Kingdom of Saudi Arabia
Ministry of Education
Qassim University
College of Computer
Information Technology Dept.

المملكة العربية السعودية
وزارة التعليم
جامعة القصيم
كلية الحاسب
قسم تقنية المعلومات

Credibility Prediction in Social Media

Students:

Asma Alrashidi 362205772
Khuzama Alsalem 362206020
Ruba Alharbi 362216679

Supervisor:

Dr. Suliman Aladhadh

A project report submitted in partial fulfilment of the requirements for B.Sc. degree in Information Technology.

Qassim-Saudi Arabia
1441/1442 (2020/2021)

Certificate

It is certified that project report has been prepared and written under my direct supervision and guidance. The project report is approved for submission for its evaluation.

Dr. Suliman Aladhadh

Dedication

We dedicate this project to everyone who taught us during our educational career, To everyone who supported us and took our steps with us, our family, friends who helped us, and to every student in search of knowledge.

Asma Alrashidi
Ruba Alharbi
Khuzama Alsalem

Acknowledgement

We thank After Allah, our great parents, and our supervisor Dr.Suliman Al Adhadh, for all the efforts and guidance made for us. We are also grateful to each member of our team.

Asma Alrashidi
Ruba Alharbi
Khuzama Alsalem

Abstract

Information credibility is a critical issue in social media. People use social media to get information on different topics. Social media is a place that people shared and exchange information and use social media as a source of information on different topics. However, information credibility becomes an issue in social media, especially with the Coronavirus Disease 2019 (COVID-19) epidemic. Misinformation has increased, which affects on peoples health, assessing the credibility of the information becomes necessary. In this project, we proposed a supervised machine learning model to assess information credibility in social media using a combination of content-based features and source-based features.

Keywords: Credibility, Information, Social media, COVID -19.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ARFF Attribute Relation File Format

COVID-19 Coronavirus Disease 2019

CSV Comma Separated Values

LIWC Linguistic Inquiry and Word Count

LR Logistic Regression

ML Machine Learning

NB Naïve Bayes

RF Random Forests

TWINT Twitter Intelligence Tool

Weka Waikato Environment for Knowledge Analysis

WHO World Health Organization

Chapter 1

INTRODUCTION

INTRODUCTION

1.1 Introduction

In early 1990, it is the start of using the internet. It became a primary global source for exchanging ideas and benefits that fit users needs, allowing them to interact with others by sending, receiving, and sharing. That led to the spread of misleading information and its lack of credibility [1].

The ease of access to information is one of the modern era features, it is necessary to think about the credibility of information on the internet. And how to assess the credibility of information on a social media platform has become an essential issue for internet users.

Twitter is considered one of the most important social media because of its wide scope for disseminating information. It is regarded as one of the fastest social media platforms to spread false rumors and an important news medium to transmit real-time events. Rumors on Twitter may reach millions of people in a short period. Therefore, we need to limit the spread of false information to not undermine its usefulness as a valuable source of information [2]. Through social media, users share information, feelings, and a general perception of a particular phenomenon or event. One example of global events currently is the coronavirus disease 2019 (COVID-19). During this period, people have increased their dependence on social media platforms. Identifying misinformation and uncovering its online spread about COVID-19 has become a critical task [3],[4].

1.2 Problem Specification and Motivation

This section explain the problem specification and motivation.

1.2.1 Problem Specification

The presence of not credible information on social media, especially in the health domain. Peoples use of medical advice or treatments spread on social media platforms can endanger a persons life due to shaded medical information. People may use this information to do something to make decisions or solve problems.

The World Health Organization (WHO) and governments have struggled with the information epidemic spreading in the social media to control the misinformation about COVID-19 by publishing instructions and providing medical advice, so obtaining credible information is very important [1]. We focus in this project on health domain specific COVID-19 in social media to prevent people from being exposed to any danger by obtaining correct information.

1.2.2 Motivation

Our works motivation is to reduce misinformation by limiting the circulation of fake news and increasing users awareness of the credibility of the information by classifying it into credible and not credible. The emerging COVID-19 is a global issue under study. Most studies focus on analyzing peoples feelings towards COVID-19 and the lack of studies on the credibility of the information. So, in our project, we focus on categorizing COVID-19 tweets for credibility.

1.3 Goals and Objectives

Credibility is one of the most important issues and plays a vital role in social media. In this project, we aim to suggest a solution to predict credibility.

1.3.1 Goals

The project general goal builds a classic machine learning (ML) model to predict the credibility of the information.

1.3.2 Objectives

Several steps must be taken to achieve the primary objectives of this project:

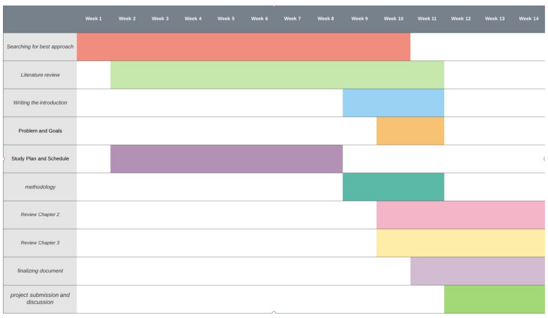
- Reviewing previous studies on the credibility of information in social media.
- Building a model that asses the health information in social media.
- Building a model that more related to COVID-19 information.

1.4 Study Scope

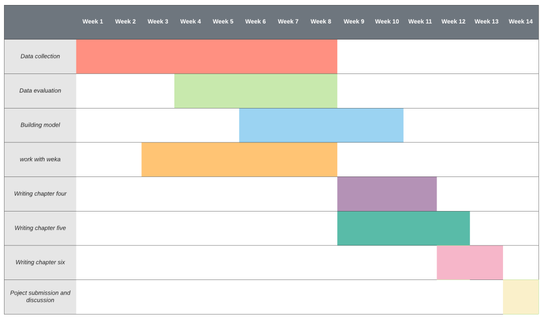
In this project, we create a supervised machine learning (ML) model. We focus on using the health information specifically related to COVID-19. To classify tweets into two classes as credible or not credible.

