

Building a Classifier to Predict Triggering or Uncourteous Speech

4/30/2024 DTSA-5511 – Introduction to Deep Learning 2024

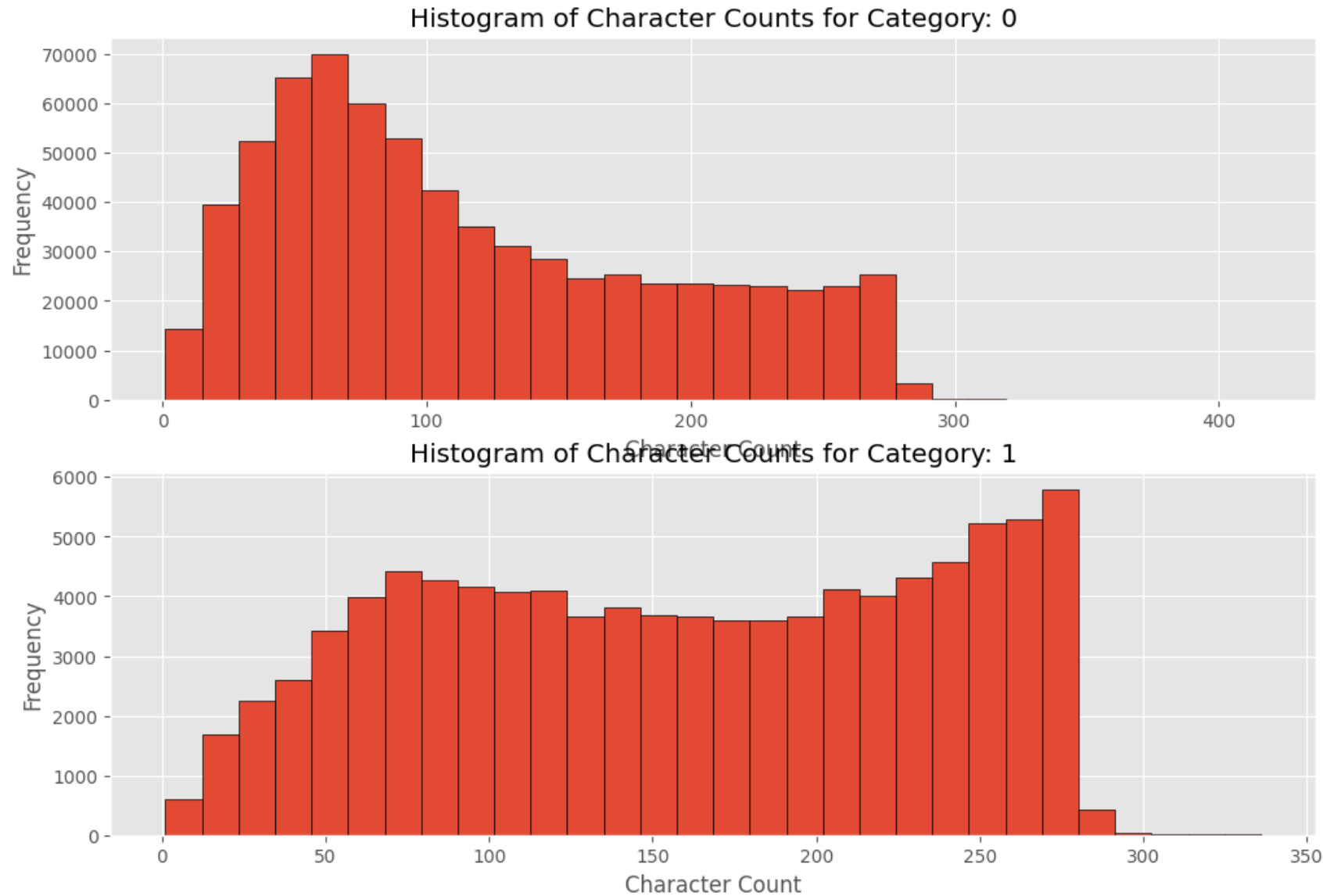
Dataset

- Total 807,174 records in total (~701,412 Non Offensive vs. 105762 Offensive)
- Twitter dataset from 2018 curated during the #MeToo movement
 - * **text** - Includes the text tweet (you can see these tweets above)
 - * **created_at** - day the tweet was created
 - * **favorite_count** - how many people have liked the tweet
 - * **retweet_count** - how many people retweeted the tweet
 - * **location** - location from where the tweet was placed
 - * **follower_count** - account follower count
 - * **friends_count** - account following count
 - * **Statuses_count** - account total tweets created
 - * **Category** - is it a mean tweet
- Will only be using the text and category columns in this analysis

