FAKE NEWS DETECTION USING AN ENHANCED SUPPORT VECTOR MACHINE WITH SENTIMENT ANALYSIS

BY

OYENIYI, Samuel Aduralere
MTech/SICT/2017/6787

DEPARTMENT OF CYBER SECURITY
FEDERAL UNIVERSITY OF TECHNOLOGY
MINNA

SEPTEMBER, 2021

FAKE NEWS DETECTION USING AN ENHANCED SUPPORT VECTOR MACHINE WITH SENTIMENT ANALYSIS

BY

OYENIYI, Samuel Aduralere
MTech/SICT/2017/6787

A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA, NIGERIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF TECHNOLOGY IN CYBER SECURITY SCIENCE

SEPTEMBER, 2021

ABSTRACT

Fake news is anything but a new idea, yet it is a usually happening wonder in current occasions. The outcome of phony news can go from being simply irritating to affecting and deceiving social orders or even countries. In previous literature, comparing Support Vector Machine (SVM) and machine learning for text categorization with Sentiment analysis suffers setbacks of low performance and lack in terms of the range of evaluated models and the diversity of the used datasets. The aim of this study is to Enhance Support Vector Machine using Sentiment Analysis for easy detection of rumour on social media platform using individual twitter account. This was achieved by collecting relevant data for performing fake news detection, using SVM and sentiment analysis for easy detection. The results obtained from the study indicate that the technique performed optimally in fake news detection with the accuracy of 98% and a false alarm rate of 0.02. This reveals that the enhancement of SVM with sentiment analysis for fake news detection enhances the performance of the detection model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

SVM	SUPPORT VECTOR MECHINE
SA	SENTIMENT ANALYSIS
NLP	NATURAL LANGUAGE PROCESSING
RNN	RECURRENT NEURAL NETWORK
EANN	EVENT ADVERSARIAL NEURAL NETWORKS
NB	NAÏVE BAYES
DT	DECISION TREE
ML	MACHINE LEARNING
TFIDF	TERM FREQUENCY INVERSE DOCUMENT FREQUENCY
CNN	CONVOLUTIONAL NEURAL NETWORK
KIM	KNOWLEDGE AND INFORMATION MANAGEMENT

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

The rise of fake news during the 2015 Presidential Election highlighted not only the dangers of the effects of fake news but also the challenges presented when attempting to separate fake news from real news Lion Gu *et al.* (2017). Fake news may be a relatively new term but it is not necessarily a new phenomenon. Fake news has technically been around at least since the appearance and popularity of one-sided, partisan newspapers in the 19th century Lion Gu *et al.* (2017). However, advances in technology and the spread of news through different types of media have increased the spread of fake news today.

As such, the effects of fake news have increased exponentially in the recent past and something must be done to prevent this from continuing in the future. The three most prevalent motivations for writing fake news has been identified and only one has been chosen as the target for this project as a means to narrow the search in a meaningful way. The first motivation for writing fake news, which dates back to the 19th century one-sided party newspapers, is to influence public opinion. The second, which requires more recent advances in technology, is the use of fake headlines as click bait to raise money. The third motivation for writing fake news, which is equally prominent yet arguably less dangerous, is satirical writing Shu. *et al.* (2017).

Detecting fake news on social media poses several new and challenging research problems. The increasing popularity of social media widely used for political purposes, the problem of fake news has gained more importance in recent years. It also imposes a great detection challenge. Manual fact-checking in many cases, is difficult, time-consuming, and expensive Cooper. (2017). Therefore, the community has been looking for various automated detection solutions that would speed up this process. In recent years, different Natural Language Processing (NLP) methods have been proposed to solve the fake news detection problem Pedregosa *et al.* (2011). This study examines the use of support vector machine (SVM) and sentiment analysis (SA) for fake news detection.

The sentiment is usually formulated as a two-class classification problem, positive and negative Al-Moslmi *et al* (2018). Sometimes, time is more precious than money, therefore instead of spending time in reading and figuring out the positivity or negativity of a review, automated techniques can be used for Sentiment Analysis. The basis of Sentiment Analysis is determining the polarity of a given text at the document, sentence or aspect level, whether the expressed opinion in a document, a sentence or an entity aspect is positive or negative. More specifically, the goals of SA are to find opinions from reviews and then classify these opinions based upon polarity. According to Weiss *et al* (2007), there are three major classifications in SA, namely: document level, sentence level, and aspect level. Hence, it is important to distinguish between the document level, sentence level, and the aspect level of an analysis process that will determine the different tasks of SA. The document level considers that a document is an opinion on its aspect, and it aims to classify an opinion document as a negative or positive opinion. The sentence level using SA aims to setup opinion stated in every sentence. The aspect level is based on the idea that an opinion consists of a sentiment (positive or negative), and its SA aims to categorize the sentiment based on specific aspects of entities.

Sentiment Analysis technique is applied to classify the documents as real positive and real negative reviews or fake positive and fake negative reviews. Fake negative and fake positive reviews by fraudsters who try to play their competitors existing systems can lead to financial gains for them. This, unfortunately, gives strong incentives to write fake reviews that attempt to intentionally mislead readers by providing unfair reviews to several products for the purpose of damaging their reputation. Detecting such fake reviews is a significant challenge. For example, fake consumer reviews in an e-commerce sector are not only affecting individual consumers but also corrupt purchaser's confidence in online shopping.

Existing methods for detecting fake news can be generally categorized into two categories based on the heterogeneity of the data, i.e., single-modal based and multi-modal based. In single-modal based methods, single type of, often textual, information such as contents, profiles and descriptions are used. For instance, Tandoc Jr, *et al.* (2018), exploits the linguistic features of misinformation by comparing real news with fake news.

Similarly, Shu *et al.* (2018), conducts fake news detection by evaluating the consistency between the body and its claim given a news article. Note that as the content type of news is not limited to only text, other data types such as images or videos could also be utilized. In particular, in social media, fake news often comes with multi-modality data including manipulated images, fake videos, or user comments, all of which provide rich information for detecting fake news. As such, multi-modality based fake news detection has gained increased attentions. For example, Zhou (2019), proposes a Recurrent Neural Network (RNN) with an attention mechanism to fuse multi-modal data from tweets for rumor detection. In addition, Wang (2018), proposes the Event Adversarial Neural Networks (EANN), which integrates multimodal features of images and texts and removes event-specific features via discriminator.

In machine learning-based techniques, algorithms such as SVM, Naive Bayes (NB), and Decision tree (DT-J48) are applied for the classification purposes. SVM is a type of learning algorithm that represents supervised machine learning approaches, and it is an excellent successful prediction approach. The SVM is also a robust classification approach Wang, *et al.* (2018), introduces a survey on different applications and algorithms for SA, but it is only focused on algorithms used in various languages, and the researchers did not focus on detecting fake reviews.

Sentiment Analysis (SA) or Opinion Mining can be described as a set of techniques used to analyze opinionated text that contains people's opinion towards different entities such as products, services, organizations or individuals, among others Gandomi and Haider, (2015). Textual data on the Internet is growing at a rapid pace and many companies and organizations are attempting to use this data stream to extract people's point of view regarding their products Sheela (2016). Notably both SA and Business Intelligence needed to follow a process to extract proper conclusions from the data.

Sentiment Classification techniques can be roughly divided into the hybrid approach, machine learning approach and the lexicon-based approach Maynard and Funk (2011). The Machine Learning (ML) approach applies ML algorithms and uses linguistic features. The lexicon-based approach relies on a sentiment lexicon, which is a collection of precompiled and known sentiment terms. More detailed, it can be divided into the dictionary-based approach and corpus-based approach which use statistical and semantic methods to find sentiment polarity, respectively. The hybrid approach combines both approaches played a critical role in the majority of the methods which is very common with sentiment lexicons Medhat *et al,* (2014).

The sentiment classification method using lexicon-based approach can be divided into the dictionary-based approach and the corpus-based approach, which depends on finding the sentiment lexicon. The dictionary-based approach begins with finding sentiment or opinion seed words and then searches the dictionary of their synonyms and antonyms. The corpus based approach depends on a seed list of opinion words and then finds other sentiment words using statistical or semantic methods in a large corpus to help in finding sentiment words with context-specific orientations.

Machine learning approaches are the dominant approaches in the sentiment analysis task Read (2005). It depends on the features of data when used to sentiment analysis. There are three approaches: unsupervised learning methods, supervised learning methods and reinforcement learning method. The supervised methods make use of a large number of labeled training documents. The unsupervised methods are used when it is difficult to find these labeled training documents when they do not exist.

The current project involves utilizing sentiment analysis and support vector machine to expose documents that are, with high probability, of fake news articles. Current automated approaches to this problem are centered on a "blacklist" of authors and sources that are known producers of fake news. But, what about when the author is unknown or when fake news is published through a generally reliable source? In these cases it is necessary to rely simply on the content of the news article to make a decision on whether or not it is fake. By collecting examples of both real and fake news and training a model, it should be possible to classify fake news articles with a certain degree of accuracy. The goal of this project is to find the effectiveness and limitations of language-based techniques for detection of fake news through the use of sentiment analysis algorithms including but not limited to convolutional neural networks and recurrent neural networks.

This type of solution is not intended to be an end-to end solution for fake news classification. Like the "blacklist" approaches mentioned, there are cases in which it fails and some for which it succeeds. Instead of being an end-to-end solution, this project is intended to be one tool that could be used to aid humans who are trying to classify fake news. Alternatively, it could be one tool used in future applications that intelligently combine multiple tools to create an end-to-end solution to automating the process of fake news classification.

1.2 Statement of the Research Problem

After the comprehensive literature review it was identified that the data used in performing the detection of fake news made use of irrelevant features for the detection. In addition the models used for the detection of fake news were not accurate enough. It was obvious that the detection accuracy of fake news could be improved if another technique could be employed.

