Detecting Fake News Using a LSTM Neural Network

Brandon Lasher

Hood College

401 Rosemont Ave

Frederick, MD 21701

ABSTRACT

Fake news has become a major issue in society today. With the social media dominating the way news spreads and the lackluster verification on any news article that is shared, fake news affected all aspects of society. It effects what news stories are covered; to the way government policies are talked about; to the way public campaigns are run; to even discussions around the dinner table. The standard way of fact checking to rule out fake news, now takes too long in the incautiously changing new cycle. In this paper we discuss building a LSTM model to detect if a text if fake or true. We find that if the text is similar enough to the training set, we are able to predict at a very high percentage that if the news is indeed fake or true.

Keywords

Machine Learning; Neural Networks; Natural Language Processing; Data Analysis; Computer Science.

1. INTRODUCTION

In the modern era, with the growing use and addictions to, social media has become a primary source of information for the majority of people. In fact, 62 percent of US adults get news on social media [1]. With social media's ease of information sharing, combined with lack of validation has allowed misinformation and fake news to spread unchecked. This has called into the question the effectiveness of the press and drastically expanded the reach of extremely partisan views. The "echo chambers" created by the social media algorithms exacerbated the problem, where it they would simply show their user more things that were likely to get shares or clicks. This is supported by the facts, the most popular fake news stories were more widely shared and many people who see fake news stories report that they believe them[2]. During the 2016 elections, the section of time which this study focusses, discussed fake news stories tended to favor Donald Trump over Hillary Clinton. That is also reflected in the data analysis section of this paper.

The effect to combat fake news has two approaches. First is the conventional approach of researching and verifying facts with various sources. This is both time and resource intensive, and in the fast-paced information age, where a news article appears on the internet seconds after the even, it is impractical. The second approach, is what the project explores, is using a machine learning model to automate the detection of fake news. This approach will use a Long short-term memory (LSTM) neural network to build a model which can be applied to detect with a high certainty if an article is fake news.

2. Approach

In this section, we present details of our approach and the framework which define it. A dataset from a Kaggle notebook [3]. will be the primary data set which we will base our model. This dataset has been around for a while, and because it deals with a very well-defined issue, there are many notebooks which attempt to build models to detect fake news. This particular notebook explores the bias of the dataset. The process we will take will be first to clean the data. We will remove any items from the text which might contribute noise to the model. Next, we will analyze the data. We will do this by looking at various features of the data as well as deriving new features based on publicly available natural language processing (NLP) models. Lastly, we will take the step necessary to build an LSTM to make our predictions. As an extra task, we will then use a second data set with information from around that same time period to test our model.

2.1 The Data

The dataset we will use can be found on Kaggle[3]. This data consists of two csv files which denote a set of fake and true news article and social media posts. Fake.csv consists of 23481 entries and True.csv consists of 21417 entries. It is laid out as follows:

Feature Name	Data Type	Description
title	object	Title of the article
text	object	Text of article, may include publisher and author
subject	object	Type of News
date	date	Date of publication

2.1.1 Data Cleaning

To optimize the learning and make the most of out the data provided we need to do our best to remove as much erroneous information as possible. Because a large portion of both datasets come form social media, so there is much to clean.

The first step to clean the data is to combine the title and the text fields to a new field called 'news'. This way when cleaning and training we will be using all the text information available about the entry. We then create and apply a function the 'news' field to clean the data which does the following.

- Removes tags, urls, and html

- Remove any additional punctuation.

- Remove any numbers

- Make all text lowercase

- Remove any stopwords

- Combine Fake and True data sets into one, but addition an additional classification column called 'type'

Stopwords are words such as 'the', 'a', 'in', etc which are commonly used but do not give much contextual meaning to the text. The removal of stopwords had a fairly large effect on the size of entries and will allow for better processing.