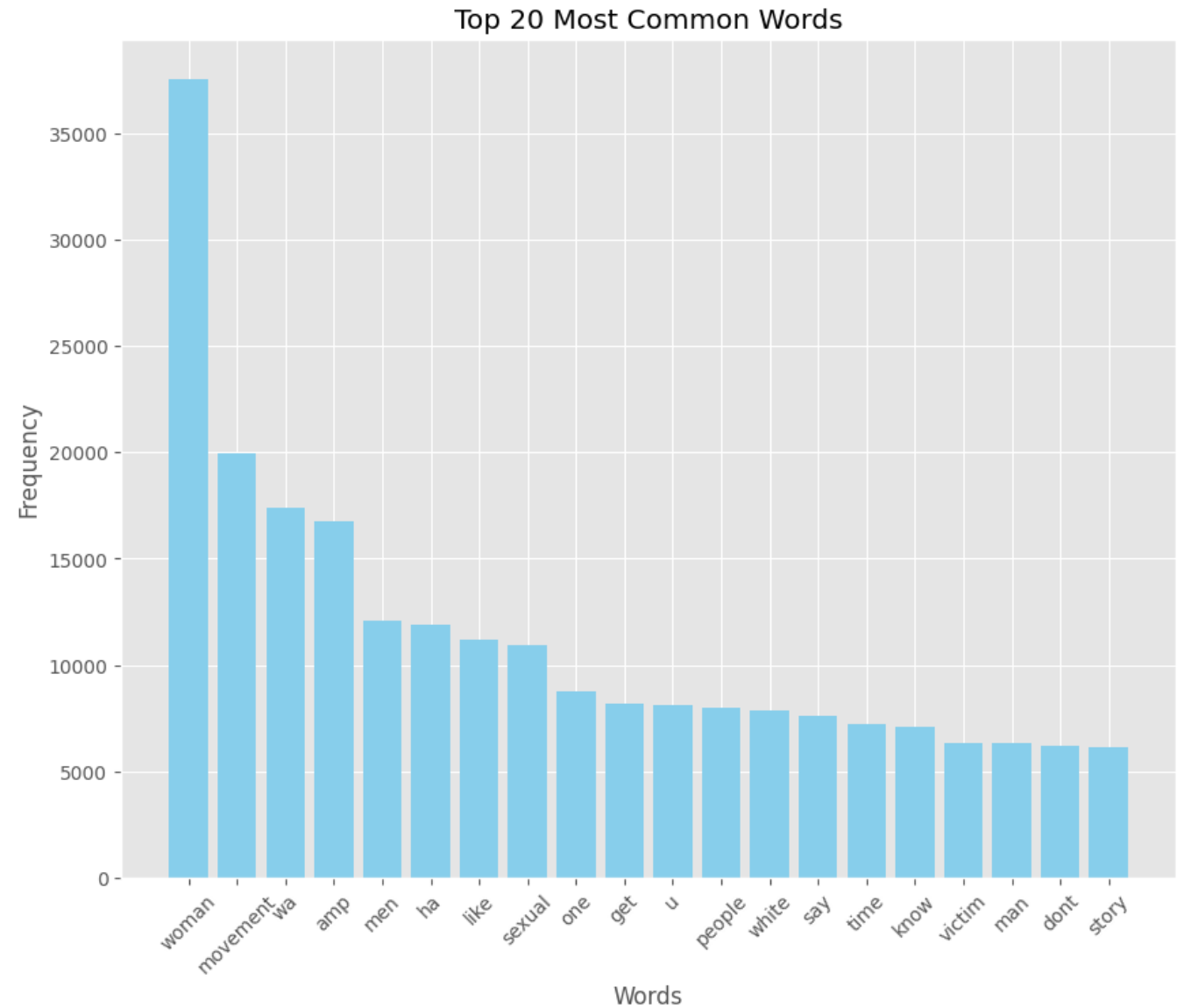
Examples of the Tweets in Dataset

- Yep just like triffeling women weaponized their poon!! Wonder if Kamala Harris ever extorted Willy Brown after throwing the poon on him, oh yeh, that how she got her first job me too is a JOKE!
(Problematic)
- Save it for Fox Fake News—watched by racists, bigots & other women-hating sexists like yourself. GOP is the party of Putin & the party of Hate—besides women, GOP hates all POC, immigrants, and all ppl not straight. You're the party of the entitled lying whiny white men.
- As a white male in this era I'm just scared that if I decide to rape or sexually assault a woman, it can be used against me years later. How am I supposed to ask a girl out?!
- I heard that received some Shush/Slush money to settle rape charges by her aide who said she just could not remember where or when. . .
- If she hasn't, it's because of people like you shaming her! How dare you tell her how she is or should be feeling! That is why I HATE the movement. Along with all of the other movements! You rape these people to further a political objective! You are despicable!
- Oh poor girl. You should have stayed disney star all your life. Because now you are being eaten by the industry. And in few years when your popularity falls, you will come with some bs like other forgotten women in hollywood.
- When was just gaining momentum, a female friend of mine who was pregnant with her daughter said "what's the big deal? I've been raped". My response was " You know I disagree with you 🤔 ". How do you respond? I was shocked!
- She sounds like one horny bitch! Why it is always men who get blamed and framed for rapes? Am not talking about REAL rapes here but those that are hoaxes like Ford's! Women can rape men too! Just look at the sick broads! How many look like normal, healthy & happy women

Tweets that are uncourteous speech are slightly longer in length



Top 20 Most Common Words



[illegible]

