DETECTION AND ANALYSIS OF FAKE REVIEWS ON ONLINE SERVICE PORTAL

BY

LIONG YONG XUAN

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEMS (HONOURS) BUSINESS INFORMATION SYSTEMS

Faculty of Information and Communication Technology

(Kampar Campus)

MAY 2022

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: Detection and Analysis of Fake Reviews on Online Service Portal

Academic Session: MAY 2022

I LIONG YONG XUAN

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

(Author's signature) (Supervisor's signature)

Address:
2-02A Pangsapuri Puteri 2,
Jalan Damai Impian 2/8,
Taman Damai Impian 2,
56000 Cheras, KL.

Dr.Ramesh Kumar Ayyasamy
Supervisor's name

Date: 8th September 2022 **Date**: 09-Sep-2022

FACULTY/INSTITUTE* OF INFORMATION AND COMMUNICATION TECHNOLOGY

UNIVERSITI TUNKU ABDUL RAHMAN

Date: 8th September 2022

SUBMISSION OF FINAL YEAR PROJECT /DISSERTATION/THESIS

It is hereby certified that **LIONG YONG XUAN** (ID No: **17ACB05552**) has completed this final year project~~/ dissertation/ thesis~~* entitled **Detection and Analysis of Fake Reviews on Online Service Portal**" under the supervision of **Dr Ramesh Kumar Ayyasamy** (Supervisor) from the Department of **Information Systems**, Faculty~~/Institute~~* of **Information and Communication Technology**.

I understand that University will upload softcopy of my final year project ~~/ dissertation/ thesis*~~ in pdf format into UTAR Institutional Repository, which may be made accessible to UTAR community and public.

Yours truly,

Liong Yong Xuan
(*Student Name*)

*Delete whichever not applicable

DECLARATION OF ORIGINALITY

I declare that this report entitled "**Detection and Analysis of Fake Reviews on Online Service Portal**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : 

Name : Liong Yong Xuan

Date : 8th September 2022

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Dr Ramesh Kumar Ayyasamy and my previous supervisor Dr Rehan Akbar who has given me this bright opportunity to engage in this research project. This project helps me in gaining more knowledge about the fake reviews detection techniques that I have never concerned before. Moreover, Dr Ramesh have being helpful in guiding me when I was facing problem and so I could complete this assignment.

To a very special person in my life, Teng Ying Ning, for her patience, unconditional support, and love, and for standing by my side during hard times. Her presence changed me a lot and made me want to be better. Finally, I must say thanks to my friends and my family for their love, support, and continuous encouragement throughout the course. I would say if without one of these people, I definitely could not complete this report.

ABSTRACT

Nowadays, the use of the World Wide Web and online service platforms has been quite popular, especially during the Covid-19 outbreak, which resulted in the implementation of lockdown, social isolation, and other preventive measures across the country. Massive amounts of products and services are offered through online platforms, leading to a significant volume of information being generated. Consumers can also provide reviews on products or services that they have purchased on online shopping platforms. In order to reach a conclusion on business strategies and product or service improvements, these reviews are beneficial to both consumers and firm alike. Some businesses, on the other hand, are recruiting writers to post fraudulent favourable impressions about their own products or services, or dishonest bad comments about their rivals' products or services, in exchange for a fee. This strategy provides incorrect information to new customers who are looking to purchase such things or services, and as a result, a system that can identify and eliminate misleading reviews are required to solve the problem. In this paper, a framework of a Machine Learning based fake review detection model has been proposed to identify which classification algorithm is the most effective with the proposed framework.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

NLP	Natural Language Processing
POS	Part-of-speech
PU	Positive Unlabelled
ASM	Author Spamicity Model
NB	Naïve Bayes
SVM	Support Vector Machine
DT	Decision Tree
RF	Random Forest
GBTs	Gradient-Boosted Trees
KNN	K-Nearest Neighbours
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Chapter 1
Introduction

Online service portals play a key role in information circulation, which is reflected as an essential asset for both sellers and customers of certain services and products in their promotional activities. Users can share service and product reviews on online service portals. Consequently, numerous individuals make decisions to buy products or services based on user reviews, and the positive reviews are more encouraging to choose services or products while negative user reviews are discouraging to choose services or products [1]. Since any user can leave a comment as a review on online service portals, spammers will deliver negative comments about the service or product, trying to mislead customer opinion. The negative reviews which are spreading online will change the user perception of a bad or good product. Therefore, the utilization of a review-cantered model to detect spam reviews and a customer-cantered model to detect spammers is crucial to detect and stop spammers and spam reviews.

1.1 Problem Statement and Motivation

Consumers have grown to rely on internet product and service reviews to help them make decisions when making online purchases. As a result, product reviews provide information that influences the purchasing decisions of customers, manufacturers, and retailers. Customers use reviews to provide word-of-mouth information about things, such as product quality, utility, and durability, and share their own experiences with others [2]. Increasing the number of online service portals has increased resources for gathering customer reviews about their service and product experience. Due to anyone can post anything and get away with it, there has been an increase in the number of false reviews. As customers increasingly communicate with one another online and share their experiences and thoughts on a variety of service interactions, businesses are allocating more resources to monitoring, analysing, and correcting, as well as boosting their online reputation [1]. There has been a rise in illusive review spam, which are fake reviews that are designed to appear genuine. Fake, strident, spam, misleading reviews are those written by those who do not have personal experiences with the topics of the reviews. Spammers spread fake reviews in order to denigrate or promote a specific brand or product, persuading consumers to purchase from that brand or not [2].

The research is important since it helps to identify spam user reviews in online service portals. The results of this study will serve as theoretical and practical contributions to the rise of spam reviews on products and services [3]. Furthermore, the research will add to the current literature on the techniques implemented by various organizations to detect spam reviews. This ensures that customers receive legitimate reviews on products they may be interested in purchasing from online portals [3].

1.2 Project Objectives

1.2.1 What are the current methods using for fake reviews detection?

The purpose of this project is to research and analyse current fake reviews detection methods. A good fake reviews detection method measures the integrity value of a review, the credibility value of the reviewers and a product or service's reliability value. The reason of researching and studying for the current methods using for fake reviews detection is to understand how effective the detection methods are, and what is the limitation of the detection methods. If there is a better fake reviews detection system in place, there will be less victims fell to prey.

1.2.2 Which classification algorithm is the most effective in the proposed Machine Learning based detection model?

Fake reviews, which provide an unreliable impression of a product's quality, limit the effectiveness of online reviews. Thus, it is important to identify fraudulent reviews. In this paper, a framework for a machine learning-based detection model is suggested. Five distinct classification algorithms—Naive Bayes (NB), K-Nearest Neighbours (KNN), Decision Tree, Support Vector Machines (SVM), and Random Forest—have been investigated in the proposed model to see which is the most effective. This study analyses the outcomes of various classification algorithms when an extraction feature from the language model—TF-IDF with bi-grams—is present. Different classification techniques could result in varying accuracy rates for spotting fraudulent reviews. Therefore, to identify the most efficient classification technique, we have demonstrated the comparison of classification results.

1.3 Project Scope and Direction

The outcome of this project is a research paper that useful for data analytics and information retrieval. This project is to target online service platform such as Agoda (hotel reservation), Amazon (products selling) et cetera. Online service platform is very popular nowadays, due to the good marketing strategies to attract the public to get their desired things through these platforms. Furthermore, these platforms usually contain the basic stakeholders such as customers, firms, and the platform administration. The administration provided the platform to firms to sell their products or services and collaborated to have good marketing strategies in order to attract the customers. Besides, the firms are mostly relying on the reviews by the consumers' previous experience to boost their sales. Hence, the relationship among these stakeholders is effective and suitable to be the targeted coverage in this project.

1.4 Contributions

This research contributes to users of online service portals. Online service portals are growing popularity day by day. The increased number of users are prone to fake reviews of subpar quality products and services or products and services that do not meet their expectations. Online service portals users' expectations are encouraged by reviews from previous consumers or experiencers that have bought the specific product or service. However, fake reviews will ruin the experiences of online service portals users as when they received the products or services that's not on par with their expectations by reading these fake reviews.

1.5 Report Organization

This report is organized into 6 chapters: Chapter 1 Introduction, Chapter 2 Literature Review, Chapter 3 System Model, Chapter 4 Experiment, Chapter 5 Conclusion, Chapter 6 Recommendation and Future Work. The first chapter is the introduction of this project which includes problem statement and motivation, project objectives, project scope and direction, project contribution, and report organization. The second chapter is the literature review carried out on several current methods using for fake reviews detection, and some analysis of existing fake review detection using Machine Learning. The third chapter is discussing the overall proposed framework of this project. The fourth chapter is regarding the experimental results that perform by the proposed model. Furthermore, the fifth chapter of this reports the conclusion, and the final chapter six is about the recommendation and future work.

Chapter 2
Literature Review

2.1 Current methods using for fake reviews detection

Analyse reviews manually is the basic method to detect fake review. This method is based on the idea that humans can identify whether other individuals are acting dishonestly. The benefit of carefully examining false reviews is that it allows for the development of understandable and interpretable heuristic rules. Costa et al. [5] established a system of rules to differentiate between benefiting and non-benefiting reviews, such as the length, opinion, and usefulness rate of the review. Filieri [6] looked into how people evaluate the reliability of an online review and discovered that aspects including the review's substance and writing pattern, as well as the existence of images, length, depth of specifics, and overwhelming positivity or negativity, these all play a significant role.

