Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Babylon
Collage of Information Technology
Software Department

Prevention and Detection Attack of Social Media Networks Using Machine Learning Methods

A Thesis
Submitted to the Council of the College of Information Technology, University of Babylon in Partial Fulfillment of the Requirements for the Degree of Master in Information Technology/Software

By

Karrar Sadiq Mohsen Jawad

Supervised by:

Prof. Dr. Wesam Sameer Abd-Ali Bhaya

٢٠٢٢ A.D. **١٤٤٤ A.H.**

بسم الله الرحمن الرحيم

﴿ عَلَّمَ ٱلْإِنسَٰنَ مَا لَمْ يَعْلَمْ ﴾

صدق الله العظيم

[سورة العلق: الآية ٥]

Dedication

To God Almighty, my creator; My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life; My homeland Iraq, The great martyrs; My family and many friends A special feeling of gratitude to my loving parents whose words of encouragement and push for tenacity ring in my ears; My sisters and brothers have never left my side and are very special; I also dedicate this dissertation to my many friends who have supported me throughout the work of the thesis.

Karar

Acknowledgments

In the Name of Allah, the Most Merciful, the Most Compassionate All praise be to Allah, the Lord of the worlds; and prayers and peace be upon Mohamed His servant and messenger.

First and foremost, I must acknowledge my limitless thanks to Allah, the Ever-Magnificent; the Ever-Thankful, for His help and blessing. I am sure that this work would have never become true, without His guidance.

I would first like to thank my supervisor, Dr. Wesam Sameer Bhaya. The door to the office was always open whenever I ran into a trouble spot or had a question about my research or writing. They consistently allowed this paper to be my work but steered me in the right direction whenever they thought I needed it.

I would like to take this opportunity to say warm thanks to all my beloved friends, who have been so supportive along the way of doing my thesis. I also would like to express my wholehearted thanks to my family for the generous support they provided me and for their unconditional love and prayers, I have the chance to complete this thesis.

Karar

Supervisor Certification

I certify that this thesis was prepared under my supervision at the Department of Software / College of Information Technology / Babylon University, by KARRAR SADIQ MOHSEN as partial fulfillment of the requirements for the degree of Master in Information Technology.

Signature:

Supervisor Name: **Prof. Dr. Wesam Sameer Bhaya**

Date: / / ٢٠٢٢

The Head of the Department Certification

In view of the available recommendations, I forward this thesis for debate by the examination committee.

Signature:

Name: Ass. Prof. Dr. Ahmed Saleem Abbass

Head of Software Department

Date: / / ٢٠٢٢

Certification of the Examination Committee

We, the undersigned, certify that (**Karrar Sadiq Mohsen**) candidate for the degree of Master in Information Technology - Software, has presented his thesis of the following title (**Detection of Social Networks Attacks using Machine Learning Methods**) as it appears on the title page and front cover of the thesis that the said thesis is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: October ١٥ , ٢٠٢٢.

Signature:
Name: **Prof. Dr. Abdul Monem S.Rahma**
Title:
Date: / / ٢٠٢٢
(**Chairman**)

Signature:
Name: **Ass. Prof. Dr. Nashwan J.Hussein**
Title:
Date: / / ٢٠٢٢
(**Member**)

Signature:
Name: **Dr. Bayadir Abbas Hussein**
Title:
Date: / / ٢٠٢٢
(**Member**)

Signature:
Name: **Prof. Dr. Wesam Sameer Bhaya**
Title:
Date: / / ٢٠٢٢
(**Member**)

Signature:
Name: **Prof. Dr. Hussein Atiya Lafta**
Title: **Professor**
Date: / / ٢٠٢٠
(**Dean of College of Information Technology**)

Abstract

Today, Online Social Networks (OSNs) is one of the important microblogging and data exchange. It has millions of users worldwide, and its users communicate together via messages and posts. Because of OSNs open architecture and behavior, it is susceptible to attacks from bogus accounts and an enormous number of automatic programs, or “bots”. Bots are viewed as harmful since they utilize the internet to send spam to social network members. Users of social networks place a high value on data security and privacy since these demands must be met if the network is to retain user interest and, ultimately, legitimacy. To address these difficulties, an effective approach for detecting and classifying bogus Twitter accounts is required.

In this thesis, a variety of feature selection techniques as well as a variety of machine learning-based methodologies are employed to identify fake accounts, Fraudulent social engineering messages that can deceive users and Text posts with sensitive information. Eight feature selection strategies were combined with five machine learning methods to preprocess the dataset and identify bogus accounts, Fraudulent messages and sensitive information. The feature selection process (which uses multiple methods to rank features) includes eight feature selection processes (Chi-Square Test Feature Selection, ANOVA Feature Selection, Mutual Information Feature Selection, Logistic Regression Feature Selection, Additional Tree Feature Selection, Embossed Feature Selection, mrMR features Selection and, Light GBM Trait Selection), all outputs are evaluated by finding similar features in the process of choosing between the outputs of the eight algorithms to find the best features to be used by the classifier for best performance. A classification method consisting of five machine learning algorithms: Support Vector Machine (SVM), Gaussian Naïve Bayes (GNB), Stochastic Gradient Descent (SGD), Random Forest (RF), and

Decision Tree (DT) was used and tested, the best of which proved to be the best in the identification process we mentioned.

Five classifiers were used for the purpose of classifying Twitter accounts, Fraudulent Social Engineering Messages and Sensitive Texts. A wide range of preprocessing operations was applied for the data cleaning and facilitating data handling, in addition to experimenting with many feature extraction techniques and then choosing the best features between their outputs. The proposed system has been tested using three datasets (Fake and Normal dataset, Social Engineering Attack using text messages and Sensitive Texts), these data were released by the Institute of Informatics and Telematics of the Italian National Research Council (IIT-CNR) for twitter to the purpose of research, and the part of sensitive text has been modified to be more suitable for this system. The results obtained indicate that the classifier Decision Tree was the best for the social Engineering texts dataset with an accuracy ٩٧%, while the Random Forst was the best classifier with an accuracy rate equal to ٩٩.٩% for the Fake and Normal dataset. As for the Sensitive Texts dataset, the classifier Decision Tree has the highest and ideal classification accuracy, which is ١٠٠٪.

Table of Contents

Subjects	**Page No.**
Abstract	I
Table of Contents	III
List of Abbreviations	V
List of Figures	VIII
List of Tables	X
List of Algorithms	XI
Chapter One: General Introduction	**Page No.**
١,١. Introduction	١
١,٢. Social Networking	٢
١,٣. Machine Learning	٣
١,٤. Literature Survey	٤
١,٥. Aim of Thesis	٥
١,٦. Thesis Layout	٦
١,٧. Objectives of Thesis	٨
١,٨. Layout of Thesis	٩
Chapter Two: Theoretical Background	**Page No.**
٢,١. Introduction	١٠
٢,٢. Online Social Networks (OSNs)	١٠
٢,٢.١. Data on Online Social Networks	١١
٢,٢.٢. Applications of Social Networks	١٤
٢,٢.٣. Various threats on online social network	١٦
٢,٣. Social Networks Data Processing of machine learning	١٦
٢,٣,١. Preprocessing data	١٧
٢,٣,١,١. Texts preprocessing	١٧

