Twitter Language Models about Kanye West

Christian Wall

Timeline

•2004. The College Dropout album was released with Label Roc-A-Fella.

•2005. Late Registration album was released with Label Def Jam.

•2006. Late Orchestration: Live at Abbey Road Studios album was released with Label Mercury.

•2007. Graduation album was released with Label Roc-A-Fella.

•2008. 808s & Heartbreak album was released with Label Roc-A-Fella.

•2010. My Beautiful Dark Twisted Fantasy album was released with Label Roc-A-Fella.

•2011. Watch the Throne album was released with Label Roc-A-Fella.

•2013. Yeezus album was released with Label Def Jam.

•2016. The Life of Pablo album was released with Label Getting Out Our Dreams.

•2018. Ye album was released with Label GOOD.

•2018. Kids See Ghosts album was released with Label Def Jam / Getting Out Our Dreams.

•2019. Jesus Is King album was released with Label Def Jam / Getting Out Our Dreams.

•2021. Donda album to be released soon.

KANYE QUICK FACTS

- **Kanye actually lived in China for a short period of time in his childhood.**
- **He is really a college dropout.**
- **Kanye used to produce Jay-Z's songs before starting his own musical career.**

Introduction

Kanye West has been an icon for what seems like my entire life, he is a fantastic artist that uses many different outlets to express himself. These things include but are not limited to, singing, dancing, rapping, producing, fashion production, theatrical performances, directing, etc.

With the anticipation of his new album 'Donda' dropping soon right after his fallout with his now ex-wife Kim Kardashian, it has led me to the choice of the subject topic 'Kanye' as I search through twitter and save tweets in order to use for my Language Models.

Data

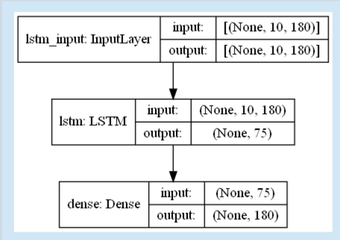
My dataset was made up of 1,000 tweets from Twitter that I gathered using Tweepy which is a module for Python which allowed me to use any keyword that I liked in order to search through tweets. I took these tweets and saved them into a text file for them to be used in both the character-based and word-based models.

```
Now playing Gold Digger by Kanye West ft. Jamie Foxx!
If I said "like Ye imma put my mom in a Mercedes" [cuz Kanye is in the
Mercedes Benz stadium playin Donda (his mom name)
Yeezus is honestly top 3 Kanye for me rn
(Mbdtf and graduation probably the other 2)
TOP 10 FAVOURITE ARTISTS (NOT RANKED)
1. BEATLES
2. KANYE
3. JPEGMAFIA
4. TYLER THE CREATOR
5. NIRVANA
6. MF DOOM
7… https://t.co/MrenZfJRDK
@HipHopNumbers Mos Def, Kendrick Lamar, Freddie Gibbs, MF DOOM, Kanye and
Tupac
@JayAree3x What happened to Kanye ? Lol
kanye needed me. he needed ME right now
they put Clifford the Big Red Dog on the Kanye album covers lmaoooo
https://t.co/l4QkEiowqT
```

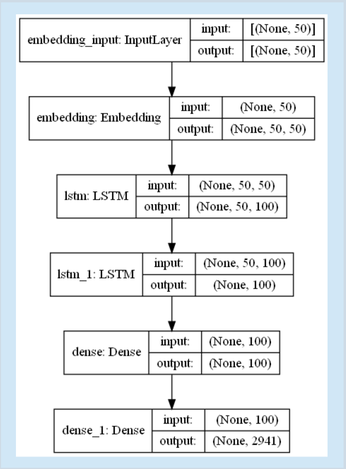
Tweets [Above]] shows some of the tweets I gathered using Tweepy

Implementation

I started generating new tweets using two different methods, a character-based language model and a word-based language model, both models allow for text generation. The problem that I dealt with the most during the implementation was dealing with 'UTF-8' encoding of files, but after that everything went smoothly. Here are the differences between the models used shown below:



Model 1 [Above] is a visualization of the character-based model's input and output



Model 2 [Above] is a visualization of the word-based model's input and output

Results

After collecting each tweet, the different models used different techniques in order to clean these tweets and gather only the data that they need in order to generate new text. They both require a starting seed text in order to generate. I generated 20 tweets for both models, here are some examples below:

```
seed text:'Kanye is'
1. Kanye is the best see @The Das Strietser anting me all there @12380Erihgo
need to Kanye West - Gollobe Younde saily be changer this is me it this all
the samm with them day Kanye should be hell me in the release kanye stinting
man me if you kanye west he likes ment if you day I'm not got
```

Example 1 [Above] is an example of a generated tweet using the character-based model

```
seed text:'Kanye is'
1.that im sensitive be a girlfriend i am so dr bs trav defina seu gosto
musical com artistasbandas sem ordem spooky bizzle conducta morandi the
smiths omg i started popular nowplaying kanye west james reid amp careless
trece dias muerto really talking of youtube tcd kayblack part for the
httpstcomviubykoxp
```

Example 2 [Above]] is an example of a generated tweet using the word-based model with the same seed text as Example 1

Conclusion

When producing tweets using both methods, it is unclear which method is exactly better than the other because there is not really a way of telling what a better tweet is. But the character-based model had a better Epoch accuracy at 67% while the word-based model was at 55%. They are both surprisingly more cohesive grammatically then I thought they would be. I assume that with even more tweets than just 1,000 that the accuracy and cohesiveness of the models would be even better.

Research

[1] Zambas, J (2018, January 28). Kanye West – 20 Crazy Facts about the Rapper. https://www.telltalesonline.com/28195/kanye-west-facts.

[2] Keras-Team (2015). Releases Keras. Github. https://github.com/keras-team/keras/releases.