Prior to cleaning example:

```
' Donald Trump Sends Out Embarrassing New Year's Eve Message; This
is Disturbing Donald Trump just couldn t wish all Americans a Happy
New Year and leave it at that. Instead, he had to give a shout out
to his enemies, haters and the very dishonest fake news media. The
former reality show star had just one job to do and he couldn t do
it. As our Country rapidly grows stronger and smarter, I want to
wish all of my friends, supporters, enemies, haters, and even the
very dishonest Fake News Media, a Happy and Healthy New Year,
President Angry Pants tweeted. 2018 will be a great year for
America! As our Country rapidly grows stronger and smarter, I want
to wish all of my friends, supporters, enemies, haters, and even the
very dishonest Fake News Media, a Happy and Healthy New Year. 2018
will be a great year for America! Donald J. Trump (@realDonaldTrump)
December 31, 2017Trump s tweet went down about as welll as you d
expect.What kind of president sends a New Year s greeting like this
despicable, petty, infantile gibberish? Only Trump! His lack of
decency won t even allow him to rise above the gutter long enough to
wish the American citizens a happy new year! Bishop Talbert Swan
(@TalbertSwan) December 31, 2017no one likes you Calvin
(@calvinstowell) December 31, 2017Your impeachment would make 2018 a
great year for America, but I ll also accept regaining control of
Congress. Miranda Yaver (@mirandayaver) December 31, 2017Do you hear
yourself talk? When you have to include that many people that hate
you you have to wonder? Why do the they all hate me? Alan Sandoval
(@AlanSandoval13) December 31, 2017Who uses the word Haters in a New
Years wish?? Marlene (@marlene399) December 31, 2017You can t just
say happy new year? Koren pollitt (@Korencarpenter) December 31,
2017Here s Trump s New Year s Eve tweet from 2016.Happy New Year to
all, including to my many enemies and those who have fought me and
lost so badly they just don t know what to do. Love! Donald J. Trump
(@realDonaldTrump) December 31, 2016This is nothing new for Trump.
He s been doing this for years.Trump has directed messages to his
enemies and haters for New Year s, Easter, Thanksgiving, and the
anniversary of 9/11. pic.twitter.com/4FPAe2KypA Daniel Dale
(@ddale8) December 31, 2017Trump s holiday tweets are clearly not
presidential.How long did he work at Hallmark before becoming
President? Steven Goodine (@SGoodine) December 31, 2017He s always
been like this . . . the only difference is that in the last few
years, his filter has been breaking down. Roy Schulze (@thbthttt)
December 31, 2017Who, apart from a teenager uses the term haters?
Wendy (@WendyWhistles) December 31, 2017he s a fucking 5 year old
Who Knows (@rainyday80) December 31, 2017So, to all the people who
voted for this a hole thinking he would change once he got into
power, you were wrong! 70-year-old men don t change and now he s a
year older.Photo by Andrew Burton/Getty Images.'
```

After Cleaning example:

```
'donald trump sends embarrassing new year's eve message disturbing
donald trump wish americans happy new year leave instead give shout
enemies haters dishonest fake news media former reality show star
one job country rapidly grows stronger smarter want wish friends
supporters enemies haters even dishonest fake news media happy
healthy new year president angry pants tweeted great year america
country rapidly grows stronger smarter want wish friends supporters
enemies haters even dishonest fake news media happy healthy new year
great year america donald j trump realdonaldtrump december tweet
went welll expect kind president sends new year greeting like
despicable petty infantile gibberish trump lack decency even allow
rise gutter long enough wish american citizens happy new year bishop
talbert swan talbertswan december one likes calvin calvinstowell
december impeachment would make great year america also accept
regaining control congress miranda yaver mirandayaver december hear
talk include many people hate wonder hate alan sandoval december
uses word haters new years wish marlene december say happy new year
koren pollitt korencarpenter december trump new year eve tweet happy
new year including many enemies fought lost badly know love donald j
trump realdonaldtrump december nothing new trump years trump
directed messages enemies haters new year easter thanksgiving
anniversary pic twitter com daniel dale december holiday tweets
clearly presidential long work hallmark becoming president steven
goodine sgoodine december always like difference last years filter
breaking roy schulze thbthttt december apart teenager uses term
haters wendy wendywhistles december fucking year old knows december
people voted hole thinking would change got power wrong year old men
change year older photo andrew burton getty images'
```

Another task we will do simply because it is informational about the data set is to reformation the date of publication field to make it easier to plot and see if the data from both sets comes from approximately the same time period.

2.1.2 Data Inspection
Before we build the model, it is a good idea to try to characterize the dataset. This way we can see trends and verify any potential issues we may encounter when building the model.