٢,٣,٢. Features Extraction	١٨
٢,٣,٣. Data Normalization	١٨
٢,٣,٤. Data splitting	١٨
٢,٣,٤,١. Cross-Validation	١٩
٢,٤. Feature Selection Methods	٢١
٢,٤,١. Chi-square	٢١
٢,٤,٢. ANOVA	٢٢
٢,٤,٣. Mutual information	٢٢
٢,٤,٤. Logistic Regression Model	٢٣
٢,٤,٥. Tree-based feature selection	٢٤
٢,٤,٦. Relief	٢٥
٢,٤,٧. The Minimum Redundancy Maximum Relevance	٢٦
٢.٤.٨. Light GBM	٢٦
٢,٥. Machine Learning (ML) Techniques	٢٧
٢,٥,١. Attacks Prediction in Online Social Networks Based on ML Techniques	٢٩
٢,٥,١,١. Support Vector Machine (SVM) Algorithm	٣١
٢,٥,١,٢. Decision Tree (DT) Algorithm	٣٢
٢,٥,١,٣. Random Forest (RF) Algorithm	٣٣
٢,٥,١,٤. Stochastic Gradient Descent (SGD) Algorithm	٣٥
٢,٥,١,٥. Gaussian Naïve Bayes (GNB) Algorithm	٣٥
٢,٦. Online Social Network Security Threats	٣٩
٢,٦,١. Classic Threat	٣٩
٢,٦,٢. Advanced Threats	٤٠
٢,٦,٣. Modern Threats	٤١

٢,٧. Online Social Network Security Solutions	٤١
٢,٧,١. Authentication Mechanisms	٤١
٢,٧,٢. Security and Privacy Settings	٤٢
٢,٧,٣. Internal Protection Mechanisms	٤٢
٢,٧,٤. Commercial Solutions	٤٣
٢,٨. Evaluation Metrics	٤٣
Chapter Three: The Proposed System	**Page No.**
٣,١. Introduction	٤٦
٣,٢. The Proposed System Architecture	٤٦
٣,٢,١. Selecting Training Mode (First Stage)	٥٠
٣,٢,٢. Data Preprocessing (Second Stage)	٥١
٣,٢,٣. Text Feature Extraction (Third Stage)	٥٦
٣,٢,٤. Normalization Data (Fourth Stage)	٥٧
٣,٢,٥. Data Splitting (Fifth Stage)	٥٨
٣,٢,٦. Feature Selection (Sixth Stage)	٥٩
٣,٢,٧. Classification Stage (Seventh Stage)	٦٠
Chapter Four: Experimental Results and Discussion	**Page No.**
٤,١. Introduction	٦٥
٤,٢. Software and Hardware Requirements	٦٥
٤,٢,١. Software Requirements	٦٥
٤,٢,٢. Hardware Requirements	٦٥
٤,٣. Datasets Description	٦٦
٤,٣,١. Social Engineering detection	٦٦
٤,٣,٢. The Fake Project dataset	٦٦

٤,٣,٣. Sensitive or Not Sensitive Text	٦٧
٤,٤. Implementation of the Proposed System	٦٩
٤,٤,١. Classification Results on Social Engineering Detection	٦٩
٤,٤,٢. Classification Results on Fake Account	٧٥
٤,٤,٣. Classification Results on Sensitive	٨١
٤,٥. Discussion of Experimental Results	٨٧
٤,٥,١. Classifiers performance with “dataset ١”	٨٧
٤,٥,٢. Classifiers performance with “dataset ٢”	٨٧
٤,٥,٣. Classifiers performance with “dataset ٣”	٨٨
٤,٦. Comparison with Previous Studies	٨٨
Chapter Five: Conclusions and Recommendations	**Page No.**
٥,١. Conclusions	٩٦
٥,٢. Recommendations and suggestions for Future Works	٩٦

List of Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
ANN	Artificial Neural Network
Acc.	Accuracy
DT	Decision Tree
FF	Feature Frequency
FP	Positive Predictive Value/Precision
FN	Negative Predictive Value
GNB	Gaussian Naïve Bayes
k-NN	The k-Nearest Neighbors
ML	Machine Learning
MLP	Multilayer Perceptron
NB	Naïve Bayes
OSN	Online Social Network
RF	Random Forest
SVM	Support Vector Machine
SNS	Social Networking Services
SGD	Stochastic Gradient Descent
SWRL	The Semantic Web Rule Language
TP	True Positive Rate/Sensitivity
TN	True Negative Rate/Specificity
WOL	Web Ontology Language

List of Figures

Figure No.	Figure Title	Page No.
(٢-١)	Constituents of online social networks	١٢
(٢-٢)	K-fold cross-validation method	٢٣
(٢-٣)	Hold-out cross-validation	٢٣
(٢-٤)	Types of machine learning	٢٩
(٢-٥)	Process of supervised learning	٣٠
(٢-٦)	SVM classifier	٣٢
(٢-٧)	DT Classifier	٣٣
(٢-٨)	RF algorithm	٣٤
(٢-٩)	Confusion matrix for the classification problem	٤٣
(٣-١)	The proposed system architecture	٤٨
(٤-١)	Confusion matrix of SVM in training stage “dataset ١”	٦٩
(٤-٢)	ROC curve of SVM in training stage “dataset ١”	٦٩
(٤-٣)	Confusion matrix of SGD in training stage “dataset ١”	٧٠
(٤-٤)	ROC curve of SGD in training phase “dataset ١”	٧٠
(٤-٥)	Confusion matrix of GNB in training phase “dataset ١”	٧٠
(٤-٦)	ROC curve of GNB in training phase “dataset ١”	٧٠
(٤-٧)	Confusion matrix of RF in training phase “dataset ١”	٧١
(٤-٨)	ROC curve of RF in training phase “dataset ١”	٧١
(٤-٩)	Confusion matrix of DT in training phase “dataset ١”	٧١
(٤-١٠)	ROC curve of DT in training phase “dataset ١”	٧١
(٤-١١)	Confusion matrix of SVM in testing phase “dataset ١”	٧٢
(٤-١٢)	ROC curve of SVM in testing phase “dataset ١”	٧٢

(٤-١٣)	Confusion matrix of SGD in testing phase “dataset ١”	٧٣
(٤-١٤)	ROC curve of SGD in testing phase “dataset ١”	٧٣
(٤-١٥)	Confusion matrix of GNB in testing phase “dataset ١”	٧٣
(٤-١٦)	ROC curve of GNB in testing phase “dataset ١”	٧٣
(٤-١٧)	Confusion matrix of RF in testing phase “dataset ١”	٧٤
(٤-١٨)	ROC curve of RF in testing phase “dataset ١”	٧٤
(٤-١٩)	Confusion matrix of DT in testing phase “dataset ١”	٧٤
(٤-٢٠)	ROC curve of DT in testing phase “dataset ١”	٧٤
(٤-٢١)	Confusion matrix of SVM in training stage “dataset ٢”	٧٥
(٤-٢٢)	ROC curve of SVM in training stage “dataset ٢”	٧٥
(٤-٢٣)	Confusion matrix of SGD in training stage “dataset ٢”	٧٦
(٤-٢٤)	ROC curve of SGD in training phase “dataset ٢”	٧٦
(٤-٢٥)	Confusion matrix of GNB in training phase “dataset ٢”	٧٦
(٤-٢٦)	ROC curve of GNB in training phase “dataset٢”	٧٦
(٤-٢٧)	Confusion matrix of RF in training phase “dataset ٢”	٧٧
(٤-٢٨)	ROC curve of RF in training phase “dataset ٢”	٧٧
(٤-٢٩)	Confusion matrix of DT in training phase “dataset ٢”	٧٧
(٤-٣٠)	ROC curve of DT in training phase “dataset ٢”	٧٧
(٤-٣١)	Confusion matrix of SVM in testing stage “dataset ٢”	٧٨
(٤-٣٢)	ROC curve of SVM in testing stage “dataset ٢”	٧٨
(٤-٣٣)	Confusion matrix of SGD in testing stage “dataset ٢”	٧٩
(٤-٣٤)	ROC curve of SGD in testing stage “dataset ٢”	٧٩
(٤-٣٥)	Confusion matrix of GNB in testing stage “dataset ٢”	٨٠
(٤-٣٦)	ROC curve of GNB in testing stage “dataset ٢”	٨٠
(٤-٣٧)	Confusion matrix of RF in testing stage “dataset ٢”	٨١
(٤-٣٨)	ROC curve of RF in testing stage “dataset ٢”	٨١

