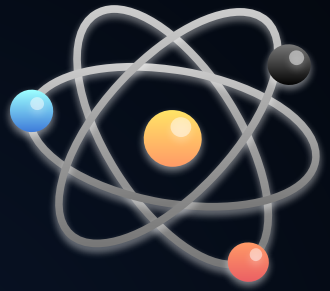
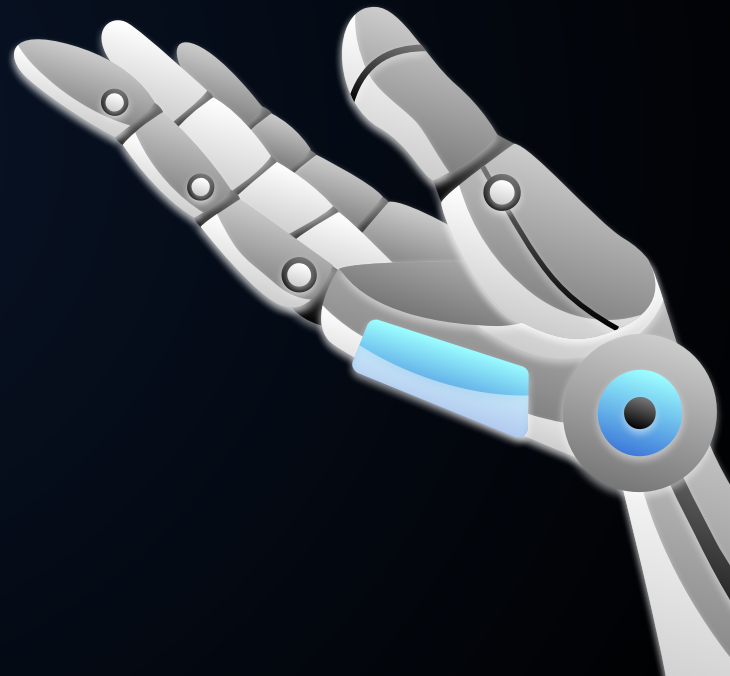
Glimpse into Twitter's Future





Team Outliers

Nanda Krishna

Dewang Vinchhi

Prayash Panda

Shishir Tripathi

INTRODUCTION

- All thanks to Elon Musk, Twitter is now the hot topic of today's World. Not only them, but we also like to be updated about other famous Celebrities
- We aim to predict how the future tweets of your favourite twitterati/celebrity might look like.
- To achieve these tasks, we needed to collect the recent tweets of different twitter users, to gain understanding of things they generally talk about.
- Using these tweets to develop a tweet generator model which predicts them depending on the user you select.
- We also tried to get the sentiment of the tweets posted by celebs and get the correct polarity of the sentence

Part 1: Future Tweet Generator

01

Data Collection

Used SNscraper to get 1000 most recent tweets for 500+ users.