However, there are some challenges that will be faced while using manual detection. The problem of employing heuristic rules seems to be that they are sometimes not precise. As an example, fraudulent reviews published by users with a low number of reviews. If the quantity of written reviews is used as an indicator, singleton spammers may go unnoticed [7]. Besides, another issue is that once spammers understand the rules of fraudulent detectors, they integrate and adjust their behaviour, rendering the rules are invalid. These difficulties may explain why humans are just not very good at predicting fraudulent reviews [4]. For example, researchers found out that the accuracy for human detection method is 21%-34% lower than for a Machine Learning model [8]-[10]. Hence, it is quite difficult for human to identify whether the review is fake or real. Apparently, the heuristic rules used to detect the genuinely of reviews are ineffective against a variety of deceptive strategies.

Another issue with manual detection would be the volume of online reviews is rapidly increasing. Manual approaches generally do not suitable for analysing a certain large quantity of reviews, assuming that a product may obtain large numbers of reviews, and reviews exist for lots of items such as firms, goods, and service providers [11]. As a result, researchers agree that using automated approaches to detect fraudulent reviews could be a better choice.

In particular for the web and text mining industries, data mining and machine learning approaches represent an intriguing commitment to fraud evaluation. Web mining, as defined

by Liu [12], is the practise of using machine learning technology and methods to identify meaningful information and relationships in web content. Web mining can be divided into three sorts of tasks: structure, usage mining, and content mining. Content mining applies data mining methods and machine learning to collect knowledge and data and categorise organisations. Content mining is evident in the appraisal of mining. Feeling mining is the process of attempting to determine the emotion of a text passage by referencing the passage's attributes. A classifier can be practised to classify new cases by breaking down the textual attributes associated with various results. Spam detection, including feeling mining, falls under the category of content mining and makes advantage of features that are just not directly related to the content [13].

Despite the fact that most current machine learning approaches are not sophisticated enough to handle spam detection, they are considered more efficient than manual detection. The main problem stated by Abbasi et al [14] is that there are no distinctions to explain how reviews are classified as genuine or fraudulent. The usage of a word package, in which single words or sets of short words are utilised as characteristics is a general text mining technique, however research indicates that this is still not enough to generate a perfectly executed spam detection classification. As a result, more functional engineering methodologies for extracting an informative set of functions to enhance spam detection must be developed. Several works in the field of literacy look at a variety of machine learning algorithms for fake review detection.

Automatic detection that uses Natural Language Processing (NLP) techniques concentrates on reviews as text information, focusing lexical features, including the keywords or -s, n-grams, punctuation, semantic consistency, latent subjects, and linguistic style signals [[8], [15]]. Non-textual predictive variables, like user IDs, location information, quantity of reviews created by a person, and several other possibly unusual actions, are the topic of another field of research. Techniques that integrate several kinds of attributes tend to be more productive at fake review detection, as is common for classification problems [8]. The important point to understand is that characteristics might include both textual and non-textual information [16].

Combinations of manual and automatic methods are theoretically workable, but they are uncommon in practise. Munzel's [17] research highlights the need of revealing not just textual but moreover contextual information with human detectors to help them identify fraudulent reviews. Harris [18] suggested a hybrid model in which human detectors were provided the data of psycholinguistic characteristics that were generated algorithmically,

together with the results of two Machine Learning classifiers. Humans could either approve or disapprove with the machine's judgement. They enhanced machine performance by 0.2 percent with adopting this hybrid strategy, and this suggests that human involvement can result in a minor improvement over a strictly machine-based method [18]. The result would be that methods for detecting fraudulent reviews vary from entirely automated to totally manual. It is worth highlighting that, even if a classification method decides if a review is real or not, a person or a cluster of people has often taken a responsibility in developing the classifier by dataset development, data pre-processing, feature engineering, and hyperparameter selection.

2.2 Machine Learning Based Fake Review Detection Method

Several strategies have been developed in previous to detecting fraudulent reviews, especially Machine Learning method. Types of data such as labelled data, unlabelled data, and partially labelled data can be easily process with the Machine Learning approach with Supervised Learning, Unsupervised Learning or Semi-Supervised Learning techniques.

Supervised Learning

Etaiwi and Naymat [19] used a supervised learning method to identify fake reviews. Before using the classification approach, a number of pre-treatment tasks must be completed. These procedures involve the following steps: stemming words, deleting punctuation, and removing stop word. They employ language features to distinguish between genuine and fraudulent reviews. Part-of-speech (POS) and bag-of-words are two language features to look out for. An individual word or a group of words that appear in a certain text are collected and stored in the bag-of-words function. Following that, several classification techniques, such as Decision Trees, Random Forests, Support Vector Machines, Naive Bayes, and Gradient Boosting Trees, are applied. In this case, Naive Bayes and Support Vector Machines produce more accurate results.

Furthermore, Rout et al. [20] used numerous criteria based on text similarity and sentiment polarity to distinguish between bogus and legitimate reviews. The researchers employed an emotion score as a characteristic in the study, which is based on the polarity of sentiments between positive and negative ratings, as well as linguistics and unigrams. The

researchers then used three algorithms, including the Support Vector Machine, the Naive Bayes method, and the Decision Tree.

Semi-Supervised Learning

To detect false reviews, Fusilier et al. [21] were the first to introduce the Positive Unlabelled (PU) learning technique. The PU-learning technique is a combination of some positive labels with unlabelled datasets, and it is described following. It is a semi-supervised approach that only employs two classifiers, one labelled as deceptive and one classified as unlabelled, and does not use a negative as a true training example. In this approach, the first unlabelled data points are treated as belonging to the negative class. Using the positive cases from the previous stage, classifiers are trained in the following step. Then classification was performed only to unlabelled instances, and labelled instances are generated as a result of the application of classifiers. Following the classification of instances into positive and negative, the positive examples that were identified as dishonest reviews are removed from the unlabelled circumstances and the remaining instances are classified as negative examples. When dealing with negative instances, classifiers are applied once more. This process is repeated until the stop criterion, which distinguishes between bogus and legitimate reviews, have been reached. Support Vector Machine and Naive Bayes are the two classifiers used in this PU-learning.

Fusilier et al. [22] conducted a comparison of the traditional PU-learning technique and the improved PU-learning technique. The researchers examined if it is capable to find a smaller number of occurrences from an unlabelled set by modifying the PU-learning approach. At the end of the process, only new negative instances that have been formed by the output of the preceding iteration are evaluated, and the classifier is only deployed to the new negative examples in that iteration. So, with each repetition, negative occurrences are decreased, and final cases are accurately identified as either phoney or authentic reviews, depending on the algorithm. The researchers of the research have discovered a way to detect both positive and negative fraudulent reviews. Their methods included Naive Bayes and a Support Vector Machine classifier that utilised both unigrams and bigrams features, and the reviews were divided into two categories: fraudulent and non-fraudulent.

Unsupervised Learning techniques

The biggest benefit of using an unsupervised learning strategy is that it allows researchers to distinguish between fraudulent and legitimate reviews without the need for a labelled dataset.

Rout et al. [20] implemented unsupervised learning technique. Based on the differences in the behavioural patterns of reviews, the researchers employed a variety of features depending on review data, reviewer data, and product information. In the research, the researchers utilized the Amazon *Cell Phones and Electronics products* reviews dataset to distinguish between fraudulent and authentic reviews.

Mukherjee et at. [23] provides an unsupervised approach for detecting opinion spam. The researchers employ a completely Bayesian method and treat sentiment spam detection as a clustering task. The Bayesian setup enables the researchers to represent spamicity of reviewers as associated with other known behavioural traits in their Author Spamicity Model (ASM). Inference in ASM leads in discovering the distributions of two groups which are spammers and non-spammers described as a set of behavioural variables. The researchers applied a range of features based on author features and review features.