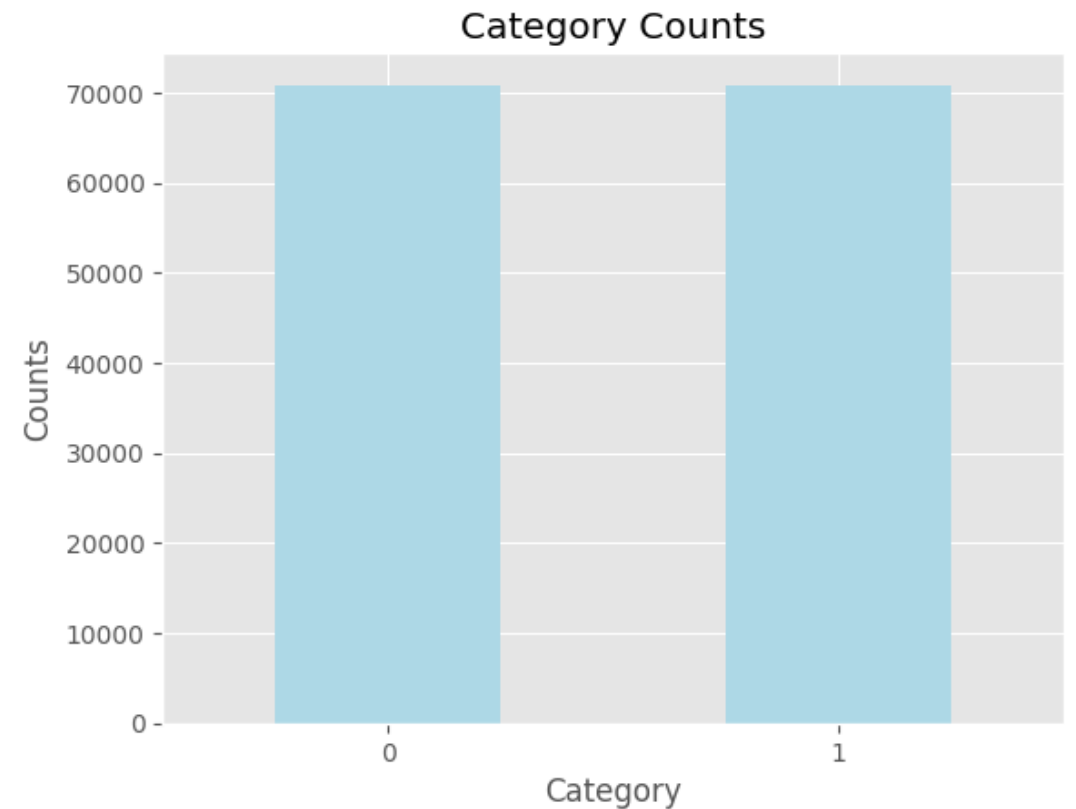
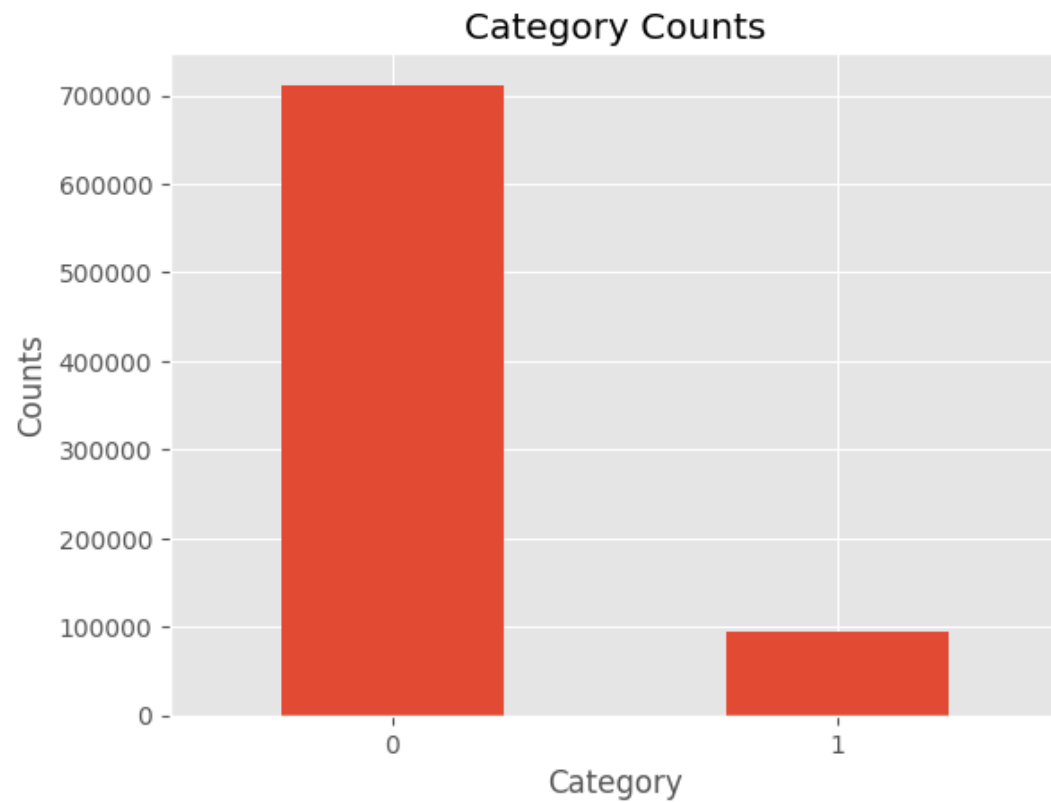
Data Cleansing

- Removed URLs and HTML tags from the tweets
- Removed all emoji characters from text
- Removed punctuation
- Split the words out with a lemmatizer

Data Preprocessing

- After the data cleansing and created a corpus of words I needed to create an encoder
- Split data into training and test datasets
- Used GloVe model to encode data
 - an unsupervised learning algorithm developed by researchers at Stanford for generating word embeddings by aggregating global word-word co-occurrence statistics from a corpus.
- Used a random under sampling approach to help balance out the dataset to ensure that the results given will not undervalue the scores (took the 800K to 140K)
- Split data and used around 20,000 tweets to build the models