1.5 Study plan and schedule

In this section, we present the project plan in phase 1 and phase 2. figure 1.1 present project gantt chart.



(a) Gantt chart phase 1.



(b) Gantt chart phase 2.

Figure 1.1: Project gantt chart.

1.6 Report Structure

This report was organized as follows:

Chapter one
Chapter one provides an introduction about the internet and social media in general and the impacts of not credible information to the people, problem specification and motivation, project goals and objectives, study scope, study plan and schedule.

Chapter two
This chapter reviews research about credibility in all aspects credibility of information on social media, credibility of health information in social media (COVID-19), and user credibility perception.

Chapter three
Describes the methodology that starts with collecting data, how to evaluate tweets related to COVID-19, the proposed model for categorizing tweets into credible and not credible.

Chapter four
Introduce implementation steps starting from the used feature, describe the implementation environment, how to convert the data format and describe used algorithms.

Chapter five
We discuss the final result for the selected algorithm and compare the performance between all algorithms.

Chapter six
In the end, we present the main contributions, challenges for the project, and future work.

Chapter 2

LITERATURE REVIEW

LITERATURE REVIEW

2.1 Introduction

Credibility assessment has become an important research area in social media platforms due to a large of information spread in these platforms without boundary. Users cannot always distinguish between fake and correct information, causing an increase in low credibility. This chapter reviews research on social media credibility, including information credibility, user perceptions, and health information credibility in social media.

2.2 Credibility of information on social media

Due to the vast amount of information, it has become important to know social media platforms credibility. In this section, we review studies that can help us to understand credibility.

In [5], classify Twitter messages into credible or non-credible messages based on the content-based features (length of the tweet, URL count, retweet count, emoticons count, hashtags count) and source-based features (Followers count, friends count, listed count). They proposed a classification model based on machine learning techniques to assess tweets credibility. Data was collected using the Twitter API stream, which contains the dataset of 5,802 tweets in English. Then, they compared the performance of five different machine learning classifiers using three feature sets: Content-based, source-based, and a combination of both. The best performing random forest is a classifier with 78.4% accuracy. Moreover, they run further analysis to compare source and content features to determine a good indicator of credibility. The results indicated that the sources characteristics are higher than the content with an accuracy of 0.778.

Also, in [6] proposed five different supervised machine learning model and compared between them to classifying tweets into credible and not credible based on text analysis using word N-grams and Term Frequency (TF) and Term Frequency Inverted Document Frequency (TF-IDF) as a feature extraction technique. The model architecture consists of two modules, offline and online modules. Offline used to train and build the classification model, and the online model is a mobile application to extract real-time tweets, the result used as input to train a model. They used two datasets English and Arabic. Zubiaga collected PHEME dataset by using Twitter streaming API. Model achieve 84.9% in English accuracy and Arabic 73.2% accuracy. To evaluate the model, they used different sizes of N-grams over the PHEME dataset. They use two experiments over a dataset to extract features using TF and TF-IDF to train and test the model using five supervised machine learning models. The best performance is achieved using a combination of both unigrams and bigrams. Linear Support Vector Machines (LSVM) achieved the best accuracy in both TF 0.844 and TF-IDF0.849.

In [7], many reviews were collected on social media about the specific news and based on these responses. They were classified into 16 criteria, and ten different news articles were taken randomly to evaluate each articles news containing three categories to assess its credibility: Red: fake, green: true, blue: half fake. A group of volunteers was identified to classify the news as true, half fake, or fake according to their opinions and experiences. They used Gaussian Mix Model (GMM) to calculate the responses and the Information Quality Index (IQI) to assess Intelligence Rate IQ. From the results, it can be concluded that IQI closely aligns with the falsification. Therefore, based on the volunteers opinions, if the news is of high credibility, the IQ is higher than 0.4 if it is fake, then the IQ will decrease dramatically.

2.2.1 Credibility of health information in Social Media (COVID-19)

Social media plays a major role in spreading health awareness if invested properly. As the Coronavirus spreads around the world, people have increased their dependence on social media platforms for news. Spreading disinformation during the pandemic also creates many problems for people's anxiety and panic.

In [8], according to the World Health Organization (WHO), noticed the spread of misinformation about the coronavirus. Accordingly, an analytical study was conducted on the Twitter platform and collected a million of tweets in two months, including these terms (corona or covid) by using tweetpy. They analyzed 288,000 users profiles, meta-data, tweet context, and excluded unique users (WHO). They built different lexicons to be used in the analytical: COVID hashtags lexicon, COVID context lexicon occupation lexicon, medical context lexicon, and virus specialty lexicon. They focused on hashtags to be analyzed regardless of the tweet content to classify hashtags into corona or non-corona based on natural language processing techniques (NLP).

Based on the analysis, they noticed that 53.5% of tweets contain corona hashtags. The content of the tweets is not related to corona, and 46% of tweets include non-corona hashtags and classify it as not corona, and the content of the tweet talks about corona. Similarly, they filtered 839.2K used to classify users profiles into medical profile and non-medical profile. It shows 2.8% tweets related to COVID by specialists and 97.2% by non-specialists. The study result shows the exploitation of the COVID-19 crisis to advertise for products or spread unauthentic medical and information. There may be some possible limitations in this study they use two hashtags (Corona or Covid) to extract information about the coronavirus. It is not comprehensive and sufficient to obtain sufficient information.