2.1.2.1 Checking the date
When we cleaned the data, one of the things we did additionally was reformat the date to see if each of the datasets came from the same time period.
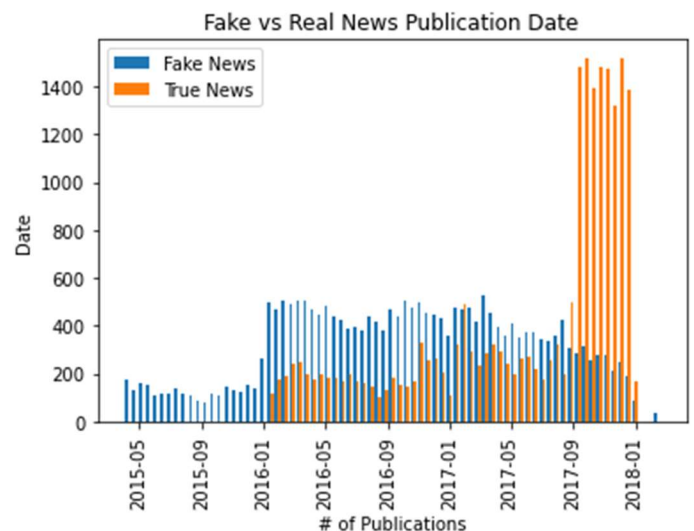

Figure 1 - Fake vs True Publication Date

The data shows that the fake news dataset is spread out and the true news is weighted heavily toward more recent dates. The effect of this is that the true news, may have a higher concentration of newer news topics than that of the fake news. But because they still roughly cover the same time period, it should be good enough for our purposes.

2.1.2.2 Checking the Subject

Next, we inspected the subject of news which was provided by the dataset. From this we found that the fake news used different subject classifiers than the true news. If we intended to use this as a primary feature we would need to merge the two and make them uniform. True news used the subject politicsNews and worldNews, while fake news used the others.
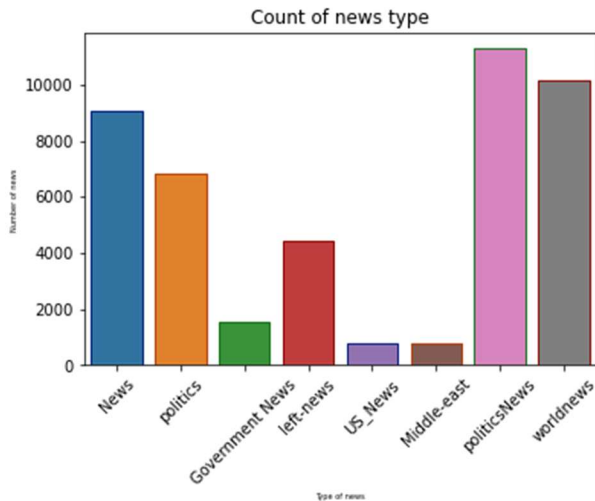

Figure 2 - News Subjects

2.1.2.3 Checking Sentiment and Subjectivity

Something that be measured in NLP is the sentiment and subjectivity of a given text. Sentiment takes a measure of the mood of the text, typically how happy or sad. This is given on a scale of -1 to +1, -1 being sad and +1 being happy. Subjectivity is a measure of how much opinion is used. 0 being totally objective and 1 being fully subjective. The way subjectivity is measured is using adverb modifiers. The more they are used the more subjective the text[4].

First, we look at the overall values for the text to get an idea of the expected shape they may take individually.
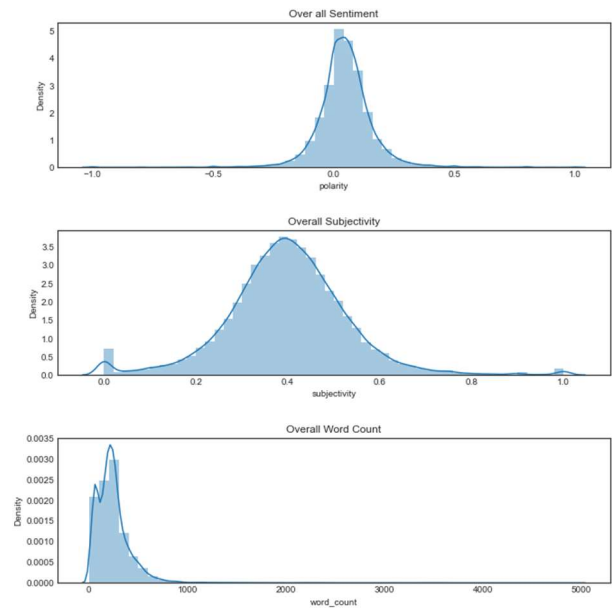