Data Imbalance

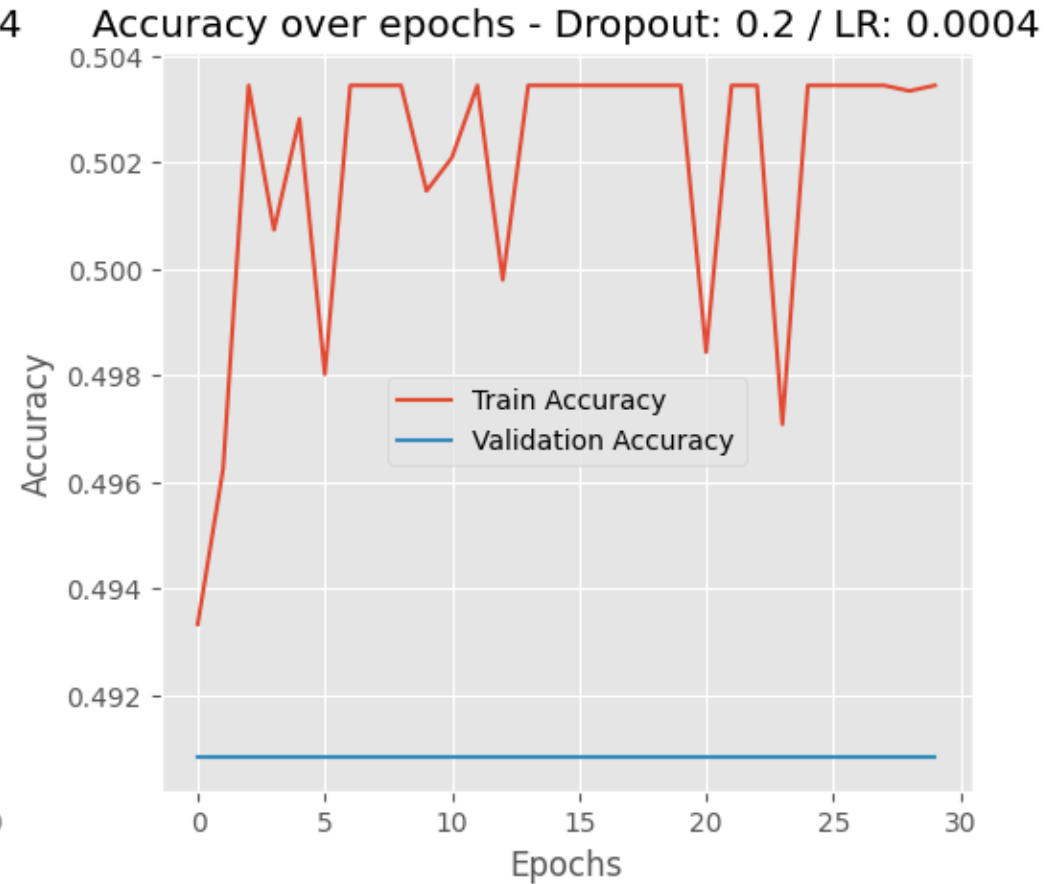
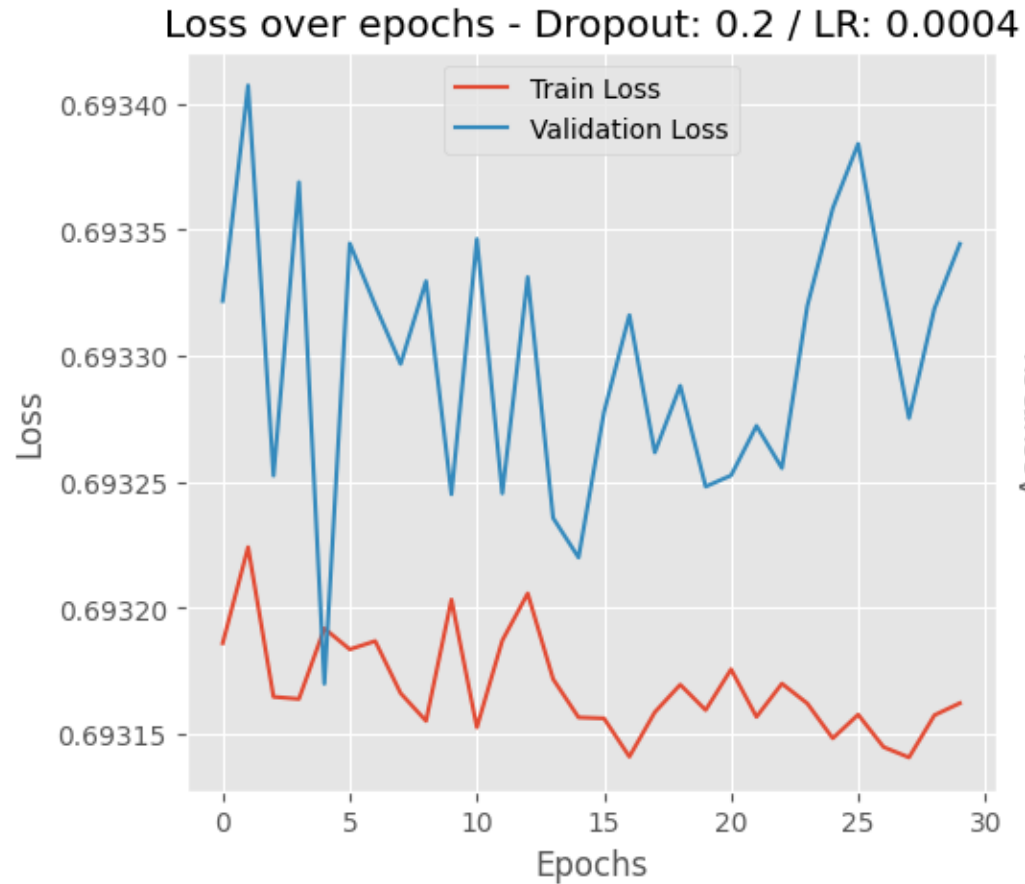


Used a random under sampling approach to help balance out the dataset to ensure that the results given will not undervalue the scores

