Deakin SIT Research Project LaTeX Template

Submitted as Research Report / Honours / Master Dissertation in
SIT723/SIT724

SUBMISSION DATE
T1-2021

First-Name Last-Name
STUDENT ID 1234567
COURSE - Master of Software Engineering Honours (S464)

Supervised by: Dr. Supervisor1, Prof, Supervisor2

Abstract

Abstract goes here...

Contents

List of Figures

List of Tables

1 Introduction

1.1 Aim & Objectives

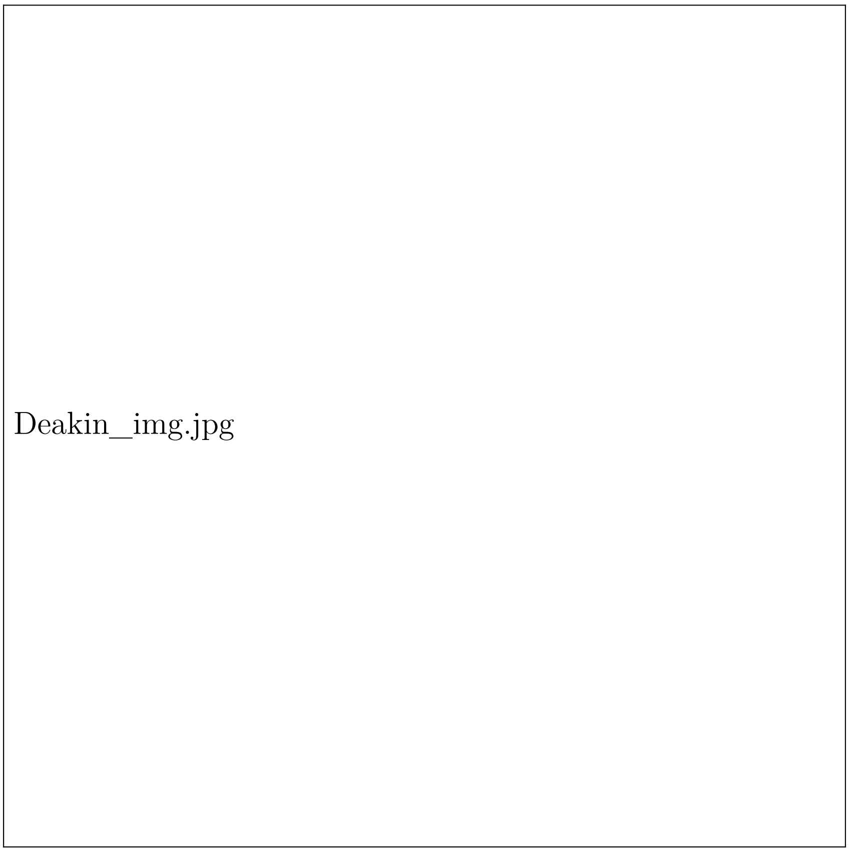

Figure 1: Deakin Campus Building.

Figure 1 shows the Deakin BC building at Burwood campus in Melbourne.

1.2 Structure

Section 2 reviews the literature. Section 3 presents the research design and methodology. Section 4 describes the approach and the technical details of artefact development. Section 5 evaluates the artefacts, on the basis of research questions

(RQs) in Section 5.1 and discusses the RQs in Section 6. Section 7 discusses threats to validity and Section 8 concludes the report.

2 Related Work

Most of COVID19 misinformation detection systems implement machine learning techniques to help public in classifying whether the news spreading in social media is reliable or not [?]. Machine learning is a subset of artificial intelligence where the main aim is to train machines by using algorithms about some statistical phenomenon to make decisions like human. It identifies the pattern of the data point based on some mathematical relation and predicts the new data point in similar way. All the ML methods are discussed in two separate subsections named Traditional ML methods and DL methods [?]. With the existing computational capabilities and a large amount of data, supervised deep learning learning model provide better performance compared to traditional machine models [?].