Figure 3 - Overall Polarity, Subjectivity and Word counts

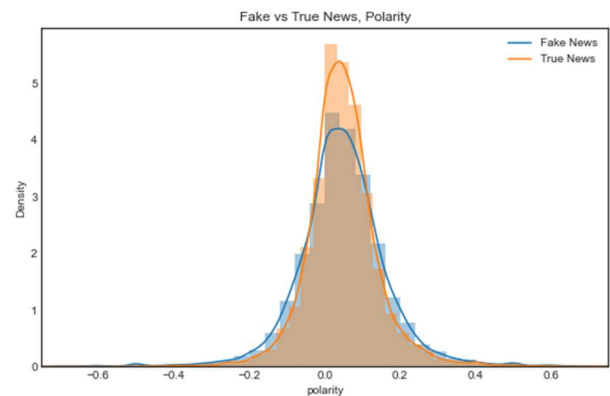Next we look at the them individually.

Figure 4 - Fake vs True Polarity

Polarity, Sentiment, is not very much between the two datasets. While true news had a smaller standard deviation than the fake news, the mean was approximately the same.
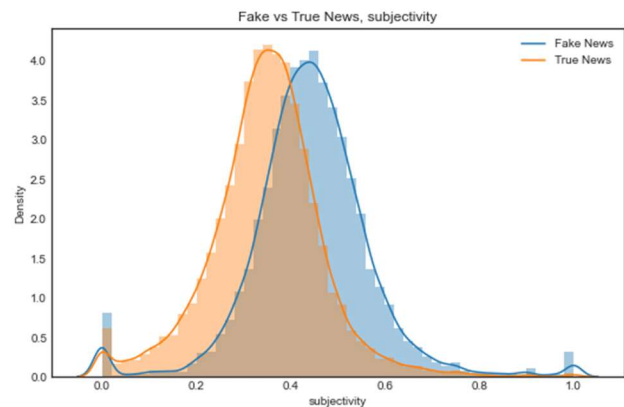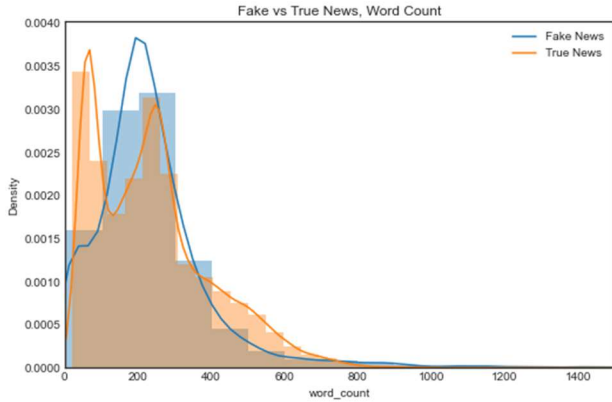

Figure 5 - Fake vs True Subjectivity

Subjectivity on the other hand was significantly different between the two data sets. Fake news had a much higher mean in the

subjectivity metric, which means it was much more subjective in its presentation.


