**ABSTRACT**

In the contemporary landscape, social media interaction, particularly the dissemination of news across networks, has become a significant source of information. Individuals are drawn to its minimal effort, easy accessibility, and rapid information dissemination, making social media a preferred platform for news consumption. Among these platforms, Twitter stands out as one of the most popular sources for real-time news and a potent medium for news propagation.

However, the ease and speed of information sharing on social media, including Twitter, also bring inherent challenges. The platform has been known to cause substantial harm by facilitating the spread of rumors and misinformation. Online users often perceive content on social media as reliable, making them susceptible to misinformation. Therefore, the automation of fake news detection becomes crucial in maintaining a robust online media and social network environment.

This paper introduces a model for identifying fake news messages within Twitter posts by employing machine learning techniques to predict accuracy ratings. The proposed model aims to automate the detection of fake news in Twitter datasets. The study involves a comparison of the performance of five well-known machine learning algorithms, namely Support Vector Machine, Naïve Bayes Method, Logistic Regression, and Recurrent Neural Network models. The evaluation demonstrates that SVM and Naïve Bayes classifiers outperform the other algorithms in terms of classification efficiency on the dataset.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
| S.NO | TOPICS |
|  | Chapter 1: INTRODUCTION |
|  | DATASET |
|  | FILE DISCRIPTION |
|  | PERFORMANCE |
|  | WORKFLOW |
|  | IMPLEMENTATION AND CODE SECTION |
|  | FUCTURE SCOPE OF PROJECT |
|  | FAKE NEWS PREDICTION MODELS |
|  | RESULT ANALYSIS |
|  | CONCLUSION |
|  | REFRENCES |

**CHAPTER 1**

**Introduction:**

* 1. **What are fake news?**
     1. **Definition:**

Fake news has quickly become a society problem, being used to propagate false or rumor information in order to change people’s behavior. An intense inspection the present tweets demonstrates that false news spreads frequently through human than a genuine news does. Lie gets traveled around us quicker, and more extensively than reality in all spheres of information, and the effects were more dangerous and horrifying. There are several kinds of tweets like issues on a government, trending topics around the world, mental abuse, urban legends, occasions in calamities. What's more notifying is that it's not just bots that is outpouring the majority of the misrepresentations, studies claimed. It was some specific individuals performing a large share of this crime. Normally general users, as well, they explained. In this case, verified users and those with numerous fans were not more often the center in spreading misinformation of the corrupted posts. Fake news on social media which got viral like a rocket in no time can cause much havoc to our society human and country.

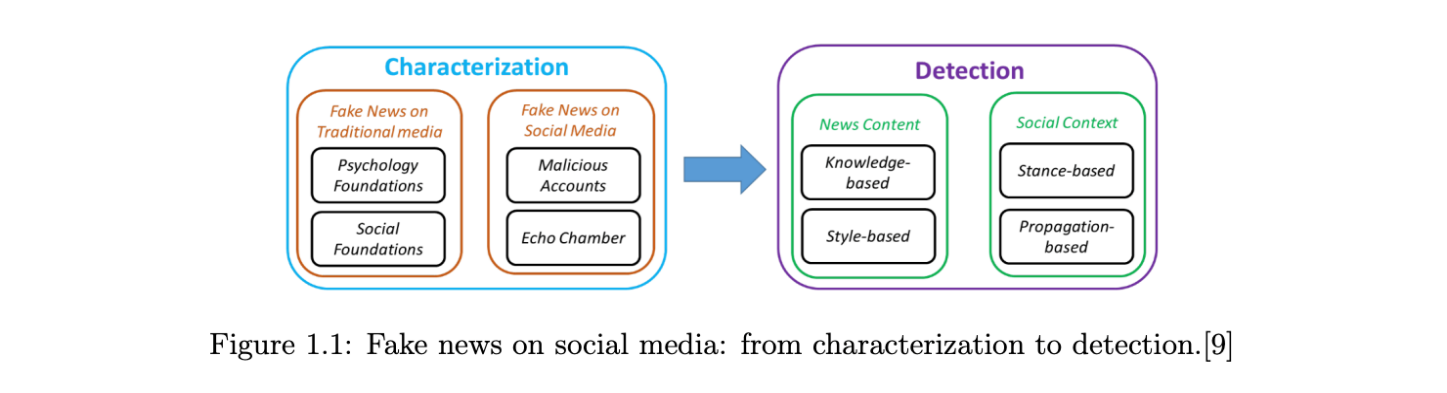
Measuring precision and validity in news contents are all around well-examined points in every sphere of daily life. The multiplication of huge scale internet-based life information and its expanding use as an essential news cue, be that as it may, is compelling a reconsideration on these specific areas. Strategies in the past that relied upon the arrangements made by the journalists “watchmen” to channel out below average substance are no increasingly proper as engagements into social networking's volume has promptly overshoot our ability to control the standard physically. Rather, stages like Facebook and Twitter have permitted sketchy and incorrect “news” substance to contact broad gatherings of people without survey. Online life clients' inclination toward accepting what their companions share and what they read despite precision enables these fake stories to proliferate broadly through and over numerous platforms. In spite of examination into gossip engendering on Twitter, gossip and unauthenticated news are getting to be progressively tricky. Computational strategies have demonstrated helpful in comparable settings where data volumes overpower human examination abilities. Moreover, regularities in bot conduct and monetarily spurred sentimentalists recommend automated machine learning methodologies could encourage to specify these issues.

The first is characterization or what is fake news and the second is detection. In order to build detection models, it is need to start by characterization, indeed, it is need to understand what is fake news before trying to detect them.

This work makes the following contribution. It has manufactured a fake news framework for news shared on Twitter utilizing five unique calculations. We consider every explanation which is posted by the clients. We utilize a rundown of highlights for a bit of news (a URL after canonicalization) the individual clients who shared the URL by tweeting or retweeting it, close by with the words that show up in the title of the news article itself. Five different types of classification models including Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, Long short-term memory, Recurrent Neural Network are implemented for this task. A combination of these classification models are also tested to further enhance the accuracy of prediction. Using scikit learn [10], these models are implemented to prepare from the training dataset using k-fold (k=2) cross-validation, and then predict using the data set.