2.3 Analysis of Existing Fake Review Detection Using Machine Learning

According to previous investigations, fraudulent reviews can be identified using a variety of methods such as classification, clustering, or a combination of the two. It is possible to accurately detect spam opinion using a variety of strategies that are dependent on features and classifiers. Tables below show several approaches which are applied to distinguish between fraudulent and legitimate reviews

2.3.1 Supervised Machine Learning

Paper Title / Author	Technique Used	Classifier	Dataset	Results
The Impact of applying Different Preprocessing Steps on Review Spam Detection [19] **Authors:** Wael Etaiwi, Ghazi Naymat	Supervised **Features used:** Linguistic feature (Stemming + remove Punctuation marks + Remove Stop words)	- Naïve Bayes (NB) - Support Vector Machine (SVM) - Decision Tree (DT) - Random Forest (RF) - Gradient-Boosted Trees (GBTs)	1600 reviews on TripAdvisor website	(Average result with all proposed preprocessing steps) **- NB:** Precision 51.8% Recall 86.8% Accuracy 85.5% **- SVM:** Precision 51.8% Recall 86.8% Accuracy 85.5% **- DT:** Precision 47.1% Recall 70.7% Accuracy 69.5% **- RF:** Precision 58.9% Recall 60.3% Accuracy 59.8% **- GBTs:** Precision 49.4% Recall 70.2% Accuracy 68.5%

Table 2.3.1 Analysis of Fake Review Detection Using Supervised Machine Learning Experimented by Previous Researchers

2.3.2 Semi-Supervised Machine Learning

Paper Title / Author	Technique Used	Classifier	Dataset	Results
Detecting positive and negative deceptive opinions using PU-learning [22] **Authors:** Donato Hernandez Fusilier, Rafael Guzman Cabrera, Manuel Montes-y-Gomez, Paolo Rosso	Semi-Supervised **Features used:** - Modified PU-learning	- Naïve Bayes (NB) - Support Vector Machine (SVM)	Ott's hotel reviews dataset	(Results for positive opinions) **- Deceptive:** Precision 85.2% Recall 72.8% F-measure 78.0% **- Truthful:** Precision 76.8% Recall 86.8% F-measure 81.1% (Results for negative opinions) **- Deceptive:** Precision 78.8% Recall 59.5% F-measure 65.7% **- Truthful:** Precision 67.2% Recall 80.3% F-measure 72.3%

Table 2.3.2 Analysis of Fake Review Detection Using Semi-Supervised Machine Learning Experimented by Previous Researchers

2.3.3 Unsupervised Machine Learning

Paper Title / Author	Technique Used	Classifier	Dataset	Results
Spotting opinion spammers using behavioral footprints [23] **Authors:** Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, Riddhiman Ghosh	Unsupervised **Features used:** Author Features + Review Fetures	Author Spamicity Model (ASM)	Amazon review dataset	(Results for spam review with ASM) **-Uninformed Priors (ASM-UP):** Precision 77.7% Recall 74.0% Accuracy 75.5% **- Informed Priors (ASM-IP):** Precision 77.9% Recall 74.8% Accuracy 75.7% **-Hyperparameter Estimation (ASM-HE):** Precision 79.6% Recall 75.1% Accuracy 77.4%

Table 2.3.3 Analysis of Fake Review Detection Using Unsupervised Machine Learning Experimented by Previous Researchers

Chapter 3

Proposed Framework

3.1 Proposed Framework

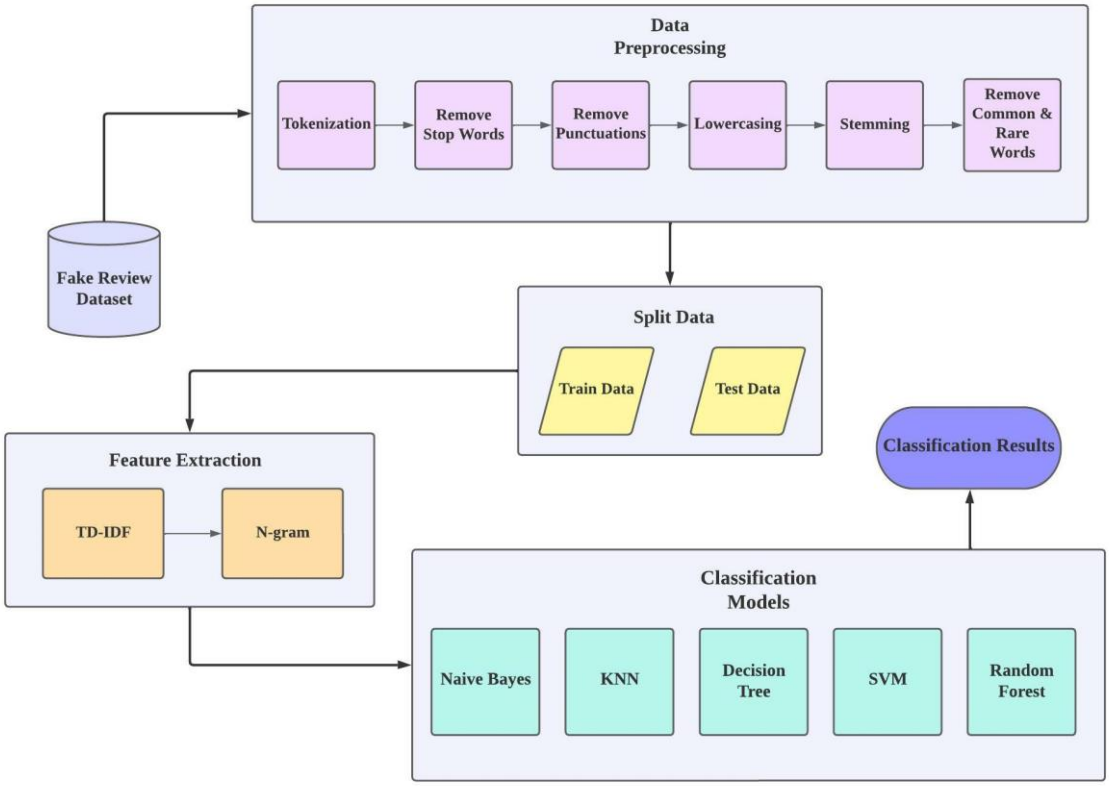


Figure 3.1 Proposed Framework for Fake Reviews Detection

The proposed framework shown in *Figure 3.1* consists of four phases to get the best classification model that will be used for fake review detection.

I. Data Pre-processing

One of the most significant phases of a machine learning technique is data pre-processing. Data pre-processing is necessary since the world's data is never suitable for use. In this study, a series of pre-processing techniques were utilized to get the dataset's raw data eligible for analysis. The following provides an explanation of the pre-processing methods utilized in the suggested framework:

a) **Tokenization:** One of the most popular methods for NLP is tokenization. Before using any other pre-processing methods, it is a fundamental step. Tokens are the individual words that make up the text. Tokenization, for instance, will separate the sentence "I love the look and feel of this pillow" into the tokens "I", "love", "the", "look", "and", "feel", "of", "this", "pillow".

b) **Removing Stop Words:** The most often used words are stop words [24], but they have no actual meaning. Typical instances of stop words are (an, a, the, this). Before moving further with the fake reviews detection approach in this study, all data are cleaned of stop words.

c) **Removing Punctuations:** Text is divided into sentences, paragraphs, and phrases using punctuation. Since punctuation marks are used often in text, it has an impact on the outcomes of any text processing approach, especially those that depend on the occurrence frequencies of words and phrases.

d) **Lowercasing:** The only pre-processing technique that significantly outperformed the baseline result was the transformation of uppercase letters into lowercase letters. Words like "Book" and "book" have the same meaning, but the models treat them differently when they are not written in lower case.

e) **Stemming:** There are numerous variations of a single phrase in the English language. When creating NLP or machine learning models, these variations in a source text led to redundant data. These models might not work well. It is required to standardize text by avoiding duplication and stemming words to their base form in order to construct a strong model.

f) **Removing Common & Rare Words:** Since the dataset's common words have high counts, most scoring systems are rewarded for identifying those words' counts more than they do for identifying the counts of other words. This makes every other word appear less frequent. Rare words are removed for an entirely different reason. Due to the uncommon, the noise overrides any associations between them and other words.

II. Split Data

A method for assessing a machine learning algorithm's effectiveness is the train-test split. It can be applied to issues involving classification or regression as well as any supervised learning algorithm.

The process includes splitting the dataset into two subsets. The train dataset is the first subset, which is used to fit the model. Instead of using the second subset to train the model, the input element of the dataset is given to it, and predictions are then made and compared to the expected values. The test dataset is the second dataset in discussion.

- **Train Dataset:** Used to fit the machine learning model.
- **Test Dataset:** Used to examine how well a machine learning model fits the data.

The purpose is to determine how well the machine learning model performs on new data which the data not used to train the model. We anticipate applying the model in this way. Specifically, to fit it to data that is already accessible and has known inputs and outputs, then to make forecasts about future cases where we won't have the target values or expected outputs. When a workable size dataset is provided, the train-test procedure is appropriate.

III. Feature Extraction

The purpose of the feature extraction is to improve the performance of either a pattern recognition system or a machine learning system. In order to provide machine learning and deep learning models with more useful data, feature extraction involves reducing the input to its key features. The essential step is to remove any unnecessary features from the data, which may actually decrease the model's accuracy [25].

a) N-Grams:

A contiguous series of n items from a given sample of text or speech makes up an n-gram. Different NLP algorithms frequently use n-grams to forecast the next potential word in a sequence.

An n-gram language model makes the assumption that a word depends only on the (n-1) words that came before it. The main objective is to compile the frequency of the n-grams in our corpus and use it to forecast the following word. A unigram language model is one in which the previous word is used to predict the following word. A bigram language model which implied in the proposed framework is one in which the previous two words are used to predict the following word.

b) TF-IDF:

The frequency of both true and false (TF) as well as the inverse document (IDF) are obtained by another textual feature method called TF-IDF. Each phrase has a unique TF and IDF score, and the sum of these two scores is referred to as the term's TF-IDF weight [26]. The reviews are categorized using a confusion matrix into the following four outcomes:

- **True Positive (TP):** Predicted real reviews are defined as real reviews.
- **True Negative (TN):** Predicted fake reviews are defined as fake reviews.
- **False Positive (FP):** Predicted real reviews are defined as fake reviews.
- **False Negative (FN):** Predicted fake reviews are defined as real reviews.