Figure 6 - Fake vs True Word Count

I also checked to see if there was a large difference in the word counts. This would be used in the LSTM training to select the vocabulary size. This was also considered because if the fake news primarily came from a different source, then the true news, such as twitter, it may show up here.
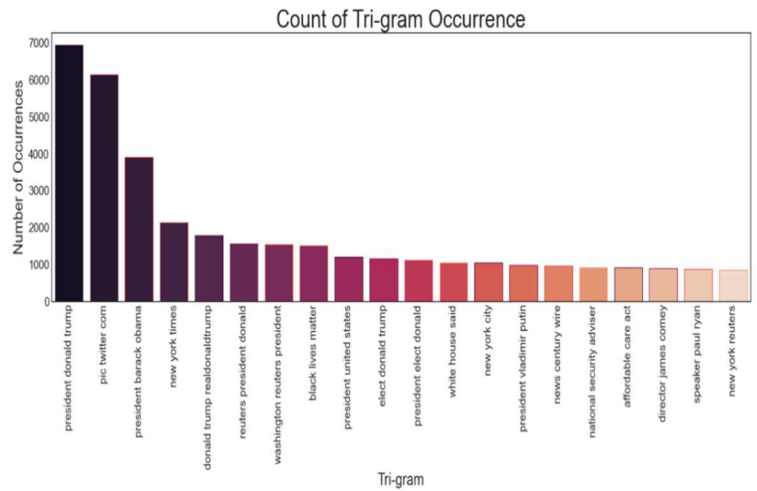
2.1.2.4 N-gram Analysis
Another inspection of the data we did was an n-gram analysis. This would allow us to see commonly used words and groups of words. This would give a good idea of the topics which are discussed in the dataset.
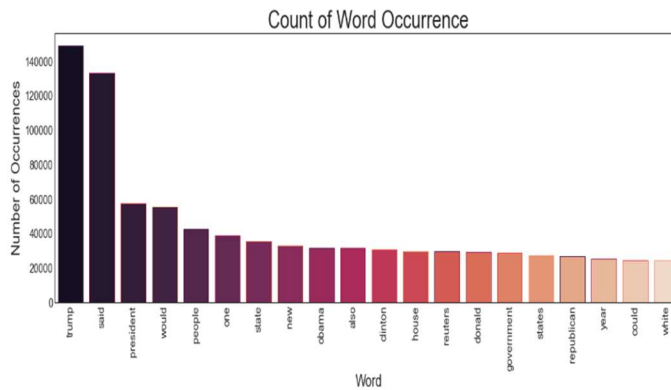
Figure 7 - Single word Occurrence
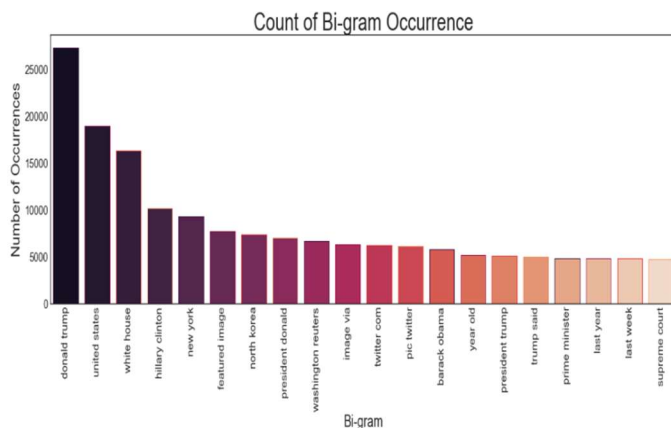

Figure 8 - Bi-gram Occurrence

Figure 9 - Tri-gram Occurrence

As expected, because this dataset was taken between and shortly after the 2016 election, the most common topics are all political related. With Donald Trump dominating the majority of the news cycle, followed closely by White house, Barack Obama and Hillary Clinton related topics. In the tri-gram chart we can start to see some the data sources, Twitter and Reuters.
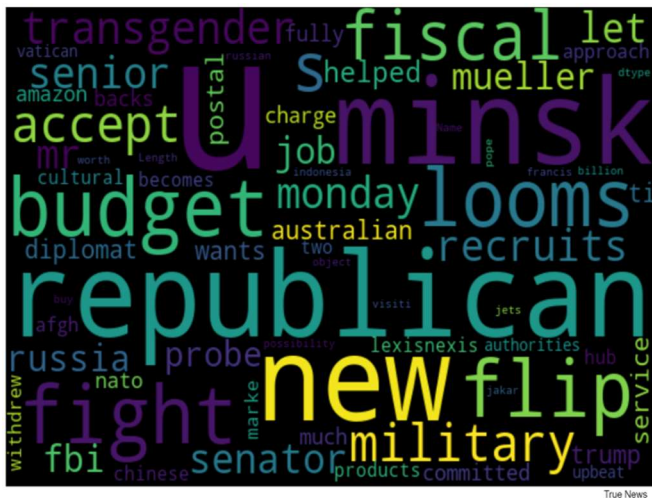
2.1.2.5 WordClouds
A common graphic that is generated for NLP studies are wordclouds. The size of the text reflects occurrence of a word.
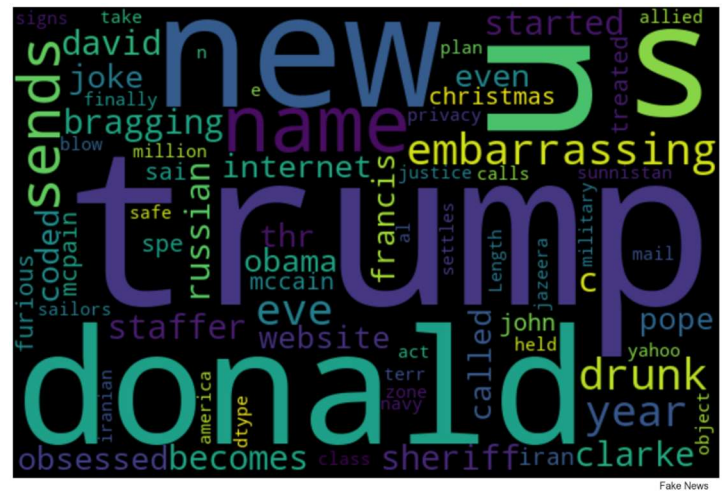