This research detect fake news using an enhanced support vector machine with sentiment analysis where relevant features were extracted for performing fake news detection and a model was developed using support vector machine and sentiment analysis for easy detection of fake news on social media and the performance was evaluated.

1.3 Aim and Objectives of the Study

The aim of this research is to detect fake news using an enhanced support vector machine with sentiment analysis.

To achieve the aim of the study the objectives of the study are to:

1. Extract relevant features for performing fake news detection
2. Develop a model using Support Vector Machine and sentiment analysis for easy detection of fake news on social media
3. Evaluate the performance of the model in 2.

1.4 Significance of the Study

Detection of fake news online is important in today's society as fresh news content is rapidly being produced as a result of the abundance of available technology. This project focuses on enhancing the use of support vector machine and sentiment analysis as light weight fact checking model which will centered on controversial events or topic in well define time window.

It is the hope that this project will provide a baseline dataset for continued research into fake news detection. The importance of combating fake news is starkly illustrated during the current COVID-19 pandemic. Social networks are stepping up in using digital fake news detection tools and educating the public towards spotting fake news. While many of these methods of detecting fake news are generally successful, they do have some limitations. This project focuses on enhancing the use of support vector machine and sentiment analysis as light weight fact checking model which will centered on controversial events or topic in well define time window.

It is the hope that this project will provide a baseline dataset for continued research into fake news detection.

1.5 Scope of the Research

The thesis is subject to developing a model using sentiment analysis to identity fake news using social media as source of information for this design. Rumors do not always have a clear distinction. The purpose of this study, focus on a deliberately narrow definition of rumor which will be defined in chapter 1 and 2.

This thesis used sentiment analysis and machine learning programme such as SVM, DT-j483 to detect fake news. This thesis used news content from twitter site, for fake news detection. The model construction process for several existing approaches. Specifically we categorize existing methods based on their main input sources as: content review v.20

CHAPTER TWO

LITERATURE REVIEW

Defining fake news is problematic. People tend to regard any news as "fake" if it does not align with their views or agenda. Tandoc Jr, *et al.* (2018), provided a typology of fake news definitions. They studied 34 different papers on fake news published between 2003 and 2017 and constructed a framework for the different types of fake news based on their definitions. The different types, which include propaganda and advertising/public relations, can all be used in information warfare to influence public opinion on a particular topic Shu *et al*., (2018).

Germishuys (2019), defines fake news as "fabricated information that mimics news media content in form but not in organizational process or intent". However, the research of Hornning (2020), suggests that there are notable differences in form especially when it comes to the titles of fake news. By the definition provided by Ma *et al*., (2017), the Wall Street Journal article that depicts Russian interference in Bulgarian elections is likely to be "fake news" because the main source that can verify the factual accuracy of the claims in the article, namely the Bulgarian secret service, is refusing to do so. This makes the story appear fabricated.

So what is the likelihood that this story is in fact fabricated? Since its factual accuracy can be neither confirmed nor denied, then it is equally likely that it is fabricated and factually accurate. The Wall Street Journal has a good reputation, so this likelihood grows slightly larger but probably not by much. After all, even reputable news outlets have produced factually inaccurate news in the past discovered that he had been fabricating facts and sources in his news reports for at least 2 years. Just because a piece of information appears in one news outlet, it does not make it factually accurate even if the outlet is a respectable news agency with rigorous organizational processes in place. Nevertheless, a generally reliable source does slightly increase the likelihood that a piece of information is factually correct especially if the opposite cannot be proven. This increase is relative to the overall reliability of the source.

What other factors can increase the likelihood? Are there similar reports that offer more concrete evidence? In the case of the Wall Street Journal article on Russian interference in foreign elections there are. Although the Russians have repeatedly denied interfering in the political affairs of foreign countries, a Czech secret service agency pointed at evidence to the contrary as early as 2008 Wang *et al*. (2018).

2.1 Psychological Foundations of Fake News

Humans are naturally not very good at differentiating between real and fake news. There are several psychological and cognitive theories that can explain this phenomenon and the influential power of fake news. Traditional fake news mainly targets consumers by exploiting their individual vulnerabilities. There are two major factors which make consumers naturally vulnerable to fake news: (i) Native Realism: consumers tend to believe that their perceptions of reality are the only accurate views, while others who disagree are regarded as uninformed, irrational, or biased; and (ii) Confirmation Bias: consumers prefer to receive information that confirms their existing views (Joachims, 1998).

Due to these cognitive biases inherent in human nature, fake news can often be perceived as real by consumers. Moreover, once the misperception is formed, it is very hard to correct it. Psychology studies show that correction of false information (e.g., fake news) by the presentation of true, factual information is not only unhelpful to reduce misperceptions, but sometimes may even increase the misperceptions, especially among ideological groups Castillo *et al.* (2011).

2.1.1 Social foundations of the fake news ecosystem.

Considering the entire news consumption ecosystem, we can also describe some of the social dynamics that contribute to the proliferation of fake news. Prospect theory describes decision making as a process by which people make choices based on the relative gains and losses as compared to their current state, Wu and Liu (2018). This desire for maximizing the reward of a decision applies to social gains as well, for instance, continued acceptance by others in a user's immediate social network.

Pang and Lee (2005), described social identity theory and normative influence theory, this preference for social acceptance and affirmation is essential to a person's identity and self-esteem, making users likely to choose "socially safe" options when consuming and disseminating news information, following the norms established in the community even if the news being shared is fake news.

Rational theory of fake news interactions can be modeled from an economic game theoretical perspective by formulating the news generation and consumption cycle as a two-player strategy game, Mathew *et al* (2019). For explaining fake news, we assume there are two kinds of key players in the information ecosystem: publisher and consumer. The process of news publishing is modeled as a mapping from original signals to resultant news report with an effect of distortion bias b, i.e., s b−→ a, where b = [−1, 0, 1] indicates [left, no, right] biases take effects on news publishing process. Intuitively, this is capturing the degree to which a news article may be biased or distorted to produce fake news. The utility for the publisher stems from two perspectives: (i) short-term utility: the incentive to maximize profit, which is positively correlated with the number of consumers reached; (ii) long-term utility: their reputation in terms of news authenticity.

Utility of consumers consists of two parts: (i) information utility: obtaining true and unbiased information (usually extra investment cost needed); (ii) psychology utility: receiving news that satisfies their prior opinions and social needs, e.g., confirmation bias and prospect theory. Both publisher and consumer try to maximize their overall utilities in this strategy game of the news consumption process. We can capture the fact that fake news happens when the short-term utility dominates a publisher's overall utility and psychology utility dominates the consumer's overall utility, and equilibrium is maintained. This explains the social dynamics that lead to an information ecosystem where fake news can thrive

2.1.2 Fake news on social media

In this subsection, we will discuss some unique characteristics of fake news on social media. Specifically, we will highlight the key features of fake news that are enabled by social media. Note that the aforementioned characteristics of traditional fake news are also applicable to social media.

2.1.3 Malicious accounts on social media for propaganda.

While many users on social media are legitimate, social media users may also be malicious, and in some cases are not even real humans. The low cost of creating social media accounts also encourages malicious user accounts, such as social bots, cyborg users, and trolls. A social bot refers to a social media account that is controlled by a computer algorithm to automatically produce content and interact with humans (or other bot users) on social media. Social bots can become malicious entities designed specifically with the purpose to do harm, such as manipulating and spreading fake news on social media. Studies shows that social bots distorted the 2016 U.S. presidential election online discussions on a large scale, and that around 19 million bot accounts tweeted in support of either Trump or Clinton in the week leading up to Election Day.

Trolls: real human users who aim to disrupt online communities and provoke consumers into an emotional response, are also playing an important role in spreading fake news on social media. **Trolling**: behaviors are highly affected by people's mood and the context of online discussions, which enables the easy dissemination of fake news among otherwise "normal" online communities Kwon *et al* (2013).

The effect of trolling is to trigger people's inner negative emotions, such as anger and fear, resulting in doubt, distrust, and irrational behavior. Finally, cyborg users can spread fake news in a way that blends automated activities with human input. Usually cyborg accounts are registered by human as a camouflage and set automated programs to perform activities in social media. The easy switch of functionalities between human and bot offers cyborg users unique opportunities to spread fake news, Zhao *et al* (2015). In a nutshell, these highly active and partisan malicious accounts on social media become the powerful sources and proliferation of fake news.

Social media provides a new paradigm of information creation and consumption for users. The information seeking and consumption process are changing from a mediated form (e.g., by journalists) to a more disinter-mediated way.

Consumers are selectively exposed to certain kinds of news because of the way news feed appear on their homepage in social media, amplifying the psychological challenges to dispelling fake news identified above. For example, users on Facebook always follow like-minded people and thus receive news that promotes their favored existing narratives. Therefore, users on social media tend to form groups containing like-minded people where they then polarize their opinions, resulting in an echo chamber effect.

The echo chamber effect facilitates the process by which people consume and believe fake news due to the following psychological factors: (1) social credibility, which means people are more likely to perceive a source as credible if others perceive the source is credible, especially when there is not enough information available to access the truthfulness of the source; and (2) frequency heuristic, which means that consumers may naturally favor information they hear frequently, even if it is fake news.

Ritter *et al* (2011), have shown that increased exposure to an idea is enough to generate a positive opinion of it, and in echo chambers, users continue to share and consume the same information. As a result, this echo chamber effect creates segmented, homogeneous communities with a very limited information ecosystem. Research shows that the homogeneous communities become the primary driver of information diffusion that further strengthens polarization

2.2 Rumour/Fake News Detection

Social media is often the root of many news stories. Social media platforms provide an easy and cheap medium to disseminate information. Traditional media platforms like newspapers, radio and television, are bound by the law to ensure the content they provide must be checked and challenged. However, for social media, there is no such regulatory oversight body, and it is left up to the discretion of the platform provider to remove false content. As a result, there is a significant volume of false content broadcast unchecked on these platforms.