For deploying machine learning techniques in building COVID19 misinformation detection models, existing studies rely on fact checking websites (i.e., Snope, Media Bias/Fact Check, Factcheck.org, etc.), where a lot of experts and annotators manually classify news into different rating categories (i.e. Reliable, Questionable, True, False, etc.). These fact-checking website can be social media or news websites, fact-checking websites, government or well-recognized authentic websites. The number of English-language fact-checks increased more than 900% from January to March 2020 [?]. Fact-checking websites such as the FactCheck.org and Poynter are the primary sources of current COVID-19 misinformation/rumor data. Furthermore, COVID19 misinformation dataset can be English text or Multilingual text.

Cui and Lee [?] present Covid-19 heAlthcare mIsinformation Dataset (CoAID), including fake news on websites and social platforms, along with related user engagements engagements (i.e., tweets and replies) about such news. They first located several fact-checking websites (i.e. WHO, WebMD, Healthline) to collect true news, and then gathering fake news from CheckYourFact, PolitiFact, etc. Similarly, Zhou et al.[?] refer to credentials rating report by NewsGuard[1] and Media Bias/Fact Check[2]

[1] https://www.newsguardtech.com/

[2] https://mediabiasfactcheck.com/

to identified 22 reliable news outlet and 38 unreliable news site, and then collected dataset from selected news websites. Recent study also focus on multilingual COVID-19 information. Shahi et al. [?] introduced a multilingual cross domain fact check COVID-19 dataset, which included 40 languages and 105 countries. While [?] indicated that not all dataset include labels, because COVID-19 dataset require well recognised fact checkers to annotating, which is time consuming due to Velocity of News.

To the date, many automated COVID19 misinformation detection system has been proposed in order to decrease harmful effect of COVID19 misinformation. For example, Constraint'21 [?] launched shared tasks to invite researchers working on COVID19 multilingual misinformation detection. Li et al.[?] assessed the performance of different pre-trained language models such as BERT, Roberta, Erbue, etc with various training strategies. And they achieved 0.9858 of weighted F1 score by transformer-based model. Additionally, Anna et al. [?] implemented COVID-Twitter-BERT (CT-BERT), transformer based model which pretrained on a large corpus of Twitter message on tweets related to COVID19, to achieved 0.9837 Weighted F1 score. Both of study training model included dataset provided by [?].

Besides the veracity labels and sources provided in the above-mentioned fact-checking platforms, other meta-information, such as sentiment and stance, is missing in most studies [?]. H et al. [?] used Latent Dirichlet Allocation(LDA) combined with Gibbs sampling in order to discovering topic of COVID19 misinformation on Reddit posts. Among top 10 topics classified by LDA, Most frequent term of "people", "virus","sympotoms","infection","cases", "diesease" are revealed. They also recognised user's emotion/sentiment based on Long short-term memory (LSTM) and SentiStrength, a free sentiment analysis method, indicated that 35.36% of user's COVID19 comments toward misinformation is postive and very positiv. In constrast, 23.16% of comments is negative and very negative and the other are neutral sentiment.

In order to get intrinsic view of research work about detecting COVID19 misinformation, it suggested to better understand COVID19 information themselves across social media. Therefore, Gabarron et al. [?] conducted a systematic review of COVID19 misinformation publication and they founded that, most studies reported that posts related to misinformation on social media included false information, jokes, rumors etc, while only four studies contribute to examining the effect of misinformation. Social media platforms provided direct access to an unprecedented amount of content and amplify rumors and misinformation. M et al. [?] perform a comparative analysis of user's activity on Gab, Reddit, YouTube, Instagram and

Twitter to study social behavior of user on topic of COVID19 misinformation. In addition, an exploratory study on misinformation spreading across social media conducted by [?] and indicated that 51.9% of the re-shares of false rumours occur after this debunking comment result from readers not reading all the comments before re-sharing. [?] Their study aim to understand the psychological impacts of misinformation on public perception. Author create online dashboard, which provides a daily list of identified misinformation tweets, along with topics, sentiments, and emerging trends in the COVID-19 Twitter discourse.