**1.1.2 Fake News Characterization**

Fake news definition is made of two parts: authenticity and intent. Authenticity means that fake news content false information that can be verified as such, which means that conspiracy theory is not included in fake news as there are difficult to be proven true or false in most cases. The second part, intent, means that the false information has been written with the goal of misleading the reader



**Definition 1** Fake news is a news article that is intentionally and verifiable false

* 1. **Feature Extraction**
     1. **News Content Features:**

Now that fake news has been defined and the target has been set, it is needed to analyze what features can be used in order to classify fake news. Starting by looking at news content, it can be seen that it is made of four principal raw components:

**• Source:** Where does the news come from, who wrote it, is this source reliable or not.

**•** **Headline**: Short summary of the news content that try to attract the reader.

**• Body Text**: The actual text content of the news.

**• Image/Video**: Usually, textual information is agreement with visual information such as images, videos or audio.

Features will be extracted from these four basic components, with the mains features being linguistic-based and visual-based. As explained before, fake news is used to influence the consumer, and in order to do that, they often use a specific language in order to attract the readers. On the other hand, non-fake news will mostly stick to a different language register, being more formal. This is linguistic-based features, to which can be added lexical features such as the total number of words, frequency of large words or unique words

* + 1. **Social Context Features**

The last features that have not been used yet are social media features. There are two approaches to use these features: stance-based and propagation-based.

**Stance-based** approaches use implicit or explicit representation. For instance, explicit representation might be positive or negative votes on social media. Implicit representation needs to be extracted from the post itself.

**Propagation-based** approaches use features related to sharing such as the number of retweet on twitter

Out of all the studies in this specific area, one study proposed a survey on the field of NLP for false news detection. This paper displays an overview of fake news recognition. Their review presents the difficulties of the methodology in the detection of news or twitters like this. They deliberately audit the data and NLP arrangements that have been produced for this whole study. They likewise examine the points of confinement of those twitters and issue definitions, our bits of knowledge, and suggested arrangements

From the above discussion we already came to know that there are two classes in a news, whether it's real or fake one. To classify a news, we need to understand the problem definition first, then we go for our model and evaluate the result. Machine Learning is replete with its algorithms but some of them are really good for “Fact or Fiction” detection and some are on an average scale.

In our model, we used 5 different types of machine learning algorithms and for the implementation work, we used Python 3.6.5 as our programmable language. The classification models that we implemented using the above- mentioned dataset are Bayesian Model, Logistic Regression & also Support Vector Machine - two most famous deep learning methods RNN Recurrent Neural Network and Long Short-Term Memory were also implemented to see how well our data fit into the model. These algorithms are good for different classifications and they got their own properties and performance based on different datasets. As we said earlier Naïve Bayes, Logistic Regression and SVM are more commonly used algorithms for classification problems.

**DATASET**

The data source used for this project is LIAR dataset which contains 3 files with .tsv format for test, train and validation. Below is some description about the data files used for this project.

“LIAR: A BENCHMARK DATASET FOR FAKE NEWS DETECTION”

William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, to appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), short paper, Vancouver, BC, Canada, July 30-August 4, ACL.

the original dataset contained 13 variables/columns for train, test and validation sets as follows:

* Column 1: the ID of the statement ([ID].json).
* Column 2: the label. (Label class contains: True, Mostly-true, Half-true, Barely-true, FALSE, Pants-fire)
* Column 3: the statement.
* Column 4: the subject(s).
* Column 5: the speaker.
* Column 6: the speaker's job title.
* Column 7: the state info.
* Column 8: the party affiliation.
* Column 9-13: the total credit history count, including the current statement.
* 9: barely true counts.
* 10: false counts.
* 11: half true counts.
* 12: mostly true counts.
* 13: pants on fire counts.
* Column 14: the context (venue / location of the speech or statement).

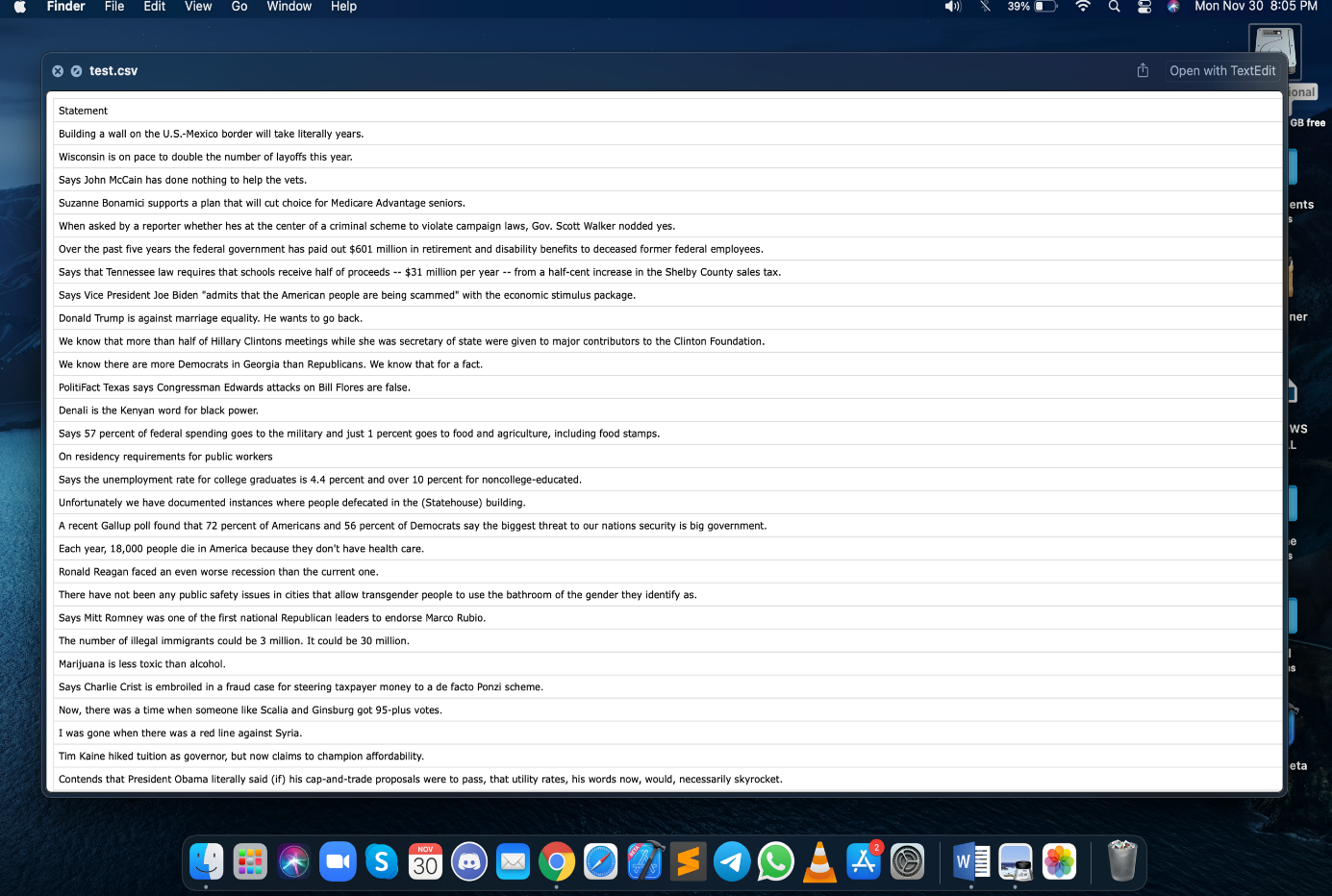
To make things simple we have chosen only 2 variables from this original dataset for this classification. The other variables can be added later to add some more complexity and enhance the features.

