Detection of Fake News Using Machine Learning

Jeevan Sai Koushik Masavarapu

Sree Kavya Ganja

Department of Computer Science and Engineering

Blekinge Institute of Technology

Karlskrona

Sweden

2020

Abstract

The rise of social media has made it easier to share information as it is cheap and easily accessible. The rapid spread of information has led people to consume news from the social media. Contrastingly, this has enabled the spread of fake news. The rapid spread of fake news has capacity for extremely negative impacts on individuals and society. Therefore, detecting fake news on social media has become an emerging area of research in the recent days.

The inability to distinguish between fake and real news calls for a proper mechanism to judge the authenticity of the news received. This study aims at detecting fake articles with the help of a classifier algorithm in Machine Learning – the Passive Aggressive classifier.

A dataset of labeled news articles is taken and are put through a Term frequency (TF)-Inverse Document Frequency (IDF) vectorizer which is used to transform text into meaningful representation of numbers. The output of the TF-IDF vectorizer is passed through the Passive-Aggressive classifier, which forms a model with the data. The model so formed can then be used to check if any article is fake or legitimate.

Once the training of the model is complete, a few potential fake articles are checked against it. The algorithm works well, with an accuracy rate of 97%.

The Passive-Aggressive classifier works remarkably well, judging by the results. Since the classifier is an online learning algorithm, it can be continuously fed data from the web and thereby prove to be an excellent real time article classifier.

**Keywords:**

Algorithm, Classifier Algorithms, Frequency, Fake news, Passive Aggressive

Contents

Abstract 3

Contents 5

List of figures 6

List of acronyms 8

1. Chapter: Introduction 10

2. Chapter: Survey of related work 12

3. Chapter: Problem statement, Objectives and Main contribution 13

4. Chapter: Solution 14

4.1 Modeling 14

4.2 Implementation 16

4.3 Verification 18

5. Chapter: Conclusion and future work 21

Reference 22

List of figures

[Figure 1 Frequency of Potential Fake Articles Encountered 11](#_Toc56933923)

[Figure 2 Model for fake news detection 16](#_Toc56933924)

[Figure 3 Data in the train set 17](#_Toc56933925)

[Figure 4 Confusion Matrix and Classification Report 18](#_Toc56933926)

[Figure 5 Overall Accuracy of Different Algorithms[6] 19](#_Toc56933927)

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

List of acronyms

|  |  |
| --- | --- |
| **Acronym** | **Unfolding** |
| TF | Term Frequency |
| IDF | Inverse Document Frequency |
| GUI | Graphical User Interface |
| SVM | Support Vector Machine |
| RNN | Recurring Neural Networks |
| LSTM | Long Short-Term Memory |
|  |  |
|  |  |

# Chapter: Introduction

Internet is the biggest source of information. Due to the abundance of data available on the internet, one must be cautious about the authenticity of the information. With the rise of social media platforms, people have more avenues to get the news other than the traditional sources such as print and broadcast media[1].

This type of false information has a significant impact on our social life. Websites deliberately publish hoaxes, propaganda and disinformation claiming to be real news to mainly affect the public opinion on some matters (mainly political)[2]. On the social network, the reach and effects of information spread occur at a fast pace, and false information has tremendous potential to cause unrest among millions of users in the present informational climate. Since fake news spreads more rapidly than real news, there is a need to classify news[3].