In [9], proposed a schema for tweets in Twitter that reflect different viewpoints with a focus on building large groups of tweets and then analyzing them. They classified tweets into seven categories to build seven questions: (1) It contains a factual claim that can be verified, (2) It is likely to contain false information, (3) It is of interest to the general public, (4) It is likely harmful to a person, company, product, or community, (5) Requires verification by a fact-checker, (6) Causes a certain type of harm to society, (7) Requires government agency attention. They used three models support vector machine (SVM), FastTaxt , and Bidirectional Encoder Representations from Transformers (BERT). They collected two data sets of 504 English tweets and 218 Arabic tweets, combined the highest number of retweets and different word keys. As a result, the best models of the English language BERT, and the Arabic model FastText. There may be some possible limitations in this study. To confront the COVID-19, samples need a comprehensive approach between different points of view from government agencies, health organizations, and the press. Many explanations are needed for each of the seven categories, not limited to two languages, but several other languages.

In [10], they analyzed social media comments to determine users opinions and perceptions about the COVID-19 pandemic. They have collected a dataset of 47 million comments from different social media platforms. Using the natural language processing (NLP) approach to extract the relevant sentences from the comments, control users negative feelings and contribute to changing them into positive emotions and categorizing related topics. The study result shows that negative topics from comments include health, psychological and social issues related to the COVID-19 pandemic, like the high mortality rate, the nature of the disease, and health concerns that affect the users negative feelings. Also, and positive topics such as public awareness, support for relief efforts, and anxiety relief.

In [11], to classify the COVID-19 tweets into fake or genuine tweets in three different languages Hindi, Bengali, and English. They propose a BERT model (Bidirectional Encoder Representations from Transformers) using a set of features (Text Features, Twitter User Features, Fact verification score, Bias score, and Source Tweet Embedding) and a deep neural network classifier. The dataset has 504 tweets in English. For English tweets using the BERT model, to classifying tweets into fake or non-fake. They use pretrained BERT embedding, with (or without) different feature combinations to classify using four different models Support Vector Machine (BERT SVM), Random Forest Classifier (BERT RFC) and, Multi-Layer Perceptron Neural Network (BERT MLP) classifier the best accuracy mBERT_NN 89.47%, 79% in Hindi and 81% or Bengal.

Also, in [12] proposed a Bidirectional Encoder Representations from Transformers (BERT) model to classify sentiment (positive, negative, neutral). The model consists of two parts BERT model and the Term Frequency-Inverse Document Frequency algorithm (TF-IDF) to analyze sentiment. They selected the dataset randomly on Weibo related to COVID-19, the number of data 999,978 posts. They use the BERT model to categorize and analyze Weibo posts sentiment, then apply TF-IDF to extract the different emotions and focus on the negative emotions related to COVID-19. The BERT model is achieving an accuracy of 75.65%. Based on the TF-IDF algorithm, people have focused on four things related to COVID-19: The virus origin (Gamey Food, Bat or Conspiracy Theory), symptom (Fever or Cough), production activity (Go to Work, Resume Work or School New Semester Beginning), and public health control (Temperature Taking, Coronavirus Cover-up, or City Shutdown).

2.2.2 User Credibility Perception

Given the different abilities and perceptions of users to distinguish between correct and wrong information, to enable users to assess the credibility of the information, they must have sufficient cognitive power and awareness to understand the Internet environment. Users credibility judgments are critical, as they use this information.

In [13], there are known metrics on social media platforms that increase your belief that the author is highly credible, for example, the number of likes, the number of followers, or the number of posts affecting the perception of user credibility. However, the author can manipulate the number of likes or shares that affect evaluating the information credibility. The study showed that trusting a friend who published the wrong post may increase rumors and that followers are affected by the publisher. It is necessary to understand and know the extent of the credibility of the information. Therefore, they used Need For Cognition (NFC), an index of information credibility that individuals are affected by cognitive activities. The result is that people with high NFC can perceive things, and people with low NFC are more susceptible to rumors than those with high NFC. A rumor spread on Facebook saying ”bees can solve mathematical problems”, to conduct a statistical study of the spread of rumors and the extent to which an individual is affected by the information credibility.

They are published from an unknown source and make it easy for experts to spot it as a fake because they have experience and knowledge in the field. A sample was conducted to study the interaction between the credibility of a content publisher on Facebook. A decision tree was used to predict the credibility of the information. The result shows that the more experience a person has, the more credible the information content will be. It was also found that the more socially linked to the name of friends, the more credible the content posted. The higher the NFC level, the more credible the published content.

Likewise, in [14] divides messages in social media into five types of true information, rumors, biases, fake news, and spams. Data was collected from two types of dataset LIAR dataset and Weibo dataset. They proposed the Bidirectional Long Short-Term Memory Networks (BiLSTM) model based on deep learning to classifying the five types of incredible messages on social media based on three factors: Sentiment features, semantic representation of information, meta-data features, and the combination of these three factors to build a model to evaluate information credibility. The proposed method focuses on emotional connotations and captures contextual text semantic information. Because social media have text data and metadata will improve the classification of unreliable messages. There may be some possible limitations in this study. First, most of the studies aim at one type of incredible message, and it has different types. Secondly, to classify incredible messages requires constructing a lot of features.

In [15], they proposed a Credibility Analysis of Arabic Content on Twitter (CAT) model to classify credibility based on a set of features content based features (sentiment, language, and text signals) and user-based features (user experience and the number of followers). They used these features to extract the credibility of the tweets and build a lexicon to extract sentiments from the words of the tweets. Data collected 9000 tweets over two conditions: 1- the tweets must be in Arabic 2- include a hashtag. They categorize the tweets credibility by six judges, on the basis of credible, non-credible, cant decide, or write annotations via two links, the first displaying the text of the tweet as it was displayed on Twitter, and the second displaying the profile on Twitter, based on the categorize of the judges committee. There was 9,000 tweets, and they found 60% credible and 40% non-credible. By using the features (user-based and content-based) to build a lexicon to analyzes each word and extract sentiment in the tweet by using lemma is a tool in ArSenL (Arabic sentiment lexicon). To assessing CAT model credibility to see if content-based features or user-based features used as deciders to credibility, using metadata and features extracted by a random forest decision tree. As the results, they cannot rely on content-based features only or user-based features. They need to both features to get a wise decision.