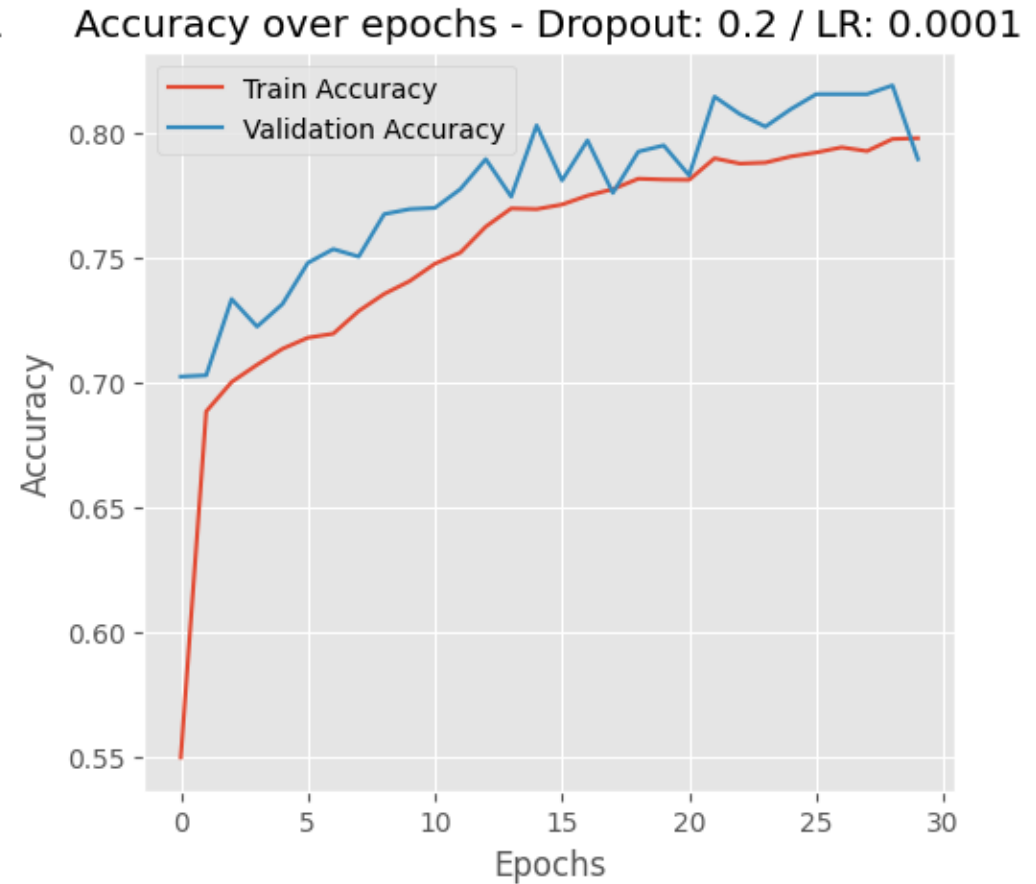
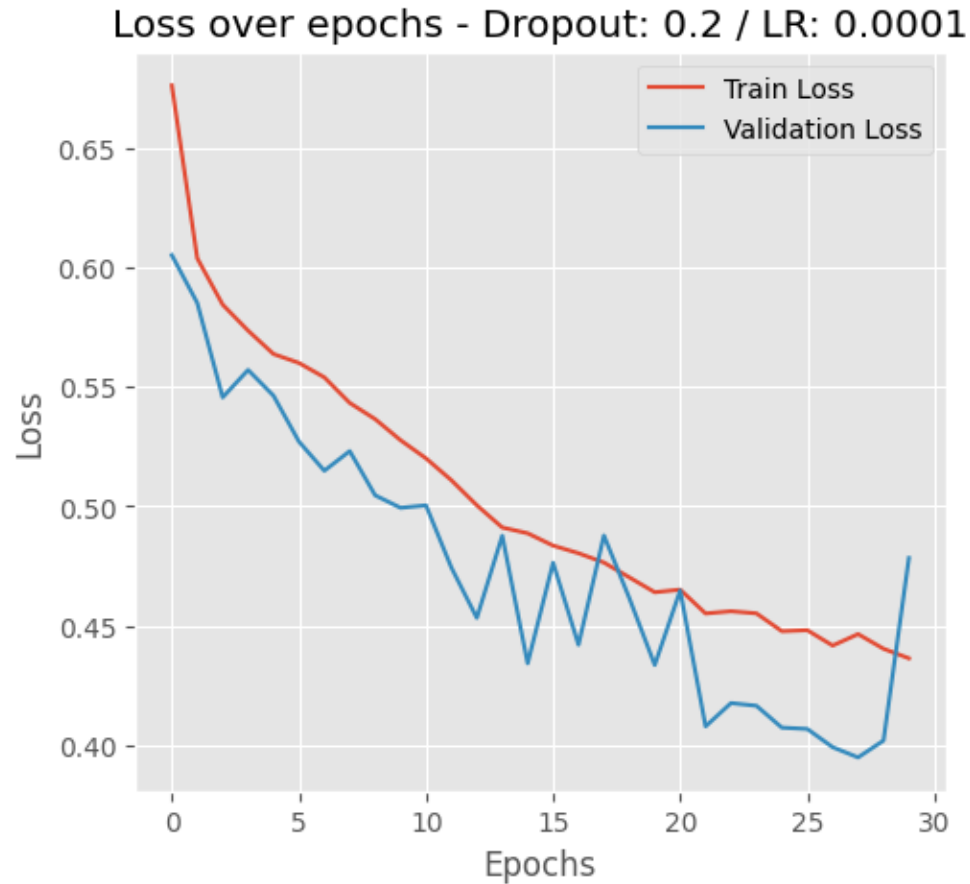
Long Short Term Memory (LSTM)

- 1.LSTMs address the vanishing gradient problem in traditional RNNs through a unique architecture that includes memory cells and three types of gates: input, output, and forget.
- 2.They are capable of learning and retaining long-term dependencies in sequence data, which is crucial for complex sequence prediction tasks.
- 3.LSTMs are extensively used across a broad range of deep learning applications, including language translation, speech recognition, and time series forecasting.
- 4.Due to their robustness in handling long sequences and complex patterns, LSTMs often require more computational resources and training time compared to simpler recurrent architectures like GRUs.

