



Fake News Detection:

Submitted by :Pramila Dalvai,

Sourav Taleda

Contents

Abstract
Introduction
Problem Definition
Literature Survey
Design Methodologies
References





Abstract

This Project comes up with the applications of NLP (Natural Language Processing) techniques for detecting the 'fake news', that is, misleading news stories that comes from the non-reputable sources. Only by building a model based on a count vectorizer (using word tallies) or a (Term Frequency Inverse Document Frequency) tfidf matrix, (word tallies relative to how often they're used in other articles in your dataset) can only get you so far. But these models do not consider the important qualities like word ordering and context.

It is very possible that two articles that are similar in their word count will be completely different in their meaning. The data science community has responded by taking actions against the problem. There is a Kaggle competition called as the “Fake News Challenge” and Facebook is employing AI to filter fake news stories out of users' feeds. Combatting the fake news is a classic text classification project with a straight forward proposition

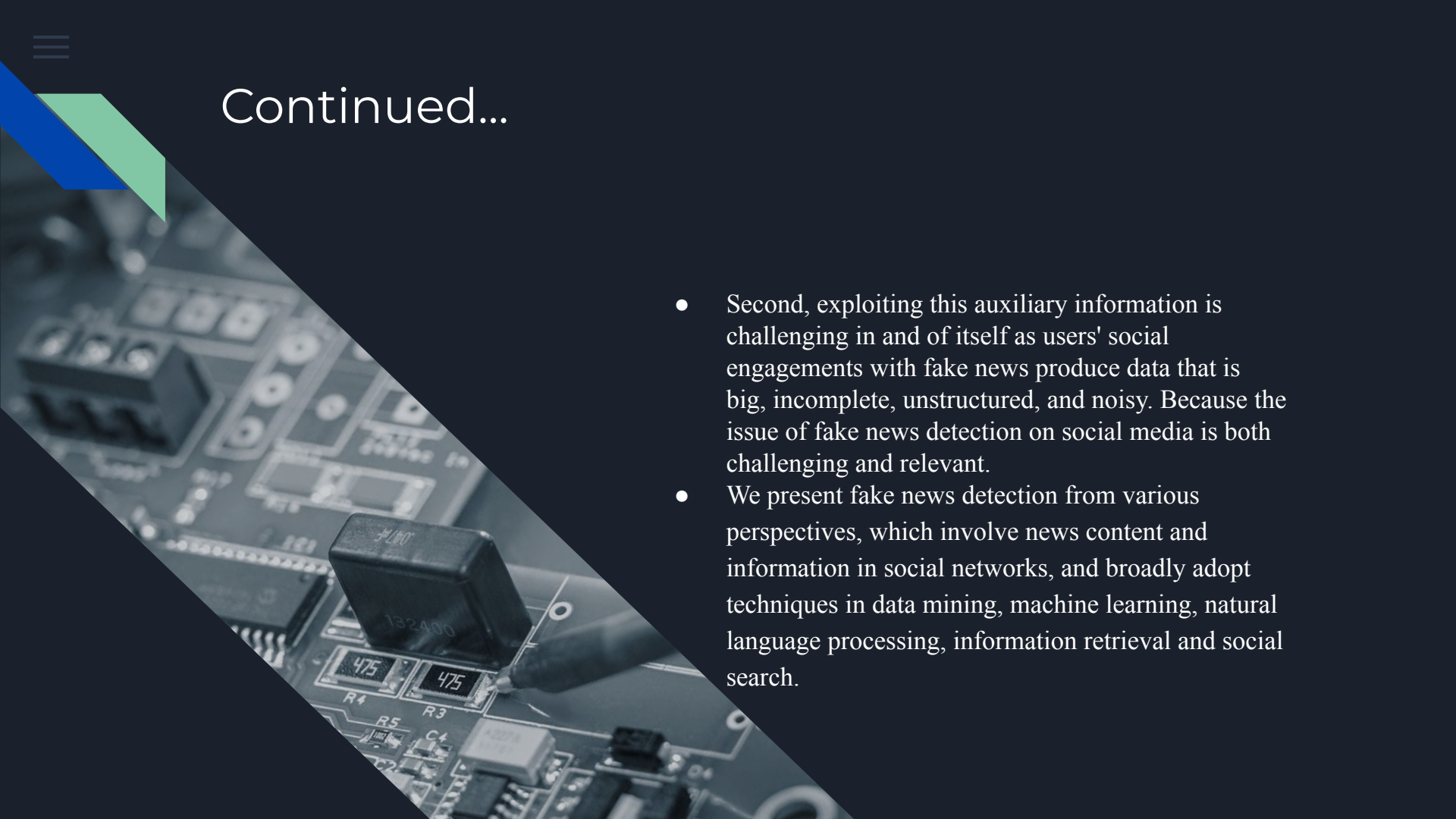


Introduction

- Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news from social media.
- On the other hand, it enables the wide spread of "fake news", i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention.
- Fake news detection on social media presents unique characteristics and challenges that make existing detection algorithms from traditional news media or not applicable. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content; therefore, we need to include auxiliary information, such as user social engagements on social media, to help make a determination.



