Fake News Detection

Shivani Sanjay Mahaddalkar

IST 736: Text Mining- Project Report

Introduction

In 2014, World Economic Forum cited fake news as one of the top ten threats to the modern society (Safieddine, Fadi, 2020). Dispelling knowledge of any kind has become easier and faster due to the advances in technology and growing usage of social media channels. Along with the convenience of spreading knowledge, social media poses a problem of widespread misinformation. It has become easier to mimic the real news articles for fake news propagation.

This project aims to understand the driving factors that distinguish the fake news from the real ones. The focus of the news articles remains on the electoral and geopolitical news from the year 2016. This was the time leading up to what can be touted as the most game changing election in the US history.

This study aims to answer the following questions: (1) Can a prediction model be built to successfully distinguish between fake and real news articles? (2) What mistakes are made during predictions? (3) Can this same framework be successfully applied to other kinds of news articles to make the distinction?

Ethics Framework

Solving such a problem, would definitely be a reason to rejoice in the modern world. News articles masking themselves as facts are propagating misinformation on a very large scale. However, such a problem comes riddled with some ethics issues that need to be thought of before making claims. Flagging a source for publishing fake news articles could be damaging to the reputation of the publishing firm, when the articles published were in fact based on facts. Hence, a human expert with domain knowledge should be consulted before tagging an article as false. However, false articles that pass the model and are predicted to be true, could be further damaging to the society. This means that publishing articles needs to pass a certain framework of ethics on behalf of the publisher.

Data Set

The data set consists of two files, one containing entries of real news articles and the other one containing entries of fake news articles. The real articles file is collected from articles published in Reuters. The fake articles are collected from different sources. These articles were flagged by Politifact, a fact-checking organization in the USA, and Wikipedia. The articles are majorly based on world news and US politics (Ahmed H, Traore I, Saad S.,2018).

The real news articles are 21417 articles and fake news articles are 23481. 0 is fake news and 1 is true news. The data set is balanced as can be visually represented below:

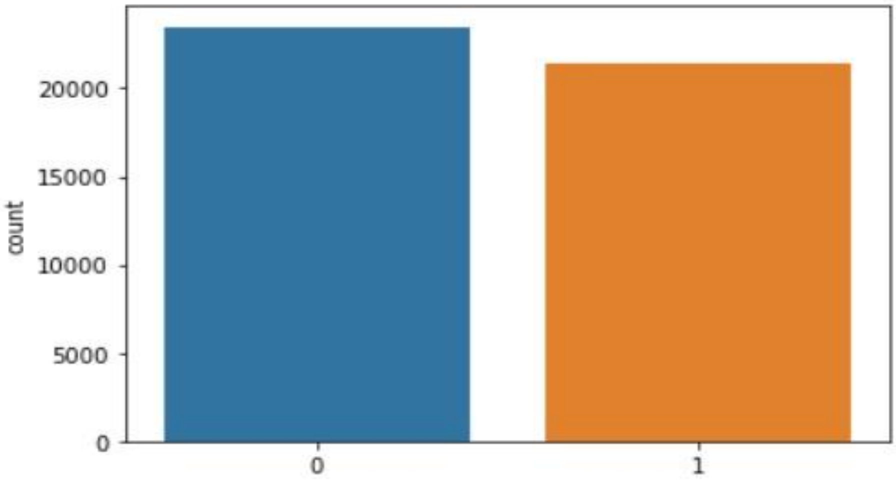

Figure 1. Distribution of fake and real news in data set.

The data set contains the title, the text, subject of the article. On merging both the data files, we also add the label attribute in the data.

Methodology

In this analysis, two algorithms are used for comparison. Using cross validation, the best algorithm and vectorizer is determined, which is further used on the testing data set. The two algorithms, multinomial Naïve Bayes model and support vector machine were used to predict the class using either the text or the title.

Multinomial Naïve Bayes is widely used in text classification. It uses parameter learning which computes the probability of a word occurring using the word frequencies. Bayesian learning approach is used which assumes that the word distribution is generated by specific word parameters and such parameters can be observed in the training data (Su, Jiang, Jelber Sayyad Shirab, and Stan Matwin, 2011).

Support vector machine first has proven to be effective for supervised machine learning. SVMs are hyperplanes that separate the training data by maximal margin (Tong, S., & Koller, D. (2001)). This is how SVMs divide the data into different classes and uses the testing data to arrange itself into the SVM space and uses the location of that testing data in the prediction.

Cross validation is a statistical method of evaluating and comparing data into two categories: train and validation set. It ensures the model does not overfit or underfit the training data and gives an idea of how the model will perform during testing.

To further validate the results, topic modeling was implemented. It aimed to segregate the articles into 5 topics. On segregation, it clubs two topics together into the training data set and clubs the rest three topics into testing data set. This allows the analysis to ensure that the model does not overfit and tries to validate the hypothesis that the prediction model could be used for other types of articles too.

The topic modeling is done using Latent Dirichlet Allocation (LDA). LDA is an algorithm that can help to analyse the latent topic representation of a given corpus or dataset. It posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics (Blei, D. et al., (2003)).