In this study [16], they analysis tweets sentiment related to COVID-19 by using a set of supervised machine learning models. The aim of this study evaluates the performance of different machine learning work using the set of proposed features. The dataset used from IEEE data port the total of data 7528 tweets. They used English tweets extracted from different hashtags: “corona”, “coronavirus”, “covid”, “pandemic”, “sarscov2”, “nCov”, “covid-19”, “ncov2019”, “2019ncov”. Then they apply five machine learning models: Random Forests (RF), eXtreme Gradient Boosting (XGBoost), Support Vector Clustering (SVC), Extra Trees Classifiers (ETC), Decision Tree (DT). To extract features they used term frequency and inverse document frequency (TF-IDF), Bag of word (BoW), and a combination of both. The Extra Trees Classifiers (ETC) achieve the highest accuracy 93% by using a combination of TF-IDF and BoW.

2.3 Summary

This chapter reviewed previous studies that are concerned with the credibility, identification of user measurement, and the credibility of the circulation of information about the COVID-19 virus, the epidemic of this era. We present studies to classify Twitter tweets to help us understand the credibility concept. And the difference in user credibility perception from one person to another to evaluate the credibility of the information. Also, The spread of fake information at the beginning of the COVID-19 epidemic about people allegation many ways of treatment and the extent to which a specific group of society was affected by this information, for example, older people and people who had a lack of awareness of using social media. These studies results can help people share on social media and improve their thinking about repost information.

Chapter 3

METHODOLOGY

METHODOLOGY

3.1 Introduction

In this chapter, we explain the proposed model, it is a supervised machine learning for validation of the credibility of information on social media. We try to classify health tweets mainly about COVID-19 into two classes credible or not credible.

This section is organized as follows: Section 3.2 discusses the dataset and section 3.3 discusses the proposed model and how it works.

3.2 Dataset

To get the needed dataset for this project, we used Twitter Intelligence Tool (TWINT) to crawl data from Twitter [17]. TWINT is an advanced Twitter scraping tool written in python that allows for scraping tweets from Twitter. We crawled English tweets related to COVID-19 by using three different hashtags COVID-19, COVID19, and coronavirus. We selected tweets randomly, and then evaluate them. We evaluate these tweets into three categories: Credible, not credible, and not sure.

The evaluation process was based on the content of the tweets including text of the tweets, links, images, etc. The final judgment was based on the majority. The total number of evaluated tweets was 648, then we removed (not sure class) as it is not part of this research. We adopted two classes, credible and not credible. Figure 3.1 shows the total number of the dataset to predict the credibility of tweets related to COVID-19.

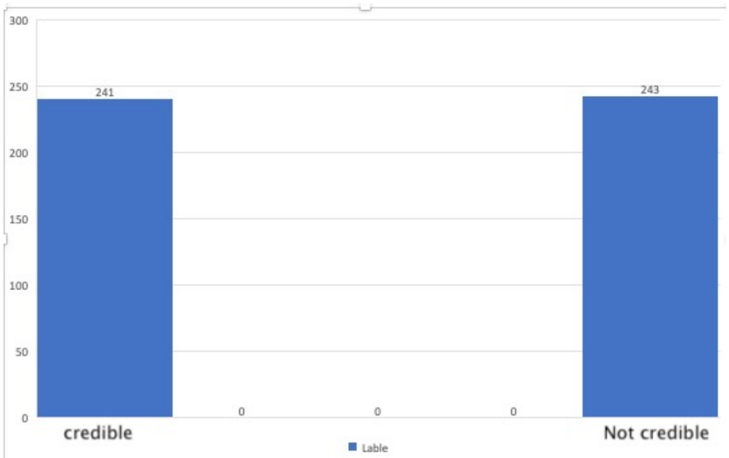


Figure 3.1: The total data of two classes

3.3 Proposed model

We aim to develop an automated model for classifying tweets into credible and not credible. Figure 3.2 shows the proposed model architecture.

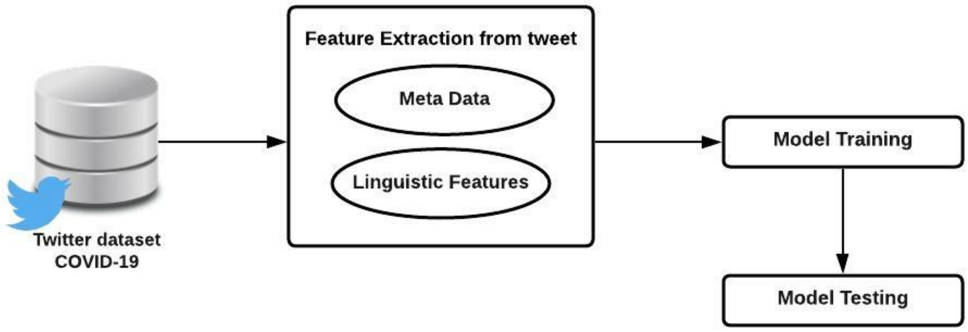


Figure 3.2: The proposed model architecture

3.3.1 Feature Extraction

To predict the credibility of tweets, features must be extracted from tweets to obtain the required features. In this project, we depend on content-based features and linguistic features we described features in detail in chapter 4.

3.3.2 Training

The training data is an initial set of data used to train a machine learning model to predict the credibility of tweets related to COVID-19. In this term. We use a scenario of 80% to train our model by using both features content and linguistic features.