Continued...

- 
- Second, exploiting this auxiliary information is challenging in and of itself as users' social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. Because the issue of fake news detection on social media is both challenging and relevant.
 - We present fake news detection from various perspectives, which involve news content and information in social networks, and broadly adopt techniques in data mining, machine learning, natural language processing, information retrieval and social search.



Problem Statement

All the news that we hear is not real sometimes. This advanced python project of detecting fake news deals with fake and real news. The main objective is to detect the fake news, which is a classic text classification problem with a straightforward proposition. It is needed to build a model that can differentiate between “Real” news and “Fake” news

LITERATURE SURVEY

1. Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature.

2. N. J. Conroy, V. L. Rubin, and J. A. H. H. "Finding fake news: A review of the proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015.

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts.



CONTINUED...

3. Fake news detection using Data mining techniques: Fake News Detection on Social Media: A Data Mining Perspective: ACM SIGKDD Explorations Newsletter: Vol 19, No 1

The issue of fake news detection on social media is both challenging and relevant, we conducted this survey to further facilitate research on the problem. In this survey, we present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. We also discuss related research areas, open problems, and future research directions for fake news detection on social media.

4. Shlok Gilda, Department of Computer Engineering, 2017 IEEE 15th Student Conference on Research and Development (SCORED) –

Evaluating Machine Learning Algorithms for Fake News Detection

5. Fake news detection using Context free Grammar:

Conroy, Rubin, and Chen outlines several approaches that seem promising towards the aim of perfectly classify the misleading articles. They note that simple content-related n-grams and shallow parts-of-speech (POS) tagging have proven insufficient for the classification task, often failing to account for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars (PCFG) have been shown to be particularly valuable in combination with n-gram methods



CONTINUED..

6. Feng, Banerjee, and Choi ^[2] are able to achieve 85%-93% accuracy in deception related classification tasks using online review corpora.

7. Rubin, Lukoianova and Tatiana analyze rhetorical structure using a vector space model with similar success. Ciampaglia et al. employ language pattern similarity networks requiring a pre-existing knowledge base.

Rather, these methods have been shown useful only in tandem with more complex methods of analysis.

8. Xia Hu, Jiliang Tang, and Huan Liu. In AAAI'14, pages 59–65, 2014 - “Spam Detection”

Online social spammer detection



CONTINUED..

9. Fake news detection using CSI models: Authors: Natali Ruchansky, Sungyong Seo, Yan Liu In this work, they propose a model that combines all three characteristics for a more accurate and automated prediction of news. Specifically, we incorporate the behavior of both parties, users and articles, and the group behavior of users who propagate fake news. Motivated by the three characteristics, They proposed a model called CSI which is composed of three modules: Capture, Score, and Integrate. The first module is based on the response and text; it uses a Recurrent Neural Network to capture the temporal pattern of user activity on a given article. The second module learns the source characteristic based on the behavior of users, and the two are integrated with the third module to classify an article as fake or not. Experimental analysis on real-world data demonstrates that CSI achieves higher accuracy than existing models, and extracts meaningful latent representations of both users and articles.