(٤-٣٩)	Confusion matrix of DT in testing stage “dataset ٢”	٨١
(٤-٤٠)	ROC curve of DT in testing stage “dataset ٢”	٨١
(٤-٤١)	Confusion matrix of SVM in training stage “dataset ٣”	٨١
(٤-٤٢)	ROC curve of SVM in training stage “dataset ٣”	٨١
(٤-٤٣)	Confusion matrix of SGD in training stage “dataset ٣”	٨٢
(٤-٤٤)	ROC curve of SGD in training stage “dataset ٣”	٨٢
(٤-٤٥)	Confusion matrix of GNB in training stage “dataset ٣”	٨٢
(٤-٤٦)	ROC curve of GNB in training stage “dataset ٣”	٨٢
(٤-٤٧)	Confusion matrix of RF in training stage “dataset ٣”	٨٣
(٤-٤٨)	ROC curve of RF in training stage “dataset ٣”	٨٣
(٤-٤٩)	Confusion matrix of DT in training stage “dataset ٣”	٨٣
(٤-٥٠)	ROC curve of DT in training stage “dataset ٣”	٨٣
(٤-٥١)	Confusion matrix of SVM in Testing stage “dataset ٣”	٨٤
(٤-٥٢)	ROC curve of SVM in testing stage “dataset ٣”	٨٤
(٤-٥٣)	Confusion matrix of SGD in training stage “dataset ٣”	٨٤
(٤-٥٤)	ROC curve of SGD in training stage “dataset ٣”	٨٤
(٤-٥٥)	Confusion matrix of GNB in training stage “dataset ٣”	٨٥
(٤-٥٦)	ROC curve of GNB in training stage “dataset ٣”	٨٥
(٤-٥٧)	Confusion matrix of RF in training stage “dataset ٣”	٨٦
(٤-٥٨)	ROC curve of RF in training stage “dataset ٣”	٨٦
(٤-٥٩)	Confusion matrix of DT in training stage “dataset ٣”	٨٦
(٤-٦٠)	ROC curve of DT in training stage “dataset ٣”	٨٦

List of Tables

Table No.	Table Title	Page No.
(١-١)	Comparison of related studies	٨
(٢-١)	Hold-out vs K-fold cross validation	٢٤
(٢-٢)	Advantages and limitations of different supervised machine learning algorithms	٣٧
(٤-١)	Social engineering sentences dataset sample	٦٦
(٤-٢)	The dataset description. It contains ٤١٥٥ accounts, ٢٨٪ fake accounts, and ٣٠٪ trust accounts.	٦٦
(٤-٣)	Sensitive dataset sample	٦٨
(٤-٤)	Results of training phase on “dataset ١”	٧٢
(٤-٥)	Results of testing phase on “dataset ١”	٧٥
(٤-٦)	Results of training phase on “dataset ٢”	٧٨
(٤-٧)	Results of testing phase on “dataset ٢”	٨١
(٤-٨)	Results of training phase on “dataset ٣”	٨٤
(٤-٩)	Results of testing phase on “dataset ٣”	٨٧
(٤-١٠)	Accuracy Comparison with related studies	٨٩

List of Algorithms

Algorithm No.	Algorithm Title	Page No.
(٣-١)	The proposed system architecture	٤٩
(٣-٢)	Text Cleaning	٥٣
(٣-٣)	Calculating of TF-IDF	٥٦
(٣-٤)	Min-Max Normalization	٥٧
(٣-٥)	Text Feature Selection	٥٨
(٣-٦)	Dataset Splitting	٥٩
(٣-٧)	Support Vector Machine (SVM) Classifier	٦٠
(٣-٨)	Stochastic Gradient Descent Classifier (SGD)	٦١
(٣-٩)	Gaussian Naïve Bayes (GNB) Classifier	٦٢
(٣-١٠)	Random Forest (RF) Classifier	٦٣
(٣-١١)	Decision Tree Classifier (DT)	٦٣

Chapter One
Introduction

Chapter One

Introduction

١,١. Overview

Online Social Networks (OSNs), such as Facebook, Twitter, and LinkedIn, have grown in popularity in recent years. OSNs are used to keep in contact, exchange information, arrange activities, and even run an e-business. Social media has grown at an exponential rate during the last two decades. Various forms of social networking have attracted many users, various events have been organized, and social networking has a tone of false profiles and fabricated news. Furthermore, the bogus accounts use their accounts for a variety of purposes, including spreading falsehoods that have an impact on a certain economy or even culture. The detection of deception news is a continuing issue [١].

Twitter, Facebook, and Instagram are significant forms of online communication that likely contains a wealth of information, opening up new avenues for text content analysis. In actuality, many people claim that the biggest obstacle to employing technology is either a “lack of IT infrastructure” or an overarching cost-benefit analysis. Despite these challenges, the technology looks to be gaining acceptance. Antifraud technology solutions have been used by more than half of the insurers studied in the previous five years, including several in the last two years [٢].

To prevent Social Network attacks and protect user privacy, Online Social Networks offers a variety of ways to report spam. A web user can detect spam on their webpage by clicking on the link. The spam profiles will be deactivated when the Social Network Application evaluates the network user reports. The Social network is attempts to quickly identify bogus posts and suspicious complaints. When blocking harmful posts and questionable profiles, for example Twitter blocks several legitimate login credentials but user cannot detect sensitive texts or social engineering text attack and detect fake accounts that follows him. As a result, there is a need to find some quick approaches to detect sensitive texts, texts attacks and discover fake accounts. In the meanwhile, these contemporary tactics have little effect on genuine user tweets [٣].

Machine Learning (ML) has gotten a lot of interest in numerous implementations and research fields, especially in cyber security. ML techniques can be used to analyze and identify threats from the vast dataset available, with hardware and computing resources becoming more accessible. Unsupervised and supervised learning are two types of machine learning algorithms and methodologies. Machine Learning is one of the most effective approaches for detecting these harmful behaviors, thus this proposed system focuses on improving the performance of classifier on detecting the three mentioned goals [٤].

This work started at ٢٠٢٠ to make a new proposed system for detections of fake and real accounts, social engineering texts attack and sensitive text and continued till the end of year ٢٠٢١.

١,٢. Social Networking

The Internet has swiftly evolved from a platform for exchanging information to one for sharing goods, ideas, and information. Social networking is a global phenomenon that has changed the way people connect with one another. It has an impact on practically every element of our lives: Education, communication, work, politics, healthcare, social relationships, and personal productivity are all important factors [٥].

A social networking service is an Internet-based platform for forming and maintaining social connections. It allows users to communicate online with people who share similar interests, whether for romantic or social reasons. Users can send and receive emails, instant messages, online comments, digital photographs, and videos, as well as write blog posts. It also allows persons with impairments to express their views and opinions in a virtual setting [٦].