3.3.3 Testing

This term used to test a machine learning algorithm after it has been trained on a primary training dataset. We tested our model used a scenario of 20% using both features content and linguistic features.

Chapter 4

IMPLEMENTATION AND TESTING

IMPLEMENTATION AND TESTING

4.1 Introduction

In this chapter, we introduce the steps of building the proposed model starting from used features and how to convert data format. And explain how the model environment works.

4.2 Used Features

In this project, we used two types of features content-based features and linguistic features.

- The first type of feature is content-based features such as number of mentions, URLs, photos, replies-count, retweets-count, likes-count, hashtags, quote-URL and video. These features generate after the user posts the tweet, and they take time to be generated.
- The second type is the linguistic features that have a well-known effect on achieving high accuracy, as well as these features always available in the tweet, while other features could be absent.

To extract the linguistic feature we used LIWC2015 [18], the total number of extracted features were 93 from four different dictionaries:

1. Linguistic Processes dictionary (total word count, words per sentence, percentage of words captured by the dictionary, and percent of words longer than six letters, etc.).
2. Psychological Processes (Social processes, Positive emotion, Negative emotion, Health, etc.).
3. Personal Concerns (Work, Achievement, Death, Money, etc.).
4. Spoken categories (Assent, Nonfluencies, Fillers, etc.).

4.2.1 LIWC

Linguistic Inquiry and Word Count (LIWC) is one of the best text analysis software programs designed by James W. Pennebaker, Roger J. Booth, and Martha E. Francis. It can analyze more than 70 languages. And contains a group of dictionaries, Linguistic Processes, Psychological Processes, Spoken categories, and Personal Concerns, for each dictionary, have a subdictionary. LIWC can read a given text and analyze each word by the text analysis module. After the targeted word is processed, the text analysis module looks for a dictionary match with the current target word. Then counts the percentage of words, then represents the degrees of different emotions, thinking styles, social concerns, and even parts of speech with its ability to analyze Twitter tweets. This technology can analyze hundreds of text files and big data and can capture negative and positive emotions. After the analysis processing targeted word, the results display on a new file.

4.3 Implementation Steps

This section will present model environment, and how to convert data format. We built a machine learning model to predict credibility in two steps: The first step of the model development was to extract the dataset that contains content features and then the feature extraction process by using LIWC2015. In the second step, we train the model using both content-based and linguistic features, with a number of machine learning classifiers to predict the credibility of English tweets related to COVID-19.

4.3.1 Implementation Environment

We used a Waikato Environment for Knowledge Analysis (Weka) environment. Weka is a data mining software named after a flightless New Zealand bird, it is a set of machine learning algorithms that can be applied to a data set directly and includes methods for the main data mining: classify, regression, classification, clustering, association rule mining, and attribute selection, it easy to apply algorithms. Weka has five graphical user interfaces (GUI) figure 4.1 shows GUI: Explorer, Experimenter, Knowledge-Flow, Workbench, and command-line interfaces (simple CLI):

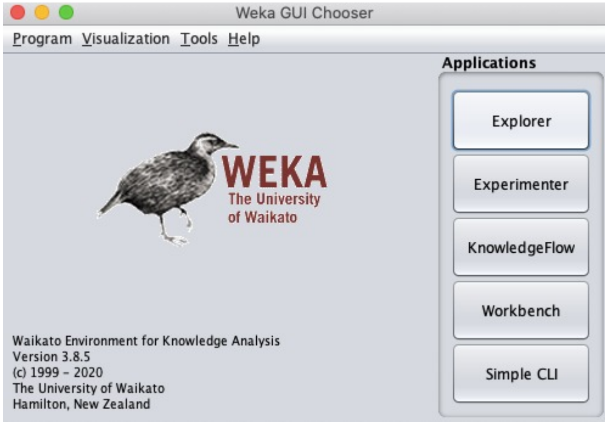
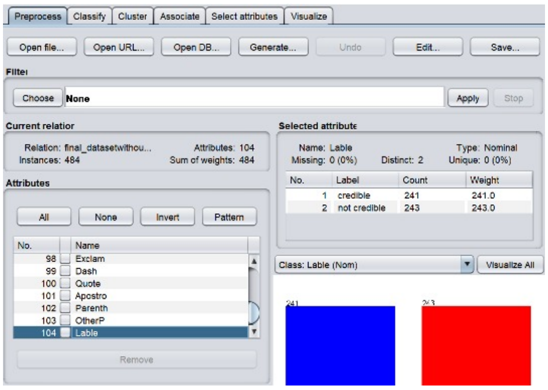
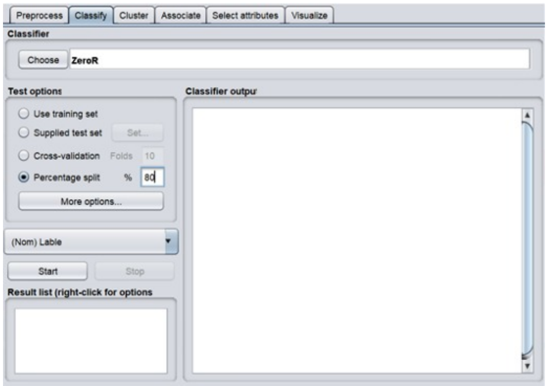


Figure 4.1: Weka graphical user interfaces.

- **Explorer** We used Weka through a graphical user interface called the Explorer figure 4.2 shows the Explorer interface:



(a) Preprocess: Choose and modify the data



(b) Classify: That allows to selecting algorithms.

Figure 4.2: Explorer interface.