Below are the columns used to create 3 datasets that have been in used in this project

* Column 1: Statement (News headline or text).
* Column 2: Label (Label class contains: True, False)

You will see that newly created dataset has only 2 classes as compared to 6 from original classes. Below is method used for reducing the number of classes.

* Original -- New
* True -- True
* Mostly-true -- True
* Half-true -- True
* Barely-true -- False
* False -- False
* Pants-fire – False

The dataset used for this project were in csv format named train.csv, test.csv and valid.csv and can be found in repo. The original datasets are in "liar" folder in tsv format. 



**File descriptions**

#### **DataPrep.py:**

This file contains all the preprocessing functions needed to process all input documents and texts. First we read the train, test and validation data files then performed some preprocessing like tokenizing, stemming etc. There are some exploratory data analysis is performed like response variable distribution and data quality checks like null or missing values etc.

#### **FeatureSelection.py:**

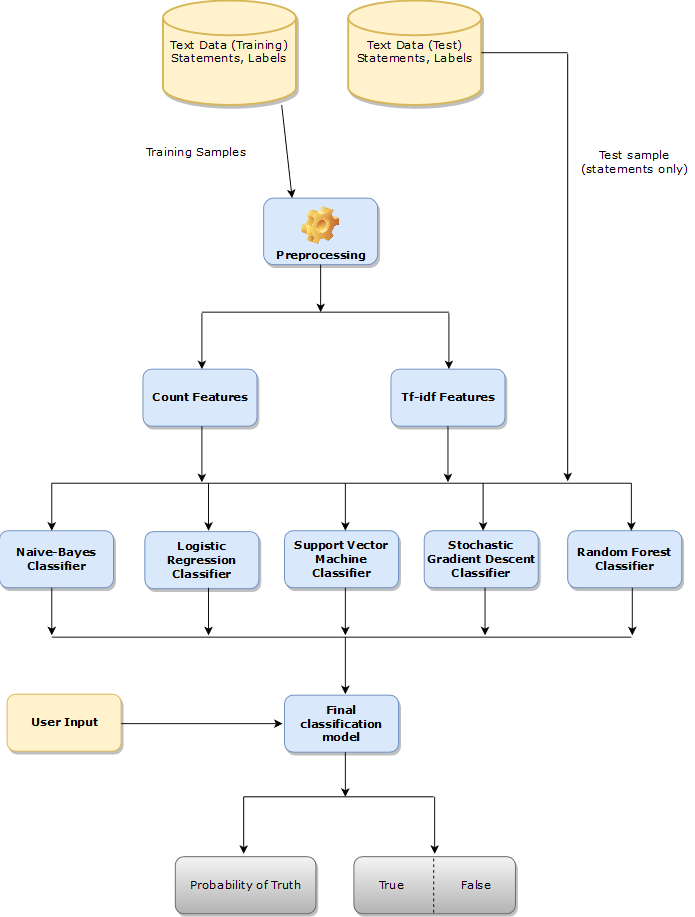
In this file we have performed feature extraction and selection methods from sci-kit learn python libraries. For feature selection, we have used methods like simple bag-of-words and n-grams and then term frequency like tf-tdf weighting. we have also used word2vec and POS tagging to extract the features, though POS tagging and word2vec has not been used at this point in the project.

#### **classifier.py:**

Here we have build all the classifiers for predicting the fake news detection. The extracted features are fed into different classifiers. We have used Naive-bayes, Logistic Regression, Linear SVM, Stochastic gradient descent and Random forest classifiers from sklearn. Each of the extracted features were used in all of the classifiers. Once fitting the model, we compared the f1 score and checked the confusion matrix. After fitting all the classifiers, 2 best performing models were selected as candidate models for fake news classification. We have performed parameter tuning by implementing GridSearchCV methods on these candidate models and chosen best performing parameters for these classifier. Finally selected model was used for fake news detection with the probability of truth. In Addition to this, We have also extracted the top 50 features from our term-frequency tfidf vectorizer to see what words are most and important in each of the classes. We have also used Precision-Recall and learning curves to see how training and test set performs when we increase the amount of data in our classifiers.

#### **prediction.py:**

Our finally selected and best performing classifier was Logistic Regression which was then saved on disk with name final\_model.sav. Once you close this repository, this model will be copied to user's machine and will be used by prediction.py file to classify the fake news. It takes an news article as input from user then model is used for final classification output that is shown to user along with probability of truth.