Social media platforms play a dual role as content producers and consumers. They allow the user to control who can see their profile. Answers to questions such as age, location, and interests are used to create a profile. Users can upload photos, add multimedia content, change the style and feel of their profiles, write blogs, comment on postings, and construct and share a contact list on some sites. Users can choose who can read their profile, contact them, add them to their contact list, and so on, to preserve their privacy [٧].

١,٣. Machine Learning

Machine learning and artificial intelligence have become so pervasive in our daily lives that it is no more a place where esoteric academics and scientists go to address a difficult research topic. Evolution is more natural than unintentional. Organizations have discovered that they may harness a massive volume of data in creating solutions with far-reaching economic value thanks to exponential development in processing speed and the emergence of smarter algorithms for tackling complicated and hard problems [٨].

In recent years, artificial intelligence (AI), particularly machine learning (ML), has expanded fast in the context of data analysis and computing, allowing applications to perform intelligently. ML gives systems the ability to learn and improve based on their experiences [٩]. In general, the effectiveness and efficiency of a machine learning solution are determined by the nature and qualities of the data as well as the learning algorithms' performance. As a result, a wide range of machine learning methods (e.g., supervised, unsupervised, semi-supervised, and reinforced) has been created to handle a wide range of data in various ML issues [١٠].

١,٤. Related Works

In this section, we will go through some of the studies on the current machine learning models for detecting attackers in social media. A social network is a site where users can use visual computer technology to share their activities, interests, backgrounds, and real-life connections. The most important areas that have been worked on in these studies are (identifying Twitter bot accounts, investigate the nature and characteristics of spam accounts in twitter, phishing prevention, spam tweets detection, and classify the URL as valid or illegitimate, categorizing normal and abnormal users in social networks, detecting terrorist attacks, identify unsuitable content and faked images) and they are arranged in chronological order, from oldest to newest as follows:

- **A. Al-Zoubi et al., (٢٠١٩) [١١]** investigate the nature and characteristics of spam accounts in a social network like Twitter to improve spam identification based on several publicly available language-independent features. Four datasets are retrieved for four different linguistic settings, and a fifth is generated by integrating them together to assess the usefulness of these features in spam detection, these datasets are collected from twitter profiles use REST API. The k-Nearest Neighbors (k-NN), Random Forest (RF), Naïve Bayes (NB), Decision Tree (DT) (J٤٨), and Multilayer Perceptron (MLP) classifiers, as well as five filter-based feature selection approaches, were used in the experiments. The results reveal that when employing the KNN model with an Arabic-language dataset, the greatest accuracy was ٩٧,٩٪.

- **Muhammet S. Başarslana, Fatih Kayaalp, (٢٠٢٠) [١٢]** this study aims to investigate the effect of types of text representation on the performance of sentiment analysis. Two datasets were used in this study. The first one is the user reviews about movies from the IMDB, which has been labeled by Kotzias, and the second one is the Twitter tweets, including the tweets of users about health topic in English in ٢٠١٩, collected using the Twitter API. They use classification models (Naïve Bayes (NB), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) algorithms) for categorizing the sentiments as positive, negative and neutral. They use Term Frequency-Inverse Document Frequency (TF-IDF) and Word٢Vec (W٢V) modeling techniques as feature extraction methods from the dataset According to the experimental results, Artificial Neural Network has the best accuracy performance in both datasets compared to the others with ٩٠% of accuracy for IMDB dataset and ٨٧٪ for twitter dataset while SVM has ٨٤٪ of accuracy for both dataset and NB has accuracy of ٨٣٪ for IMDB dataset and ٧٤٪ for twitter dataset.

- **M. Jabardi and A. Hadi, (٢٠٢٠) [١٣]** presents a novel method that performs two tasks at once: it identifies and classifies Twitter bots using rules from Semantic Web Rule Language (SWRL) and ontological engineering. This study uses Fake Project dataset released by the Institute of Informatics and Telematics of the Italian National Research Council (IIT-CNR). The Web Ontology Language (OWL), Semantic Web Rule Language (SWRL) rules, and reasoners were employed to classify fake accounts as spam bots or fake followers. These techniques are used to inductively learn the criteria that separate a bot from a genuine account. After accurately detecting the false accounts with (٩٧٪) accuracy in the first step, their algorithm correctly classified the fake accounts into spam or fake follower bots with an accuracy score of (٩٤,٩٪). Additionally, it has been shown that the ontology classifier offers straightforward and intelligible decision criteria, making it a more interpretable model than other machine learning classifiers.

- **A. Sarker et al., (٢٠٢٠) [١٤]** develops a new methodology for analyzing Twitter data and detecting terrorist attacks. This study collect dataset from twitter using Twitter ٤j which is a Java library for the Twitter API. The Aho-Corasick algorithm is used to do pattern matching and assign the weight in this model, which uses a ternary search to identify the weights of predefined keywords. The weights are divided into three categories: terror attack, severe terror attack, and normal data, and are utilized as classification attributes. Two machine learning techniques, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are used to predict whether a terror incident occurred. The results of the KNN algorithm give an accuracy of ٩٧,١٪ while the SVM algorithm gives an accuracy of ٩٧,٧٪.
- **A Ramalingaiah , S Hussaini , S Chaudhari, (٢٠٢١) [١٥]** Detection of malicious bots using supervised Machine learning techniques such as Decision tree, K nearest neighbors, Logistic regression, and Naïve Bayes and calculate their accuracies and compare it with the proposed classifier which uses Bag of bots' word model to detect Twitter bots from a given training data set which was get from Kaggle under the name(charvijain٢٧/detecting-twitter-bot-data). This dataset contains various attributes like URL, description, friends, several followers, a screen name (used to communicate online), location, id, verified (if the user is authenticated), favorite (used for liked tweets), listed count. The proposed model performs the best as it uses a bag of words model, which only searches for particular words in users, screen names, and tweets. With such a simple algorithm, it efficiently detected those words, vectorize them, and easily

classify them as bots or non-bots as two binary values. The training and the test accuracies as exclusively different i.e., the features are trained only on the training data which is a separate file and experimented on test data which is a separate test file. This proposed classifier has ٩٩,٢% of accuracy against another compared classifier which has (DT ٨٧,٢٪, KNN ٨٢,٦, LR ٦٧,٤٪ and NB with ٣٩,٧٪).

- **S. Rahman et al., (٢٠٢١) [١٦]** propose a hybrid anomaly detection approach that cascades multiple machine learning methods (DT-SVMNB), such as: decision tree (C٥,٠), Support Vector Machine (SVM), and Naïve Bayesian (NB) classifier for categorizing normal and abnormal users in social networks. They compiled a list of distinct traits based on user profiles and content. The suggested machine learning model, DT-SVMNB, is trained using two types of datasets with the selected features. In the social network, this approach categorizes users as depressed or suicidal. An experiment with their approach utilizing synthetic and actual social network datasets. The accuracy of the performance analysis is around ٩٨%, demonstrating the usefulness and efficiency of the proposed solution.

- **Asma A. Alsufyani, Sabah M. Alzahrani, (٢٠٢١) [١٧]** proposed to use natural language processing (NLP) along with machine learning techniques for text phishing detection, they started with ٦،٢٢٤ emails from an existing dataset that contains both phishing and legitimate emails. Four different machine learning algorithms were used for training which are k-nearest neighbors (KNN), Multinomial Naive Bayes (MNB), Decision Tree and AdaBoost. The developed models had to classify the text messages into two categories, which are phishing and legitimate. They used dataset collected by the researcher [El Aassal, A., Moraes, L., Baki, S., Das, A., Verma, R., and Verma, A. D. R.