This problem has been around for a long time, however, owing to the current political climate, and the controversy surrounding the Nigeria election in 2015, it has garnered tremendous interest in the literature, and as a result various approaches to identifying fake news have been explored. The following section outlines the different approaches found in the literature to detect false information.

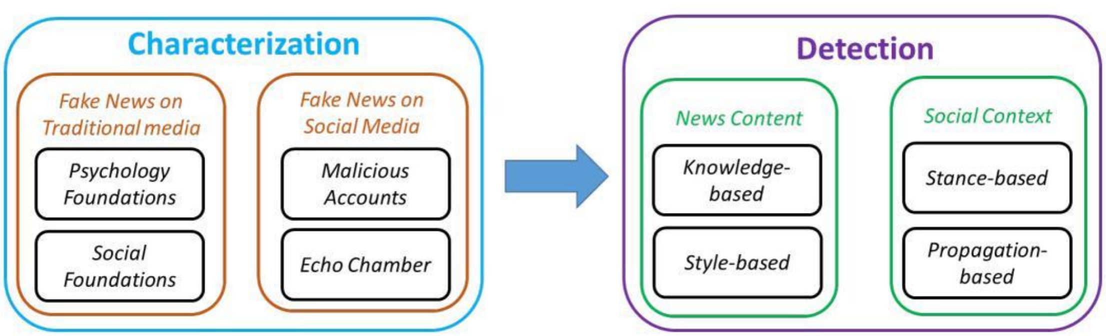

Figure 2.1 Fake News: From Characterization to Detection (Shu *et al*., (2017))

Figure 2.1 outlines the main area of focus that has been addressed in the literature, summarized by Shu, *et al*., (2017). The first task is to find the features that make a rumour, followed by a detection phase.

2.2.1 Characterization of Fake News

Studies concerned with rumours in social media, often begin by defining what constitutes as a rumour Zhao, *et al*. (2015), Zubiaga, *et al.*, (2018). Some are defined from social psychology literature, as in the case of, Qazvinian, *et al*., (2011), "a rumour is defined as a statement whose truth-value is unverifiable or deliberately false".

Zubiaga, A. *et al.*, (2018), base their definition on major dictionary definitions. The Oxford dictionary describes a rumour as: "A currently circulating story or report of uncertain or doubtful truths. " If there is a tweet about a potential politician goes to another party, there is no way to verify the veracity of this rumour before an actual signing takes place. Hence, this research is based on Qazvinian, *et al*., (2011), definition and has formulated in the context of rumours. A rumourous tweet about a political candidate moving to another party which is not known at the time when it was tweeted.

2.2.2 Fake news on traditional news media

The psychology foundation tries to answer why people are strongly influenced by fake news. The intention of this is to exploit the individuals Shu, *et al*., (2017), making consumers of the content believe what they see and obstruct other rational thoughts. In addition, affirming individual's perceptions by showing content to people who have preconceived knowledge about a topic.

People who tend to believe their perceptions of reality as only accurate view can believe fake news as true. They think that those who disagree with them are biased and irrational. Also, people who prefer to receive news that confirm their existing belief and views are mostly biased, while others are people who are socially conscious and choose a safer side while consuming and discriminating news following the norms of the community, even if the news shared is Fake. These psychological and social human behavioral patterns are the two main foundations of Fake news in the Traditional media. Along with these two factors, malicious twitter bots serves as the foundations of Fake news in Social media, Yang, *et al.*, (2018).

Another aspect of fake news on traditional media is the social foundation, this they describe as a system with two key players, the publisher and the consumer. They both have different goals, for the publisher the primary goal is to maximize profit, which is linked to number of subscribers, referred to as short-term utility, and their reputation, being a long-term utility. Whereas consumers want to acquire unbiased information referred to as information utility and news content that affirms preconceived knowledge - psychology utility, Shu, *et al*., (2017), posits that fake news thrives when the: "short-term utility dominates a publisher's overall utility [long-term utility], and psychology utility dominates the consumer's overall utility, and an equilibrium is maintained" Interestingly, even though, Shu, *et al*., (2017), coined this in context of traditional media, this aspect directly applies to social media which is the focus of this research. In this context the publishers are accounts and consumers are mostly fans. Here the goal of the accounts would be to increase their follower counts, and the result of this would possibly mean more revenue and recognition.

2.2.3 Fake news on social media

Fake news in social media largely falls into two categories, malicious accounts used for propaganda purposes and the "echo chamber" effect. With the first category, the primary objective of malicious accounts is to be medium a for malicious activity which includes trolls and social bots. A social bot is an account which is algorithmically controlled to publish certain types of messages. For instance, researchers from Fire Eye have established that thousands of Twitter accounts that campaigned against Hillary Clinton likely will be controlled by automated social bots.

The echo chamber effect, while not a fake-news phenomenon in and of itself, helps to exacerbate the problem of fake news. The echo chamber, or filter bubble, effect is described as a situation where people are only exposed to like-minded content or people. This can result in the dramatic polarization of opinions. The echo chamber effect allows people to believe fake news due to the psychological factors such as confirmation bias and social credibility. People tend to be convinced that a source is credible if others, particularly those whom we respect, perceive it as credible. The frequency with which people encounter content also induces people into believing the fake news.

2.3 Detection

The first task in detection is to create a feature set that allows models to identify rumors. Survey on rumour detection on social media by, Shu, *et al*., (2017), has identified the following approaches for detection:

1. Style-based: determine the style content in the language
   i. linguistic-based which include the lexical features such word counts, the frequency of words and unique words.
   ii. Syntactic features such as bag-of-words parts-of-speech tags.
2. Knowledge-based where content is checked against an external knowledge base fact-check the content.

However, all the above works primarily focus on detecting rumours about topics related to current affairs Zhao, *et al*., (2015).

2.3.1 Rumour classification

In most system architectures found in the literature, a rumour classification phase follows the rumour detection phase. This task aims to predict the veracity of a given rumour. Various approaches can be seen in the literature using rule-based systems, machine learning, Kwon, *et al*., (2013), and probabilistic approaches.

The critical factor for this task is to find the right set of features to enable us to infer the veracity of a rumour. According to Zubiaga, *et al*., (2018), Castillo, *et al.*, (2011)'s research has been influential on this topic. The goal of their work was to determine how accurate the authors of tweets are. In their work, they used two classifiers first to distinguish news content from conversational tweets using decision trees followed by another classifier to assess credibility.

They used four categories of features: message, user, topic and propagation features. Where the message feature includes the length of the tweet, sentiment score; user-based features include whether or not it is a verified account; the topic features include the length of tweets; and finally, propagation features include an indication of the initial number of tweets on a topic.

Building on from the feature set introduced by Castillo, *et al*., (2011), later researched by Kwon, *et al*., (2013), has used temporal, structural and linguistic features. The temporal features try to capture how rumours change over time. Structural features indicated network and the linguistic features. These features will be shown to have performed better than Castillo, *et al*., (2011) models.

2.4 Data Processing Pipeline

Information extraction is one of the most important parts of the project. Identifying entities and comparing them with existing knowledge bases is how this project will identify the entities and subsequently label the stated claim around a transfer to be true or false. The existing knowledge bases, in this case, are transfers listed on the official English Premier League website for 2017/18 Summer Transfer Window.

Nguyen, *et al*., (2014), Nguyen and Cao, (2015) have built systems to automatically store player transfer information by extracting information from the articles on Sky Sports and uses semantic web technologies to represent the transfer information. The overall goal was to create a data structure, where users can search for related content. Even though representing transfer rumours using the semantic web is not the concern for this project, methods described in the data pipeline have been hugely beneficial for this project.

They proposed crawling data from the Sky Sports website, then using a pre-processing step, an entity recognition step using KIM API, Popov, *et al* (2003), followed by rules to detect relations related to football transfer. The use of Twitter data, rather the use of news articles, has been less researched. One such study is Ireson, *et al* (2017). Both Nguyen, *et al*. (2014) and Nguyen, and Cao, (2015), have used articles written by journalists from Sky Sports. Forming a multi-stage pipeline would ensure components are modular and extra components could be added if necessary. A similar approach was followed in this project where we start by building a corpus followed by two annotation processes.

2.5 Sentiment Analysis

Sentiment Analysis is one the most challenging topics of Natural Language Processing (NLP). This aspect of the tweet will help identify whether there is any correlation between the sentiment expressed and accuracy.

Sentiment Analysis is one of the most popular fields in NLP, as it has many useful applications and data Twitter is one of most commonly explored in this field. Primarily due to the accessibility of tweets, easy to use tools and have shown to have useful applications, for example, Starbucks using sentiment expressed about their products to make informed decisions

As mentioned earlier Twitter data is informal and conversational. Stanford's Core NLP are trained on content reviews corpus introduced by Pang, and Lee, (2005), might lose out on some signals example capitalization. However, there are tools that will be specifically designed for this extracting sentiment from tweets and accommodate irregularity the language constructs in the tweet. Examples include VADER which is specifically tuned to capture sentiment expressed in social media.

The most common approaches are Lexicon-based and Machine Learning based Wei, (2012).

Lexicon: Identifies words that best describe the sentiment

Advantages include: once they are built there is no need to train.

Disadvantages include: It is often built using WordNet corpus, which does not contain colloquial expressions. Also, it performs poorly when certain words can be either positive or negative depending on the context. A.Moreo, *et al*., (2012)

```
SENTIMENT

POSITIVE="POSITIVE"

NEGATIVE="NEGATIVE"

NEUTRAL="NEUTRAL"

SENTIMENT_THRESHOLDS=(0.4,0.7)

# EXPORT

KERAS_MODEL="model.h5"

WORD2VEC_MODEL="model.w2v"

TOKENIZER_MODEL="tokenizer.pkl"

ENCODER_MODEL="encoder.pkl"
```

Source; www.programmableweb.com

Figure 2.2 Sentimental Model Formulation

Machine learning (ML): Using machine learning algorithms that learn characteristics based on data that is labeled as either positive or negative. According to Mäntylä, *et al*., (2016), Pang, *et al*., (2002), has been an influential study on sentiment analysis on Twitter. They used bag of words and Support Vector Machine to classify the sentiment of the tweets.