Figure 10 - Fake News WordCloud

Figure 11 - True News WordCloud

2.2 Building the Model

To approach this problem, we decided to build an LSTM. An LSTM is a type of recurrent neural network (RNN) that is capable of learning and remembering over long periods of time. This makes it well-suited for NLP tasks that require the model to remember and use past information in order to make predictions or decisions.

LSTMs work by incorporating memory cells, which are able to store and access information over extended periods of time. These cells are connected to input, output, and forget gates, which control the flow of information into and out of the cell. The gates can be trained to learn what information to remember and what to forget, allowing the LSTM to focus on the most relevant information and avoid being overwhelmed by irrelevant details[5]. And thus, it can avoid the vanishing gradient problem which was a primary problem is RNNs.

To build the embedding vector, which will help the model focus on relevant information, we used the Word2Vec library. This is a technique which builds a vocabulary vector of associated words using an existing neural network model. It uses a consume similarity function to indicated semantically similar words[6]. Because most of our articles have less than 1000 words, we will limit our text to 1000 words each. Then we tokenize the words and use them to build a vocabulary list which has the word associations given by Word2Vec. An example output from this list would be:
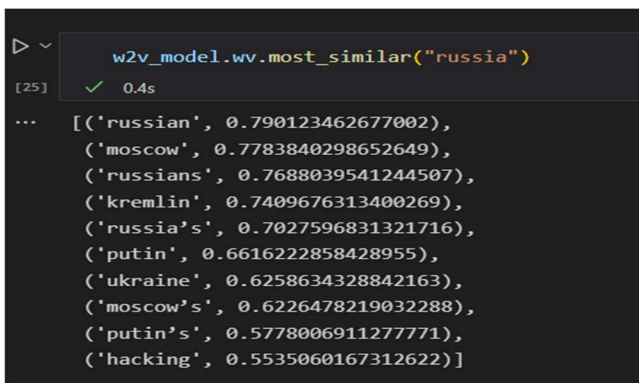
Figure 12 - Word2Vec output example "russia"

Now that we have made our entry size uniform and created the embedding vector we can now make and train the model. For this model we choose to make it 128 nodes and used 75% of the data for training and validation it.
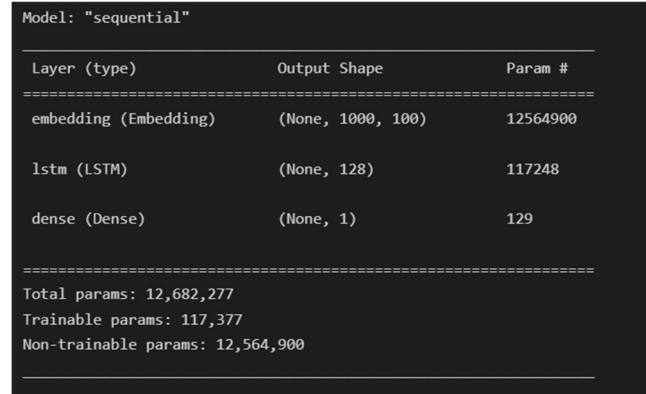
Figure 13 - Model Summary

Once the model was trained using .7 of the training and validation set(52.5% of the overall data), it used .3 of training and validation set (22.5% of the overall data) to check to see how the model training did. Once the training and validation test was complete, we used the 25% we kept aside to see how accurate we made our model. Below is a summary of the results

	Loss Value	Accuracy
Training data	.0948	.9693
Validation data	.0284	.9926
Test Data	-	.9926

Figure 14 - Results from model training and testing

Because the building of this model is highly dependent on the dataset, we decided to try it on a different dataset which was supposably from the same time period, though no date information was given. This dataset was from a Kaggle competition to build a model to detect fake news[7]. This data had to be reformatted to match how our model was trained, but we were able reuse the cleaning functions once it was reformatted. With the new dataset we only ended up with a 73.9% accuracy score. The likely cause of this was that a large enough number of words fell into the bucket which we did not train on, thus it was not able to know the associated semantics of the word.