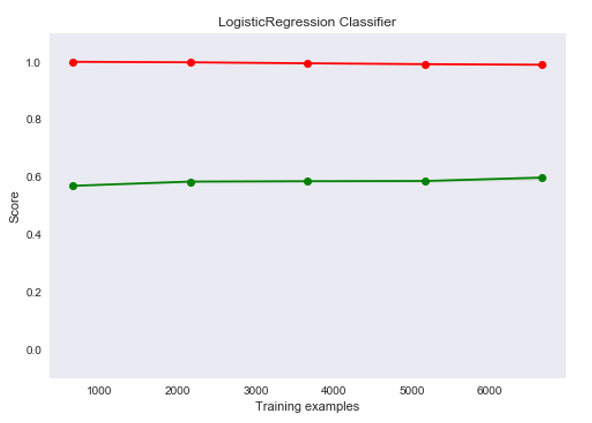


*Above is the Process Flow of the project:*

### Performance

Below is the learning curves for our candidate models.

**Logistic Regression Classifier:**

****

Logistic regression is one of the most common and useful classification algorithms in machine learning. If you wish to become a better machine learning practitioner, you’ll definitely want to familiarize yourself with logistic regression.

Classification and regression tasks are both types of *supervised learning*, but the output variables of the two tasks are different. In a [regression](https://medium.com/datadriveninvestor/regression-in-machine-learning-296caae933ec)task, the output variable is a numerical value that exists on a continuous scale, or to put that another way the output of a regression task is an integer or a floating point value.

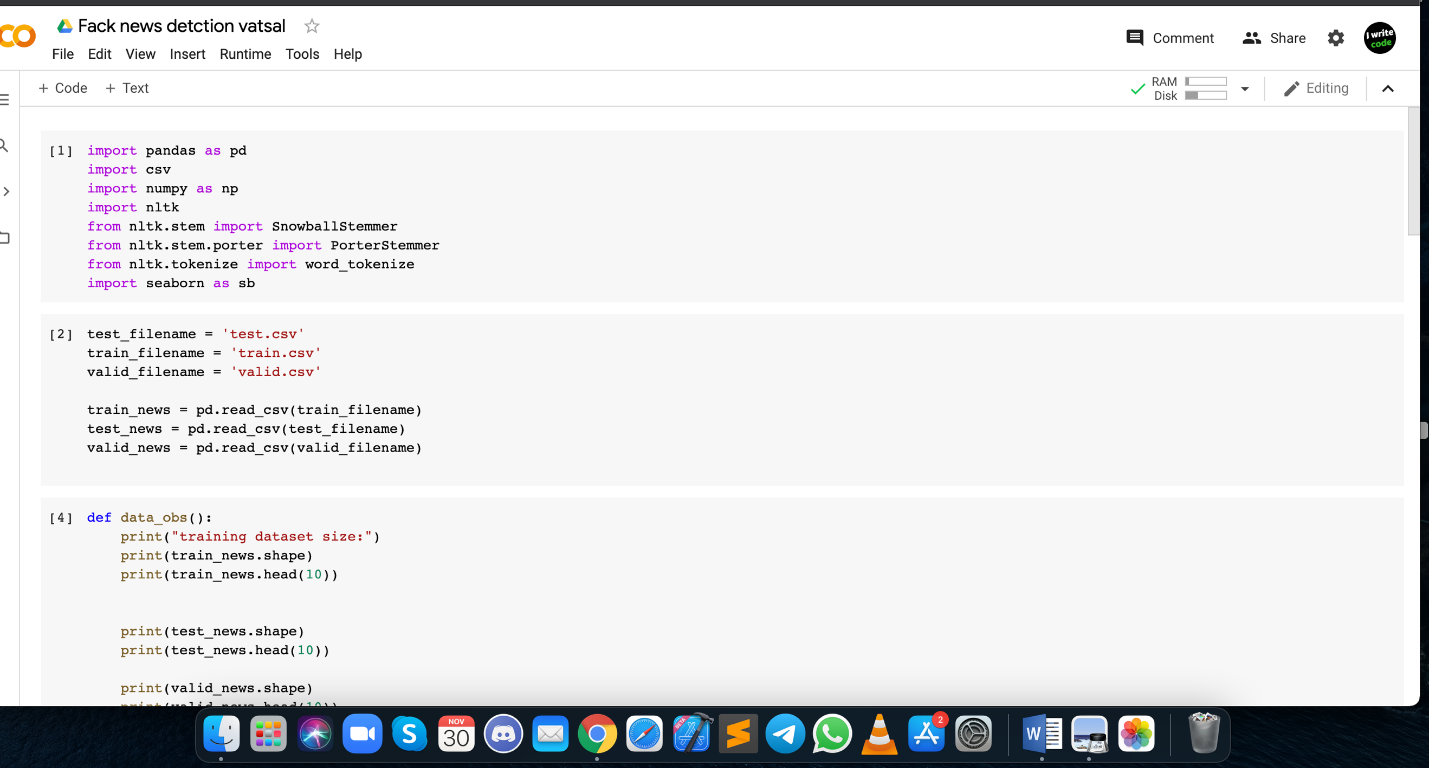
In contrast, in a classification task, the outputs of an algorithm fall into one of [various pre-chosen categories](https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/). The classification model attempts to predict the output value when given several input variables, placing the example into the correct category.