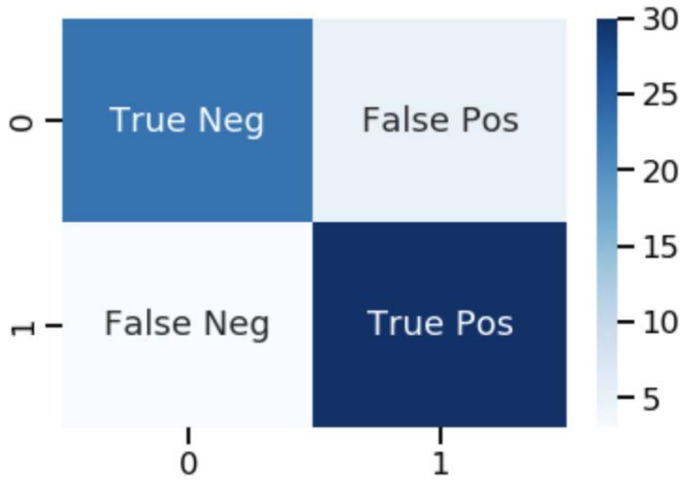


Figure 3.2 Confusion Matrix in Machine Learning Algorithm

IV. Classification Models

a. Naïve Bayes (NB):

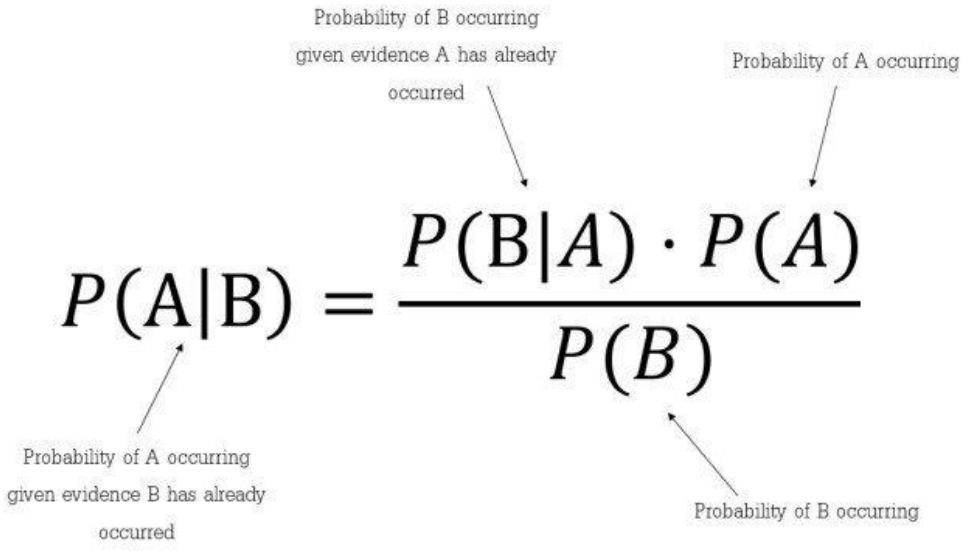


Figure 3.3 Formula of Bayes Theorem

The core concept of NB is based on the Bayes theorem, which stated in the *Figure 3.2*. By counting the frequency and total values in a dataset, NB determines a set of probabilities. Numerous application fields, including text classification, spam filtering, and recommendation systems, have effectively used NB.

b. K-Nearest Neighbours (KNN):

One of the most basic yet effective classification methods is KNN. Statistical estimation and pattern recognition have seen the largest use of KNN [27]. KNN's primary purpose is to categorize instance queries based on the votes of a collection of similarly classed cases. Typically, the distance function is used to calculate similarity [28].

c. Decision Tree:

Another machine learning classifier that focuses on creating a tree to represent a judgment of training data is called Decision-Tree [29]. Based on the optimal feature split, the algorithm begins to iteratively build the tree. A predetermined function, such as entropy, information gain, gain ratio, or Gini index, is used to select the best features.

d. Support Vector Machines (SVM):

By identifying the best separable hyper-plane that classifies the provided training data, SVM is a discriminating classifier that, in essence, divides the given data into classes [31].

e. Random Forest:

Successful solutions to the overfitting issues that arise in the decision tree include Random Forest [30]. Making a bag of trees from various dataset samples is the fundamental principle of random forest. When building each tree in the forest, Random Forest selects a tiny random number of features rather than building the tree from all features.

Chapter 4

Experiment

4.1 Experimental Results

An Amazon Review dataset (2018) which is publicly released has been utilized to evaluate the proposed framework. This dataset contained 4,055 reviews of Home and Kitchen products. 2,028 of the reviews are categorized as "REAL" and 2,027 as "FAKE," respectively. The pie chart of the target labelled reviews count is displayed in *Figure 4.1*. Since the number of samples is balanced, the classification can be fair without considering the factors class imbalance while choosing an algorithm or adjusting the data.

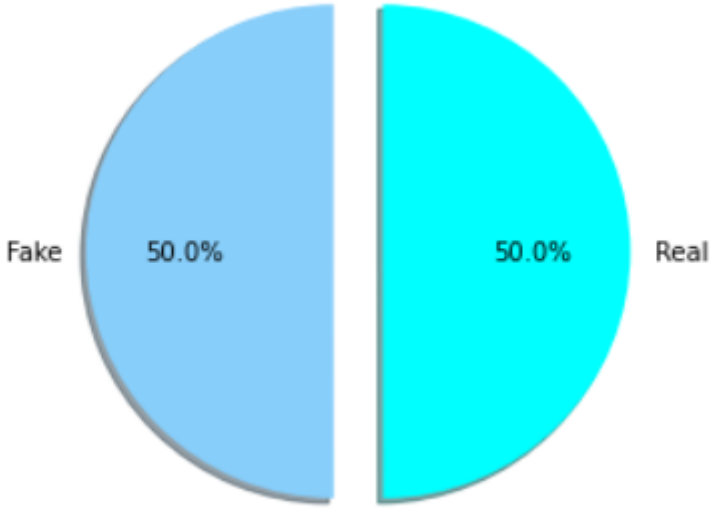


Figure 4.1 Target Labelled Reviews Count – Pie Chart

The confusion matrix from the testing with Naïve Bayes (NB) classifier is given in *Figure 4.2*. The total 1,339 test examples have been classified into 660 TN, 31 FP, 335 FN and 313 TP. The accuracy of fake reviews detection that we obtained with NB classifier is 72.66% with 0.79 average precision, 0.72 average recall, and 0.71 average F1-score. The classification report for this NB classifier is given in *Table 4.1*.

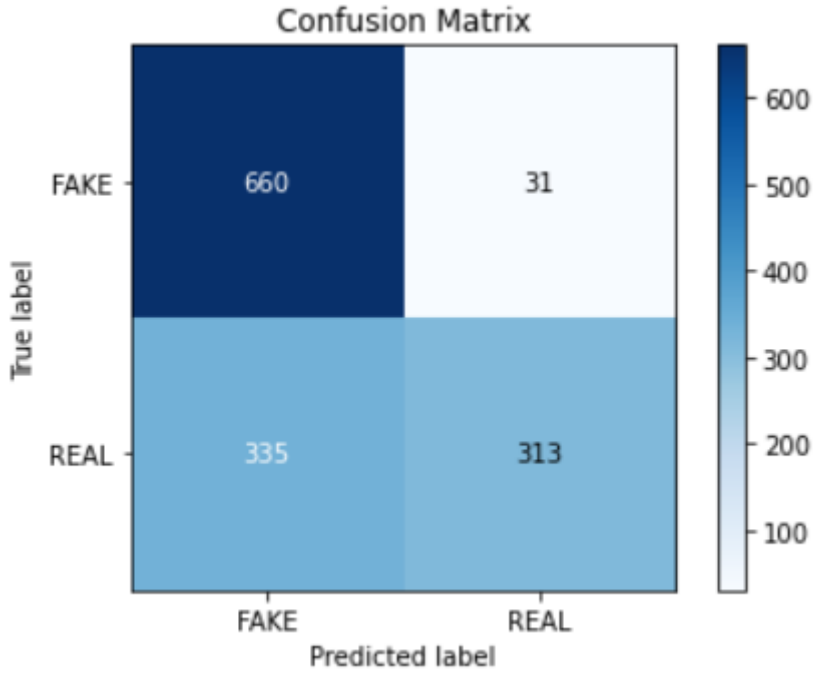


Figure 4.2 Confusion Matrix for Naïve Bayes (NB) Classifier

	precision	recall	f1-score	support
FAKE	0.66	0.96	0.78	691
REAL	0.91	0.48	0.63	648
accuracy			0.73	1339
macro avg	0.79	0.72	0.71	1339
weighted avg	0.78	0.73	0.71	1339

Table 4.1 Classification Report for Naïve Bayes (NB) Classifier

Figure 4.3 shows the confusion matrix from the testing with the K-Nearest Neighbours (KNN) algorithm. The 1,339 test examples in total have been broken down into 676 TN, 15 FP, 609 FN, and 39 TP categories. Using the KNN algorithm, we were able to detect fraudulent reviews with a 53.39% accuracy rate, 0.62 average precision, 0.52 average recall, and 0.40 average F1-score. The classification report for this KNN algorithm is given in *Table 4.2*.