10. Fake news detection using automation: <https://arxiv.org/abs/1708.07104>:

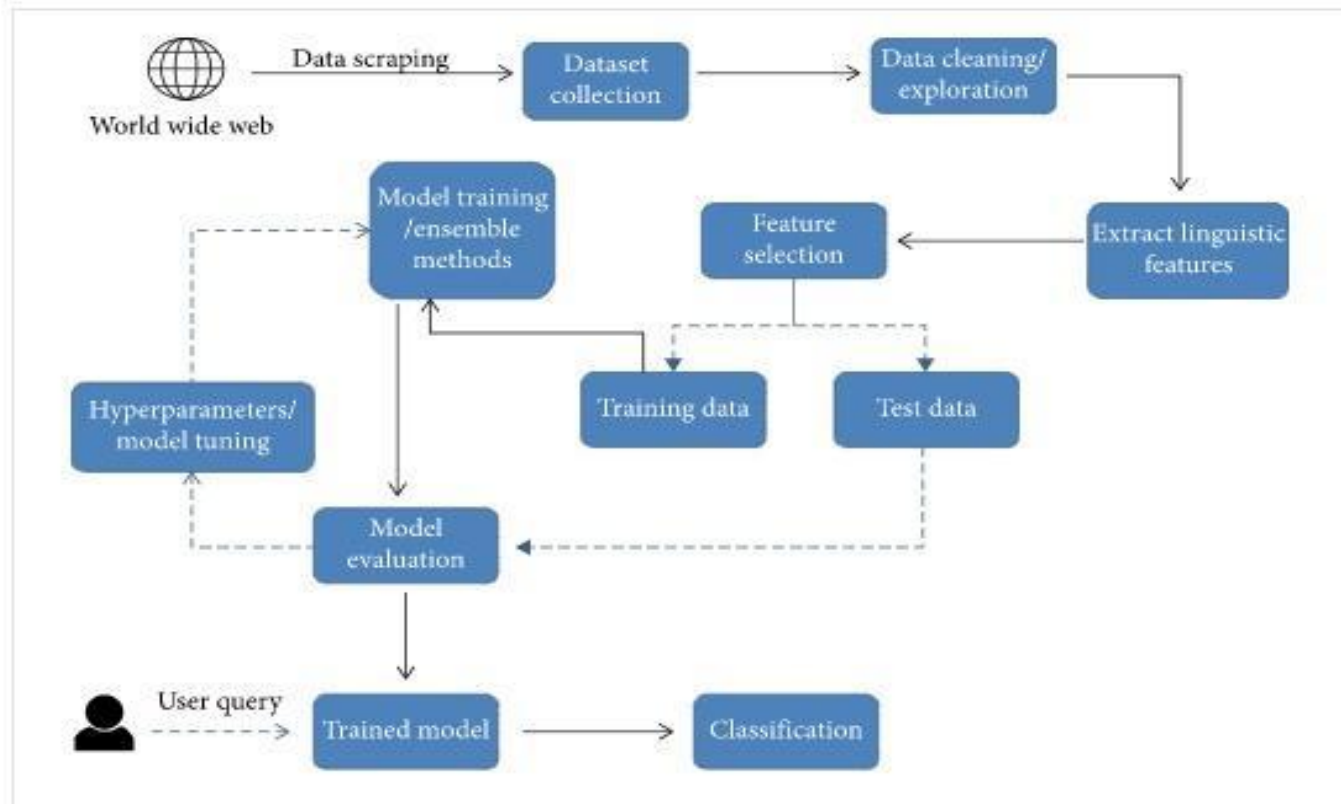
They introduce two novel datasets for the task of fake news detection, covering seven different news domains. We describe the collection, annotation, and validation process in detail and present several exploratory analysis on the identification of linguistic differences in fake and legitimate news content. Second, They conduct a set of learning experiments to build accurate fake news detectors. In addition, They provide comparative analyses of the automatic and manual identification of fake news.

Design Methodologies

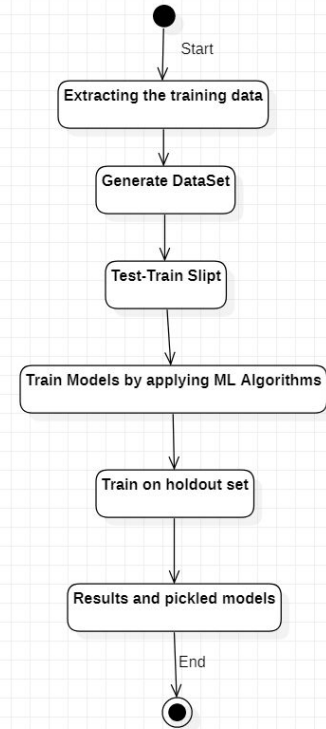
1. High Level Design – Project Structure