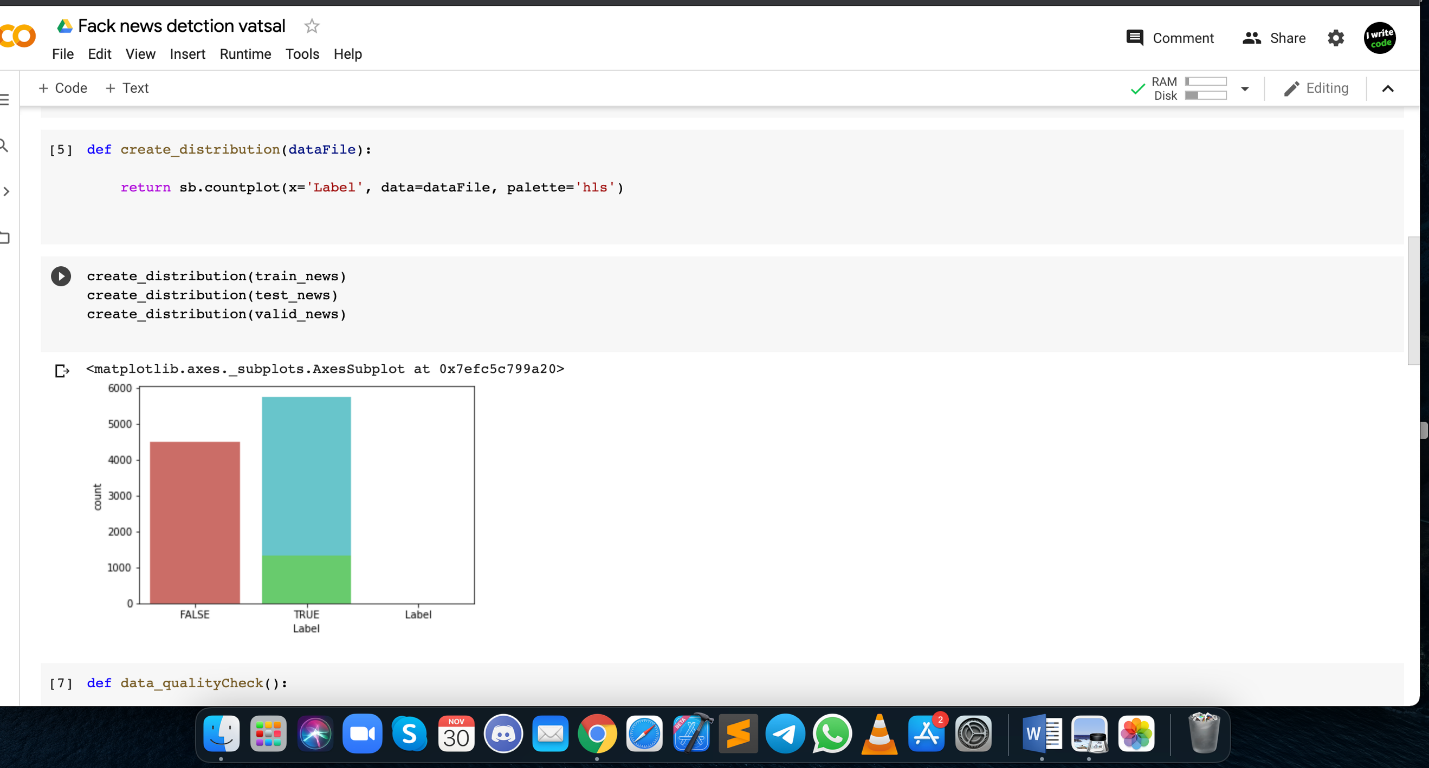
**Random Forest Classifier**

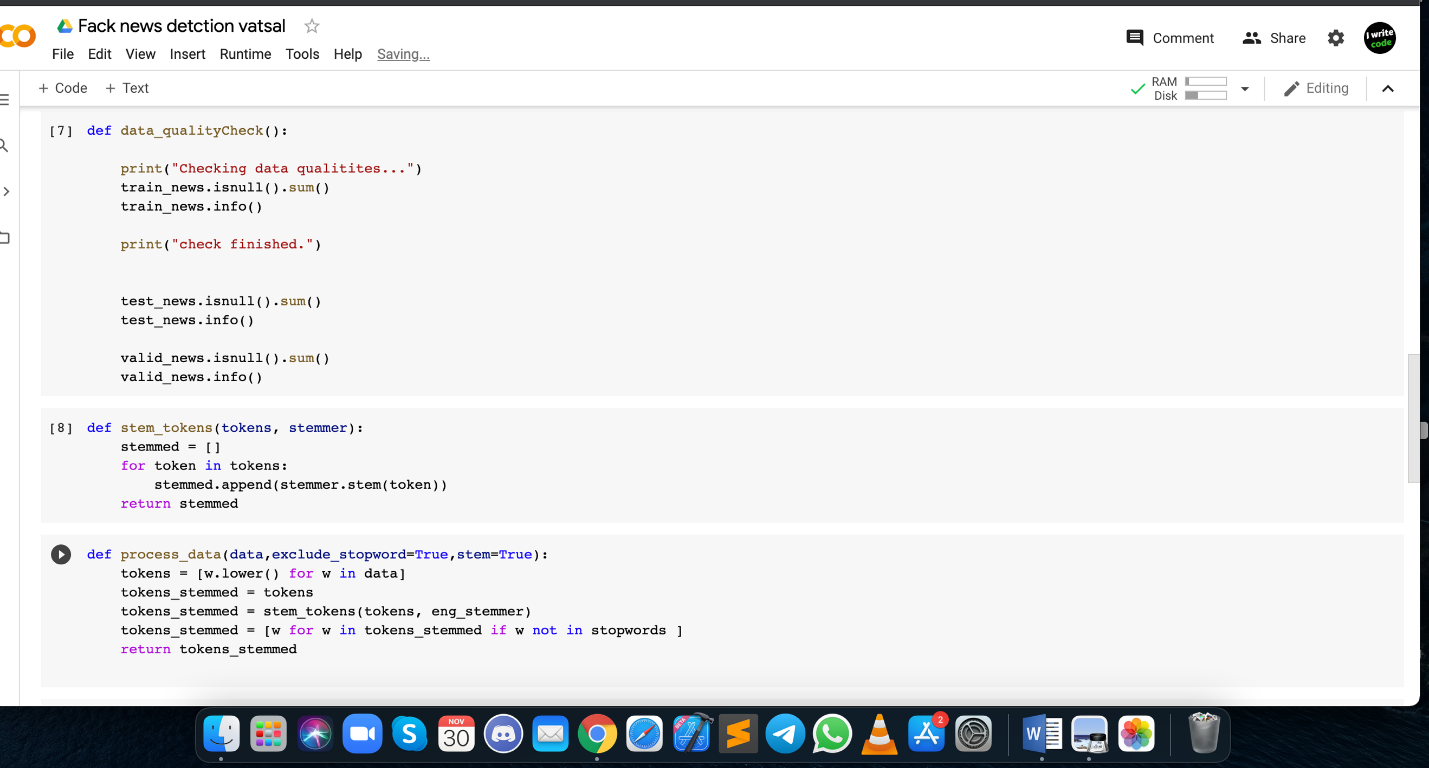
### /Users/vibhavsharma/Downloads/Fake_News_Detection-master 2/images/RF_LCurve.png