Figure 1 shows how often people see news articles or new reports on the internet that they perceive to be invented. The research was conducted in Sweden by Statistica in 2017. As can be inferred from the figure, 19% believe that they saw invented news several times per week, 25% believe it was sometime per week, 18% believe it was sometime per month and another 18% believe it was even less. Only 4% of the respondents were sure that they had never seen fake news on the Internet.[12]

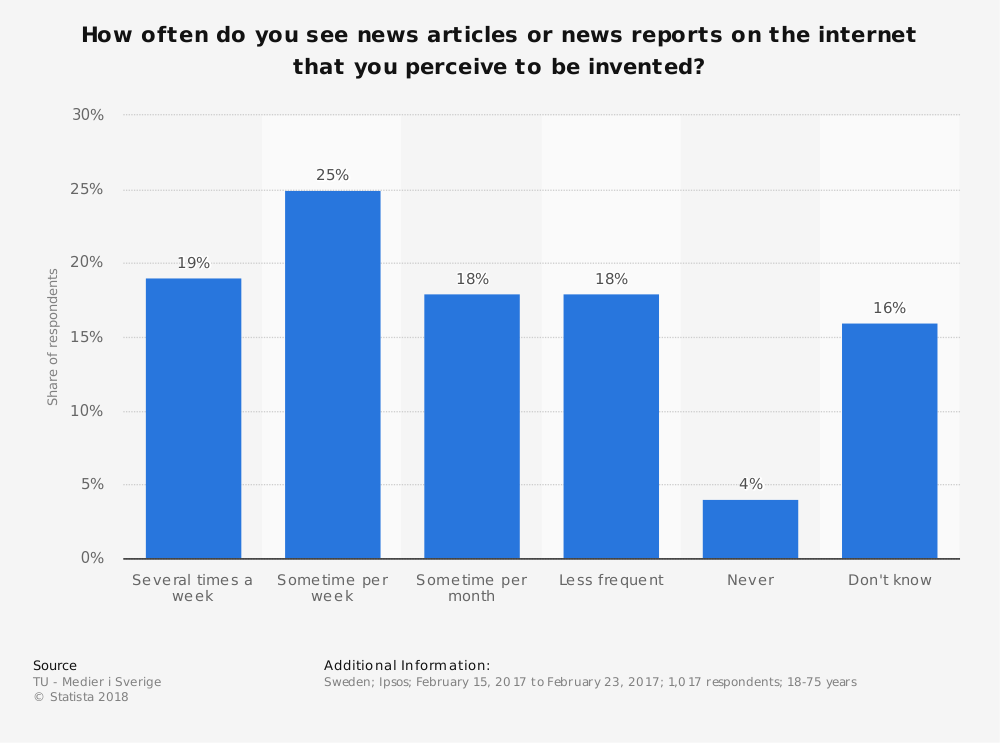


Figure 1 Frequency of Potential Fake Articles Encountered

Inaccurate news articles recurrently push a single point of view, have a livid tone, or make wildly exaggerated claims. In such news articles there can be two types of data:

1. Visual based
2. Linguistics based

Visual based type of fake articles use content that acculturates graphical representation of photoshopped images or videos that grab attention on social media applications which are used frequently to post and share information like Facebook, Instagram, and Twitter.

Linguistics based type of fake news are mainly text and string content. Text is used as a communication system and it consists of features like tone, grammar, and pragmatics for discourse analysis. Some example of linguistics based fake news can be spam or junk emails, blogsites, even popular websites can create their own content and attract users with their authentic presence. In this paper, we are mainly focussed on this linguistics-based type of fake news. In this approach, individual words are treated as single and significant units. Frequency of each word is obtained which are then analysed and aggregated. This approach is language dependent.

A classifier to identify fake news is discussed here. A dataset of news articles labelled true or false is run through the TF-IDF vectorizer, a tool for converting text to numbers, which are then passed through the Passive-Aggressive classifier. Then, the trained model is put under tests and the accuracy is calculated.

# Chapter: Survey of related work

As detection of fake news is classification problem many scientists address this issue by the means of machine learning and artificial intelligence. The algorithms in machine learning and artificial intelligence have started to work much better in the recent days and the hardware is cheaper and bigger datasets are available. There are several studies that discuss about fake news detection. Many journals and articles explained the core of the problem and how automatic detection of fake news has become an urgent problem in recent years [4].