(٢٠١٨)], they found that KNN, Decision Tree and AdaBoost models are effective against text phishing attack with accurzcy of ٩٣٪ for KNN, ٩٠٪ for DT and ٩٢٪ for ada boost.

- **Xiaoyu Luo, (٢٠٢١) [١٨]** made two analytical experiments in weka to check Support Vector Machine, Naive Bayes and Logistic Regression classifiers in English text classification using available dataset on the UCI library which contain English documents. The experimental results performed on a set of ١٠٣٣ text document present that the SVM classifier outperforms the rest of the machine learning technique with ٠,٨٨٪ of accuracy.
- **Maryam Heidari, James H Jr Jones, Ozlem Uzuner, (٢٠٢١)[١٩]** using Cresci ٢٠١٧ labeled data set of bots and human users on Twitter proposed a new way to classify bots by identifying new features to be used with the evaluation of the machine learning models such as Random Forest, Support Vector Machine, Logistic Regression, and feed-forward neural network. This experiment identifies Random Forest as powerful models for social media bot detection with accuracy of ٨٨٪.
- **B. Kavin et al., (٢٠٢٢) [٢٠]** presents an artificial intelligence technique for detecting spam on Twitter social networks. To construct a model, they used a Support Vector Machine (SVM), Artificial Neural Network (ANN), and the Random Forest (RF) technique. The findings show that the proposed support vector machine algorithm has the best precision, recall, and F-measure when compared to RF and ANN algorithms. The findings might be used to monitor and track shared photos on social media to identify unsuitable content and faked images, as well as to protect social media from digital threats and attacks. For SVM, RF, and ANN, the accuracy is ٩٧,٤٥٪, ٩٥,٥٦٪, and ٩٣,٢١٪, respectively. Table (١,١) presents the comparison of the related works.

Table (١،١). Comparison of related studies

Ref. No.	Year	Author	Techniques	Accuracy
[١١]	٢٠١٩	A. Al-Zoubi et al.	KNN+ RF+ DT+MLP	٩٧,١٣٪
[١٢]	٢٠٢٠	Muhammet S. Başarslana	SVM, NB, NN	٩٠٪
[١٣]	٢٠٢٠	M. Jabardi and A. Hadi	SWRL+ OWL	٩٤,٩٪
[١٤]	٢٠٢٠	A. Sarker et al.	SVM+KNN	٩٧,٧٪
[١٥]	٢٠٢١	A Ramalingaiah , S Hussaini	Bag of bots' word model	٩٩,٢٪
[١٦]	٢٠٢١	S. Rahman et al.	DT-SVMNB	٩٨٪
[١٧]	٢٠٢١	Asma A. Alsufyani	DT	٩٢%
[١٨]	٢٠٢١	Xiaoyu Luo	SVM	٨٨٪
[١٩]	٢٠٢١	Maryam Heidari	RF	٨٨٪
[٢٠]	٢٠٢٢	B. Kavin et al.	SVM+RF+ANN	٩٧,٤٥٪

١,٥. Problem Statement

The fundamental issue addressed by this thesis is that there is currently no codified approach or model for individuals to use, particularly for teaching themselves to be more attentive against social networking attacks, particularly OSNs attacks [٢١]. The presence of fake profiles may lead to major problems for users of social networking sites, the most important of which is phishing or blackmail. And because social networks are available to all people and of all ages, such problems threaten an entire society and must be addressed. The lack of high detection accuracy, especially when the diversity of data and attacks is one of the main problems in Online Social Networks (OSNs).

١,٦. Aim of Thesis

The purpose of this work is to construct a system for predicting of sensitive tweets, social engineering text messages and fake profiles into predefined classes based on contents of tweets, messages, and followers' profile so that the textual information in the datasets is considered as features used for classification and the account appearance characteristics are considered as features.

Find efficient techniques for sensitive tweets, social engineering messages and fake Twitter followers’ detection and evaluate their performance.

١,٧. Objectives of Thesis

This thesis sought to detect attacks in OSNs and provide a prediction model that can help in fake accounts detection and prevent attacks. To achieve these aims, the following objectives would be met:

١) Modifying a dataset based on specific texts to be analyzed in identifying the sensitive texts and attack texts on social media platform.

٢) Feature selection methods are used to make the processing more accurate. It also increases the prediction power of the algorithms by selecting the most critical variables and eliminating the redundant and irrelevant ones.

٣) Training a classifier to categorize the modified tweets into sensitive and not sensitive information. In comparison to earlier research, it is hoped that our chose features will improve the accuracy level of identifying sensitive information.

١,٨. Thesis Layout

Aside from the current chapter, there are four more chapters in this study, which are organized as follows:

- **Chapter Two, “Theoretical Background”,** deals with the fundamental of social networks, types of Machine Learning (ML) algorithms, attacks on social networks, and so on.
- **Chapter Three, “The Proposed System Architecture”,** is concerned with describing the design of the presented system.
- **Chapter Four, “Results and Discussion”,** The experimental results of each of the measures in this study are shown by the reasons.
- **Chapter Five, “Conclusions and Recommendations”,** This chapter offers a collection of suggestions for potential directions for future works in addition to the findings from the research that was done.

Chapter Two

Theoretical Background

Chapter Two
Theoretical Background

٢,١. Introduction

Nowadays, social networks are becoming an integral element of modern life, enabling widespread communication and interaction amongst individuals who share common beliefs, interests, and worldviews. People use social media sites like Facebook, Myspace, and Twitter to build personal and professional networks, get information that is pertinent to them, and share a lot of sensitive information with others. Even while Online Social Networks (OSNs) are becoming more and more popular, spammers and attackers are increasingly targeting various social media by uploading messages that include dangerous information.

In this chapter, the basic concepts of OSNs especially Twitter will be discussed in some detail. The basic knowledge of machine learning for classification is described, including the main procedures and role of machine learning, the most used pre-processing and statistical techniques for feature extraction and feature reduction, and the basic theory of some machine learning techniques used in this thesis. The chapter concludes with a set of performance measures of classifiers.

٢,٢. Online Social Networks (OSNs)

Information could now be shared in ways that had never been possible before once the internet started to gain popularity in the middle of the ١٩٩٠s. But the exchange of information still lacked a human touch. Later, social networking sites introduced a personal preference to online information exchange in the early ٢٠٠٠s, which was well-liked by the general public. Online Social Networks (OSNs) like Facebook, Twitter, and Instagram have grown to be quite important in today’s society.

OSNs are utilized not only as a means of communication but also for company promotion and popularity. At first look, indicators such as follower count or characteristics of the shared material, such as the number of likes, comments, or views, are used to gauge an account’s popularity. As a result, users of any social network may have a propensity to artificially boost its metrics in order to gain greater advantages from OSNs [٢١].

Social networks play a dual function in content consumption and production. They provide the user the option to control who can see their profile. Answers to inquiries about things like age, location, interests, etc. are used to create a profile. On certain websites, users may write blogs, and multimedia material, change the profile’s appearance and feel, upload photos, add contacts, and create and share contact lists. Social networks often contain restrictions that let users decide who may read their profile, contact them, add them to their contact list, and other functions in order to preserve user privacy [٢٢].

People often take the need to protect the data contained on these social networking sites for granted since they view social media as a personal communication tool. People are gradually posting more and more information on social networks in a variety of formats, which might result in unparalleled access to personal and corporate data. For opponents looking to hurt someone, the wealth of information saved on social networks is immensely alluring [٢٣].

With this enormous quantity of knowledge at their disposal, they can wreak devastation on the entire planet. Additionally, social media has developed into a fantastic platform for marketers to promote, and if they do not take social media security concerns seriously enough, they leave themselves up to a number of dangers and endanger the security of their sensitive data. The fundamental components of social networks are depicted in Fig. (٢,١), along with the industries in which they are most prevalent [٢٤].