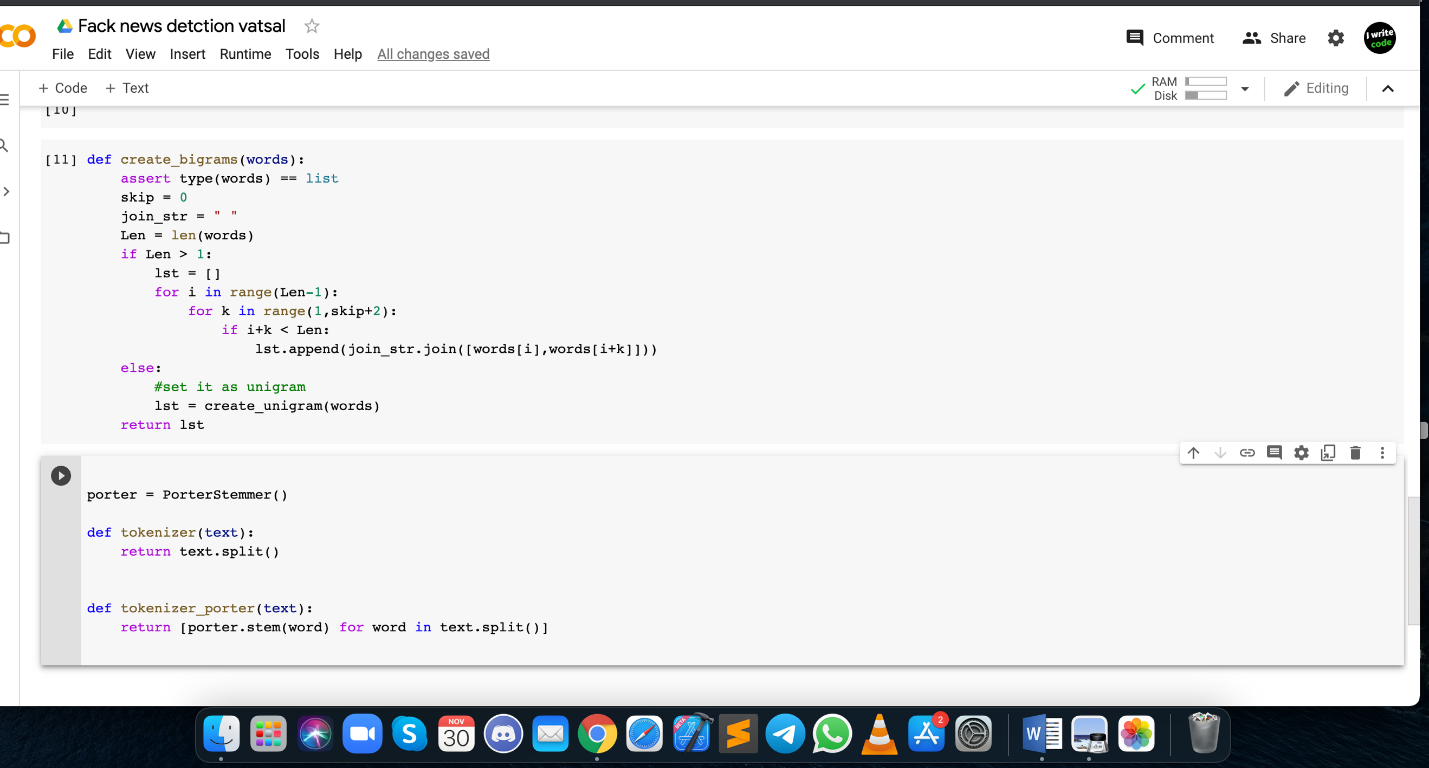
A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled with the max\_sample parameter if bootstrap==true (default), otherwise the whole dataset is used to build each tree.