In [5], the authors discuss about general methods that are available for this type of problem. In the paper, two major categories of methods were discussed

* linguistic approaches where the content of the message is extracted and analysed
* Network approaches in which the metadata or structured knowledge network queries can be harnessed.

Some authors discussed about including fake news characterisation on psychology, social theories, existing algorithms from data mining perspective, metrics and representative datasets [6]. This [7] paper proposed a multi-domain visual neural network to fuse pixel domains for detecting fake news which capture patterns of fake news images and detect. There were papers that tackle this problem by probabilistic and geometric machine learning models and compare them to see which model is the best model to detect fake news[8] [6].

This article [2] describes how a particular method works for a particular labelled dataset and support the idea of artificial intelligence for this issue. Some authors used the stance detection technique which predicts whether the document agrees or disagrees with respect to the target claim[9].

Out of all the papers there is one paper which discusses about the field of natural language processing (NLP) for fake news detection where the authors review task formulation and NLP solutions that have developed for this task. They examine the point of restraint and issue definitions, highlighting the importance of NLP for fake news detection [10].

# Chapter: Problem statement, Objectives and Main contribution

The world of internet is a cesspool of fake news. Fake news aims to persuade the reader to trust false information which deems these articles difficult to perceive. The rate of producing digital news is large and quick, running daily at every second[11]. Such news creates a lot of confusion, may trigger riots between people and sometimes even war between two nations. The lack of proper mechanism to distinguish fake news from actual news has caused a lot of problems to the common man.

The main objective of this report is how to ascertain the degree of authenticity of a news article. Given a dataset of news articles how many of them would be real and how many of them would be false. Using python and machine learning algorithms a model can be built that can detect fake articles.

The main contribution of this paper is a model that can differentiate between fake news and real news by applying classification techniques and frequency vectorizer on a dataset of public news. Then, the model is tested for accuracy. We are using the TF-IDF vectorizer which converts text into meaningful numerical information and apply Passive- Aggressive classifier to fit the model which tells us if the classification is correct or wrong.

# Chapter: Solution

From the above discussion, we already came to know that news can be either fake or real. To classify some news, we need to understand the problem definition first, then we go for our model and evaluate the result. Machine Learning provides many algorithms with which we can detect fake news some of them are good and some of them work on an average scale.

In this paper, we are using a classifier algorithm called Passive-Aggressive classifier. Passive-Aggressive Algorithms are a family of algorithms which are used for large scale learning. Passive indicates if the classification is correct and aggressive signifies that classification is incorrect.

To process the dataset, we are using the TF-IDF vectorizer which is a vectorization technique in machine learning. Vectorization is a technique in NLP to convert text into numerical data as computers do not understand our language, it only understands numbers. TF-IDF is a measure that is used to score the importance of a word based on how often it appears in a document or a given collection of documents.

## Modeling

In this paper, we are using a dataset from Kaggle website which should first be pre-processed, which includes cleaning and filtering data. The next step after data pre-processing will be using the TF-IDF vectorizer to convert text into numerical data and initialize a Passive-Aggressive classifier to fit the model.

**Term Frequency (TF) – Inverse Document Frequency (IDF) Vectorizer**

**TF-IDF is a vectorization technique in machine learning to convert text into a meaningful representation of numbers. It assigns a metric to represent the presence of that particular term. The Intuition of TF-IDF is if a word appears frequently in a document it should be given a higher priority. If a word appears in too many other documents it is probably not a unique identifier, therefore we should assign a lower priority.**

**We assume that a higher number of repetitions of a word would mean greater importance in the given text. We standardize the occurrence of the word with the size of the document hence term frequency.**

**Tf(w) =**

**While calculating the term frequency each word is given equal priority. There may be a word that has high frequency across the documents and therefore it would contribute less in deriving the meaning of the document. Words like a, an, the, Etc might suppress the weights of more meaningful words. To minimize this effect, term frequency is marked down by a factor called Inverse document frequency.**