Text Mining Models:

In order to maximise accuracy, the project tried different vectorizers. The three different vectorizers used were term frequency, term frequency inverse document frequency and Boolean vectorizer.

With 5 fold cross validation, the cross validation accuracy of the models is given in the table below.

	Model based on title	Model based on text
MNB with TFIDF	91.14%	89.42%
MNB with TF	91.27%	90.05%
MNB with Boolean	91.64%	92.35%
SVM with TFIDF	91.89%	97.47%
SVM with TF	91.84%	97.50%
SWM with Boolean	91.79%	97.47%

The support vector machine models work better in both cases. However, they show immense improvement in accuracy in models that use text for prediction.

We use SVM for further creating our model with term frequency vectorizer.

Splitting the data, 60% for training and 40% for testing we conduct further error analysis. One point to pay attention here is to see whether the model is specializing in predicting for certain topics or it could be used for a more general scope.

To examine that we use LDA topic modeling to separate the articles into 5 topics. The top words for each topic are given below:

Topic 0: International News

trump north korea north korea court eu iran senate minister government brexit tax republican turkey uk campaign war nuclear plan america

Topic 1: Middle east News

house russia clinton white trump white house hillary obama military law syria russian myanmar probe foreign tells media say attack election

Topic 2: Election News

says trump obama video president china watch deal talks news breaking factbox election putin trump says urges hillary poll cnn france

Topic 3: International News

vote police video pm south black new saudi islamic leader border opposition presidential islamic state killed mexico man support gun race

Topic 4: Election News

trump video state party anti security congress donald ban donald trump governor speech judge democrats trade sanders gop aid help obamacare

Results

When the train and test split is done randomly, the results are given below.

For title:

	Predicted False	Predicted True
Ground Truth False	8860	480
Ground Truth True	455	8165

Top Features that help in prediction:

Features for Real News	Features for Fake News
Egypt	Video
Brazil	Breaking
Ukraine	GOP
Kremlin	Watch
Rohingya	Just
Exclusive	Lied
Myanmar	Hillary
Urges	Racist
Factbox	Illegals

The accuracy of the model is 94.79%. The precision and recall are both 0.95 for the model.

For text:

	Predicted False	Predicted True
Ground Truth False	9204	136
Ground Truth True	146	8474

Features for Real News	Features for Fake News
EDT	Read
Monday	Featured
Friday	GOP
Said	Image
Tuesday	Sen
Wednesday	Getty
Nov	Mr.
Reuters	Wire
Representatives	Rep

The accuracy of the model is 98.43%. The precision and recall both are 0.98 for the model.

From topic modeling, we train the data on the election news topic derived from the LDA topic modeling and test it on the international news.

Using SVM classifier with term frequency, we found the following results.

For title:

	Predicted False	Predicted True
Ground Truth False	7376	2409
Ground Truth True	265	8454

The accuracy goes down to 85.62%. The precision is 0.87 whereas the recall is 0.86.

For text:

	Predicted False	Predicted True
Ground Truth False	9366	419
Ground Truth True	107	8612

The accuracy is 97.16%. The precision and recall are 0.97 for the training data.

Inference from Results

The high accuracy in the model with random train and test data split indicated that this problem needed further inspection. The initial model indicates that words that sound more professional in the title have a strong indication of being true news articles. Whereas more attention-grabbing words are used in the headlines for fake articles.

For the body of the articles, the features are not easily decipherable and do not seem to follow an intuitive path to demystify the features.

The concern that the model is becoming highly specialized in the same kinds of news articles and would not be scalable to a wider scope proves to be true to a certain extent.

On training and testing on two different topics of news articles, the accuracy in both the cases drops. For title of the articles, the test accuracy goes down from 94.79% to 85.62%. More articles are wrongly labeled as true news articles than before.

For the text of the articles, the accuracy goes down but not by much. It goes from 98.43% to 97.16%, which is not significantly low.

It indicates that the model requires a bigger text data for every entry like the text of the article, as opposed to the title of the article. This happens to humans when an article has a sensational headline, and on clicking one finds it to be false based on the style of writing. The result from the model points to a similar phenomenon.

Conclusion

Fake news detection is a difficult task that requires significant domain knowledge. However, this project was successful in building a model that has a high accuracy for prediction. It needs to be seen whether such a model can be used in a completely different type of news. As this data set consisted of mainly US politics and its international relations, such a model could also be used for a specialized predictor.

References:

1. Safieddine, Fadi. (2020) "Political and Social Impact of Digital Fake News in an Era of Social Media." *Fake News in an Era of Social Media: Tracking Viral Contagion* :43.
2. Ahmed H, Traore I, Saad S. (January/February 2018) “Detecting opinion spams and fake news using text classification”, Journal of Security and Privacy, Volume 1, Issue 1, Wiley.
3. Ahmed H, Traore I, Saad S. (2017) “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127- 138).
4. Su, Jiang, Jelber Sayyad Shirab, and Stan Matwin. (2011) "Large scale text classification using semisupervised multinomial naive bayes." *ICML*.
5. Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of machine learning research, 2*(Nov), 45-66.
6. Blei, D. et al., (2003) Latent dirichlet allocation. Journal of machine learning research 3(4- 5):993-1022