- **Experimenter** The experimenter an environment makes enables the user to create, run, modify, and analyse experiments and statistical tests.
- **Knowledge-Flow** The knowledge flow data can be processed and analyzed through Weka components and placed in a layout to connect them to form a ”knowledge flow”.
- **Workbench** The workbench is very similar to the explorer. It is an environment that reduces or removes the need for tools such as data manipulation, result visualization,

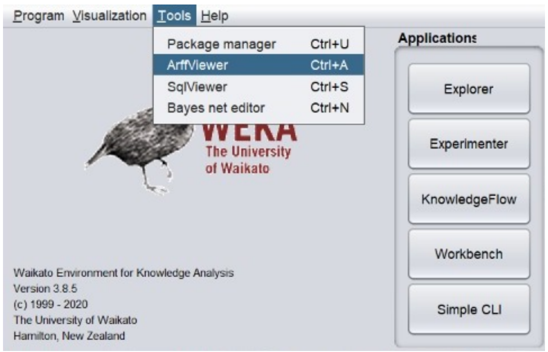
database linkage, cross-validation and comparison of rule sets, data manipulation, result visualization, database linkage to complement the basic machine learning tools.

- **Command-line interfaces (simple CLI)**

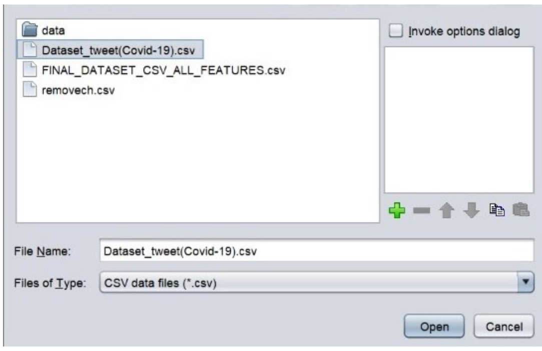
  A simple command line interface provided to run Weka functions directly.

4.3.2 Convert CSV file into ARFF file

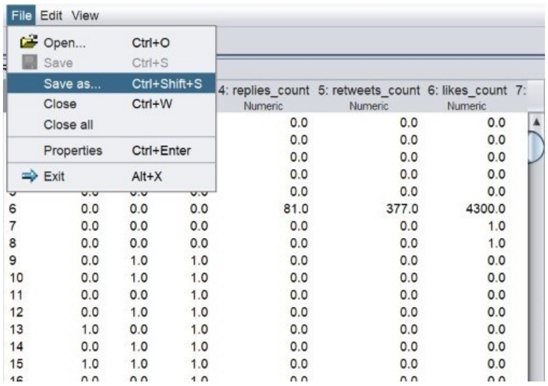
Attribute Relation File Format (ARFF) is a default way to represent data in Weka. We extracted data into Comma Separated Values (CSV) file format figure 4.3 shows how to convert the dataset into an ARFF file format.



(a) open the ARFF-Viewer to change the file format CSV into ARFF files format.

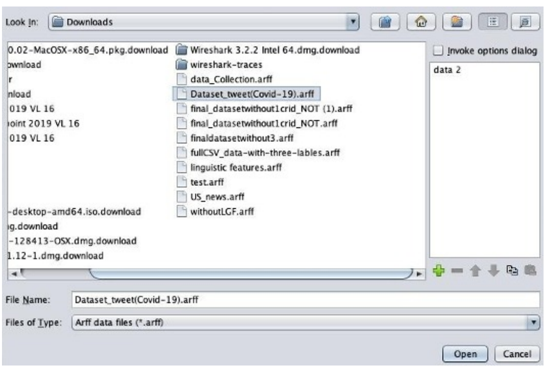


(b) By clicking on open button the data will display.

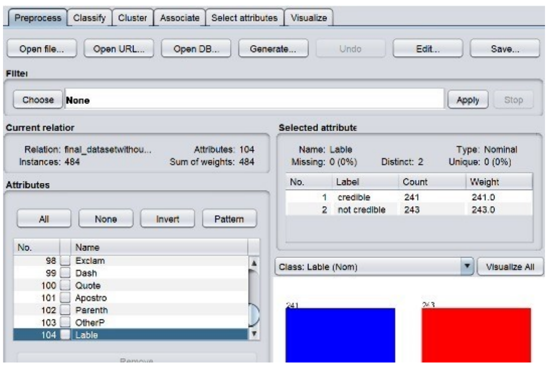


(c) Data will displayed. By clicking the file menu click save as to convert data.

(d) Selecting save as enter a filename with ARFF.



(e) To open data in Weka by select the "Explorer" and "Open file" choose ARFF file format.



(f) The file ARFF is ready to use.

Figure 4.3: Explorer interface.

4.4 Machine learning Algorithm

In this section, we present four classifiers that we used in this project. All these classifiers have been selected based on popularity and effectiveness in the credibility prediction domain [5].

We tested four supervised algorithms Random Forest (RF), J48, Naive Bayes (NB), and Logistic Regression(LR) by using both linguistic and content features. We split the dataset into 80%-20%, 80% for testing, and 20% for training.

1. **Random Forest**
   Is a supervised learning algorithm used for both classifications as well as regression. It is combined with a series of tree classifiers to get a more accurate and stable prediction.

2. **Naïve Bayes**
   Is a classification algorithm based on Bayes Theorem and it is a simple learning algorithm , a set of probabilistic classifiers that aim to process, analyze, and categorize data.

3. **J48**
   A Decision tree is of the popular algorithms for machine learning to examine data, which is a support tool with a tree-like structure that designs the possible outcomes of the data.

4. **Logistic Regression**
   Is a supervised machine learning algorithm that is used for classification problems, it is a predictive analysis algorithm, and based on the concept of probability.

4.5 Summary

In this chapter, we describe used featuers content based featuer and linguistic feature. We presented implemntaion steps for proposed model and describe the model environment, how to convert CSV file into ARFF and split data to build a model to predict tweets credibility in chapter five we will more discuss and compare between models.