**Idf(w) =**

**Tf-Idf(w) = Tf(w) \* Idf(w)**

**Passive – Aggressive Classifier**

**Passive – Aggressive algorithms are a family of algorithms that can be used for large-scale learning. It is an online learning algorithm, where the input data comes in sequential order and machine learning model is updated step by step. Batch learning algorithms, as opposed to online learning algorithms, require entire dataset to be input at once, which is very inconvenient when there is a large amount of data and it is also impractical to train the entire dataset because of the amount of data. In such cases, these online learning algorithms come handy because new data is being added every second and it would be easy to dynamically read the data.**

Intuitively, passive signifies that if the classification is correct, we should keep the model, and, aggressive signifies that if the classification is not correct, update the model to adjust to more misclassified examples. Unlike other algorithms, it does not converge, rather it makes updates to correct the loss.

Figure 2 Model for fake news detection

## ****Implementation****

Python and machine learning algorithms are used to build a model to detect fake news. Since this is a classification problem, we require a set of news articles. In this paper, a dataset which contains a training set and a testing set is used. The training data contains the following attributes:

Id: Uniqued id for every news article.

Title: Title of the news article.

Author: Author of the article.

Text: Text in the article.

Label: Marks the article whether itis true or false.

1: Unreliable/Fake

0: Reliable/True

The test data also contains the same attributes but it does not contain label. Below is the image of how the data looks.

image


Figure 3 Data in the train set

The most important step of creating ML model is to prepare the data which includes cleaning, filtering and removing outiliers in the data. As this data is from an online source there can be few empty articles as well which should be removed else a lot of noise would be created in our model. As mentioned above, we are using a training set and a testing set, this is to make suke sure that this model performs well on any new dataset.

Then we initialize a TF-IDF vectorizer with maximum document frequency 0.7 and stop words. The words with frequeny greater than 0.7 will be discarded as those words might not be unique identifiers and should be given lower priority. This TF-IDF changes the collection of raw documents into a matrix of TF-IDF features. We set the Passive-Aggressive classifier on the trainset and predict on test set with TF-IDF vectorizer. We can tell how well the model works based on performance metrics like accuracy, presicion, recall and f1-score.

## Verification

Firstly, let us look at the confusion matrix we obtained after implementing the algorithm.

Chart, waterfall chart

Description automatically generated

Figure 4 Confusion Matrix and Classification Report

From the above image, we can say that considering accuracy this model is one of the most effective ways to detect fake news articles. But accuracy is not end-all model metric to use when selecting the best model. The other metrics we can use are precision, recall and f1-score.

Precision talks about how accurate the model is, out of the predicted positive how many are actual positive. Precision is a good measure to design when the false positives can change facts to fake.

Precision =

Recall calculates how many of the actual positives our model capture through labelling it as True Positive. Applying the same understanding, we know that Recall shall be the model metric we use to select our best model when there is a high cost associated with false negative.

Recall =

 F1 which is a function of Precision and Recall. F1 Score is needed when you want a balance between Precision and Recall. So, the difference between F1 Score and Accuracy is, from the previous discussion we have seen that accuracy can be largely contributed by a large number of True Negatives which in most of the circumstances, we do not focus on much whereas False Negative and False Positive usually has business costs (tangible & intangible) thus F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall.

F1 =

From the above classification report and based on the performance metrics we can say that this algorithm can effectively judge a news article whether it is a fake or a real article.

Chart, bar chart

Description automatically generated

Figure 5 Overall Accuracy of Different Algorithms[6]

According to the paper [6], where different classification algorithms were used to detect fake news and compare them with each other to see which algorithm works the best. The authors worked on algorithms like Naïve bayes, SVM, logistic regression and RNN, LSTM. The graph in figure 5 depicts the accuracy of the algorithms. From the graph we can see that the accuracy of these algorithms is average so, in conclusion the classifier algorithm used in this paper can effectively detect fake news.