LSTM Model (lr = 0.0004, epoch – 30, batch=32) w/ Stop Words



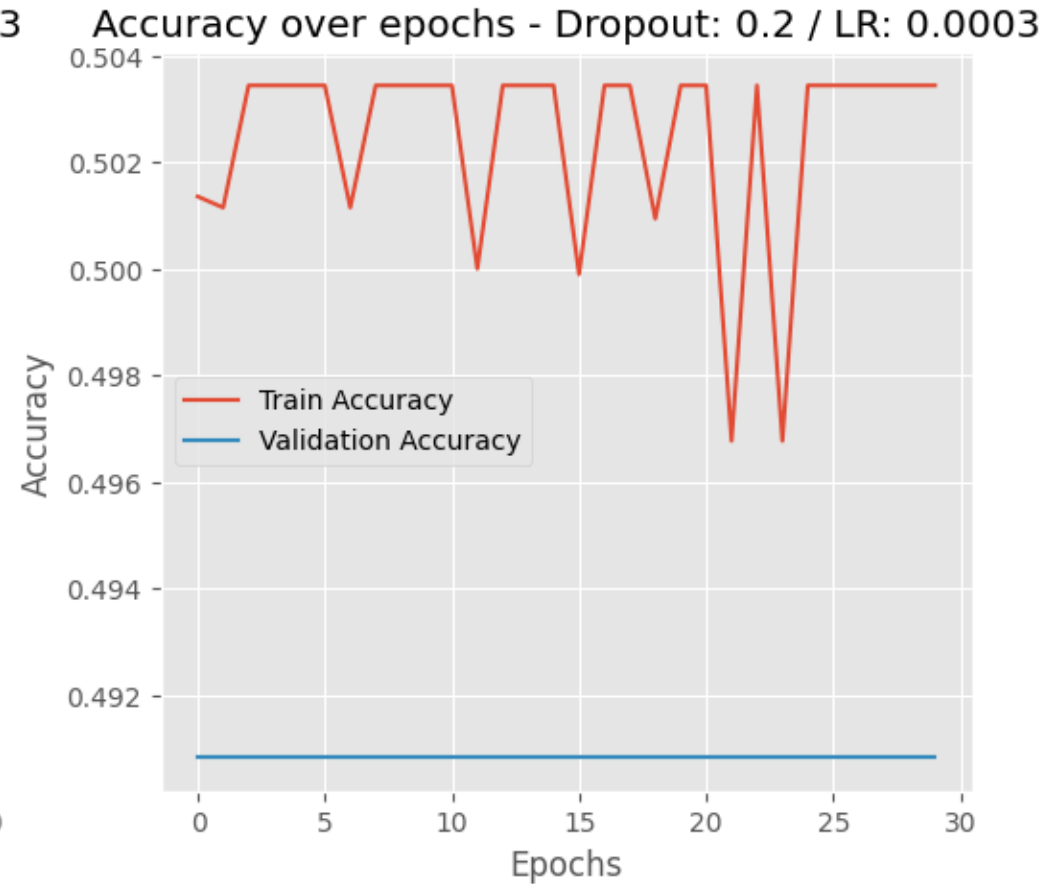
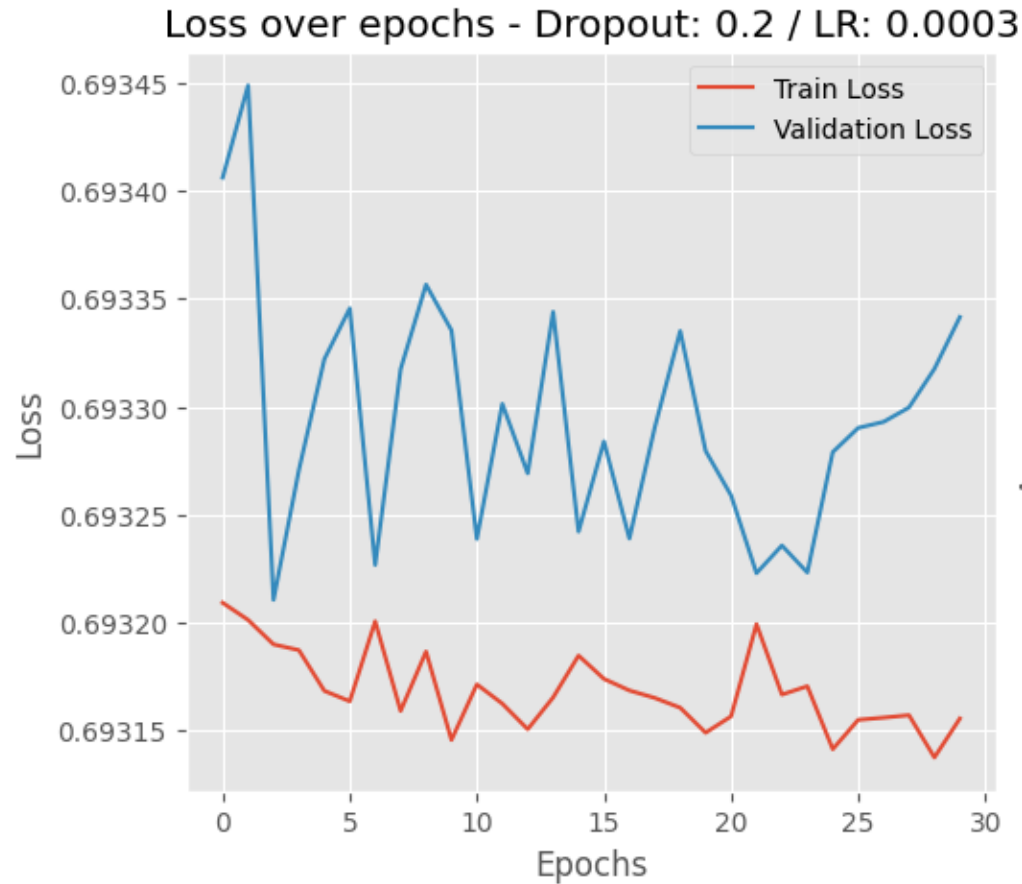
LSTM Model (lr = 0.0001, epoch – 30, batch=32) w/ Stop Words