Chapter 5

RESULTS AND DISCUSSIONS

RESULTS AND DISCUSSIONS

5.1 Introduction

In this chapter, we present the major findings of the machine learning models that we used to predict the credibility of tweets related to COVID-19. We are presenting the confusion matrix, accuracy of the model, precision and recall.

- **Confusion Matrix**

  It is a popular technique for measuring the performance of a machine learning classification algorithm. It is a table that helps to know the algorithm performance on a set of data to find out the predicted values. It contains a column and a row the column representing expected values, and the row the actual values table 5.1 shows the confusion matrix of the binary classification.

	P (Predicted)	**N (Predicted)**
P (Actual)	True Positive	False Negative
N (Actual)	False Positive	True Negative

Table 5.1: Confusion matrix of the binary classification.

  Where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives).

- **Accuracy**

  Accuracy is the measure of evaluating correctly classify instances. To calculate the accuracy number of correct predictions divided by the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

- **Precision**

  Focus to correctly classify positive labels, to calculate precision the number of true positives divided by the sum of the true positives and false positives.

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

- **Recall**

  The recall measures the number of correctly predicted positive labels. To calculate recall the number of true positives divided by the sum of true positives and false negative.

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

5.2 Major Findings

This section compares the performance of four different machine learning classifiers: Random Forests (RF), Naïve Bayes (NB), Logistic Regression (LR), and J48 using both content-based features and linguistic features. Random Forests is achieved the best performance compares with other algorithms with an accuracy of 79.3% .

5.2.1 Result

- **Random Forests**
  Random Forests is a supervised machine learning algorithm used in many researches. We are applying Random Forests as a classifier the model can correctly predict the credibility of the tweets with an accuracy 79.3%, precision 79.4%, and recall 79.4%.

- **J48**
  J48 is a supervised machine learning, and it is a popular algorithm. We applied J48 by using content-based features and linguistic features. The model can correctly predict the credibility of the tweets with an accuracy 71.1%, precision 71.8%, and 71.1% recall.

- **Logistic Regression**
  Logistic Regression is a supervised machine learning algorithm. We applied Logistic Regression as a classifier by using both content-based features and linguistic features. The model can correctly predict the credibility of the tweets with an accuracy of 64.9%, recall 64.9%, and precision 65.3%.

- **Naive Bayes**
  Naive Bayes is a supervised machine learning algorithm. We applying Naive Bayes as a classifier by using both content-based and linguistic features it gets an accuracy 63.9%, precision 67%, and recall 63.9%.

5.3 Discussion

This project aims to classify tweets related to COVID-19 into two classes credible and not credible by using a supervised machine learning technique. The benefit of classifying tweets to reduce not credible health information related to COVID-19 that has a negative impact on people. This section discusses the result of the four machine learning algorithms and previous studies.

We compare the performance between four machine learning algorithms when training the classifiers by using a set of features content-based and linguistic features and compared with the features effective in the models. After the training process, the Random Forests achieved high accuracy 79.3%, precision and recall get the same result 79.4% compared with other algorithms. While J48 got less than Random Forests in terms of accuracy 71.1%, precision 71.8%, and recall 71.1%. And the logistic regression gets less results compared with Random Forests and J48 with an accuracy 64.9%, precision 65.3%, and recall 64.9%. Naive Bayes get the worst result with accuracy and recall 63.9% compared with other algorithms, but in the precision get higher than Logistic Regression with a 67% table 5.2 shows the proposed models performance.

The dataset contains a number of features, but not all features are effective in the classification process. In Weka [19], we used the selection attributes technique to ranking importance features. If features have a value between 0 and 1, that means they have higher effectiveness, whereas a value that is equal to zero has a lower level of effectiveness. After comparing the

	Accuracy	Precision	Recall
Random Forests	0.793	0.794	0.794
J48	0.711	0.718	0.711
Logistic Regression	0.649	0.653	0.649
Naive Bayes	0.639	0.670	0.639

Table 5.2: The models performance.

importance of features between four algorithms we note in all algorithms URL has a high impact feature, and they have the same importance of features figure 5.1 displays feature importance in four algorithms.

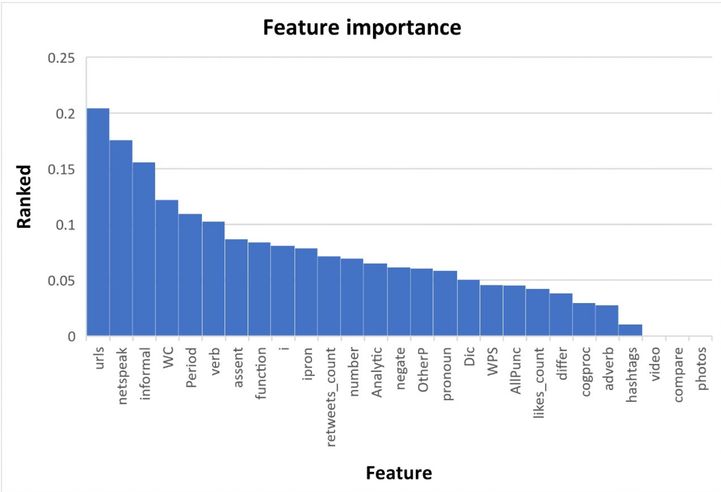


Figure 5.1: Feature importance.

Compared with the new study [16] they used IEEE data port. To extract features they used the term frequency-inverse document frequency (TIF-IDF) and bag-of-words (BoW) to analyze the sentiment of tweets related to COVID-19 and they get an accuracy 93%. Also, in [12] proposed a Bidirectional Encoder Representations from Transformers (BERT) model to classify sentiment related to COVID-19 to extract the different emotions they used Term Frequency-Inverse Document Frequency algorithm (TF-IDF) the model is achieving an accuracy of 75.65%. While in this project, the data was collected by using TWINT tools, we used both content-based features and the Linguistic Inquiry and Word Count (LIWC) tool to extracted linguistic features to predict the credibility of tweets related to COVID-19 and we get an accuracy 79.3%. The power in our model we used all the features based on the text features and metadata of the tweets in a short time.