# Chapter: Conclusion and future work

Fake News is a powerful instrument in swaying public opinion. By tackling it with the right model a lot of damage can be prevented. The algorithm discussed in this paper can effectively detect fake news.

For fake news detection the dataset keeps growing as the time passes thus the passive aggressive classifier is the best choice of algorithm because it is an online algorithm, and it can take data continuously one after other. The accuracy of this algorithm is 97%

There is scope for future work. We intend to develop a GUI (Graphical User Interface) where one can paste a piece of news and get its classification result.

Reference

[1] ‘Tackling The Fake News Phenomenon - Legislation And Other Factors - Free Essay Example’, *EDUZAURUS*. https://eduzaurus.com/free-essay-samples/tackling-the-fake-news-phenomenon-legislation-and-other-factors/ (accessed Oct. 13, 2020).

[2] M. Granik and V. Mesyura, ‘Fake news detection using naive Bayes classifier’, in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, May 2017, pp. 900–903, doi: 10.1109/UKRCON.2017.8100379.

[3] N. Smitha and R. Bharath, ‘Performance Comparison of Machine Learning Classifiers for Fake News Detection’, in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Jul. 2020, pp. 696–700, doi: 10.1109/ICIRCA48905.2020.9183072.

[4] S. Kumar and N. Shah, ‘False Information on Web and Social Media: A Survey’, *ArXiv180408559 Cs*, Apr. 2018, Accessed: Oct. 14, 2020. [Online]. Available: http://arxiv.org/abs/1804.08559.

[5] N. K. Conroy, V. L. Rubin, and Y. Chen, ‘Automatic deception detection: Methods for finding fake news’, *Proc. Assoc. Inf. Sci. Technol.*, vol. 52, no. 1, pp. 1–4, 2015, doi: 10.1002/pra2.2015.145052010082.

[6] Abdullah-All-Tanvir, E. M. Mahir, S. Akhter, and M. R. Huq, ‘Detecting Fake News using Machine Learning and Deep Learning Algorithms’, in *2019 7th International Conference on Smart Computing Communications (ICSCC)*, Jun. 2019, pp. 1–5, doi: 10.1109/ICSCC.2019.8843612.

[7] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, ‘Exploiting Multi-domain Visual Information for Fake News Detection’, in *2019 IEEE International Conference on Data Mining (ICDM)*, Nov. 2019, pp. 518–527, doi: 10.1109/ICDM.2019.00062.

[8] K. Poddar, G. B. Amali D., and K. S. Umadevi, ‘Comparison of Various Machine Learning Models for Accurate Detection of Fake News’, in *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Mar. 2019, vol. 1, pp. 1–5, doi: 10.1109/i-PACT44901.2019.8960044.

[9] M. Mohtarami, R. Baly, J. Glass, P. Nakov, L. Marquez, and A. Moschitti, ‘Automatic Stance Detection Using End-to-End Memory Networks’, *ArXiv180407581 Cs*, Apr. 2018, Accessed: Oct. 14, 2020. [Online]. Available: http://arxiv.org/abs/1804.07581.

[10] R. Oshikawa, J. Qian, and W. Y. Wang, ‘A Survey on Natural Language Processing for Fake News Detection’, *ArXiv181100770 Cs*, Mar. 2020, Accessed: Oct. 14, 2020. [Online]. Available: http://arxiv.org/abs/1811.00770.

[11] ‘Detecting fake news online’, *Wikipedia*. Oct. 04, 2020, Accessed: Oct. 24, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Detecting\_fake\_news\_online&oldid=981832499.

1. “ Sweden: Reading Fake News on the Internet 2017 | Statista.” Accessed November 22, 2020. https://www.statista.com/statistics/682020/survey-on-reading-fake-news-on-the-internet-in-sweden-by-frequency/.

.