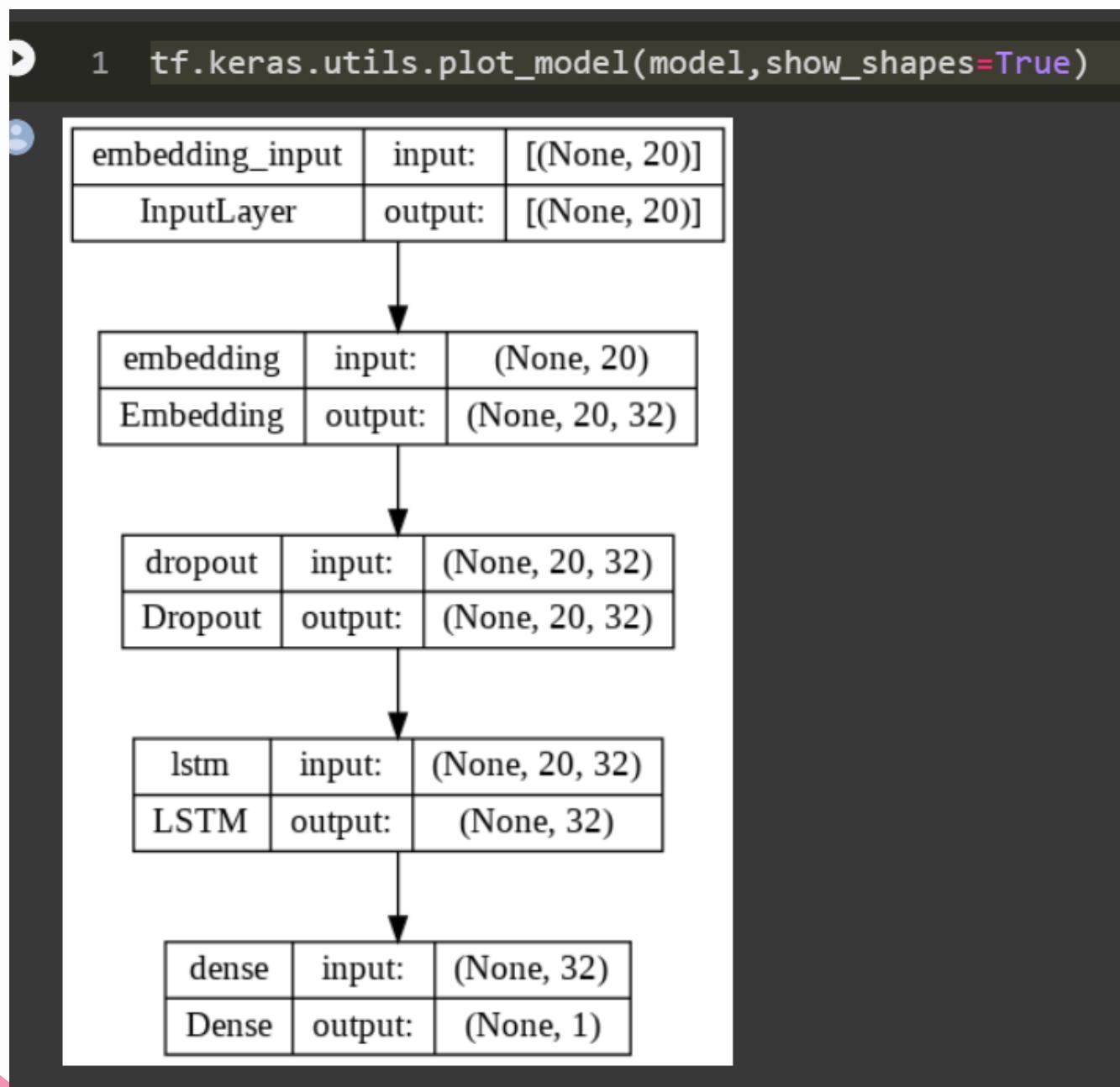


NOW ON
TO THE
CODE

mm



RESULTS



Cannot connect to TPU backend

You cannot currently connect to a TPU due to usage limits in Colab. [Learn more](#)

To get more access to TPUs, consider purchasing Colab compute units with [Pay As You Go](#).

[Close](#)

[Connect without TPU](#)

Cannot connect to GPU backend

You cannot currently connect to a GPU due to usage limits in Colab. [Learn more](#)

To get more access to GPUs, consider purchasing Colab compute units with [Pay As You Go](#).

[Close](#)

[Connect without GPU](#)

RESULTS

Best observed learning rate was 0.001

Training accuracy without callbacks=92.53%

with callbacks =96.37%

PROBELMS FACED

- Combining both test and train datasets for merged dataset creation
- Removing not stop words
- Removing emojis
- WordnetTagger
- Slow GPU
- Deciding the maximum features
-

THANK YOU FOR YOUR PATIENCE

I032- Brshank Singh Negi

I033- Vishan Oberoi

SOURCES

<https://www.kaggle.com/competitions/nlp-getting-started>

https://www.tensorflow.org/api_docs/python/tf/keras

<https://www.javatpoint.com/precision-and-recall-in-machine-learning>

<https://stackoverflow.com/questions/61760508/how-to-choose-num-words-parameter-for-keras-tokenizer>

<https://stackoverflow.com/questions/42943291/what-does-keras-io-preprocessing-sequence-pad-sequences-do>

<https://stackoverflow.com/questions/71787993/keras-pad-sequence-and-tokenizer>

<https://www.kaggle.com/code/kmkarakaya/lstm-output-types-return-sequences-state/notebook>

Dropout paper : <https://arxiv.org/pdf/1512.05287.pdf>

[**Link to Notebook**](#)

https://colab.research.google.com/drive/1v_zGr9hBDVY-hY6plGPHEQtaVOioFn8#scrollTo=1bPlVJjVBdEB