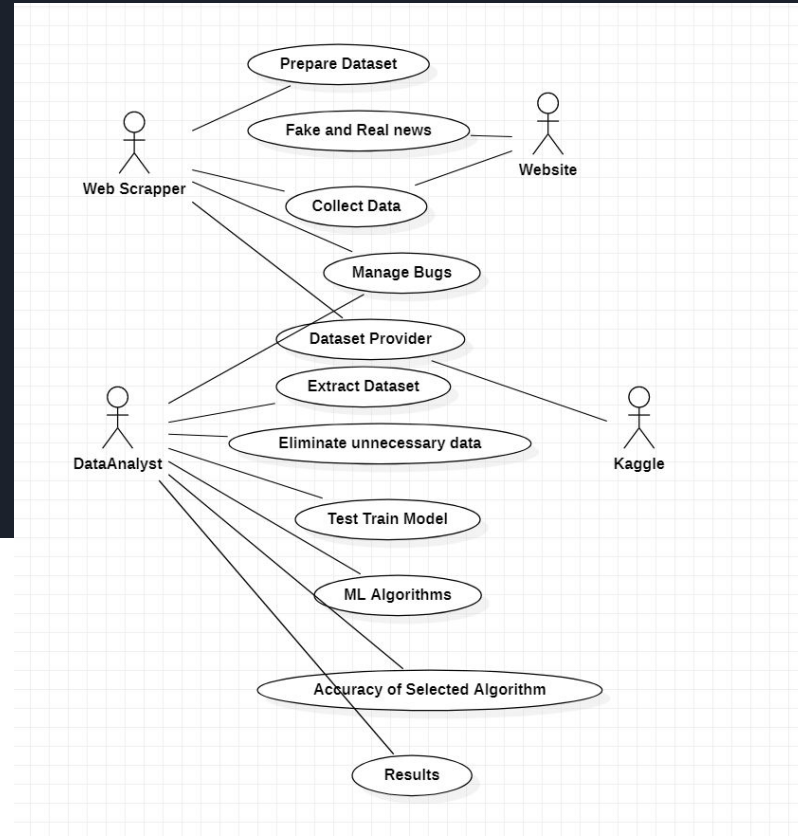
1. Data Flow Diagram



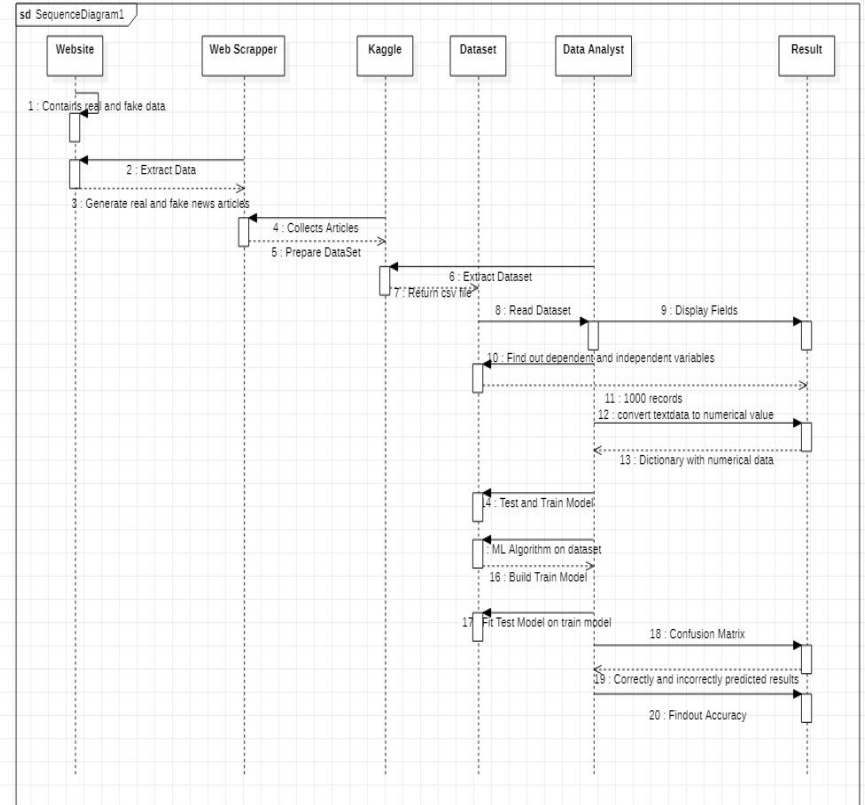
ACTIVITY DIAGRAM:



USECASEDIAGRAM:



SEQUENCE DIAGRAM





REFERENCES

1. Blake E Ashforth and Fred Mael. 1989. Social identity theory and the organization. *Academy of management review* 14, 1 (1989), 20–39.
2. Xin Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 601–610.
3. S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*, Association for Computational Linguistics, 2012, pp. 171–175.
4. Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4, 2015.
5. Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: Detecting and preventing clickbaits in online news media. In *ASONAM’16*