02

Preprocessing Tweets

Filtered the data to remove URLs, emoticons and Non-english tweets

03

Model Pipeline

Created a pipeline for tokenization, dataset and fine-tuning of the Model

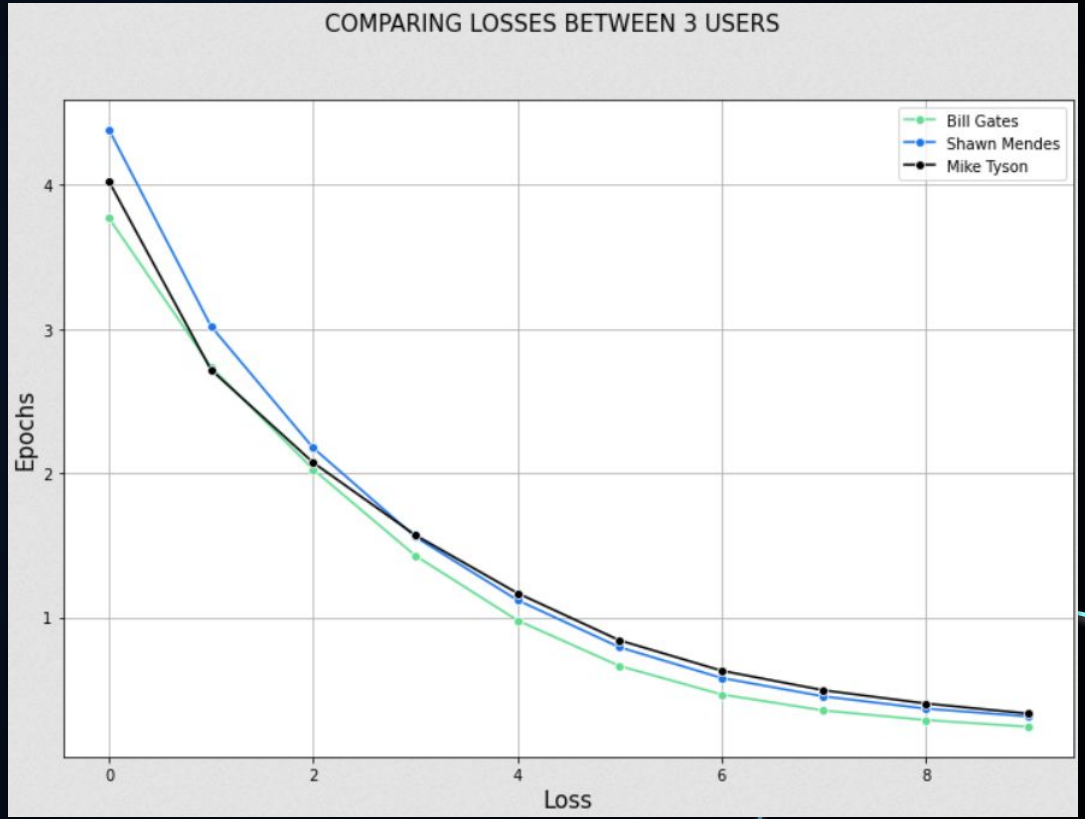
04

Tweet Generation

Finally, tweets were generated based on a user and given prompt

Evaluating Model Losses

Models were trained for 10 epochs each with Early Stopping callbacks monitoring loss with a patience of 2.



A look at the Generated Tweets

Bill Gates

I want to build safe worlds for all people not just the lucky few who get their genes to play a role in a healthy productive and equitable world i hope others in positions of great wealth and privilege will step up in this moment too

Shawn Mendes

I would like to create a more sustainably sourced and merch line for the Summer of love thank you so much.

Mike Tyson

My life is fight tomorrow 9am pst only on fight nextday rosenbergradio vs. bettymwhite mgmgrand tickets

Part 2: Language Model and Text Classification

01

Tweets Generation

Generated about 170k tweets for 10 twitter handles

02

Preprocessing

Removed all the non-alphanumeric characters and clean the tweets

03

Language Model

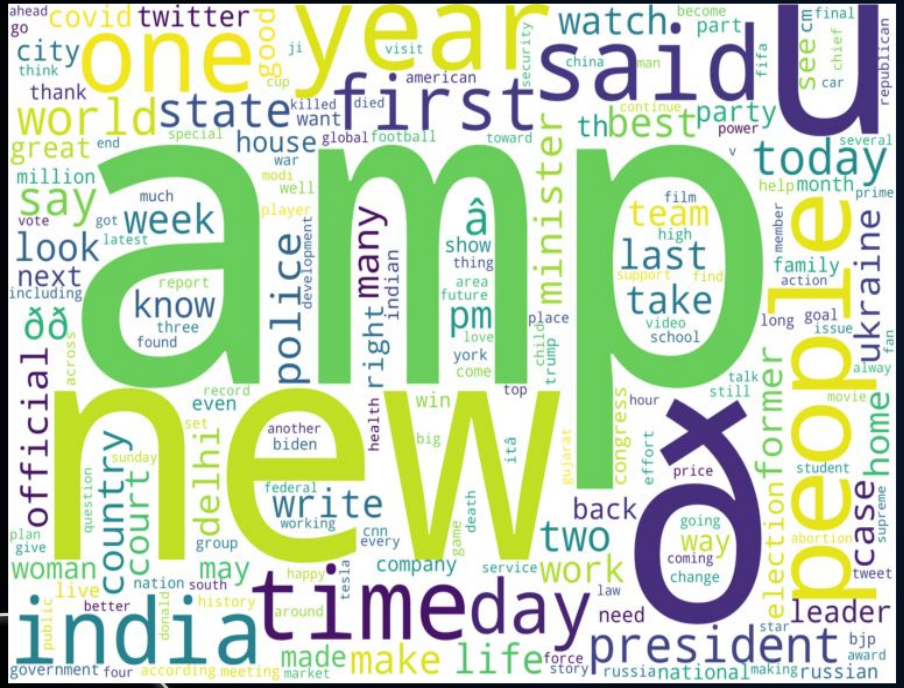
Using ULMFiT, created a Language Model.