In [5], they used metadata content-based features and source-based features and then applying Random Forests as a classifier with an accuracy 78.4%. They used a dataset collected by Zubiaga. In our project, we used content-based and linguistic features, and then applying Random Forests we get accuracy 79.3% higher than [5]. Because linguistic features have a higher effect on the model.

Chapter 6

CONCLUSIONS AND FUTURE WORK

CONCLUSIONS AND FUTURE WORK

6.1 Main contributions

This work tries to assess information credibility in social media related to COVID-19. We built a machine learning model that determines tweets credibility prediction related to COVID-19 into two classes credible or not credible. We achieve a good result with an accuracy 79.3%. Based on our knowledge, this project is the first model to assess the credibility of tweets related to COVID-19.

6.2 Challenging for project

In this project, there are a number of limitations::

- Lack of an available dataset because of privacy and policy imposed by Twitter.
- Need to evaluate the dataset, and we evaluated data manually, which took time and effort to assess each tweet.
- Due to lack of time, we were unable to increase the size of dataset.
- Lack of previous studies related to the credibility prediction of tweets related to COVID-19, to compare with our model.

6.3 Future work

There are a number of possible directions to improve the results and achieve higher accuracy by increasing the size of the dataset. We may improve the used methodology using different hashtags, languages, and various new features such as TIF-IDF and bag-of-words. One of the future research directions is to study the unclear tweet.

References

[1] S. Y. Rieh, "Credibility assessment of online information in context," 2014.

[2] X. Hu, "Assessing source credibility on social media——an electronic word-of-mouth communication perspective," Ph.D. dissertation, Bowling Green State University, 2015.

[3] K. Sharma, S. Seo, C. Meng, S. Rambhatla, and Y. Liu, "Covid-19 on social media: Analyzing misinformation in twitter conversations," *arXiv preprint arXiv:2003.12309*, 2020.

[4] S. Tasnim, M. M. Hossain, and H. Mazumder, "Impact of rumors and misinformation on covid-19 in social media," *Journal of preventive medicine and public health*, vol. 53, no. 3, pp. 171–174, 2020.

[5] N. Y. Hassan, W. H. Gomaa, G. A. Khoriba, and M. H. Haggag, "Supervised learning approach for twitter credibility detection," in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. IEEE, 2018, pp. 196–201.

[6] N. Hassan, W. Gomaa, G. Khoriba, and M. Haggag, "Credibility detection in twitter using word n-gram analysis and supervised machine learning techniques," *Int. J. Intell. Eng. Syst*, vol. 13, pp. 291–300, 2020.

[7] P. Rawat and C. Gupta, "User response based information quality assessment of social media news posts," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*. IEEE, 2018, pp. 1–6.

[8] A. Mourad, A. Srour, H. Harmanani, C. Jenainatiy, and M. Arafeh, "Critical impact of social networks infodemic on defeating coronavirus covid-19 pandemic: Twitter-based study and research directions," *arXiv preprint arXiv:2005.08820*, 2020.

[9] F. Alam, F. Dalvi, S. Shaar, N. Durrani, H. Mubarak, A. Nikolov, G. D. S. Martino, A. Abdelali, H. Sajjad, K. Darwish *et al.*, "Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms," *arXiv preprint arXiv:2007.07996*, 2020.

[10] O. Oyebode, C. Ndulue, A. Adib, D. Mulchandani, B. Suruliraj, F. A. Orji, C. Chambers, S. Meier, and R. Orji, "Health, psychosocial, and social issues emanating from covid-19 pandemic based on social media comments using natural language processing," *arXiv preprint arXiv:2007.12144*, 2020.

[11] D. Kar, M. Bhardwaj, S. Samanta, and A. P. Azad, "No rumours please! a multi-indic-lingual approach for covid fake-tweet detection," *arXiv preprint arXiv:2010.06906*, 2020.

[12] T. Wang, K. Lu, K. P. Chow, and Q. Zhu, "Covid-19 sensing: Negative sentiment analysis on social media in china via bert model," *Ieee Access*, vol. 8, pp. 138 162–138 169, 2020.

[13] A. Hershkovitz and Z. Hayat, "The role of tie strength in assessing credibility of scientific content on facebook," *Technology in Society*, p. 101261, 2020.

[14] L. Wu, Y. Rao, H. Yu, Y. Wang, and N. Ambreen, "A multi-semantics classification method based on deep learning for incredible messages on social media," *Chinese Journal of Electronics*, vol. 28, no. 4, pp. 754–763, 2019.

[15] R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, and K. Shaban, "Cat: Credibility analysis of arabic content on twitter," in *Proceedings of the Third Arabic Natural Language Processing Workshop*, 2017, pp. 62–71.

[16] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for covid-19 tweets sentiment analysis," *Plos one*, vol. 16, no. 2, p. e0245909, 2021.

[17] C. Zacharias. (2019) twint tool. Accessed: 15.4.2021. [Online]. Available: https://github.com/twintproject/twint

[18] R. J. B. James W. Pennebaker and M. E. Francis. (2015) Liwc2015. Accessed: 15.4.2021. [Online]. Available: https://liwc.wpengine.com/

[19] S. Gnanambal, M. Thangaraj, V. Meenatchi, and V. Gayathri, "Classification algorithms with attribute selection: an evaluation study using weka," *International Journal of Advanced Networking and Applications*, vol. 9, no. 6, pp. 3640–3644, 2018.