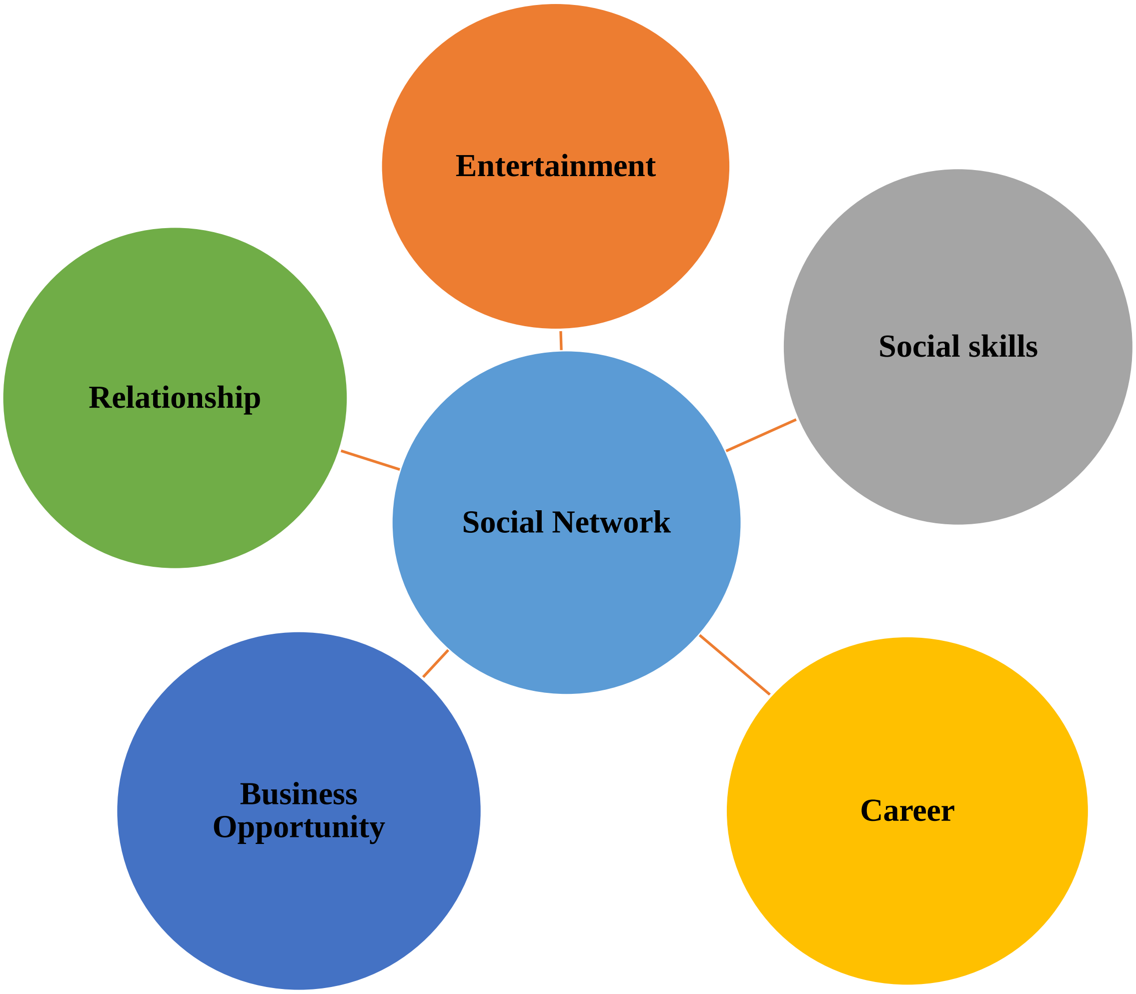


Figure (٢,١). Constituents of online social networks [٢٤]

٢,٢,١. Data on Online Social Networks

Boyd and Ellison’s definition already implies that there are two categories of user-related data on which OSNs depend [٣٣]:

A. Profiles: A user’s profile serves as their representation of the outside world and is linked to them. Typically, this is a description of oneself or one’s alter ego (pseudonym, avatar).

B. **Connections**: A link between two users exists and can take many different forms, such as a friend, coworker, fan, etc. A graph can be used to depict a set of connections. Other sorts of user information, however, are frequently implicated depending on the extra services the OSN provides.

C. **Messages**: messages are taken in their broadest sense. Multimedia may be a part of any data that is transmitted between individuals or between groups of users. The foundation for new OSN functionality is this. Even more so than friendship graphs, user interaction has been shown to be a valuable source of data about the underlying social network.

D. **Multi-media**: Users can send each other pieces of information, as well as publish them to either public or private data spaces (like picture albums, blogs, or Facebook "Walls"). Examples include written blog posts, visual images, audio recordings of music or voices, and video clips (video).

E. **Tags**: A user-defined tag is a keyword (meta-data) added to material (either the uploader or other users). When a user recognizes the individuals in a photo and tags the image with their names, this is referred to as "tagging" on Facebook. This directly links the identified individuals to the image.

F. **Preferences**: Many OSNs provide their users with some sort of peer or content matching or recommendation capabilities. Users frequently express preferences openly, which may or may not be made public. Preferences can occasionally be inferred indirectly from user behavior.

G. **Groups**: a number of users, the collaborative document, shared tastes or histories, or access to a common area are just a few examples of the resources, characteristics, or privileges that groups frequently share.

H. Behavioral information: browsing patterns and activities made by a user when using the OSN to complete tasks This kind of meta-data is especially comprehensive. It may and be used to infer information like interests, friendships, or even implicit information like physical location. Even if the activity there is unrelated to using a social network, behavioral data may also be discovered on conventional websites.

I. Login credentials: Most OSNs ask for or let users' login in order to access the service. The login credentials include this login information. The same item may be obtained on conventional websites.

٢,٢,٢. Applications of Social Networks

Applications for social networking have developed into significant businesses that provide consumers with online social networking platforms. Common uses include crowdsourcing, education, business, finance, healthcare, politics, and social contact mediated by computers, as follows [٣٤]:

A. Social Interaction: Social networking websites make it possible to link people who share interests and activities beyond political, economic, and geographic boundaries. They also allow computer-mediated social contact. They offer a contemporary sort of amusement. People use them to make new friends, reconnect with existing friends, find others who share their interests and maintain contact with long-lost acquaintances. Additionally, they offer a platform for online interaction and the sharing of private data for dating. Social networking sites are used by some job searchers to improve their chances of getting job offers and landing lucrative jobs.

B. Education: The manner that students and instructors interact in learning is being impacted by social networks. These days, they are utilized for information sharing, educator professional development, and learning. Social media is used by scientific groups for information exchange. Social networks are often used by researchers and librarians to discuss ideas and maintain professional connections. Networks for learning and research can emerge from social media. Many colleges use social networking sites like Facebook, Twitter, and Instagram often, and every university has at least one page on each of these sites. Education through social networking has difficulties with privacy, true friendship, time consumption, and misunderstanding. On the other hand, the main advantages are flexibility, reproducibility, convenience, and accessibility.

C. **Business:** Another great use is social networking among businesses. It may be a powerful marketing tool for companies, company owners, actors, musicians, or artists. Social networking sites are used by businesses primarily for five purposes: building brand recognition, managing their online reputation, recruiting, learning about new technology and rivals, and snooping on prospective customers. Social networking sites assist businesses in marketing their goods, identifying the requirements of customers, and gathering feedback from a variety of sources. Use of virtual money in social networks opens up the new financial potential for the world. Customers may share their own experiences on social media, which lowers the risk of purchasing a new product and assists early adopters in making an educated selection.

