THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

COMP562 — Machine Learning

Final Project Report

Submitted by:
Maanav Singh

Contents

1 Overview of the problem

The problem of unreliable news being spread rapidly digitally with next to no fact-checking is a immense problem and results in the rampant spread of misinformation. The negative effects of fake news range from uncertainty in the fairness of elections, to damage to the reputation and profitability of individuals and corporations. Market research indicates that as much as $39 billion is lost annually as a result of uncontrolled fake news[2]. This large issues establishes a large monetary and societal motivation for the development of effective techniques to detect this issue.

Natural Language Processing techniques have been of great interest to mitigating the effects of fake-news [4]. The development of such solutions, provide at times super-human levels of fake-news detection with the scalability to scan large portions of the internet. In this paper, we will introduce a novel technique to detect fake news built upon existing work in NLP.

2 Overview of data and pre-processing

In our pursuit of developing a machine learning model to detect fake-news, we've identified a dataset[1] found on the online Data Science site Kaggle, which provides an aggregated and labeled collections of articles scraped from fake and reputable news sources. This dataset contains data from around 45,000 articles which will allow us to spare a large number of samples for validation as well as counteract the effects of sampling bias from limited sources.

2a Data format and cleaning

The data is originally presented in the form of two comma separated tables of real and fake articles. We will combine these two tables into one by adding a *is_fake* column to the table. Each row in the table represents and article, with the columns describing the article being: article title (*title*), article text (*text*), subject, date, and the newly introduced *is_fake*. Some samples are shown in the below figure.

After inspecting the dataset, a few key issues are aparent. Namely: there are discrepancies between the subject classes in the true and false news sources. Some date fields are malformed, with completely incorrect and unparsable data.

These concerns introduce the following preprocessing steps:(1) Map subject labels in Fake news set to corresponding ones in the Real News set.(2) Add a binary field is fake to both tables before concatenating.(3) Parse dates and drop malformed rows.(4*) Compute sentiment of article title and body and add another field sentiment

	title	text	subject	date	is_fake	subject_num
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	worldnews	0.952652	1	1
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	worldnews	0.952652	1	1
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	worldnews	0.951705	1	1
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	worldnews	0.950758	1	1
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	worldnews	0.946970	1	1
...	...	...	...	...	...	...
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	0.828598	0	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	0.828598	0	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	0.828598	0	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	0.828598	0	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	0.828598	0	1

44898 rows × 6 columns

2b Feature selection

While preparing our dataset for the ultimate goal of developing a machine learning classifier, we will need to ensure our data is both correctly formatted and appropriate for the model.

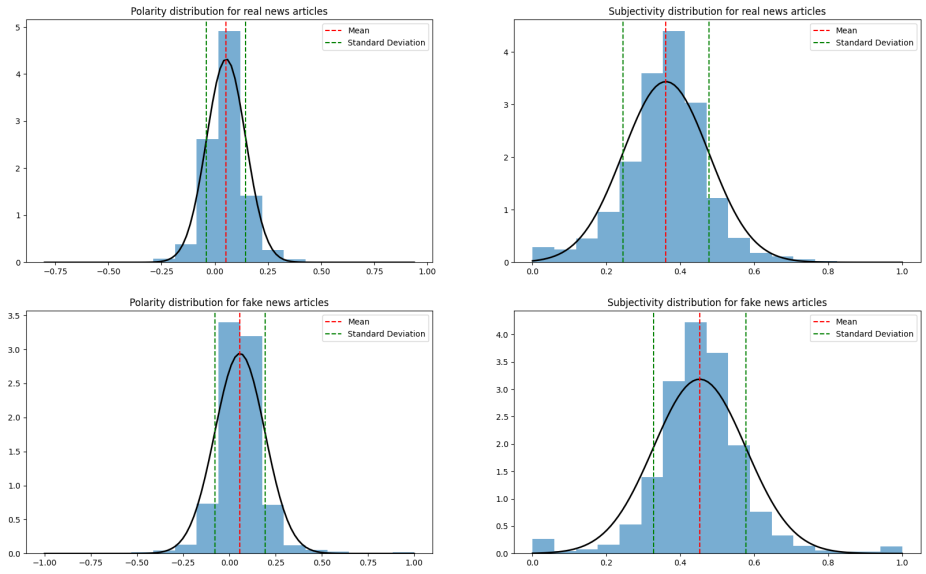
Currently our available features are:

1. Title and Text (sequence of words)
2. Subject (category represented as a binary number)
3. Date (normalized epoch time between 0 and 1)
4. Polarity and Subjectivity (Tone and Presence of Bias from Sentiment Analysis)