Advantages include: Generalizing better.

Disadvantages include: it requires considerable time and effort to label the data and train.

2.5.1 Comparison between google cloud NLP, stanford core NLP and vader.

Google Cloud NLP (GCNLP) and Stanford Core NLP (SCNLP), and VADER all offer sentiment analysis functionality. However, VADER, Hutto, C. and Gilbert, E. (2014) is a rule-based system which is specifically designed to take into account social media constructs whereas SCNLP is trained on content reviews. Meanwhile, GCNLP is a black box its workings are unknown to the public.

2.5.2 Google cloud NLP (GCNLP)

This system is based on Google's complex deep learning models

1. For a given text it gives a score, represented by numerical score and magnitude values.
2. Scores are then aggregated into an overall sentiment score and magnitude for an entity
3. Magnitude can be used to disambiguate

The Treebank was built on the corpus introduced by Pang and Lee (2005). The corpus consists of sentences from content reviews, which was parsed with Stanford parser and has unique phrases from those and trees annotated by human judges.

Their methods can be seen to be performing better at capturing the sentiment and scope of negation than bag of words.

2.5.3 Vader

VADER tool is a sentiment analyzer that uses the lexical approach to map words to sentiment. It computes the scores by doing a dictionary lookup of the sentiment of phrases and sentences. Since it is adapted to social media content, it works well in detecting emoticons and internet slang. It also uses text constructs such as punctuation and capitalization. It produces a score between -1 and 1. Where -1 is negative, 0 neutral and 1 being positive.

2.6 Language Modeling

Language models are a medium that can be used to represent text as numerical vectors. This section outlines the need for language models. Two approaches have been studied out here in this project one is count based, and the other is predictive models.

The traditional methods used for sentiment analysis that used BOW methods which ignore word ordering and may use hand-designed negation feature will not capture all the details. As a result, there has been a shift in literature towards using word embeddings and neural networks, which will be discussed in the following sections. The count-based methods compute the statistics of word occurrences, whereas the predictive models give a probabilistic interpretation.

2.6.1 Bag of words

The basic model, Bag of words also known as unigram model Kampaki and Adamides (2014), used this method to generate feature set for the machine learning models. Moreover, has been widely used in the literature while working with classification on twitter data, Culotta (2020). Also, for rumour detection tasks, ngrams have been used by Qazvinian, (2011), to represent textual features.

The bag of words representation contains all unique words in the corpus. For given corpus of 2 documents.

(1) I love ice-creams too

(2) Mary likes ice-creams too.

2.6.2 Term-frequency inverse-document frequency (TFIDF)

Term-Frequency Inverse-Document Frequency (TFIDF) has two components: term frequency (TF), which reflects the importance of a word in document; and inverse document frequency (IDF) which reflects the importance of the word in the whole corpus Manning, *et al*., (2008), Together it will describe a word's importance to document in a collection.

It is calculated using this formula Tfidf = tf x idf.

The advantages of count-based such bag of words and TFIDF methods includes:

i. Easy to compute
ii. Sophisticated smoothing techniques can be used to improve the distribution

The disadvantages are:

i. BOW can be sparse and could find it hard to capture long dependencies

ii. It does not take in to account the morphological aspects

Given the short text nature of the tweets, these models show to be quite useful in the modeling text.

2.6.3 Word embeddings

Another popular method in recent times in the literature is word embeddings. It is also known as context-predicting Baroni, *et al*., (2014). These techniques describes have a close connection to the Distributional Hypothesis , which states that words which occur in same contexts tend have similar meanings. This was popularized by Firth, (1957), Word2vec and its variations try to capture the word similarities. It does this by predicting surrounding words of each word. In recent times, word embeddings are popular technique when it comes to twitter data. As noted by Nakov, *et al*. (2019), for the task of Sentiment analysis on Twitter text, significant number of high performing teams has used word embeddings. However, an investigation of performance between BOW and embedding on social media rumour veracity by Ma, *et al.*, (2017), concluded that BOW performs better. However, this research was done on Chinese text; therefore applying to English language text could have different result.

2.7 Learning Algorithms

In recent times Machine learning has been a popular tool to answer research questions in the literature. In machine learning there are supervised, unsupervised and reinforcement learning methods Bishop, (2006). Supervised machine learning algorithms a set of inputs and desired outputs also known as labels, the algorithm will try and learn by minimizing the difference between the predicted and the desired output, example of algorithms include Support Vector Machines, and Random Forest. Whereas un-supervised do not require labels and it tries to find the structures itself, example of algorithms include k-means. Finally, reinforcement learning where the algorithm tries to learn by trial and error.

In machine learning literature the term “features” is often used, features are characteristics of a particular observation passed into the learning algorithm Bishop, (2006). These features help the algorithms to learn patterns. For instance, in text classification, textual data converted into bag of words model is one of the sets of features that could be used.

2.7.1 Support vector machine

Support Vector Machine (SVM) is a machine learning algorithm that can be used for classification and regression Vapnik, (1995). In an SVM classifier, a separating hyper plane is drawn so that it separates the data into different classes.

2.8 Benchmark Studies

While most of the existing researches have focused on defining the types of fake news and suggesting different approaches to detect them, very few studies are carried out to compare such approaches independently on different datasets. Among the categories, the benchmark- based studies are the most similar to our study. This thesis compares other research with the previous studies along three themes:

(1) Experimental setup and results,

(2) Dataset length and diversity, and

(3) Range of models explored. We discuss the related work below.

Gerimhuys, (2019), compared the performance of SVM, LR, and CNN models on their proposed dataset ‘‘knowledge graph’’ Cui, et *al* (2018), compared deep machine learning models and artificial Neural network for fake news detection on different datasets . Hassan, *et al*. (2019). Comparing SVM and tradional machine learning for text categorization with Sentiment analsysis (i.e., k-NN, Decision Tree, Naive Bayes, SVM, AdaBoost, Bag- ging) for fake news detection on different datasets. In summary, these few existing comparative studies lack in terms of the range of evaluated models and the diversity of the used datasets. Moreover, a complete exploration of the advanced pre-trained language models for fake news detection and comparison among them and with other models (i.e., traditional and deep learning) were missing in previous works. The benchmark study presented in this thesis is focused on dealing with the above issues. We extend the state-of-the-art research in fake news detection by offering the use of sentiment analysis and deep learning with traditional learning Programme models).

REFERENCES	Title	Problem	Techniques	Model	Accuracy	Specificity
Juri Gerimhuys, (2019)	Investigating Content-based Fake News Detection using Knowledge Graphs	Fails to deal with long sequences containing multiple verbs, many entities are not recognized as named entities.	Knowledge Graphs	B-transE Model	95%	0.867
Limeng Cui et *al* 2018	Sentiment-Aware Multi-Modal Embedding for Detecting Fake News	Unable to mitigate the problem of fake news better, SAME was unable to do early detection (due to the usage of irrelevant features)	Deep Learning and Artificial Neural Network (ANN)	SAME Model	93.4%,	0.829
S. Hassan, M. Rafi, and M. S. Shaikh, 2019	Comparing svm and naive bayes classifiers for text categorization with Sentiment analysis	Computational burden while performing document clustering	Knowledge enrichment in Multitopic Conference(INMIC),	DKV	96%	Yes

Table 2.1 Overview of benchmarked paper

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Research Design

The Table 3.1 represents the research design of the enhancement of support vector machine using sentiment analysis for fake news detection. Furthermore, the research design encompasses the following Stages; Stage 1: Data Collection Stage 2: Combination of svm and sentiment analysis in fake news detection Stage 3: Testing and Evaluation

Table 3.1 Research Design

STAGES	PROCESS
1. DATA COLLECTION	• Data-set Description and Collection • Data-set Pre-processing
2. COMBINATION OF SVM AND SENTIMENT ANALYSIS IN FAKE NEWS DETECTION	• Experimental Flow chart • Model Formulation
3. TESTING AND EVALUATION	• Evaluation and Validation

To accomplish this goal, dataset of twitter content was analyzed using the machine learning tool for text classification and a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization. An open source software issued under the GNU General Public License, GNU is a recursive acronym for “GNU’s Not Unix”.

The methodology as shown in Figure 3.1, we followed some steps that are involved in SA using the approaches described below.