3. Supporting Code and Documents

GitHub Repository that contains following items can be found at

https://github.com/blasher565/dm_cs522/tree/main/final%20project

Located in Code folder

- Dataset from Kaggle
- Jupyter Notebook

Located in Paper folder

- Project Paper (Word and PDF)

Located in Powerpoint folder

- Presentation Slides (PPT and PDF)
- Poster (PPT and PDF)

4. Conclusions

In this paper, we see how an LSTM can be used to build a NLP model which can determine if a piece of text is fake or true. However, that model is limited to the training vocabulary and training semantics. We saw that when using a subset of our training data to evaluate our model we ended with a very high rate of accuracy. However, if we used a dataset which we didn't train on and likely had different words and semantics our accuracy rate dropped significantly. 73.9% is still much better than randomly guessing, but it still is not enough to be able to reliably use this model for practical purposes.

The data itself had several key properties. Frist, Donald Trump dominated the topics. This is seen in the n-gram analysis in Figure 7. Fake news had much more subjectivity. This is seen in the subjectivity plots in Figure 5.

Several avenues could be explored to improve this project. First, by training on a wider dataset to increase the vocabulary. It could also be improved by utilizing the polarity and subjectivity information. Lastly, is to change the model into a bidirectional LSTM.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] J. Gottfried, "News Use Across Social Media Platforms 2016," Pew Research Center's Journalism Project, May 26, 2016.https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/ (accessed Dec. 10, 2022).

[2] C. S. Singer-Vine Jeremy, "Most Americans Who See Fake News Believe It, New Survey Says," *BuzzFeed News*. https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey (accessed Dec. 10, 2022).

[3] "Only one word 99.2%." https://kaggle.com/code/josutk/only-one-word-99-2 (accessed Dec. 11, 2022).

[4] "Tutorial: Quickstart — TextBlob 0.16.0 documentation." https://textblob.readthedocs.io/en/dev/quickstart.html (accessed Dec. 11, 2022).

[5] "Long Short-Term Memory Networks With Python," *MachineLearningMastery.com*. https://machinelearningmastery.com/lstms-with-python/ (accessed Dec. 11, 2022).

[6] "A Beginner's Guide to Word2Vec and Neural Word Embeddings," *Pathmind*. http://wiki.pathmind.com/word2vec (accessed Dec. 11, 2022).

[7] "Fake News." https://kaggle.com/competitions/fake-news (accessed Dec. 11, 2022).

[8] F. Alsuliman, S. Bhattacharyya, K. Slhoub, N. Nur, and C. N. Chambers, "Social Media vs. News Platforms: A Cross-analysis for Fake News Detection Using Web Scraping and NLP," in *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, Jul. 2022, pp. 190–196. doi: 10.1145/3529190.3534755.

[9] "A Closer Look at Fake News Detection | Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence." https://dl.acm.org/doi/abs/10.1145/3369114.3369149 (accessed Dec. 11, 2022).

[10] Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User Preference-aware Fake News Detection," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, Jul. 2021, pp. 2051–2055. doi: 10.1145/3404835.3462990.

[11] F. Bogale Gereme and W. Zhu, "Fighting Fake News Using Deep Learning: Pre-trained Word Embeddings and the Embedding Layer Investigated," in *2020 The 3rd International Conference on Computational Intelligence and Intelligent Systems*, New York, NY, USA, Feb. 2021, pp. 24–29. doi: 10.1145/3440840.3440847.

[12] Q. Sheng, X. Zhang, J. Cao, and L. Zhong, "Integrating Pattern- and Fact-based Fake News Detection via Model Preference Learning," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Oct. 2021, pp. 1640–1650. doi: 10.1145/3459637.3482440.

[13] W. Xu, J. Wu, Q. Liu, S. Wu, and L. Wang, "Evidence-aware Fake News Detection with Graph Neural Networks," in *Proceedings of the ACM Web Conference 2022*, New York, NY, USA, Apr. 2022, pp. 2501–2510. doi: 10.1145/3485447.3512122.

[14] J. Cui, K. Kim, S. H. Na, and S. Shin, "Meta-Path-based Fake News Detection Leveraging Multi-level Social Context Information," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, New York, NY, USA, Oct. 2022, pp. 325–334. doi: 10.1145/3511808.3557394.