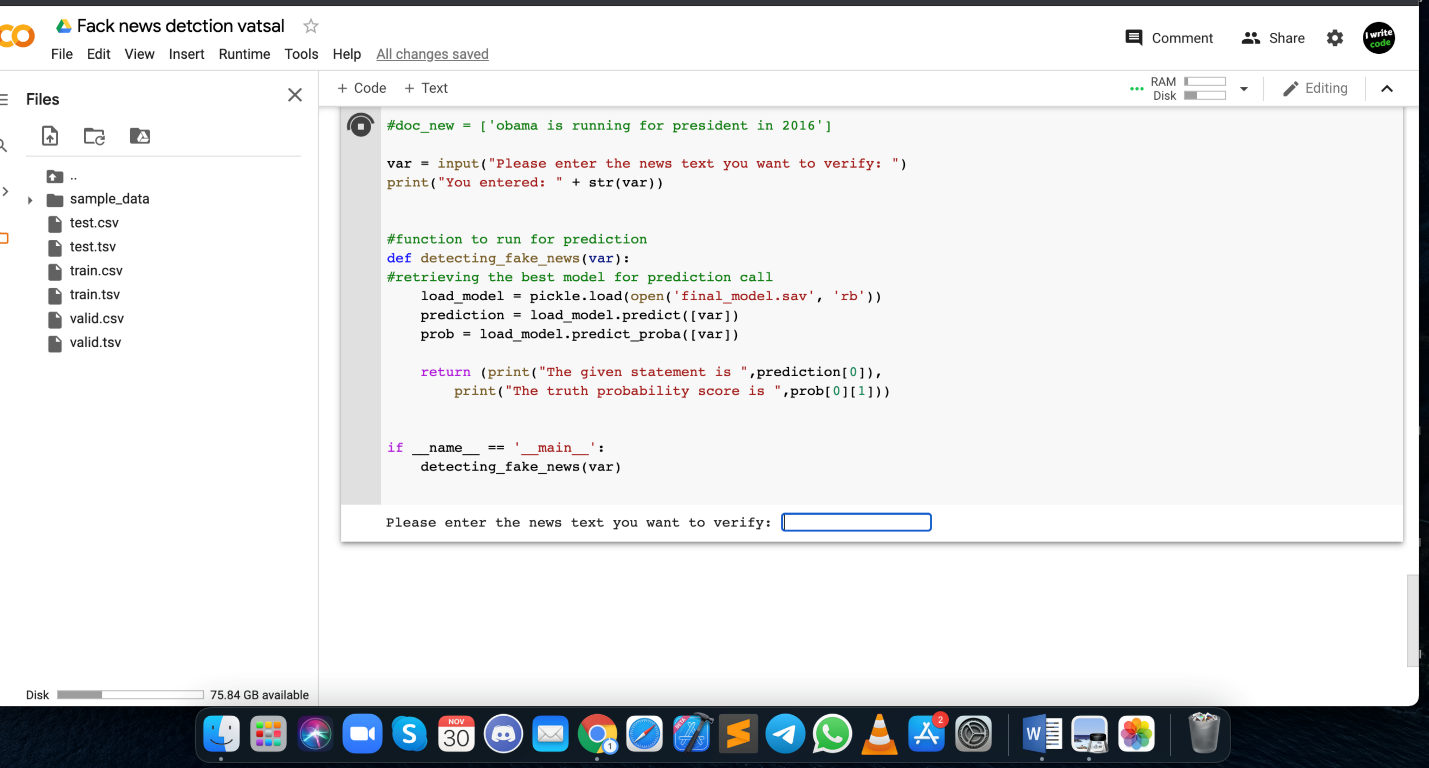
**CODE SECTION:**

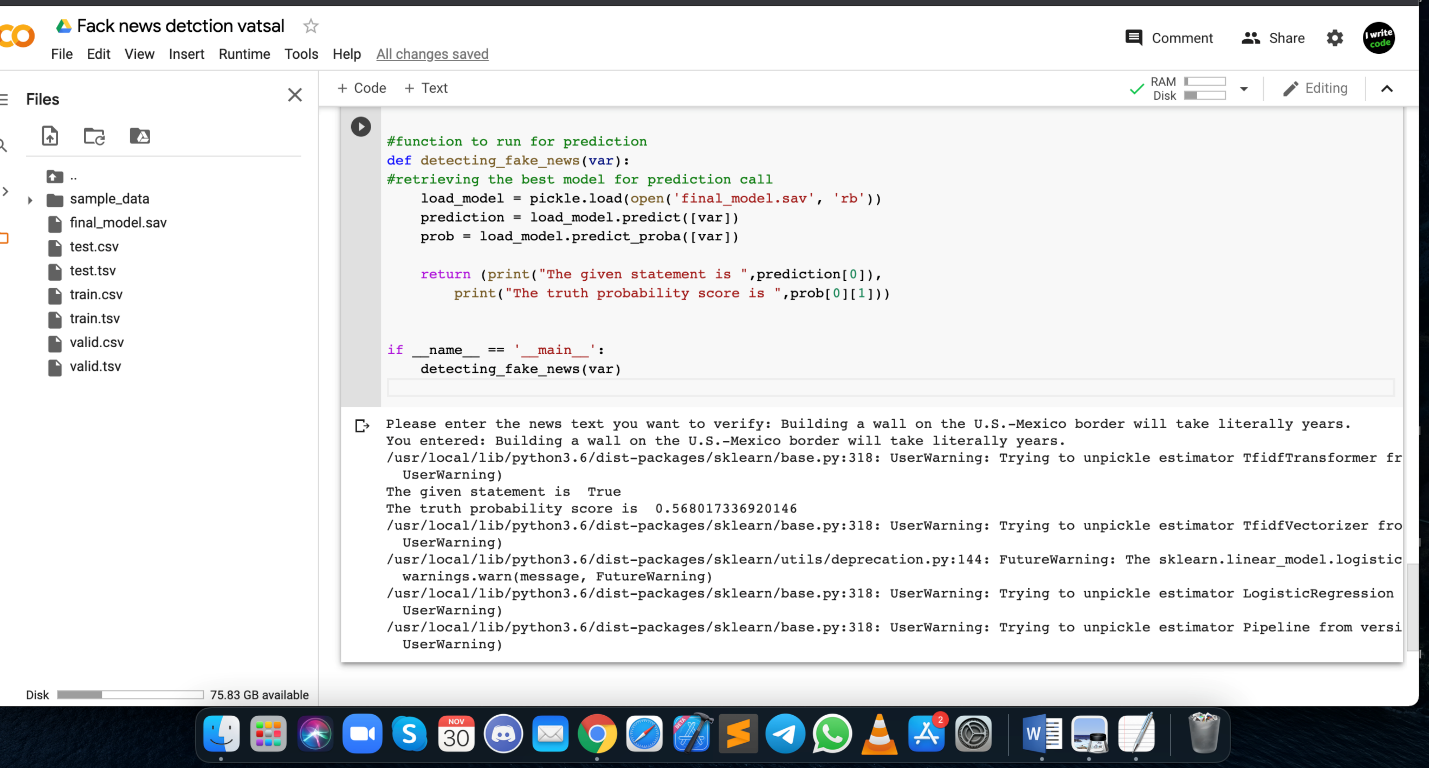






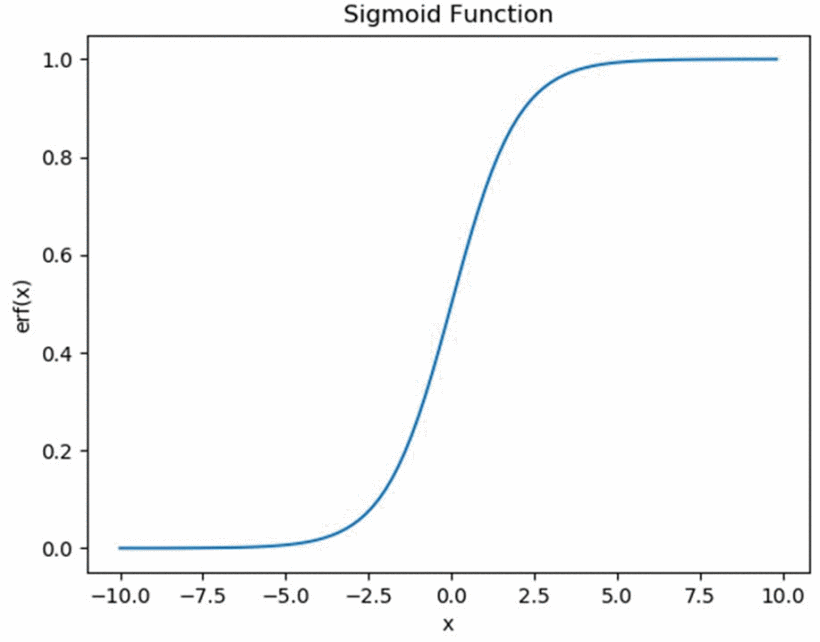






#### **Sigmoid Activation:**

The Sigmoid curve looks like a function in S shape as shown in Figure 1. Sigmoid is used because it exists between 0 and 1. It is therefore particularly used for models in which the probability as an output must be predicted. Since there's only a chance of anything from 0 to 1, Sigmoid's the right decision.



We have implemented inter and intra comparison on five different classification model. Based on the comparison we have also considered the precision and recall value of the models. A combination of five different models is tested in accuracy using N-Gram and Character level vectors in this order to improve the accuracy of prediction. Using scikit learn, these models are implemented to learn from the training data using k- fold (k=2) cross-validation, and then predict using the data set. Then we have evaluated the performance of the models.

**Future Scope of Project**

Basing fake news detection only on supervised models on text have shown not to be enough in all the cases. In order to solve this problem, most of the research focus on additional information such as author information. I think the most successful approach would be automatic fact checking model, that is, compelling the model with some kind of knowledge base, the purpose of the model would then be to extract information for the text and verify the information in the database. The problem with this approach would be that the knowledge base would need to be constantly and manually update to stay up to date.

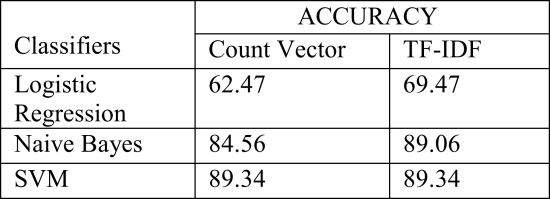
As we can see that our best performing models had an f1 score in the range of 70's. This is due to less number of data that we have used for training purposes and simplicity of our models. For the future implementations, we could introduce some more feature selection methods such as POS tagging, word2vec and topic modeling. In addition, we could also increase the training data size. We will extend this project to implement these techniques in future to increase the accuracy and performance of our models.

**Result analysis:**

Some hypotheses can be made on why same models works very well on one dataset and does not work well on the other one. The first thing we can think of is that the original hypothesis on different styles of writing between fake and reliable news is only verified in one dataset, the Fake News Corpus, and it is the most logical one, as these texts are coming from online newspapers (or pretending to be), and thus capitalize on advertisements for making money. The second dataset, Liar-Liar Corpus is described by its authors as a collection a short sentence coming from various contexts such as political debate, interviews, TV ads and so on, thus it induces a lot of variety in writing style. For instance, it contains a transcription of vocal messages, which have in essence a different style from written one. The data exploration chapter had already given an insight about this fact, as 2D data projection of the Liar-Liar Corpus shows no clear sign of separation, when Fake News Corpus shows one at the first look.