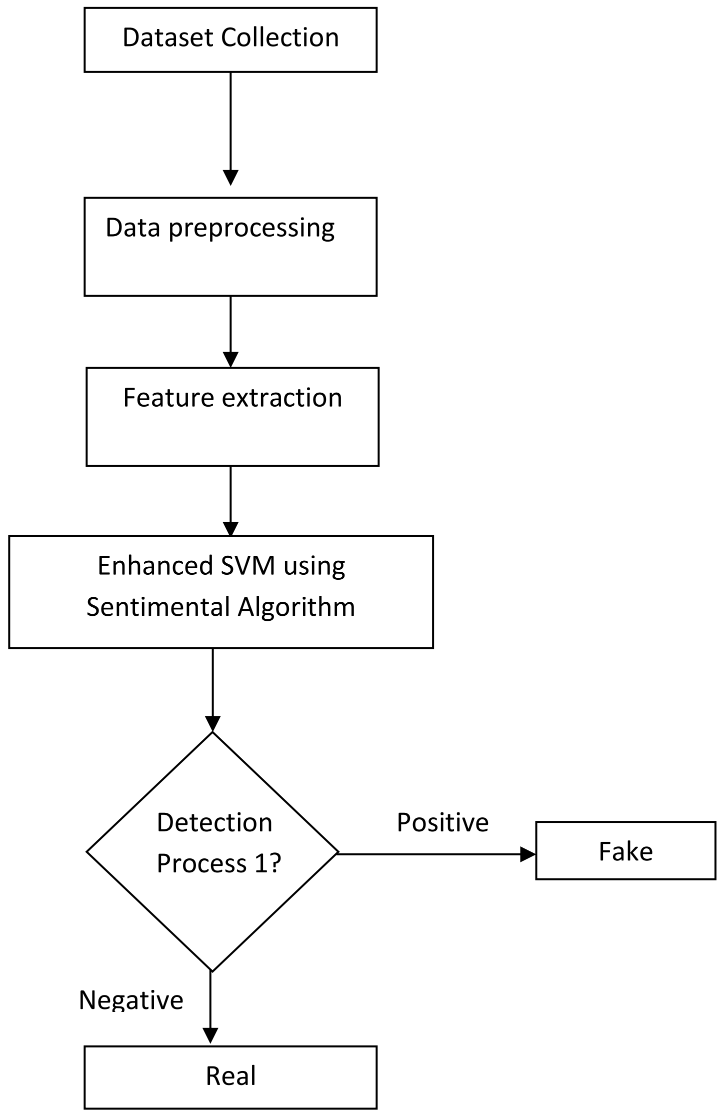

Figure 3.1: Experimental Flow Chart

Step 1: Content reviews collection

To provide an exhaustive study of machine learning algorithms, the experiment is based on analyzing the sentiment value of the standard dataset. The original dataset of the content reviews was used to test our methods of reviews classification. The dataset is available and has been used in V.K. Singh, *et al.*, (2018), which is frequently conceded as the standard gold dataset for the researchers working in the field of the Sentiment Analysis. The first dataset is known as content reviews dataset V2.0 which consists of 5000 content reviews out of which 3000 reviews are positive, and 2000 reviews are negative. A summary of the datasets collected is described in Table 3.2.

Table 3.2: Description of Dataset

Dataset	Content of Dataset
Content Review 2.0	5000 content review (3000 + 2000)

DATASET

DATASET_COLUMNS = ["target", "ids", "date", "flag", "user", "text", "statement", "subject", "speaker", "state", "statement id"]

DATASET_ENCODING = "ISO-8859-1"

TRAIN_SIZE = 0.8

 # TEXT CLEANING

TEXT_CLEANING_RE = "@\S+|https?:\S+|http?:\S|[^A-Za-z0-9]+"

 # WORD2VEC

W2V_SIZE = 300

W2V_WINDOW = 7

W2V_EPOCH = 32

W2V_MIN_COUNT = 10

 # KERAS

SEQUENCE_LENGTH = 300

EPOCHS = 8

BATCH_SIZE = 1024

Figure 3.2 Dataset Description

Dataset details

i. Target: the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)

ii. Ids: The id of the tweet (5000)

iii. Date: the date of the tweet (Sat May 16 23:58:44 UTC 2009)

iv. Flag: The query (lyx). If there is no query, then this value is NO_QUERY.

v. User: the user that tweeted (robotickilldozr)

vi. Text: the text of the tweet (Lyx is cool)

vii. Statement: words typed down

viii. Subject: Text heading/highlight

ix. Speaker: Username of the owner of the content

x. State: Location of the tweet

xi. Statement ID: ID of the text

Figure 3.3 Dataset Details

Step 2: Data preprocessing

The preprocessing phases include preliminary operations, which help in transforming the data before the actual SA task. Data preprocessing plays a significant role in many supervised learning algorithms. We divided data preprocessing as follows:

1. String to Word Vector.

To prepare the dataset for learning involves transforming the data by using the String to Word Vector filter, which is the main tool for text analysis in Weka. The String To Word Vector filter makes the attribute value in the transformed datasets Positive or Negative for all single- words, depending on whether the word appears in the document or not. This filtration process is used for configuring the different steps of the term extraction. The filtration process comprises the following two sub-processes:

i. Configure the tokenizer

This sub-process makes the provided document classifiable by converting the content into a set of features using machine learning.

ii Specify a stop words list

The stop words are the words to be filtered out, eliminate, before training the classifier. Some of those words are commonly used (e.g., "a," "the," "of," "I," "you," "it," "and") but do not give any substantial information to our labeling scheme, but instead they introduce confusion to our classifier. In this study, we used a 630 English stop words list with content reviews dataset V2.0. Stop words removal helps to reduce the memory requirements while classifying the reviews.

2 Attribute Selection

Removing the poorly describing attributes can significantly increase the classification accuracy, in order to maintain a better classification accuracy, because not all attributes are relevant to the classification work, and the irrelevant attributes can decrease the performance of the used analysis algorithms, an attribute selection scheme was used for training the classifier.

Step 3: Feature Selection

Feature selection is an approach which is used to identify a subset of features which are mostly related to the target model, and the goal of feature selection is to increase the level of accuracy. In this study, few feature selection methods were implemented and were widely used for the classification task of Sentiment analysis with stop words methods and removal of other columns which tends to have similar meaning to each other which indicates the aspect foe the classification. The columns used in the dataset for this project includes:

Dataset details

i. Target: the polarity of the tweet (0 = negative, 2 = neutral, 4 = positive)

ii. Ids: The id of the tweet (5000)

iii. Date: the date of the tweet (Sat May 16 23:58:44 UTC 2009)

iv. Flag: The query (lyx). If there is no query, then this value is NO_QUERY.

v. User: the user that tweeted (robotickilldozr)

vi. Text: the text of the tweet (Lyx is cool)

Step 4: Sentiment Classification algorithms

In this step, sentiment classification algorithms were used, and they have been applied in many domains such as commerce, medicine, media, biology, etc. There are many different techniques in classification method like NB, DT-J48, SVM, K-NN, Neural Networks, and Genetic Algorithm. In this study, two popular classifiers were used which are: SVM and DT-J483

1. Support Vector Machine (SVM)

SVM in machine learning is a supervised learning model with the related learning algorithm, which examines data and identifies patterns, which is used for regression and classification analysis. Recently, many classification algorithms have been proposed, but SVM is still one of the most widely and most popular used classifiers.

2. Decision Tree (DT-J48)

The DT-J48 approach is useful in the classification problem. In the testing option, percentage split was used as the preferred method.

Sentimental Model Formulation

SENTIMENT

POSITIVE="POSITIVE"

NEGATIVE="NEGATIVE"

NEUTRAL="NEUTRAL"

SENTIMENT_THRESHOLDS=(0.4,0.7)

MACHINE LANGUAGE (SVI) =“DETECTION”

 # EXPORT

KERAS_MODEL="model.h5"

```
WORD2VEC_MODEL="model.w2v"

TOKENIZER_MODEL="tokenizer.pkl"

ENCODER_MODEL="encoder.pkl"
```

Source; www.programmableweb.com

Figure 3.4 Enhanced Sentiment Model Formulation

Procedure of sentimental Analysis

Inputs : The multi-model input $\{m_i\}_{i=1}^{N}$,Label $\{n_i\}_{i=1}^{N}$ and the learning rateη

Begin

1. Apply sentiment analysis (S_A) for text and segmentation process for image (S_p);
2. Create feature matrix of text (f_t) and image (f_i);
3. Training set $\{f_t, f_i, i = 1, 2, \ldots n\}$ and weight matrix q_i,i=1,2,...n passed to Sentiment for feature optimization (γ);
4. Calculate training set (S), error set (E) ,and remaining set (R);
5. Apply kernel classification (K);
6. Based on test data classify and update it with iteration.
7. **end**

Here, a set of m news article containing the text and image information, we can represent the data as a collection as a set of text-image tuples denoted as $A = (A^T{}_i, A^I{}_i)_i{}^m$. ($S_A$) is process of sentiment analysis for textual part of news article and features obtain from it is represented by (f_t).Similarly, process of visual feature extraction is represented by (S_p) and all the features obtained from images are represented by (f_i). All features from text and image are fused together by simple concatenate method. This combined representation is taken input to Sentiment for feature optimization denoted by (γ). Optimal features are further used for classification.

Step 5: Detection Processes

After training, the next step is to predict the output of the model on the testing dataset, and then a confusion matrix is generated which classifies the reviews as positive or negative. The results involve the following attributes:

True negative (TN) are events which are real and are effectively labeled as real, True Positive (TP) are events which are fake and are effectively labeled as fake. Respectively, False Positives (FP) refers to Real events being classified as fakes; False Negatives (FN) are fake events incorrectly classified as Real events. The confusion matrix, (1)-(6) shows numerical parameters that could be applied following measures to evaluate the Detection Process (DP) performance. In Table III, the confusion matrix shows the counts of real and fake predictions obtained with known data, and for each algorithm used in this study there is a different performance evaluation and confusion matrix.