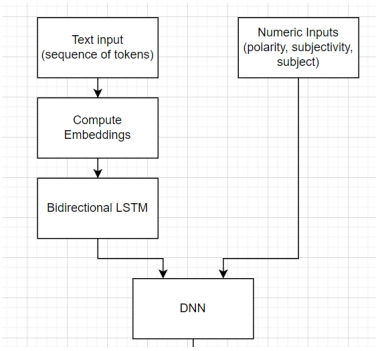
Polarity and Subjectivity were computed using the popular open-source sentiment analysis python library TextBlob which uses the VADER algorithm for sentiment analysis [3]. Using a pretrained model for evaluating these characteristics means our final model can have less parameters and complexity while still maintaining the same or better performance due to the absence of regions of the network computing similar quantities as VADER.

The subject field will likely yield information to the model regarding class: So this will be used as an input to the model. By the same token, the date field in it's normalized epoch time form will not generalize well to articles outside of the time-range of the training set. For this reason the date feature will not be used in the training of the model. In order to explore the relevance of sentiment, the polarity and subjectivity distributions are plotted. From the figure below, it's clear there are stark differences in the distribution of the two sentiment variables for real and fake news. Namely there is more variance in polarity and a higher average subjectivity for fake news. These observable differences indicate these variables will make good input features for the model.

3 Model Design and Training



With a solid foundation of clean and relevant predicting features, the next decision is designing an appropriate model architecture and hyperparamaters. Due to the sequentual and inter-related nature of textual data, a Recurrent Neural Network with Long Short Term Memory (LSTM) has been shown to perform very well for article text [5]. Prior to training we tokenize the text, by converting words into numbers based on frequency and pad the sequences to provide uniform lengths to the model. Additionally, the model needs to accept numeric inputs for the sentiment analysis results and subject results. This was achieved by added a separate input layer and using a Tensorflow merge layer to concatenate the output of the LSTM layers to the inputs in a way where the gradients can still be computed. This architecture is shown below

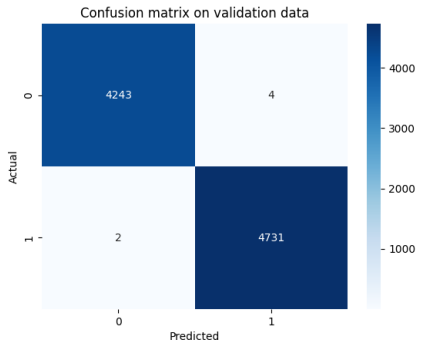


Prior to training, the 44k records were partitioned into a 80-20% training-validation split. This enables the bulk of the data to be used for training, while having around 9k records for validating the performance of the model. While training both the training and validation

accuracy increased in-step until leveling out: This indicates that learning has concluded and any further training would result in over-fitting.

4 Evaluation

In evaluating our model, we primarily used a Confusion matrix on the validation set. From the Confusion Matrix, we see that we had 4243 True Negatives, 4 False Positives, 2 False Negatives, 4731 True Positives. From this, we found that our model had a precision of 99.92 percent, a recall of 99.96 percent, and an accuracy of 99.93 percent. However, there may have been data leakage in our model from direct unintended markers in the text identifying the news source rather than the actual content. Which may inflate the accuracy on the validation set.



Furthermore, The model was evaluated on two current day articles outside the training and validation set. One being an actual article from CNN discussing the death of a politician and another being a satire article defaming King Charles III published by the Onion. The model was able to correctly predict that the CNN article is real news, and that the Onion article is fake news. This indicates that the model is able to generalize to articles outside of it's training corpus.

5 Conclusion and Other Works

This model can be applied in real world scenarios to help better guide people towards identifying which news may be fake and political. For example, during a major presidential election, using this model will help people to identify which news is spreading false information about presidential candidates and what news to believe. This model can also be applied to other prominent figures, such as politicians and celebrities. In terms of fake news detection, other models perform at best around 96-98%[6] compared to 99% accuracy from our model on the validation thus outperforming other popular although more general models such as BERT.[4]

6 Citations

[1] Cleacutement Bisaillon. Fake and real news dataset, Mar 2020.

[2] Roberto Cavazos. The economic cost of bad actors on the internet. *CHEQ*, 2019.

[3] C. Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225, May 2014.

[4] Junaed Younus Khan, Md. Tawkat Islam Khondaker, Sadia Afroz, Gias Uddin, and Anindya Iqbal. A benchmark study of machine learning models for online fake news detection. *Machine Learning with Applications*, 4:100032, 2021.

[5] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. August 2018.

[6] Pawlicki M. Kozik R. et al. Szczepański, M. New explainability method for bert-based model in fake news detection. *Sci Rep*, 11(23705), 2021.