Then Logistic regression model was performed. This time the performance was slightly optimized than before, predicting 74% and 76% respectively on count and word level vectors. Whereas 75% and 76% was the results of Logistic Regression stage.

After testing the data with several supervised methods, deep learning model was chosen as it always gives an efficient output layer based on the weight of input layers. For this task of classifying fake news, Recurrent Neural Network and Long Short-Term Memory Model was taken as a method at this point.



## **Conclusion**

We analyzed a computerized model for checking the verification of news extracted from Twitter which gives general answers for information accumulation and expository demonstration towards fake news recognition. After having an idea from the supervised models, a deep learning-based model is proposed to identify fake news.

The accuracy metric presumably would be altogether improved by methods for utilizing progressively complex model. It is worth noting, that even with the given dataset, only part of the information was used. The current project did not include domain knowledge related features, such as entity-relationships. Future studies could extract name entities from each pair of news headline and news body and analyze their relationships through a knowledge base.

This work exhibits a programmed model for identifying fake news in well-known Twitter strings. Such a model could be important to a huge number of social media users by expanding their own credibility decisions.

The dataset in this examination is relied upon to be utilized for arrangements which utilized machine learning based statistical calculations, for example, Support Vector Machines (SVM), Naive Bayes (NB), Recurrent Neural Network (RNN), Logistic Regression (LR), Long Short Term Memory (LSTM). In this investigation, SVM performs best for characterization technique.

**REFRENCES**

**[1]** Meyer, The Grim Conclusions of the Largest-Ever Study of Fake News.

Show Context[Google Scholar](https://scholar.google.com/scholar?as_q=The+Grim+Conclusions+of+the+Largest-Ever+Study+of+Fake+News&as_occt=title&hl=en&as_sdt=0%2C31)

**[2]** Tandoc Lim. Ling, Defining “Fake News” A typology of scholarly definitions, August 2017.

Show Context[Google Scholar](https://scholar.google.com/scholar?as_q=Defining+%E2%80%9CFake+News%E2%80%9D+A+typology+of+scholarly+definitions&as_occt=title&hl=en&as_sdt=0%2C31)

**[3**] DATASETS- <https://www.aclweb.org/anthology/P17-2067/>

<https://www.aclweb.org/anthology/P17-2067/>, <https://www.aclweb.org/anthology/venues/acl/>

[4] Jeffrey Gottfried and Elisa Shearer. News Use Across Social Medial Platforms 2016. Pew Research Center, 2016

[5] Harry Zhang. The Optimality of Naive Bayes. page 6.

[6] WWF. Wwf 10yearschallenge, 2019.

[7] Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. O’Reilly Media, 2009.