Table 3.3 The Confusion Matrix

	Real	Fake
Real	True Negative Reviews (TN)	False Positive Reviews (FP)
Fake	False Negative Reviews (FN)	True Positive Reviews (TP)

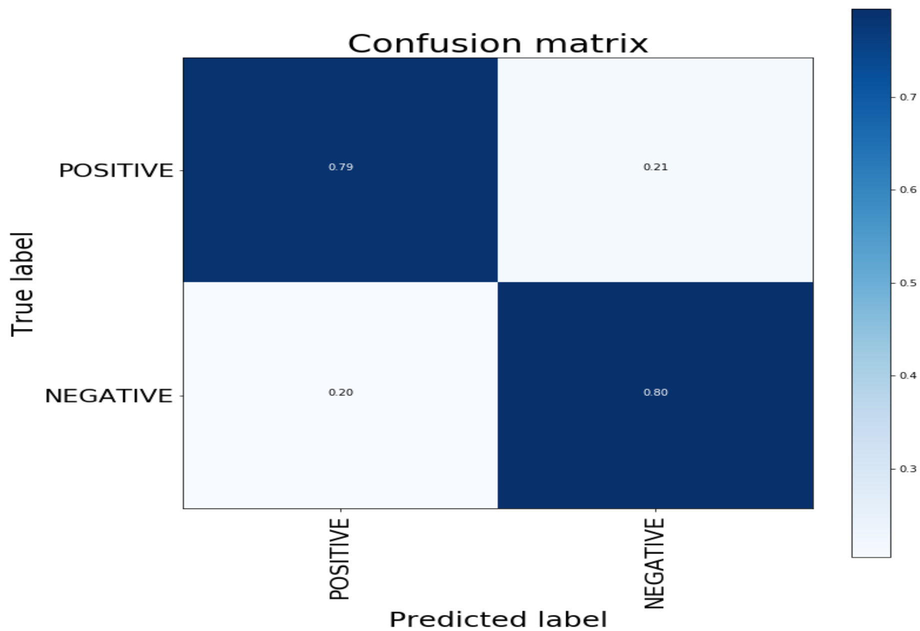

Figure 3.5 Confusion matrix

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Results

In this part, the test results from two machine learning approaches to classifying sentiment of datasets were presented and compared with content review dataset V2.0 and content review dataset ISOT gotten from Kaggle. This data were gotten from a freely available datasets, which are accessible on the websites of Kaggle (https://www.kaggle.com/datasets).

Information in both datasets contains news published on websites. The ISOT collection is dominated by the vast majority of political information and news from around the world. The dataset contains two files (true and fake) in csv file format. The real information database was created based on the websites of a reliable Reuter's news agency, and the fake information was collected from the pages marked as unreliable by Politifact. Experimental result on dataset V 2.0

Table 4.1 Confusion Matrix for All Methods

Machine learning Programme	**Sentiment A**	**Real**	**Fake**
SVM	Real	2109	1281
	Fake	792	818
DT-J48	Real	2662	530
	Fake	338	1470

The number of real and fake predictions made by the classification model compared with the actual results in the test data is shown in the confusion matrix. The confusion matrix is obtained after implementing SVM and DT-J48 algorithms. Table 4.2 displays the results for confusion matrix for V2.0 dataset. The columns represent the number of predicted classifications made by the model. The rows display the number of real classifications in the test data.

Table 4.2 Evaluation Parameters and Accuracy For All Methods.

Classification algorithms	Fake Positive Reviews %	Fake Negative Reviews %	Real Positive Reviews %	Real Negative Reviews %	Precision %	Accuracy %
SVM+ sentiment Analysis	19.1	18.2	81.8	87.95	91.72%	98.35
DT-J48 + Sentiment Analysis	23.8	33	67	76.2	73.8	81.1

Two main performance evaluation measures have been introduced for Machine learning programme. These include Fake Positive Reviews predictive value, Fake Negative Reviews predictive value, Real Positive Reviews predictive value, Real Negative Reviews predictive value, accuracy and Precision. Table 4.2 shows the results of evaluation parameters for all methods and provides a summary of recordings obtained from the experiment. SVM surpasses as the best accuracy with 98.35 and DT-j48 has 81.1, the tabulated observations list the readings as well as accuracies obtained for a specific supervised learning algorithm on a dataset of a content review. Part of the results have been published as shown in Appendix C

4.2 Discussion

Table 4.2 present the summary of the experiments. Two supervised machine learning algorithms: SVM and DT-J48 have been applied content gotten from social media site (twitter). We observed that well-trained machine learning algorithms could perform very useful classifications on the sentiment polarities of reviews. In terms of accuracy, SVM is the best algorithm for all tests since it correctly classified 98% of the reviews in dataset V2.0. SVM tends to be more accurate than DT-J48s.

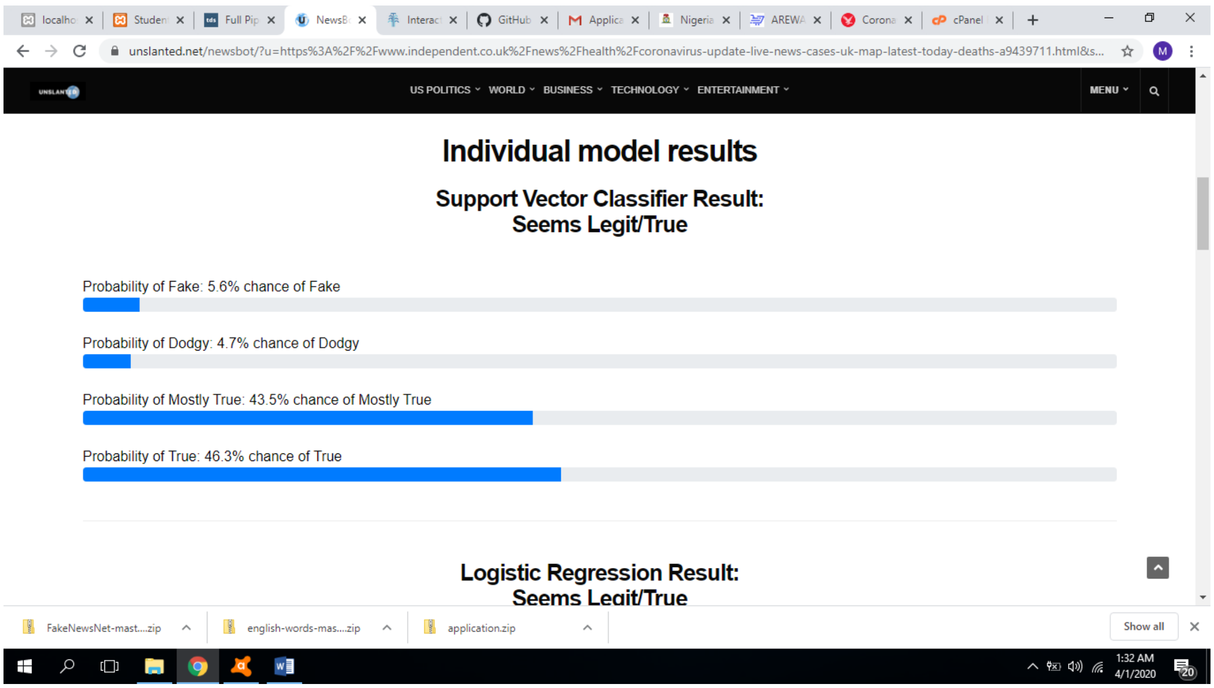

Figure 4.1 SVM classifier for fake news detection

For both the above-described tests, the model worked correctly and detected true and false information. Tests were carried out repeatedly, confirming the validity, robustness and credibility of the model. Examples of correct model operation are shown in Fig. 4.1. All entered texts were subjected to the procedure of eliminating irrelevant elements from texts, before submitting them to the model. The procedure is identical to the one carried out in the pre-processing stage on the raw data.

```
# create example sentence
sentence = Sentence('The space station staff liked her when they interviewed her she seemed
polite and quiet and incurious. That was important. One of the astronauts, a bearded
Russian with kind eyes, asked her a question: Will you be lonely in space? She looked at
the faint lines scrawled around his eyes and forehead, and she supposed he had a family
somewhere, maybe small children. Yes, she said, but I have always been lonely. The
astronaut nodded, and she could see he understood. She could see his aquiline profile as he
turned to someone off screen, and she knew she would get the job.')


# predict tags and print
classifier.predict(sentence)

print(sentence.labels)

__label__fake (0.86463862657547)
```

Figure 4.2 Result of fake news

Figure. 4.2 shows the screenshot of the launched model, which recognizes that the information from the Sci-Fi Short Stories website is fake

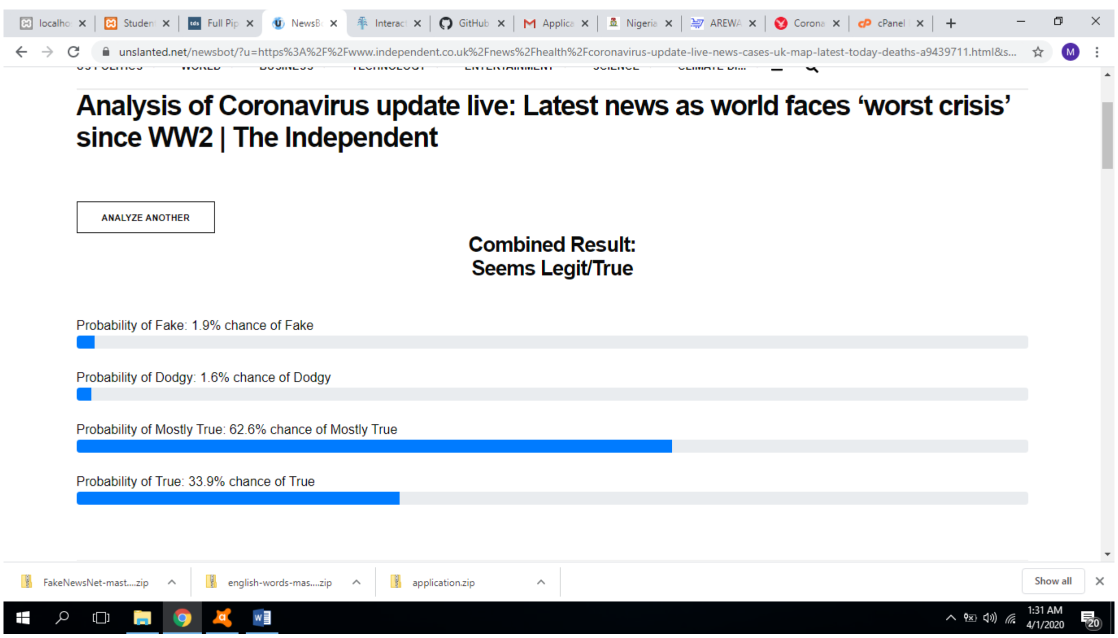

Figure 4.3 DT-J48s screenshot sample

DT-J48s can be used in real-time solutions because its execution time is short. The results indicated that DT-J48s has an impact of word embedding techniques with a accuracy of 81.1% The process of using machine language like DT-J48s on sentiment analysis, shows that one of the crucial elements to obtain better results is by pre-processing of raw data.