To the date, misinformation spreads and changes very quickly, often unpredictably. Universal language models may perform weakly in these recent misinformation detection due to the lack of large-scale annotated data and adequate semantic understanding of domain-specific knowledge [?]. Furthermore, in the case of fact checking related to COVID-19 claims, both understanding and trust are necessary for the adoption of the predictions [?]. Only a few studies focus on how explanations and predictions from machine learning model can be harness to improve human decision [?]. And model interpretability is a major challenge to applications of ML methods, which has not been given enough attention in the field of machine learning research [?], especially COVID-19 misinformation detection system. Therefore, it suggested that in addition to improving the performance of the model in task of COVID-19 infodemic, we should also improve the interpretability of the model so that it can communicate with decision maker or public. In this study, we use SHAP [?] to explain model in CoAID, ReCOVery and our own multilingual COVID-19 dataset.

3 Research Design & Methodology

3.1 COVID-19 Misinformation on News Sites

We identified three reliable fact website sources collect data. Poynter[3], Snope[4], reliable media outlets evaluated by Media BiasFact check[5].

[3] https://www.poynter.org/ifcn-covid-19-misinformation/

[4] https://www.snopes.com/fact-check/

[5] https://mediabiasfactcheck.com/

Poynter is a non-profit making institute of journalists. In COVID-19 crisis, Poynter came forward to inform and educate to avoid the circulation of the fake news. Poynter maintains an International Fact-Checking Network(IFCN), the institute also started a hashtag #CoronaVirusFacts and #DatosCoronaVirus to gather the misinformation about COVID-19. Poynter maintains a database which unites more than 100 fact-checkers around the world and includes 70+ countries and 40+ languages.

Snope is an independent publication owned by Snopes Media Group. Snopes verifies the correctness of misinformation spread across several topics. As for the fact-checking process, they manually verify the authenticity of the news article and performs a contextual analysis. In response to the COVID-19 infodemic, Snopes provides a collection of a fact-checked news article in different categories based on the topic of the article.

Reliable Media Outlets

3.2 Data Collection

This subsection explains the steps followed to collect data from different-checking websites.

Misinformation Dataset We collected data from an online fact-checker website called Poynter [20]. Poynter has a specific COVID-19 related misinforma- tion detection program named 'CoronaVirusFacts/DatosCoronaVirus Alliance Database12'. This database contains thousands of labelled social media information such as news, posts, claims, articles about COVID-19, which were manually verified and annotated by human volunteers (fact-checkers) from all around the globe.

3.3 Define classes for misinformation

We collected fake news from several fact checking websites, and original classes of misinformation vary depend on how fact checking rating system. For example, "Pants on Fire", a simple rhyme known by Children all over the United States, they say it when someone gets caught in a lie. In other words, when someone gets busted for lying[?]. which be used by PolitiFact to rated news as false. And Factchek[6] used “Manipulation” to annotate news, which contain claims that are beyond misleading or are based on methods that can be easily manipulated or framed in a manipulative way. Furthermore, data crawled from Snopes even contains 98 misinformation check websites cross the world. For an instance, “faux” be used in french fact checking organisation which means fake, and ’fałsz’ is Polish word and equal to false in English. Overall, we manually checked rating system provided by different fact checking website and normalised theses classed by mapping them into 4 labels (i.e. “Flase”, “mixed”,“True”,“Others”). We also provided a table to overview of verdict categories that we and original fact check websites defined misinformation. Details of definition of different type of fact-check articles can be found in this study[?].

Table 1: Normalisation of original categorisation by the fact checking web sites.

Normalisation	Original Classed	Definition
True	First solution	The rated statements are demonstrably true and no significant details are missing. Data from selected reliable news organisation discussed in Section3.1
False	Second solution	Data 1 Data 2 Data 3
Mixed	Third solution	Data 1 Data 2 Data 3
Others	Third solution	Data 1 Data 2 Data 3

3.4 Data Cleaning & Processing

[6] www.factcheck.kz

3.5 Model Setup

3.6 Model Explanation

4 Artefact Development Approach

5 Empirical Evaluation

5.1 Research Questions (RQs)

6 Results & Discussion

7 Threats to Validity

8 Conclusion & Future Work

8.1 Future Work