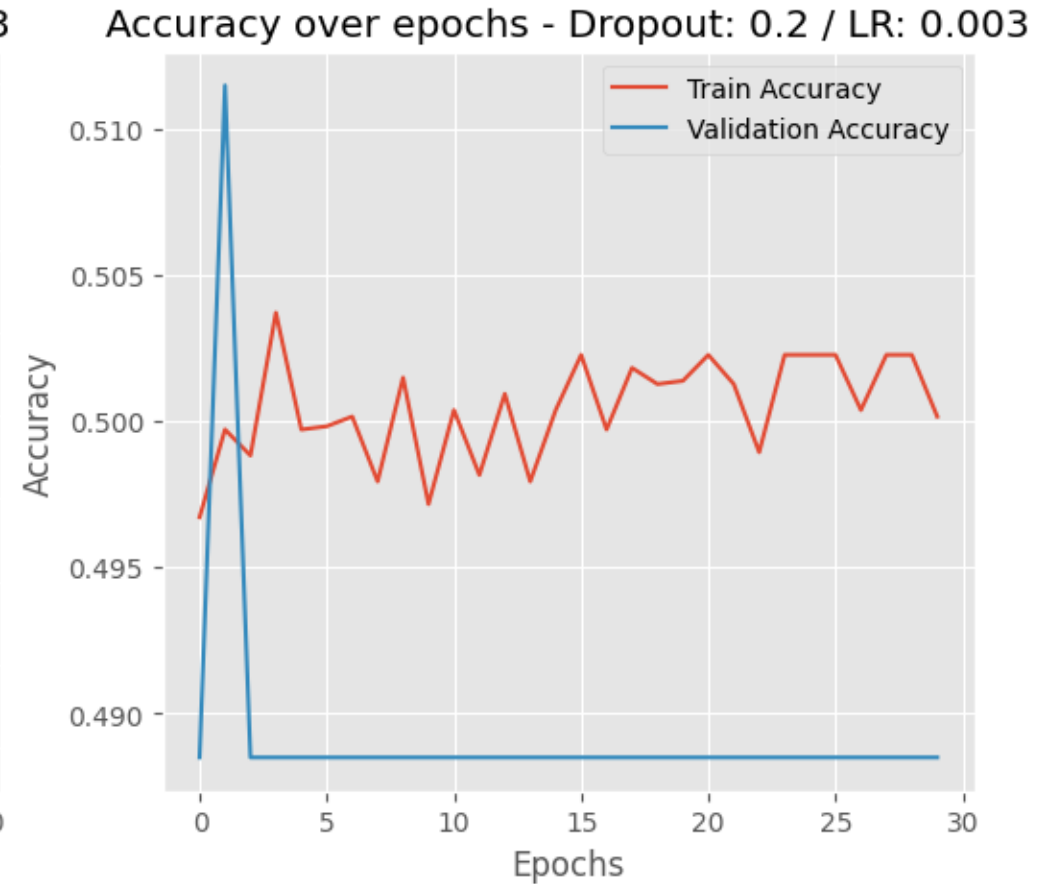
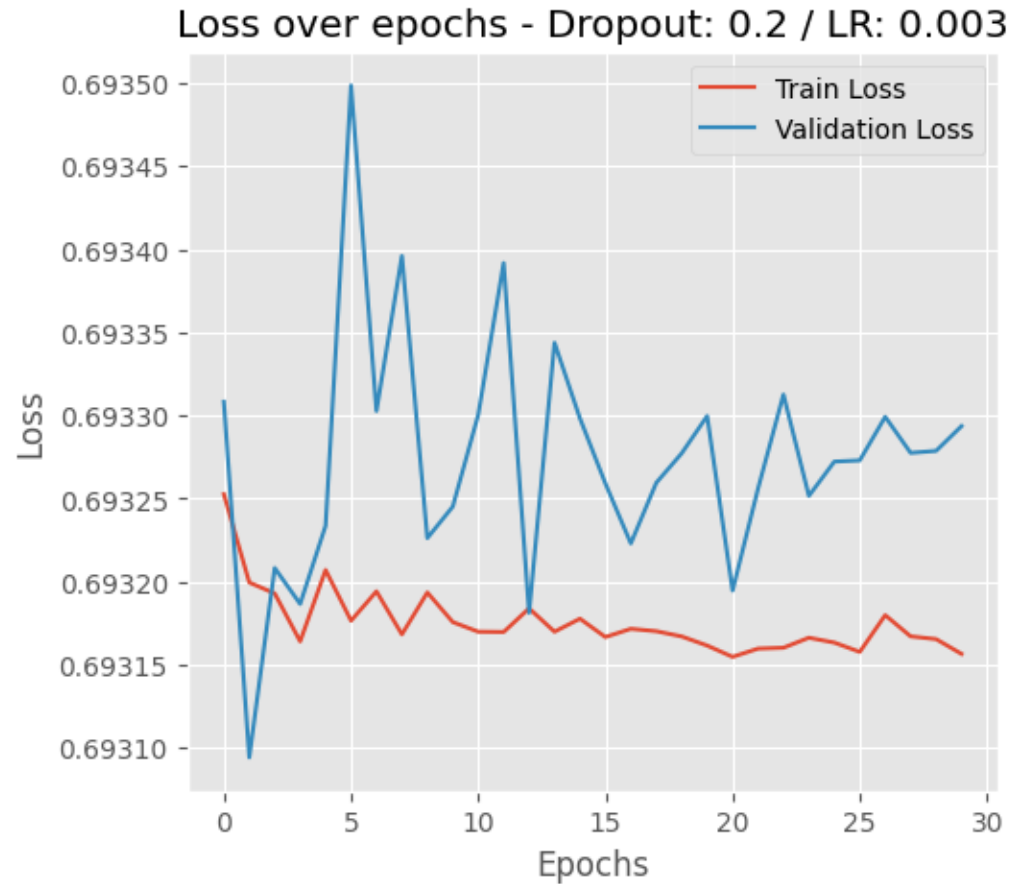
Gated Recurrent Unit (GRU)

1. GRUs simplify traditional RNN architecture by merging the forget and input gates into a single "update gate," reducing complexity and training time.
2. They efficiently learn dependencies from sequences and manage memory effectively through a unique gating mechanism that mitigates the vanishing gradient problem.
3. GRUs are widely used in deep learning applications like speech recognition, language modeling, and time series analysis due to their efficiency and effectiveness.
4. Despite their simpler structure compared to LSTMs, GRUs often achieve competitive performance in many tasks with faster training and fewer parameters.