D. **Healthcare:** Different forms of social connectedness among various stakeholders, including physicians, patients, and caregivers, are made possible by social media. Social networking sites (SNS) are used to share fresh knowledge from the research and help physicians and nurses give high-quality treatment to their patients, making them useful teaching and learning tool. These technologies have the potential to directly impact almost every area of healthcare.

E. Politics: Social media appears to be having an effect on political activity and movements all around the world. All around the world, it has affected voting and sparked societal transformations, turmoil, upheavals, and revolutions. Citizens will be able to exercise their right to free expression thanks to social networking. Additionally, it is beneficial to interest the younger generations in politics and to encourage participation in the democratic process. For instance, Barack Obama used social media effectively in his ٢٠٠٨ campaign to mobilize supporters, empower volunteers, and greatly increase donations. Obama was the first US president to properly appreciate social media’s power.

٢,٢,٣. Various threats on online social network

Due to the availability of the internet and our technologically advanced world. The assaults that people have been tracking since social media first became popular are listed below. The following three groups are used to categorize threats [٣٥]:

A. Conventional threats: contain dangers that users have encountered since the social network’s inception.

B. Modern threats: involve assaults that target user accounts using sophisticated methods.

C. Targeted attacks: involve assaults that are directed at a specific user and can be carried out by any user out of a variety of personal grudges.

٢,٣. Social Networks Data Processing of machine learning

The following preprocessing and feature extraction techniques are used for data and are described in this section:

٢,٣,١. Preprocessing data

The data is subjected to a number of operations in this step to get it ready for the classification process, including text cleaning procedures such as eliminating stop words, URLs, numerals, lower case, stemming words, lemmatizing words, symbols, removing non-English terms, and normalizing data.

٢,٣,١,١. Texts preprocessing

The process of cleaning and preparing the text for classification is known as pre-processing the data. Online writings frequently have distracting elements like HTML tags, JavaScript, and ads. Additionally, many words in the text have little bearing on its overall direction on a word-by-word basis. Keeping those words increases the problem's high dimensionality and makes classification more challenging because each word in the text is considered as one dimension. Here is the idea behind having the data appropriately pre-processed: cutting down on text noise should help the classifier perform better and classify data more quickly, which will help with real-time sentiment analysis. Online text cleaning, white space removal, expanding abbreviation, stemming, stop words removal, negation handling, and feature selection are some of the phases that make up the entire process. The remaining phases are referred to as transformations, while the last step—which applies various functions to choose the necessary patterns is referred to as filtering [٣٦].

٢,٣,٢. Features Extraction

The words, concepts, or phrases that powerfully indicate an opinion as good or negative are known as features in the context of opinion mining. This indicates that they have a greater influence on the text’s orientation than other words do. There are various techniques for selecting features, including some that are syntactic (based on the syntactic position of the word, such as adjectives), some that are univariate (based on the relationship between each feature and a single category, such as chi-squared ($x^٢$) and information gain), and some that use genetic algorithms and decision trees depending on feature subsets.

By giving the text a particular amount of weight, it is possible to evaluate the significance of each feature in a variety of ways. Feature frequency (FF), Term Frequency Inverse Document Frequency (TF-IDF), and feature presence are the most widely used ones (FP). The document has tf instances in total. Eq. (٢،١) yields the TF-IDF as follows:

$$TF - IDF = tf(t,d) * Log\ (N/DF(d,t)) \qquad \ldots (٢،١)$$

where N represents the total number of documents and DF represents the total number of documents that include this feature and tf represents the frequency of term in document. [٣٧].

٢،٣،٣. Data Normalization

Data normalization is the process of changing raw data values into another form with qualities that are more suited for modeling and analysis. Normalization seeks to guarantee that all genes are measured in the same unit of measurement. As a consequence, it is utilized to prevent the discrepancy between the effect of tiny values and big values that dominate the outcomes. Many methods for data normalizing exist, including min-max and z-score normalization. Eq. (٢،٢) is used in min-max normalization to determine the value [٣٨].

$$\hat{V}= \frac{V - min_a}{max\ _a - min_a} * new\_max_a \text{ - } new\_min_a \qquad (٢،٢)$$

Where **V:** represents the feature value, $\boldsymbol{min_a}$ **:** is the minimum original value for any feature, $\boldsymbol{max_a}$**:** is the maximum original value for any feature, $\boldsymbol{new\_max_a}$**,** and $\boldsymbol{new\_min_a}$: are the maximum and minimum interval of values.

٢,٣,٤. Data splitting

Data splitting is the process of dividing accessible data into two halves, typically for cross-validation reasons. A predictive model is created using one set of data, while the effectiveness of the model is assessed using another set of data.

Where the data has been split using hold out cross validation with ٧٠٪ training for each dataset and ٣٠٪ for testing to avoid overfitting.

٢,٣,٤,١. Cross-Validation

One of the strategies for ensuring correct generalization and preventing overtraining is the use of cross-validation methodologies. The fundamental idea is to split data set T into two subsets: one for training and one for estimating the results of the final model. Cross-main validation’s goal is to approximate the model's results in a way that is reliable and consistent [٣٩].

- **K-fold Cross-Validation**

The initial sample will be randomly divided into k equal-sized sub-samples for the k-fold cross-validation which is shown in Fig. (٢,٢). The remaining k sub-samples will be used for data training while one sub-sample from the total of k sub-samples will be used as validation data to test the model. Additionally, k repetitions of cross-validation are performed, with k subsamples serving as validation data each time. Additionally, one estimate might be produced by multiplying the k findings. This method has the advantage of repeated random subsampling in that all observations were used for training and validation, with each observation only being used once [٤٠].

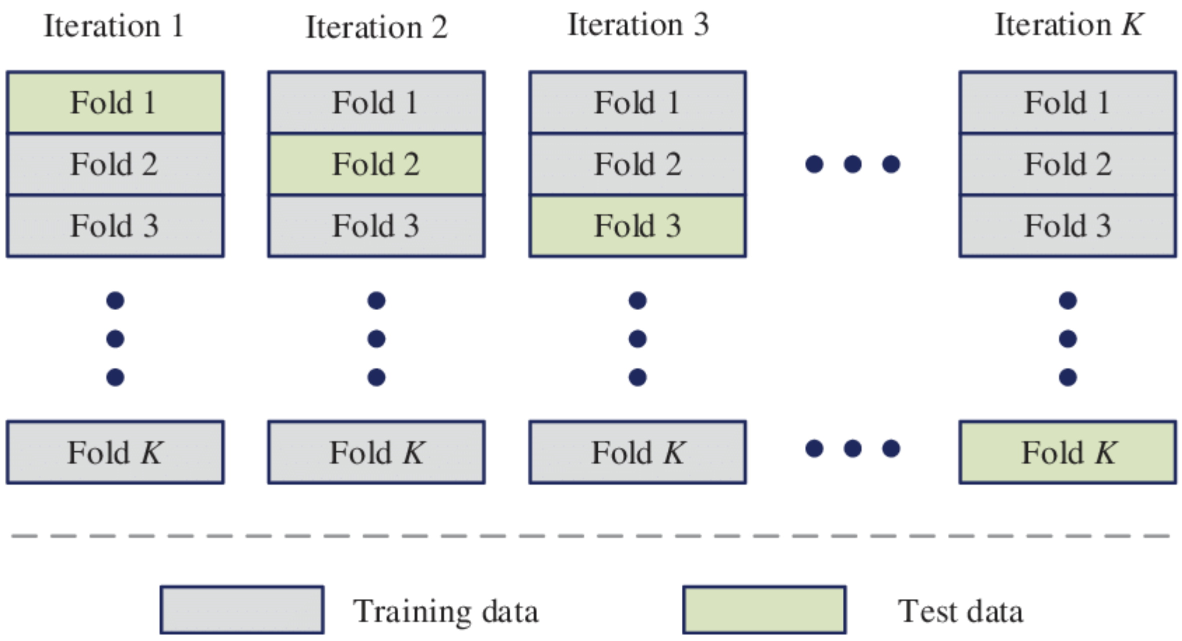


Figure (٢،٢). K-fold cross-validation method [٤٥].