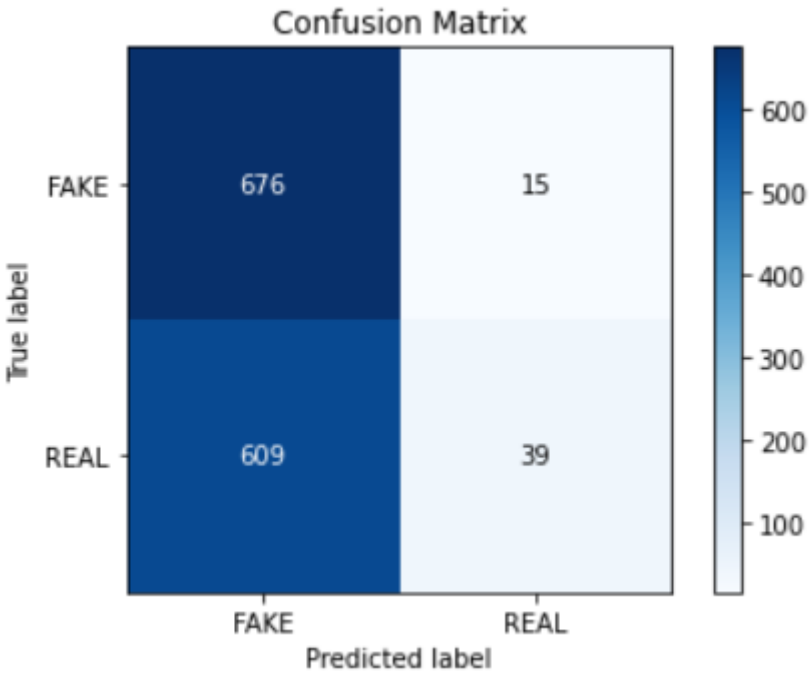


Figure 4.3 Confusion Matrix for K-Nearest Neighbours (KNN) Algorithm

	precision	recall	f1-score	support
FAKE	0.53	0.98	0.68	691
REAL	0.72	0.06	0.11	648
accuracy			0.53	1339
macro avg	0.62	0.52	0.40	1339
weighted avg	0.62	0.53	0.41	1339

Table 4.2 Classification Report for K-Nearest Neighbours (KNN) Algorithm

In *Figure 4.4*, the confusion matrix from the testing using Decision Tree learning is presented. There were 1,339 test samples in all, and they were divided into 517 TN, 174 FP, 94 FN, and 554 TP. With 0.80 average precision, 0.80 average recall, and 0.80 average F1-score, we were able to detect fraudulent reviews with a 79.98% accuracy using Decision Tree learning. In *Table 4.3*, the classification report for this Decision Tree learning is provided.

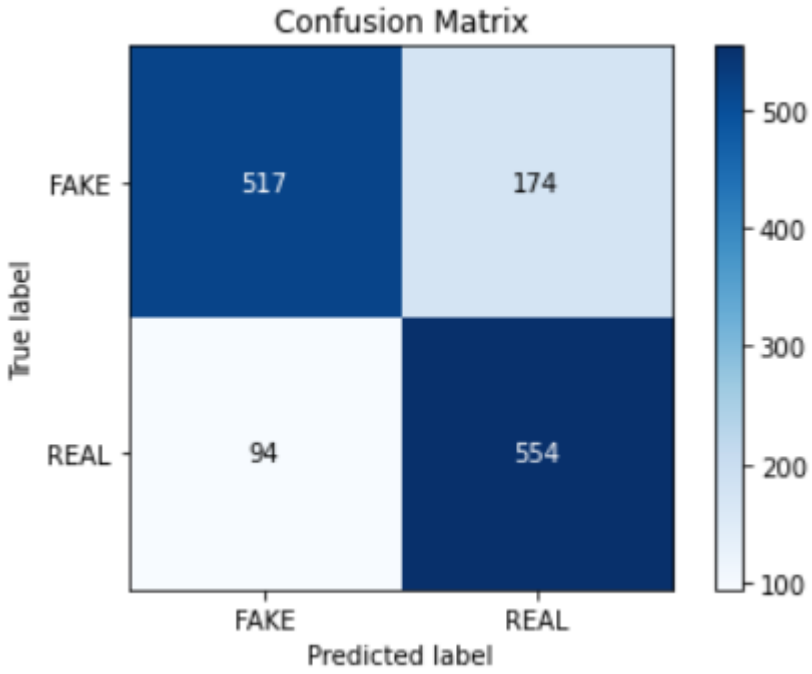


Figure 4.4 Confusion Matrix Decision Tree Learning

	precision	recall	f1-score	support
FAKE	0.85	0.75	0.79	691
REAL	0.76	0.85	0.81	648
accuracy			0.80	1339
macro avg	0.80	0.80	0.80	1339
weighted avg	0.80	0.80	0.80	1339

Table 4.3 Classification Report for Decision Tree Learning

Figure 4.5 shows the confusion matrix from the testing using the Support Vector Machine (SVM). A total of 1,339 test cases were categorized as 488 TN, 203 FP, 62 FN, and 586 TP. We reached 80.20% accuracy in detecting false reviews using SVM, with 0.81 average precision, 0.81 average recall, and 0.80 average F1-score. In *Table 4.4*, the classification report for this SVM method is provided.

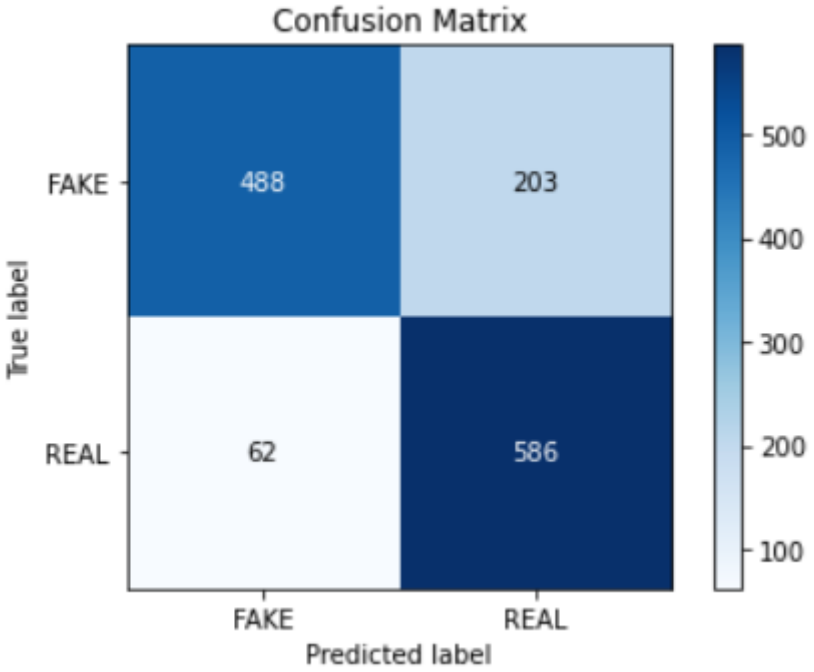


Figure 4.5 Confusion Matrix Support Vector Machine (SVM)

	precision	recall	f1-score	support
FAKE	0.89	0.71	0.79	691
REAL	0.74	0.90	0.82	648
accuracy			0.80	1339
macro avg	0.81	0.81	0.80	1339
weighted avg	0.82	0.80	0.80	1339

Table 4.4 Classification Report for Support Vector Machine (SVM)

Last but not least, *Figure 4.6* shows the confusion matrix from the testing with Random Forest. A total of 1,339 test samples have been categorized as 500 TN, 191 FP, 48 FN, and 600 TP. The accuracy of fake review detection using Random Forest is 82.15%, with average precision, recall, and F1-score values of 0.84, 0.82, and 0.82, respectively. *Table 4.5* contains the classification report for this Random Forest.