04

Text Classification

Based on the Language model, created a Text classification for Polarity analysis

Model Selection

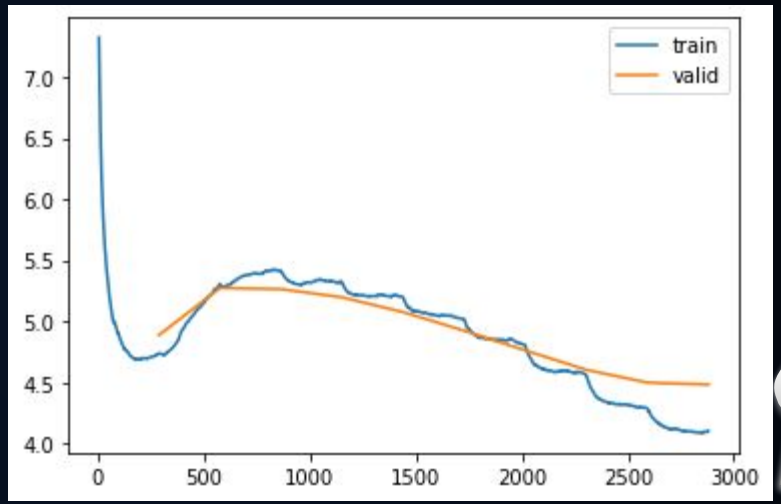


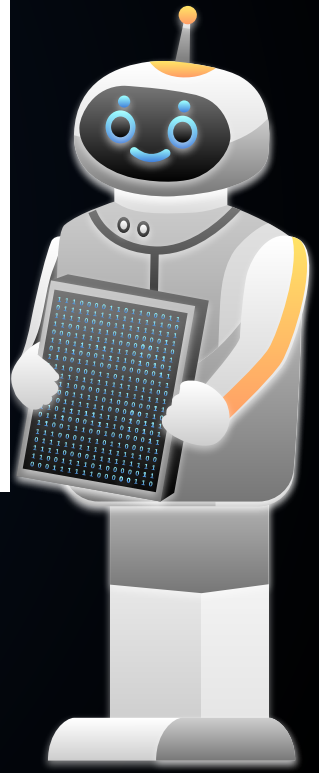
- For the Text Classification, we used ULMFiT which involves 3 processes:
1. Pretrained General Language Model
2. Fine-tuned language model
3. Fine-tuning the classifier

- The model used for this is AWD LSTM

Observations

- We were able to successfully generate a text classification model.
- The initial encoder model (language model) accuracy is ~28%
- The final model accuracy for text classification is ~90%





Models : Limitation And Biases

Tweet Generator

- The original tweets can contain phrases and slangs that are native to users from certain regions and can't be incorporated with the fixed vocabulary.
- The expressions conveyed by emoticons are not taken into the model as well as words with repeated end characters for exaggerating emotions.
- It would also contain GPT-2 model's original biases towards certain topics and sections of society.

Text Classifier Model

- The model could be fine tuned further more and even better accuracy can be generated.
- We could try with different model architecture other than AWD LSTM, like Transformer

Conclusions and Future Scope

Tweet Generation

- Fine-tuned GPT-2-medium model with 355M parameters on our tweets from 400 + users.
- Finally, generated tweets for selected users to access the model performance.

Text Classification

- Text classifier model accuracy - 93%
- Model is able to get the required sentiment when a test tweet is passed.

Future Scope

- More prioritised dataset (containing info about one field)
- Create Individual bots for users to generate tweets incorporating more data as well as for people in focused areas of interest

THANK YOU