Table 4.3: Comparison between New Benchmark Study (Enhanced SVM with SA) And Prior Benchmark Studies

REFERENCES	Title	Techniques	Model	Accuracy	Specificity
Juri Gerimhuys, (2019)	Investigating Content-based Fake News Detection using Knowledge Graphs	Knowledge Graphs	B-transE Model	95%	0.867
Limeng Cui *et al* 2018	Sentiment-Aware Multi-Modal Embedding for Detecting Fake News	Deep Learning and Artificial Neural Network (ANN)	SAME Model	93.4%,	0.829
S. Hassan, M. Rafi, and M. S. Shaikh, 2019	Comparing svm and naive bayes classifiers for text categorization with Sentiment analysis	Knowledge enrichment in Multitopic Conference(INMIC),	DKV	96%	Yes
Samuel, 2020	Fake news detection using Sentiment Analysis and Machine Language	Sentiment Analysis	SVM, DT-J48	98%	Yes

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has strong negative impacts on individual users and broader society. In this thesis explored the fake news problem by reviewing existing literature on fake news detection approaches from a data mining perspective, including feature extraction and model construction. We also further discussed the datasets, evaluation metrics, and promising future directions in fake news detection research and expand the field to other applications. The novelty of the thesis is the application of the Sentiment Analysis in detecting which content is false on twitter handle. The model fulfills its tasks and allows for the analysis of texts with high accuracy. During the fake news process, the accuracy was up to 98.35%.

The current work concerned the distinction between label fake and label true. However, there are many additional subcategories under the fake news category; future work will concern the creation of a model to distinguish those sub-categories.

5.2 Recommendations

Based on the discussion, on fake news and how technology has changed over the last years enabling us to develop tools that can be used in the fight against fake news. These thesis explored the importance of identifying fake news using sentiment analysis, the influence that misinformation can have on the public's decision making and which approaches exist to combat fake news. The current battle against fake news on COVID-19 and the uncertainty surrounding it shows that a sentiment approach towards fake news detection is needed. Human wisdom as well as digital tools needs to be harnessed in this process. Hopefully this thesis has put in measures model that will stay in place to detected fake news on social medial and other digital media platform. Owners and the general public should take responsibility and work together in detecting and combating fake news.

5.3 Contributions to Knowledge

The contributions of this study are as follows:

i. The extraction of the best features for the detection of fake news - Page 34.

ii. The use of sentiment analysis and support vector machine algorithms is used for the easy detection of fake news - Page 35.

iii. The model developed yield better accuracy of 98.35% for the detection of fake news – Page 39.

5.4 Suggestion for Further Study

Due to time constrain and system configuration, this study was limited to using only support vector machine and Decisions Tree J48 on sentiment analysis in detecting fake news.

This study suggest that further research should be carried on other machine language and its effectiveness on sentiment analysis in detecting fake news, rumors on other social media sites.

This study also suggest that more research should be carried on the application of sentiment analysis and other machine language

REFERENCES

Al-Moslmi, T., Albared, M., Al-Shabi, A., Omar, N., & Abdullah, S. (2018). Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. *Journal of information science*, *44*(3), 345-362.

Baroni, M., Dinu, G., & Kruszewski, G. (2014, June). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 238-247).

Bishop, C. M. (2006). Pattern recognition. *Machine learning*,*128*(9).

Castillo, C., Mendoza, M., & Poblete, B. (2011, March). Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 675-684).

Cooper, G., (2017). Named Entity Recognition for Twitter. [Online] Available at: http://blog.thehumangeo.com/twitter-ner.html [Accessed 16 May 2018].

Cui, L., Wang, S., & Lee, D. (2019). Same: sentiment-aware multi-modal embedding for detecting fake news. In *Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining* (pp. 41-48).

Culotta, A. (2010, July). Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics* (pp. 115-122).

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Germishuys, J. (2019). Investigating Content-based Fake News Detection using Knowledge Graphs.

Hassan, S., Rafi, M., & Shaikh, M. S. (2019). Comparing SVM and Naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment. In*2011 IEEE 14th International Multitopic Conference* (pp. 31-34). IEEE.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1).

Ireson, N., & Ciravegna, F. (2017, October). Football Whispers: Transfer rumour detection. In *CEUR Workshop Proceedings* (Vol. 1963). CEUR Workshop Proceedings.

Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

Kampaki, S. & Adamides, A., (2014). Using Twitter to predict football outcomes. ArXiv e-prints.

Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013, December). Prominent features of rumor propagation in online social media. In *2013 IEEE 13th international conference on data mining* (pp. 1103-1108). IEEE.

Lion Gu, V. K., Yarochkin, F., Leopando, J. &Estialbo, J., (2017). Fake News and Cyber Propaganda: The Use and Abuse of Social Media. [Online] Available at: https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/fake-news-cyber-propaganda-the-abuse-of-social-media [Accessed 15 May 2018].

Ma, B., Lin, D., & Cao, D. (2017). Content representation for microblog rumor detection. In *Advances in Computational Intelligence Systems* (pp. 245-251). Springer, Cham.

Manning, C. D., Raghavan, P. & HinrichSchütze, (2008). Introduction to Information Retrieval.s.l.:Cambridge University Press.

Mäntylä, M. V., Graziotin, D. & M. K., (2016). The Evolution of Sentiment Analysis - A Review of Research Topics, Venues,. [Online] Available at: https://arxiv.org/ftp/arxiv/papers/1612/1612.01556.pdf.

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. In*Proceedings of the 10th ACM conference on web science* (pp. 173-182).

Maynard, D., & Funk, A. (2011). Automatic detection of political opinions in tweets. In *Extended Semantic Web Conference* (pp. 88-99). Springer, Berlin, Heidelberg.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113.

Moreo, A., Romero, M., Castro, J. L., & Zurita, J. M. (2012). Lexicon-based comments-oriented news sentiment analyzer system. *Expert Systems with Applications*, *39*(10), 9166-9180.

Nakov, P. Ritter, A., Rosenthal, S., Sebastiani, F., & Stoyanov, V. (2019). SemEval-2016 Task 4: Sentiment Analysis in Twitter. Proceedings of the 10th International Workshop on Semantic Evaluation , pp. 1-18.

Nguyen, Q.-M. & Cao, T.-D., (2015). A novel approach for automatic extraction of semantic data about football transfer in sport news. International Journal of Pervasive Computing and Communications, Volume 11, pp. 233-252.

Nguyen, Q. M., Cao, T. D., & Nguyen, T. T. (2014, December). Automatic creation of semantic data about football transfer in sport news. In *Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services* (pp. 356-364).

Pang, B., & Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *arXiv preprint cs/0506075*.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques.*arXiv preprint cs/0205070*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D., & Goranov, M. (2003, October). KIM–semantic annotation platform. In *International Semantic Web Conference* (pp. 834-849). Springer, Berlin, Heidelberg.

Qazvinian, V., Rosengren, E., Radev, D., & Mei, Q. (2011, July). Rumor has it: Identifying misinformation in microblogs. In*Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (pp. 1589-1599).

Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop* (pp. 43-48).

Ritter, A., Clark, S., Mausam & Etzioni, O., (2011). Named Entity Recognition in Tweets: An Experimental Study}. EMNLP.

Sheela, L. J. (2016). A review of sentiment analysis in twitter data using Hadoop. *International Journal of Database Theory and Application*, *9*(1), 77-86.

Shu, K. Sliva, A., Wang, S., Tang, J., & Liu, H., (2017). Fake News Detection on Social Media: A Data Mining Perspective. IGKDD Explor. Newsl., Volume 19, pp. 22-36.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and spatialtemporal information for studying fake news on social media. *arXiv preprint arXiv:1809.01286*.

Singh, V. K., Piryani, R., Uddin, A., & Waila, P. (2013, March). Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. In *2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)* (pp. 712-717). IEEE.

Tandoc Jr, E. C., Ling, R., Westlund, O., Duffy, A., Goh, D., & Zheng Wei, L. (2018). Audiences' acts of authentication in the age of fake news: A conceptual framework. *new media & society*, *20*(8), 2745-2763.

Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital journalism*, *6*(2), 137-153.

Vapnik, V., (1995). The Nature of Statistical Learning Theory. 1 ed. S.1.:Springer-Verlag New York

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018). Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 849-857).

Wei, F., (2012). Sentiment Analysis and Opinion Mining. [Online] Available at: http://web.nchu.edu.tw/~jodytsao/MarkegingG/IIR10-Sentiment%20Analysis.pdf [Accessed 14 January 2018].

Weiss, G. M., McCarthy, K. L. & Zabar, B., (2007). Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs?. DMIN. *7*(35-41), 24

Wu, L., and Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. In*Proceedings of the eleventh ACM international conference on Web Search and Data Mining* (pp. 637-645).

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., & Yu, P. S. (2018). TI-CNN: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*.

Zhao, Z., Resnick, P. & Mei, Q., (2015). Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. Proceedings of the 24th International Conference on World Wide Web, pp. 139-145.

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*, *51*(2), 1-36.