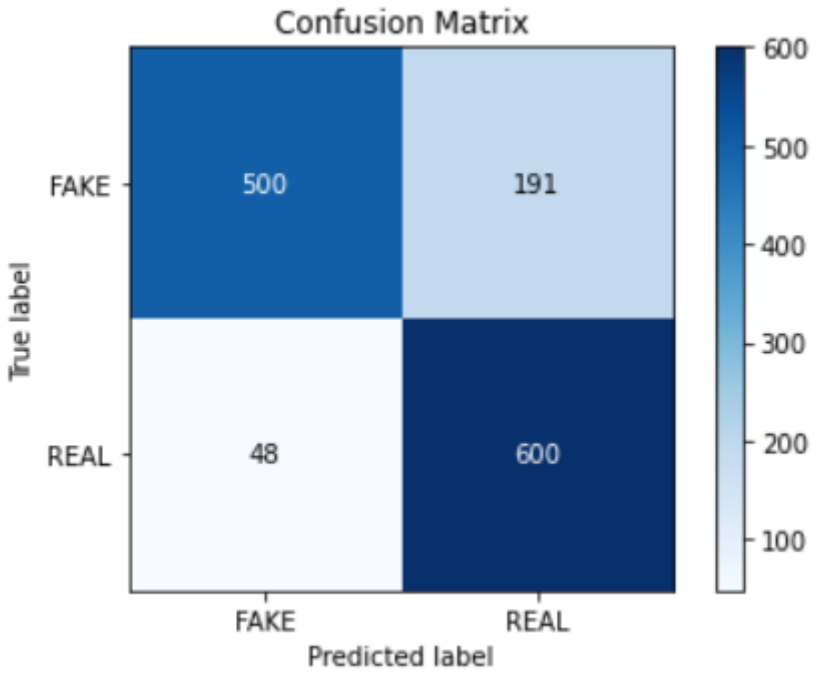


Figure 4.6 Confusion Matrix Random Forest

	precision	recall	f1-score	support
FAKE	0.91	0.72	0.81	691
REAL	0.76	0.93	0.83	648
accuracy			0.82	1339
macro avg	0.84	0.82	0.82	1339
weighted avg	0.84	0.82	0.82	1339

Table 4.5 Classification Report for Random Forest

Chapter 5
Conclusion

5.1 Conclusion

In this paper, we have done an experiment to determine the effectiveness of the proposed framework for the fake reviews detection using Machine Learning technique.

Figure 5.1 shows the comparison of the accuracy of all classifiers. The experiment results show that Random Forest algorithm got the highest accuracy which is 82.15% with the implementation of the proposed data processing method.

	Algorithms	Percentage
0	Naive Bayes	72.666169
1	K-Nearest Neighbors algorithm	53.398058
2	Decision Tree learning	79.985063
3	Support Vector Machine (SVM)	80.209111
4	Random Forest	82.150859

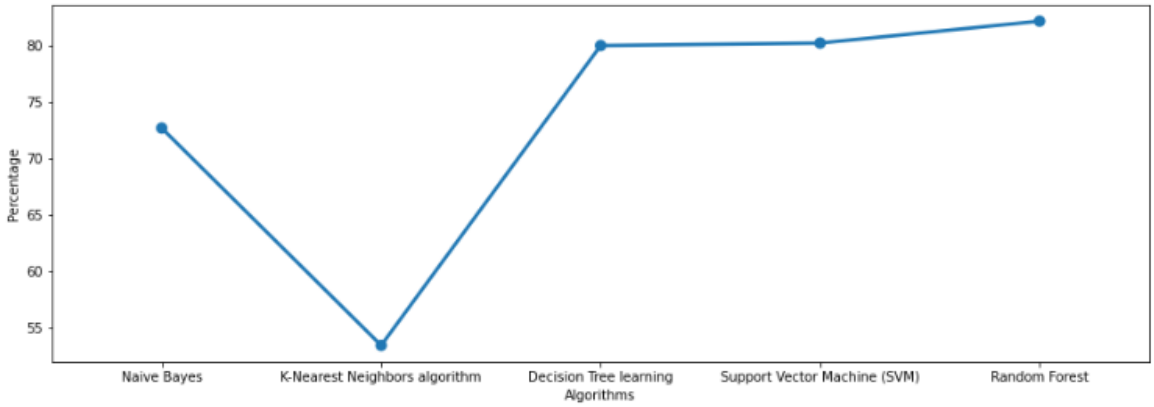


Figure 5.1 Comparison of Accuracy Percentage of All the Classifier Algorithms

In terms of positive observations, precision is the probability of accurately anticipated observations to all predicted positive observations. How many reviews that are categorized as true are actually real according to this metric? High precision and low false positive rate are related. By using the Random Forest technique, we were able to get an average precision of 0.84, which is rather good.

Moreover, recall can also be referred to as sensitivity or true positive rate. The ideal recall for a good classifier is 1 (high). This is a rather effective classifier for identifying fraudulent reviews because we were able to achieve an average recall of 0.82 by using the Random Forest algorithm, which is close to 1. Besides, the F1-score is a metric that considers both recall and precision. Only when recall and precision are both high can F1-score increase. It is more useful to use the F1-score, which is the harmonic mean of recall and precision. Using the Random Forest algorithm, we were able to obtain an average F1-score of 0.82, which indicates that the proposed model with this classifier algorithm is more accurate than other models in the experiment. *Table 5.1* shows the summary of precision, recall, F1-score for all the examined classifier algorithms.

	NB			KNN			DT			SVM			RF		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
FAKE	0.66	0.96	0.78	0.53	0.98	0.68	0.85	0.75	0.79	0.89	0.71	0.79	0.91	0.72	0.81
REAL	0.91	0.48	0.63	0.72	0.06	0.11	0.76	0.85	0.81	0.74	0.90	0.82	0.76	0.93	0.83
AVG	0.79	0.72	0.71	0.62	0.52	0.40	0.80	0.80	0.80	0.81	0.81	0.80	0.84	0.82	0.82

Table 5.1 Summary of Precision, Recall, F1-score for All the Classifier Algorithms

Chapter 6

Recommendation and Future Work

6.1 Recommendation and Future Work

Reviews are very important for people's decision-making. Therefore, detecting false reviews is a continuing and active study field. This paper presents a method for detecting fraudulent reviews using machine learning. Additional features targeted toward reviewers could be implemented to it. Additionally, feature sets can be used to evaluate content and rating activity. Furthermore, the reviews we used for our investigation were in English. Hence, other languages are potential of being used in fake reviews detection system.

REFERENCES

[1] D. Martens and W. Maalej, "Towards understanding and detecting fake reviews in App Stores," *Empirical Software Engineering*, vol. 24, no. 6, pp. 3316–3355, 2019.

[2] L. Li, B. Qin, W. Ren, and T. Liu, "Document representation and feature combination for deceptive spam review detection," *Neurocomputing*, vol. 254, pp. 33–41, 2017.

[3] S. Saumya, J. P. Singh, and Y. K. Dwivedi, "Predicting the helpfulness score of online reviews using Convolutional Neural Network," *Soft Computing*, vol. 24, no. 15, pp. 10989–11005, 2019.

[4] J. Salminen, C. Kandpal, A. M. Kamel, S.-gyo Jung, and B. J. Jansen, "Creating and detecting fake reviews of online products," *Journal of Retailing and Consumer Services*, vol. 64, p. 102771, 2022.

[5] A. Costa, J. Guerreiro, S. Moro, and R. Henriques, "Unfolding the characteristics of incentivized online reviews," *Journal of Retailing and Consumer Services*, vol. 47, pp. 272–281, 2019.

[6] R. Filieri, "What makes an online consumer review trustworthy?," *Annals of Tourism Research*, vol. 58, pp. 46–64, 2016.

[7] V. Sandulescu and M. Ester, "Detecting singleton review spammers using semantic similarity," *Proceedings of the 24th International Conference on World Wide Web*, 2015.

[8] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding Deceptive Opinion Spam by Any Stretch of the Imagination," pp. 309–319, 2011.

[9] D. Plotkina, A. Munzel, and J. Pallud, "Illusions of truth—experimental insights into human and algorithmic detections of fake online reviews," *Journal of Business Research*, vol. 109, pp. 511–523, 2020.

[10] A. Morales, H. Sun, and X. Yan, "Synthetic review spamming and Defense," *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*, 2013.

[11] S. P. Algur, A. P. Patil, P. S. Hiremath, and S. Shivashankar, "Conceptual level similarity measure based review spam detection," *2010 International Conference on Signal and Image Processing*, 2010.

[12] L. Bing, *Web Data Mining*. Berlin Heidelberg, New York: Springer, 2008.

[13] R. V. Bandakkanavar, M. Ramesh and H. Geeta, A survey on detection of reviews using sentiment classification of methods, IJRITCC, vol. 2, no. 2, pp. 310-314, 2014.

[14] Abbasi, Zhang, Zimbra, Chen, and Nunamaker, "Detecting fake websites: The contribution of Statistical Learning theory," *MIS Quarterly*, vol. 34, no. 3, p. 435, 2010.

[15] N. Jindal and B. Liu, "Opinion spam and analysis," *Proceedings of the international conference on Web search and web data mining - WSDM '08*, 2008.

[16] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *Journal of Big Data*, vol. 2, no. 1, 2015.

[17] A. Munzel, "Assisting consumers in detecting fake reviews: The role of Identity Information Disclosure and consensus," *Journal of Retailing and Consumer Services*, vol. 32, pp. 96–108, 2016.

[18] C. G. Harris, "Comparing human computation, machine, and hybrid methods for Detecting Hotel Review spam," *Lecture Notes in Computer Science*, pp. 75–86, 2019.

[19] W. Etaiwi and G. Naymat, "The impact of applying different preprocessing steps on review spam detection," *Procedia Computer Science*, vol. 113, pp. 273–279, 2017.

[20] J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabeled data," *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 3187–3211, 2016.