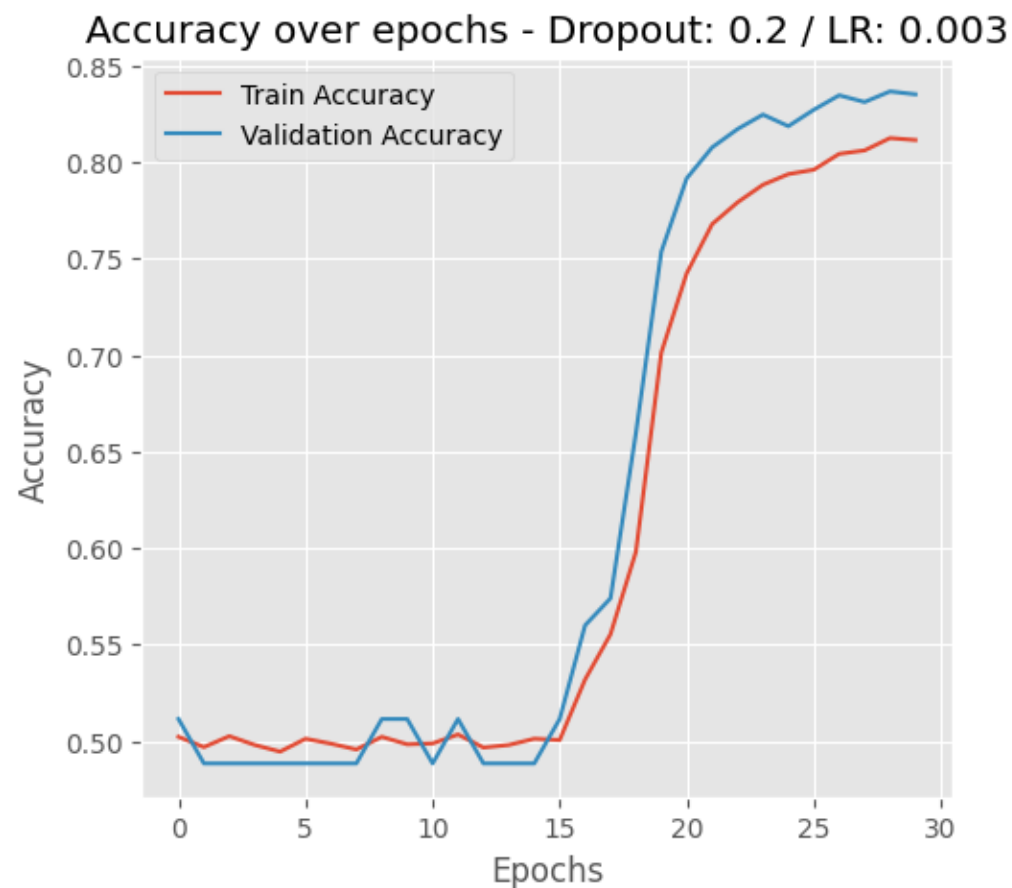
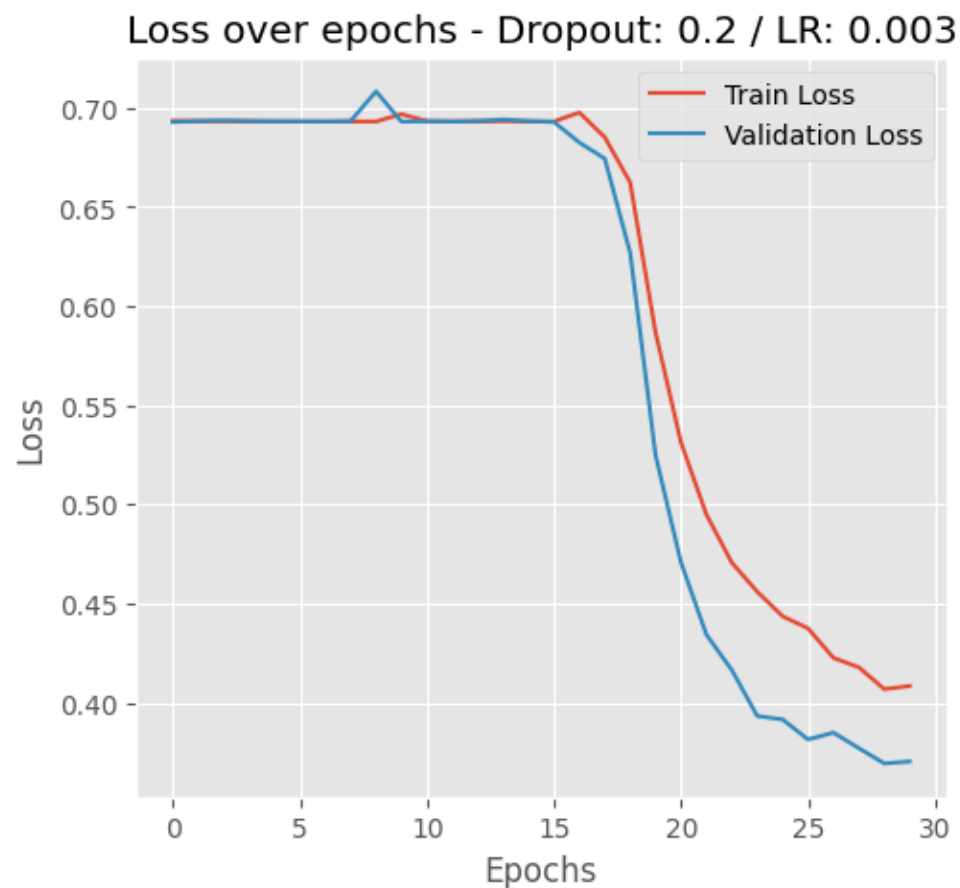
GRU Model (lr = 0.003, dropout=0.2, epoch=30) w/ Stop Words



GRU Model (lr = 0.001, dropout=0.2, epoch=30, batch=32) w/o Stop Words



GRU Model (lr = 0.003, epoch=30, batch=32) w/o Stop Words



Results

| Model | Learning Rate | Epoch | Dropout | Batch | Validation Accuracy |
|---------------------------|----------------|-----------|------------|-----------|---------------------|
| LSTM w/ Stop Words | 0.0001 | 30 | 0.2 | 32 | 0.819 |
| LSTM w/ Stop Words | 0.0004 | 30 | 0.2 | 32 | 0.4908 |
| LSTM w/ Stop Words | 0.0003 | 30 | 0.2 | 64 | 0.4908 |
| LSTM w/o Stop Words | 0.00003 | 30 | 0.2 | 32 | 0.7340 |
| GRU w/ Stop Words | 0.00002 | 4 | 0.2 | 15 | 0.4885 |
| GRU w/ Stop Words | 0.00002 | 4 | 0.1 | 15 | 0.4885 |
| GRU w/ Stop Words | 0.0004 | 4 | 0.2 | 30 | 0.4885 |
| GRU w/ Stop Words | 0.00001 | 60 | 0.2 | 64 | 0.4885 |
| GRU w/o Stop Words | 0.00001 | 30 | 0.3 | 32 | 0.7780 |

Predicted Tweet Results

- category: 0 - Entitled obnoxious defensive lying weasel This thing is to make decisions that may affect my life He is a women hater pred - [0.39418933]
- category: 0 - Thank you and for what you did for the women and survivors this week pred - [0.33509126]
- category: 0 - Knitting s amp getting ready for January 19 2019 pred - [0.31650913]
- category: 1 - Yep just like triffeling women weaponized their poon Wonder if Kamala Harris ever extorted Willy Brown after throwing the poon on him oh yeh that how she got her first job me too is a JOKE pred - [0.74771756]
- category: 0 - No the President wants to end movement posing as the movement pred - [0.2526919]
- category: 0 - Lock Justin Timberlake up he raped JANET JACKSON pred - [0.30892724]
- category: 0 - Over a decade for before it was a hashtag Earlier today I was on Albertas Legislature ground as Edmontonians spoke out at the rally

Learnings

- Teasing out the hyperparameters can be tricky (especially when trying to activate the dataset) and **time consuming**
- Better way to keep track of models
- Next steps
 - Incorporate more data to the training (Used around 20% of the total data)
 - Try other data encoders (such as Word2Vec, TikToken, etc)
 - Test other model architectures (Bidirectional RNN, Transformer networks)
 - Integrate other data (add user count, user followers, etc)