Appendix A

MACHINE LANGUAGE (SVI)

```
>>> import nltk
>>> nltk.download([
...     "names",
...     "stopwords",
...     "state_union",
...     "twitter_samples",
...     "content_reviews",
...     "averaged_perceptron_tagger",
...     "vader_lexicon",
...     "punkt",
... ])
[nltk_data] Downloading package names to /home/user/nltk_data...
[nltk_data]   Unzipping corpora/names.zip.
[nltk_data] Downloading package stopwords to /home/user/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package state_union to
[nltk_data]     /home/user/nltk_data...
[nltk_data]   Unzipping corpora/state_union.zip.
[nltk_data] Downloading package twitter_samples to
[nltk_data]     /home/user/nltk_data...
[nltk_data]   Unzipping corpora/twitter_samples.zip.
[nltk_data] Downloading package movie_reviews to
[nltk_data]     /home/user/nltk_data...
[nltk_data]   Unzipping corpora/movie_reviews.zip.
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /home/user/nltk_data...
[nltk_data]   Unzipping taggers/averaged_perceptron_tagger.zip.
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     /home/user/nltk_data...
[nltk_data] Downloading package punkt to /home/user/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
[nltk_data]   matches the id in the PolitiFact website API (unique for each
sample)
```

```
[nltk_data]   date: The time each article was published in the PolitiFact
website
[nltk_data]   speaker: The person or organization to whom the Statement relates
[nltk_data]   statement: A claim published in the media by a person
[nltk_data]   sources: The sources used to analyze each Statement
[nltk_data]   paragraph_based_content: content stored as paragraphed in a list
[nltk_data]   fullText_based_content: Full text using pasted paragraphs
True
```

Appendix B

Enhance SVM and Sentiment analysis

Data loading and cleaning

In [1]:

```
linkcode
%matplotlib inline
%config InlineBackend.figure_format = 'retina'

import numpy as np
import pandas as pd
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt
import seaborn as sns

import nltk
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer
from nltk.tokenize import TweetTokenizer

from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.model_selection import train_test_split, StratifiedKFold, cross_val_sco
re
from sklearn.pipeline import make_pipeline, Pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import make_scorer, accuracy_score, f1_score
from sklearn.metrics import roc_curve, auc
from sklearn.metrics import confusion_matrix, roc_auc_score, recall_score, precisio
n_score
```

In [2]:

```
data = pd.read_csv("../input/Tweets.csv")
```

We take only the tweets we are very confident with. We use the BeautifulSoup library to process html encoding present in some tweets because scrapping.

In [3]:

```
data_clean = data.copy()
data_clean = data_clean[data_clean['airline_sentiment_confidence'] > 0.65]
data_clean['sentiment'] = data_clean['airline_sentiment'].\
    apply(lambda x: 1 if x=='negative' else 0)
```

data_clean['text_clean'] = data_clean['text'].apply(lambda x: BeautifulSoup(x, "lxml").text)

We are going to distinguish two cases: tweets with negative sentiment and tweets with non-negative sentiment

In [4]:

data_clean['sentiment'] = data_clean['airline_sentiment'].apply(lambda x: 1 if x=='negative' else 0)

In [5]:

data_clean = data_clean.loc[:, ['text_clean', 'sentiment']]

In [6]:

data_clean.head()

Out[6]:

	text_clean	sentiment
0	@WHO What @dhepburn COVID19.	0
2	@UNICEF covid is increase in Africa...	0
3	@WHO it's really going to fast...	1
4	@WHO it's has a really big impact	1
5	@VirginAmerica seriously would pay $30 a fligh...	1

Machine Learning Model

We split the data into training and testing set:

In [7]:
```
train, test = train_test_split(data_clean, test_size=0.2, random_state=1)
X_train = train['text_clean'].values
X_test = test['text_clean'].values
y_train = train['sentiment']
y_test = test['sentiment']
```
In [8]:
```
def tokenize(text):
    tknzr = TweetTokenizer()
    return tknzr.tokenize(text)

def stem(doc):
    return (stemmer.stem(w) for w in analyzer(doc))

en_stopwords = set(stopwords.words("english"))

vectorizer = CountVectorizer(
    analyzer = 'word',
    tokenizer = tokenize,
    lowercase = True,
    ngram_range=(1, 1),
    stop_words = en_stopwords)
```
We are going to use cross validation and grid search to find good hyperparameters for our SVM model. We need to build a pipeline to don't get features from the validation folds when building each training model.

In [9]:
```
kfolds = StratifiedKFold(n_splits=5, shuffle=True, random_state=1)
```
In [10]:
```
np.random.seed(1)

pipeline_svm = make_pipeline(vectorizer,
                            SVC(probability=True, kernel="linear", class_weight="balanced")
)

grid_svm = GridSearchCV(pipeline_svm,
                    param_grid = {'svc__C': [0.01, 0.1, 1]},
                    cv = kfolds,
                    scoring="roc_auc",
                    verbose=1,
                    n_jobs=-1)
```

```
grid_svm.fit(X_train, y_train)
grid_svm.score(X_test, y_test)
Fitting 5 folds for each of 3 candidates, totalling 15 fits
[Parallel(n_jobs=-1)]: Done  15 out of  15 | elapsed:  3.6min finished
```

Out[10]:

```
0.92026612299784205
```

In [11]:

```
grid_svm.best_params_
```

Out[11]:

```
{'svc__C': 0.1}
```

In [12]:

```
grid_svm.best_score_
```

Out[12]:

```
0.90562337806870141
```

In [13]:

```
def report_results(model, X, y):
    pred_proba = model.predict_proba(X)[:, 1]
    pred = model.predict(X)

    auc = roc_auc_score(y, pred_proba)
    acc = accuracy_score(y, pred)
    f1 = f1_score(y, pred)
    prec = precision_score(y, pred)
    rec = recall_score(y, pred)
    result = {'auc': auc, 'f1': f1, 'acc': acc, 'precision': prec, 'recall': rec}
    return result
```

Let's see how the model (with the best hyperparameters) works on the test data:

In [14]:

```
report_results(grid_svm.best_estimator_, X_test, y_test)
```

Out[14]:

```
acc': 0.83632369095569392,
 'auc': 0.92027395510635412,
 'f1': 0.87114442202363784,
 'precision': 0.91520290732889154,
 'recall': 0.83113311331133111}
```

In [15]:

```
def get_roc_curve(model, X, y):
    pred_proba = model.predict_proba(X)[:, 1]
    fpr, tpr, _ = roc_curve(y, pred_proba)
    return fpr, tpr
```

In [16]:

```
roc_svm = get_roc_curve(grid_svm.best_estimator_, X_test, y_test)
```

In [17]:

```
fpr, tpr = roc_svm
plt.figure(figsize=(14,8))
plt.plot(fpr, tpr, color="red")
plt.plot([0, 1], [0, 1], color='black', lw=2, linestyle='--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Roc curve')
plt.show()
```

Let's see if our model has some bias or variance problem ploting its learning curve:

In [18]:

```
from sklearn.model_selection import learning_curve

train_sizes, train_scores, test_scores = \
   learning_curve(grid_svm.best_estimator_, X_train, y_train, cv=5, n_jobs=-1,
                  scoring="roc_auc", train_sizes=np.linspace(.1, 1.0, 10), random_state=1
)
```

In [19]:

```
def plot_learning_curve(X, y, train_sizes, train_scores, test_scores, title='', ylim=No
ne, figsize=(14,8)):

   plt.figure(figsize=figsize)
   plt.title(title)
   if ylim is not None:
      plt.ylim(*ylim)
   plt.xlabel("Training examples")
   plt.ylabel("Score")

   train_scores_mean = np.mean(train_scores, axis=1)
   train_scores_std = np.std(train_scores, axis=1)
   test_scores_mean = np.mean(test_scores, axis=1)
   test_scores_std = np.std(test_scores, axis=1)
   plt.grid()

   plt.fill_between(train_sizes, train_scores_mean - train_scores_std,
                    train_scores_mean + train_scores_std, alpha=0.1,
                    color="r")
```

```
    plt.fill_between(train_sizes, test_scores_mean - test_scores_std,
                     test_scores_mean + test_scores_std, alpha=0.1, color="g")
    plt.plot(train_sizes, train_scores_mean, 'o-', color="r",
             label="Training score")
    plt.plot(train_sizes, test_scores_mean, 'o-', color="g",
             label="Cross-validation score")

    plt.legend(loc="lower right")
    return plt
```

In [20]:

```
plot_learning_curve(X_train, y_train, train_sizes,
                    train_scores, test_scores, ylim=(0.7, 1.01), figsize=(14,6))
plt.show()
```

It looks like there isn't a big bias or variance problem, but it is clear that our model would work better with more data:. if we can get more labeled data the model performance will increase.

Examples

We are going to apply the obtained machine learning model to some example text. If the output is **1** it means that the text has a negative sentiment associated:

In [21]:

```
grid_svm.predict(["flying with @united is always a great experience"])
```

Out[21]:

```
array([0])
```

In [22]:

```
grid_svm.predict(["flying with @united is always a great experience. If you don't lose your luggage"])
```

Out[22]:

```
array([1])
```

In [23]:

```
grid_svm.predict(["I love @united. Sorry, just kidding!"])
```

Out[23]:

```
array([0])
```

In [24]:

```
grid_svm.predict(["@united very bad experience!"])
```

Out[24]:

```
array([1])
```

In [25]:

```
grid_svm.predict(["@united very bad experience!"])
```

Out[25]:

```
array([1])
```

Appendix C

LIST OF PUBLISHED ARTICLES

Olusanjo Fasola, Joseph Ojeniyi & Samuel Oyeniyi., (2020). Intelligent based Framework for Detection of Fake News in the Social Network Platforms. Proceedings of the 15th international Conference on Cyber Warfare and Security, Doi:10.34190/ICCWS.20.116 (Scopus-index conference)

Samuel A. Oyeniyi, Joseph A. Ojeniyi,. (2021). Development of a conceptual framework and a measurement model for the detection of fake news. International Journal of Innovative Research in Advanced Engineering (IJRAE) Vol:VIII,138-147, doi: https://doi.org/10.26562/ijira0e2021.v0807.001