[21] D. H. Fusilier, R. G. Cabrera, M. M. Gomez, and P. Rosso, "Using PU-Learning to Detect Deceptive Opinion Spam," 2013.

[22] D. H. Fusilier, M. M. Gómez, P. Rosso, and R. G. Cabrera, "Detecting positive and negative deceptive opinions using pu-learning," *Information Processing & Management*, vol. 51, no. 4, pp. 433–443, 2015.

[23] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 2013

[24] C. Silva and B. Ribeiro, "The importance of stop word removal on recall values in text categorization," in Neural Networks, 2003. Proceedings of the International Joint Conference on, vol. 3. IEEE, 2003, pp. 1661–1666.

[25] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 4, pp. 388–400, 1993.

[26] J. Ramos et al., “Using tf-idf to determine word relevance in document queries,” in Proceedings of the first instructional conference on machine learning, vol. 242, 2003, pp. 133–142.

[27] M.-L. Zhang and Z.-H. Zhou, “Ml-knn: A lazy learning approach to multi-label learning,” Pattern recognition, vol. 40, no. 7, pp. 2038–2048, 2007.

[28] N. Suguna and K. Thanushkodi, “An improved k-nearest neighbor classification using genetic algorithm,” International Journal of Computer Science Issues, vol. 7, no. 2, pp. 18–21, 2010.

[29] M. A. Friedl and C. E. Brodley, “Decision tree classification of land cover from remotely sensed data,” Remote sensing of environment, vol. 61, no. 3, pp. 399–409, 1997.

[30] A. Liaw, M. Wiener et al., “Classification and regression by randomforest,” R news, vol. 2, no. 3, pp. 18–22, 2002.

[31] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features.” 1998.

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3T3	**Study week no.: 2**
Student Name & ID: Liong Yong Xuan 17ACB05552	
Supervisor: Dr Ramesh Kumar Ayyasamy	
Project Title: Detection and Analysis of Fake Reviews on Online Service Portal	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Plan on the topic that proposed for the Final Year Project 1

2. WORK TO BE DONE

Catch up on the rest task of Final Year Project 2.

3. PROBLEMS ENCOUNTERED

Distribution of the section in the report

4. SELF EVALUATION OF THE PROGRESS

My progression is going smooth in this current situation

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3T3	**Study week no.: 4**
Student Name & ID: Liong Yong Xuan 17ACB05552	
Supervisor: Dr Ramesh Kumar Ayyasamy	
Project Title: Detection and Analysis of Fake Reviews on Online Service Portal	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Completed report writing for Chapter 1 and 2.

2. WORK TO BE DONE

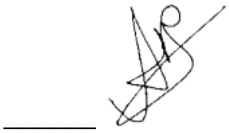
Design a fake review detection system framework and find a dataset to examine the proposed framework.

3. PROBLEMS ENCOUNTERED

Have to find a dataset, which all the attributes fulfil my requirement

4. SELF EVALUATION OF THE PROGRESS

All the progression is still on the track.



Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3T3	**Study week no.: 6**
Student Name & ID: Liong Yong Xuan 17ACB05552	
Supervisor: Dr Ramesh Kumar Ayyasamy	
Project Title: Detection and Analysis of Fake Reviews on Online Service Portal	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

System design diagram for the proposed framework has been done.

2. WORK TO BE DONE

Develop the system with Jupyter Notebook.

3. PROBLEMS ENCOUNTERED

Due to first time developing a Machine Learning system, some of the codes are not functioning well as expectation.

4. SELF EVALUATION OF THE PROGRESS

All the progression is still on the track, self-assigned tasks are completed.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3T3	**Study week no.: 8**
Student Name & ID: Liong Yong Xuan 17ACB05552	
Supervisor: Dr Ramesh Kumar Ayyasamy	
Project Title: Detection and Analysis of Fake Reviews on Online Service Portal	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Able to do the Data Pre-processing for the targeted data frame, now proceeding to the detection part.

2. WORK TO BE DONE

Complete the system development and obtain the experiment results.

3. PROBLEMS ENCOUNTERED

Due to first time developing a Machine Learning system, some of the codes are not functioning well as expectation.

4. SELF EVALUATION OF THE PROGRESS

Self-assigned tasks are completed, have to speed up to complete the development of the model.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3T3	**Study week no.: 10**
Student Name & ID: Liong Yong Xuan 17ACB05552	
Supervisor: Dr Ramesh Kumar Ayyasamy	
Project Title: Detection and Analysis of Fake Reviews on Online Service Portal	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Complete developed the model and the experiment results obtained.

2. WORK TO BE DONE

Interpret the experiment results in the report and determine which classification method is the most effective.

3. PROBLEMS ENCOUNTERED

I need to understand and learn the relationship between precision, recall, F1-socre and accuracy rate, to determine which classification algorithm is the best in the proposed framework.

4. SELF EVALUATION OF THE PROGRESS

Report writing of Chapter 1-3 are done. Due to some personal issues, the progression is a bit delayed.

Supervisor's signature

Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

(Project II)

Trimester, Year: Y3T3	**Study week no.: 12**
Student Name & ID: Liong Yong Xuan 17ACB05552	
Supervisor: Dr Ramesh Kumar Ayyasamy	
Project Title: Detection and Analysis of Fake Reviews on Online Service Portal	

1. WORK DONE

[Please write the details of the work done in the last fortnight.]

Report writing of Chapter 1-4 are done. Proceeding to conclusion and recommendation.

2. WORK TO BE DONE

Based on the experiment results, determine which classification algorithm is the best in the proposed framework. Provide some recommendation for the future work.

3. PROBLEMS ENCOUNTERED

The proposed framework maybe is not so effective because of the lacking knowledge of the new learning programming language.

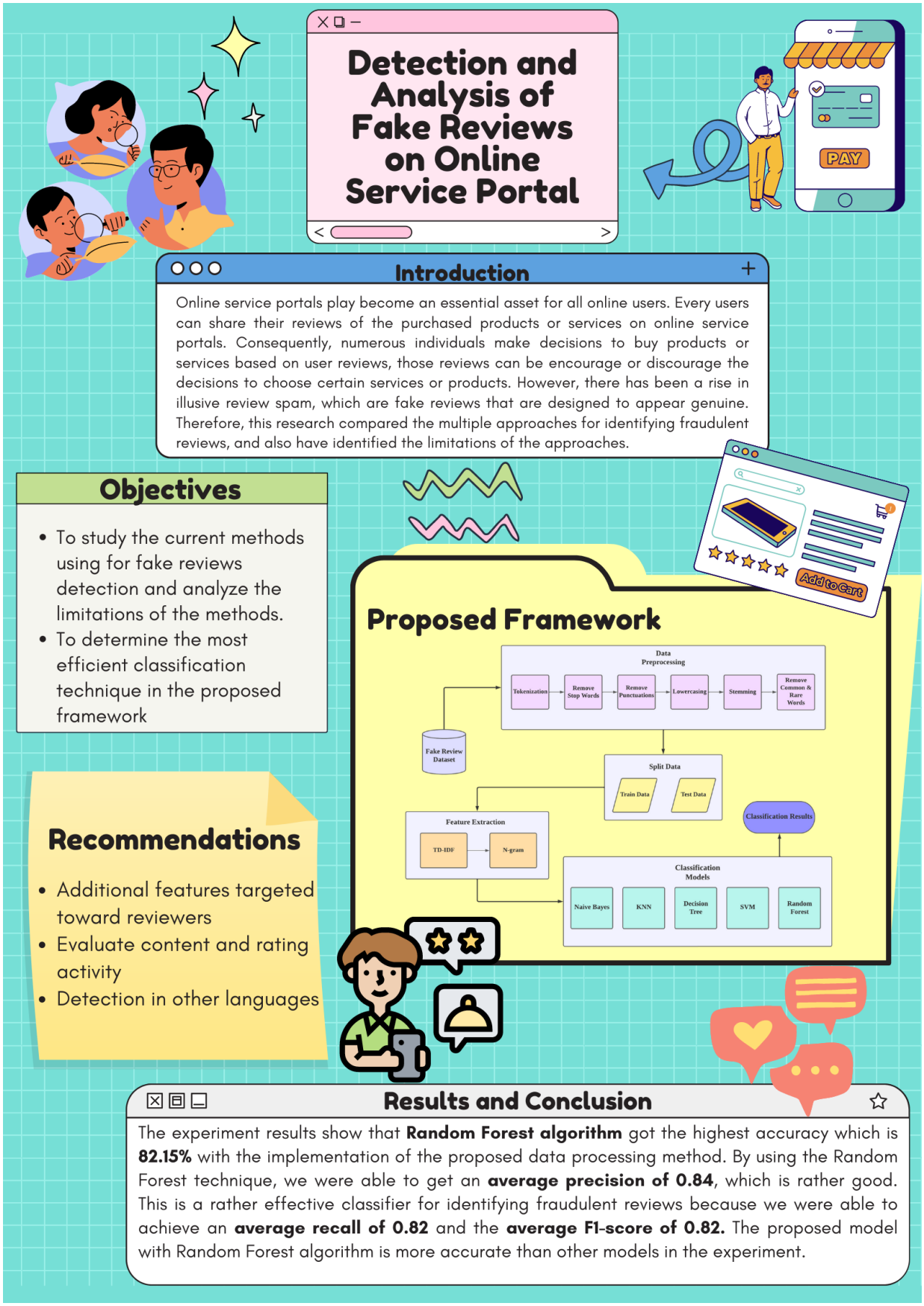
4. SELF EVALUATION OF THE PROGRESS

The report is almost done. Soon can be finalize and submit by the deadline.

Supervisor's signature

Student's signature

POSTER


Detection and Analysis of Fake Reviews on Online Service Portal
PAY
Introduction
Online service portals play become an essential asset for all online users. Every users can share their reviews of the purchased products or services on online service portals. Consequently, numerous individuals make decisions to buy products or services based on user reviews, those reviews can be encourage or discourage the decisions to choose certain services or products. However, there has been a rise in illusive review spam, which are fake reviews that are designed to appear genuine. Therefore, this research compared the multiple approaches for identifying fraudulent reviews, and also have identified the limitations of the approaches.
Objectives
To study the current methods using for fake reviews detection and analyze the limitations of the methods.
To determine the most efficient classification technique in the proposed framework
Add to Cart
Proposed Framework
Data Preprocessing
Tokenization
Remove Stop Words
Remove Punctuations
Lowercasing
Stemming
Remove Common & Rare Words
Fake Review Dataset
Split Data
Train Data
Test Data
Feature Extraction
TD-IDF
N-gram
Classification Results
Classification Models
Naive Bayes
KNN
Decision Tree
SVM
Random Forest
Recommendations
Additional features targeted toward reviewers
Evaluate content and rating activity
Detection in other languages
Results and Conclusion
The experiment results show that Random Forest algorithm got the highest accuracy which is 82.15% with the implementation of the proposed data processing method. By using the Random Forest technique, we were able to get an average precision of 0.84, which is rather good. This is a rather effective classifier for identifying fraudulent reviews because we were able to achieve an average recall of 0.82 and the average F1-score of 0.82. The proposed model with Random Forest algorithm is more accurate than other models in the experiment.

PLAGIARISM CHECK RESULT

Turnitin Originality Report

Processed on: 08-Sep-2022 14:19 +08
ID: 1893866383
Word Count: 5477
Submitted: 4

fyp2 By Liong Yong Xuan

Similarity Index	Similarity by Source	
12%	Internet Sources:	8%
	Publications:	4%
	Student Papers:	7%

include quoted include bibliography excluding matches < 8 words mode:
quickview (classic) report Change mode print download

1% match (student papers from 28-Aug-2022)
Submitted to National College of Ireland on 2022-08-28

1% match (student papers from 01-Sep-2022)
Submitted to Southampton Solent University on 2022-09-01

1% match (Internet from 29-Aug-2022)
https://www2.cs.uic.edu/~liub/publications/KDD-2013-Arjun-spam.pdf

1% match (student papers from 20-Aug-2022)
Submitted to Athlone Institute of Technology on 2022-08-20

1% match (publications)
Wael Etaiwi, Ghazi Naymat. "The Impact of applying Different Preprocessing Steps on Review Spam Detection", Procedia Computer Science, 2017

1% match (student papers from 24-Aug-2022)
Submitted to University of Essex on 2022-08-24

1% match (Internet from 30-May-2022) https://mdpi-res.com/d_attachment/entropy/entropy-24-00705/article_deploy/entropy-24-00705.epub
1% match (Internet from 11-Jun-2022) https://academic-accelerator.com/Manuscript-Generator/Naive-Bayes/support-vector-machine
<1% match (student papers from 14-Aug-2022) Submitted to National College of Ireland on 2022-08-14
<1% match (Internet from 02-Aug-2022) http://www.scienceschola r.us
<1% match (Internet from 29-Jun-2021) https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/
<1% match () Ong, Chien Young. "Order and supporting system for Lido Enterprise", 2012
<1% match (Internet from 06-Aug-2022) http://dspace.daffodilvarsity.edu.bd:8080
<1% match (student papers from 10-Nov-2021) Submitted to UC, Irvine on 2021-11-10
<1% match () Andre Lamurias, Sofia Jesus, Vanessa Neveu, Reza M. Salek, Francisco M. Couto. "Information Retrieval Using Machine Learning for Biomarker Curation in the Exposome-Explorer", Frontiers in Research Metrics and Analytics
<1% match (publications) Rami Mohawesh, Shuxiang Xu, Son N. Tran, Robert Ollington, Matthew Springer, Yaser Jararweh, Sumbal Maqsood. "Fake Reviews Detection: A survey", IEEE Access, 2021
<1% match (student papers from 27-Jan-2022) Submitted to The British College on 2022-01-27

<1% match (publications) Neeraj Kumar Singh*, Arun Kumar Sunaniya. "High Density Video Impulse Noise Reduction Scheme Based on Spatially Growing Modified Median Filter", International Journal of Innovative Technology and Exploring Engineering, 2020
<1% match (Internet from 08-Nov-2019) https://journals.sagepub.com/doi/10.1177/0142331215587568?icid=int.sj-full-text.similar-articles.3
<1% match (Internet from 21-Dec-2021) http://www2.cs.uh.edu
<1% match (Internet from 17-Jul-2020) https://textbooks.elsevier.com/manualsprotectedtextbooks/9780123814791/Instructor%
<1% match (Internet from 17-Feb-2020) https://worldscientific.com/doi/10.1142/S1469026819500056
<1% match (Internet from 14-Nov-2020) https://www.mdpi.com/2304-8158/9/11/1622
<1% match (publications) Rani .T.P, Suganthi .K, Magilan Saravanan, Ashish Kumar Sahu, K. Martin Sagayam, Ahmed A. Elngar. "Predicting Online Fraudulent Transactions Using Machine Learning", Research Square Platform LLC, 2022
<1% match (Internet from 14-Jan-2022) http://eprints.utar.edu.my
<1% match (publications) Farid Ayeche, Adel Alti. "Facial Expressions Recognition Based on Delaunay Triangulation of Landmark and Machine Learning", Traitement du Signal, 2021
<1% match (student papers from 05-Jun-2022) Submitted to Liverpool John Moores University on 2022-06-05

<1% match (student papers from 17-May-2019) Submitted to University of Central England in Birmingham on 2019-05-17
<1% match (Internet from 13-Jul-2016) http://dblp.dagstuhl.de
<1% match (Internet from 13-Jul-2020) https://docplayer.net/1133870-Adaptive-real-time-anomaly-based-intrusion-detection-using-data-mining-and-machine-learning-techniques.html
<1% match (Internet from 22-May-2018) http://shura.shu.ac.uk
<1% match (Internet from 28-Jan-2020) http://www.ijettjournal.org
<1% match (publications) "Advances in Natural Language Processing", Springer Science and Business Media LLC, 2008
<1% match (publications) Jacob W. Kamminga, Duc V. Le, Jan Pieter Meijers, Helena Bisby, Nirvana Meratnia, Paul J.M. Havinga. "Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags", Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018
<1% match (publications) Welly Naptali, Masatoshi Tsuchiya, Seiichi Nakagawa. "Topic-Dependent Language Model with Voting on Noun History", ACM Transactions on Asian Language Information Processing, 2010
<1% match (Internet from 23-Aug-2022) https://www.cancerimagingarchive.net/Publications.html

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	Liong Yong Xuan
ID Number(s)	17ACB05552
Programme / Course	BACHELOR OF INFORMATION SYSTEMS (HONOURS) BUSINESS INFORMATION SYSTEMS
Title of Final Year Project	Detection and Analysis of Fake Reviews on Online Service Portal

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 12 % **Similarity by source** Internet Sources: 8 % Publications: 4 % Student Papers: 7 %	
Number of individual sources listed of more than 3% similarity: 0	

Parameters of originality required and limits approved by UTAR are as Follows:

(i) Overall similarity index is 20% and below, and

(ii) Matching of individual sources listed must be less than 3% each, and

(iii) Matching texts in continuous block must not exceed 8 words

Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor

Name: Dr.Ramesh Kumar Ayyasamy

Date: 09-Sep-2022

Signature of Co-Supervisor

Name: _____

Date: _____

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

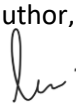
CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	17ACB05552
Student Name	Liong Yong Xuan
Supervisor Name	Dr Ramesh Kumar Ayyasamy

TICK (√)	DOCUMENT ITEMS Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
	Front Plastic Cover (for hardcopy)
√	Title Page
√	Signed Report Status Declaration Form
√	Signed FYP Thesis Submission Form
√	Signed form of the Declaration of Originality
√	Acknowledgement
√	Abstract
√	Table of Contents
√	List of Figures (if applicable)
√	List of Tables (if applicable)
	List of Symbols (if applicable)
√	List of Abbreviations (if applicable)
√	Chapters / Content
√	Bibliography (or References)
√	All references in bibliography are cited in the thesis, especially in the chapter of literature review
	Appendices (if applicable)
√	Weekly Log
√	Poster
√	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)
√	I agree 5 marks will be deducted due to incorrect format, declare wrongly the ticked of these items, and/or any dispute happening for these items in this report.

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report.



(Signature of Student)

Date: 8th September 2022