➢ Hold-out Cross-Validation

The simplest form of cross-validation method, hold-out cross-validation in Fig. (٢،٣) is known for its simplicity and effectiveness. It divides the data set T (of size n) into three mutually disjoint subsets: testing T_t, validation T_v, and training $T_{tr}$of nt, nv, and ntr, respectively. One of the advantages of this method is that the proportion of these three data sub-sets is not strictly restricted. Furthermore, the model was trained on the training subset T_{tr}, while the validation subset T_v was used to evaluate the model's success during training to prevent overtraining. When the output on T_v is good enough, or when it can no longer be improved, training will be discontinued. When comparing the performance of m > ١ computational models $L_١$,…, L_m against each other, the testing subset T_t will be used to obtain a confident estimate of the models' performance. The problem with this method is that the three subgroups must be efficiently divided; as a result, the model's output would remain constant because the model's accuracy is dependent on the test group [٤١].

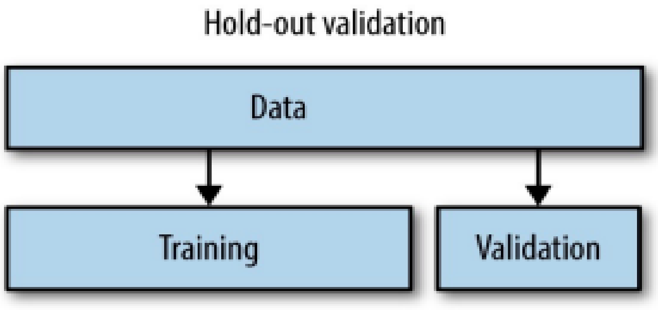


Figure (٢،٣). Hold-out cross-validation [٤١].

Table (٢،١) explain the main difference between these two types.

Table (٢،١). Hold-out vs K-fold cross validation [٤٢].

Hold-out	K-fold
١. Holdout method partitions the full set of data into two sets, namely the training set and the test set. It is common to hold out two-thirds of the data for training (learning phase) and the remaining one-third of the data are for testing (classification phase). ٢. Each set must be chosen independently and randomly.	١. k-fold Cross Validation method randomly partitions the document into K mutually exclusive subsets (called fold) $s_{١}, s_{٢},\ldots,$ s_k each of approximately equal size. ٢. Model training and testing is performed k times, iteratively. The i^{th}, a subset s_i is reserved as the test documents while the remaining subsets are used as training documents. For example, at first, iteration s_i will be the testing documents, and $s_{٢}$, $s_{٣},\ldots,$ $s_k$will be the training documents. Accuracy equals the ratio of total appropriate classification from k iterations with a total number of documents.

٢,٤. Feature Selection Methods

Feature reduction technology is used to obtain the best features that can contribute to obtaining higher accuracy in detection. Common ways to reduce features are as follows:

٢,٤,١. Chi-square

The chi-square ($\chi^{٢}$) statistic is a measure for determining how well a model fits real data. A chi-square statistic needs data that is random, raw, mutually exclusive, and derived from a large enough sample. It is used in statistics to determine if two events are independent. From the data of two variables, we may get the observed count O and anticipated count E. The Chi-Square test reveals the difference between expected count E and actual count O. The chi-squared formula is as follows:

$$x_C^2 = \text{sum}\ \frac{(O_i - E_i)^{٢}}{E_i} \qquad \dots (٢,٣)$$

Where c is the number of degrees of freedom.

O stands for observed value (s)

E stands for anticipated value (s)

It was used to pick features that are heavily dependent on the reaction while we’re selecting features [٤٣].

٢,٤,٢. ANOVA

It is a statistical analysis technique that differentiates observed aggregate variability within a data set into systematic and random variables. Random variables have no statistical significance in the data set supplied, but systematic impacts do. The ANOVA test is used in regression research to investigate the influence of independent variables on the dependent variable. The ANOVA test is used in regression research to investigate the influence of independent variables on the dependent variable. The ANOVA test is used in regression research to investigate the influence of independent variables on the dependent variable. In regression research, the ANOVA test is used to explore the impact of independent factors on the dependent variable. The F-Value is calculated by multiplying two Chi-distributions by the degrees of freedom of each.

$$F = \frac{\left(\frac{x_{١}^{٢}}{n١-١}\right)}{\frac{x_{٢}^{٢}}{n٢-١}} \qquad \ldots (٢,٤)$$

Where: $x_{١}$ and $x_{٢}$ are chi-distribution and $n_{١}, n_{٢}$ are its respective degrees of freedom [٤٤].

٢,٤,٣.Mutual information

Mutual information is a measure between two (possibly multi-dimensional) random variables X and Y that quantifies the amount of knowledge received about one random variable through the other random variable. The reciprocal information comes from:

$$I(X;Y) = \int_{X} \int_{Y} \frac{p(x.y)\log p(x.y)}{p(x)p(y)} dxdy \qquad \ldots (٢,٥)$$

The combined probability density function of X and Y is p (x, y), while the marginal density functions are p(x) and p(y). This is a useful feature when the combined mutual information for a chosen feature and the target variable is maximized [٤٥].

٢,٤,٤. Regularized Logistic Regression Feature selection

Logistic regression is a standard statistical technique addressing binary classification problems [٥]. However logistic regression models tend to over-fit the learning sample when the number p of features, or input variables, largely exceeds the number n of samples. This is referred to as the small n large p setting, commonly found in biomedical problems such as gene selection from microarray data.

A typical solution to prevent over-fitting considers an l٢ norm penalty on the regression weight values, as in ridge regression [١٠], or an l١ norm penalty for the (Generalized) LASSO [٢٠،١٦], possibly a combination of both, as in Elastic Net [٢٢]. The l١ penalty has the additional advantage of forcing the solution to be sparse, hence performing feature selection jointly with the classifier estimation. This classifier is chosen because, if well regularized, it tends to offer good predictive performances and its probabilistic output helps assigning a confidence level to the predicted class.

Let x ∈ Rp denote an observation made of p feature values and let y ∈ {−١, +١} denote the corresponding binary output or class label. A logistic regression models the conditional probability distribution of the class label y, given a feature vector x as follows.

$$P(x;y) = \frac{١}{١+\exp(-y(w^t\, x+v))} dxdy \qquad \dots (٢,٦)$$

where the weight vector w ∈ R^p and intercept v ∈ R are the parameters of the logistic regression model. The equation w^t x+v = ٠ defines an hyperplane in feature space, which is the decision boundary on which the conditional probability of each possible output value is equal to ١/٢.[٤٦].

٢,٤,٥. Tree-based feature selection

a Decision tree induction is used for selecting relevant features. Decision tree induction is the learning of decision tree classifiers. It constructs a tree structure where each internal node (non-leaf node) denotes the test on the attribute. Each branch represents the outcome of the test, and each external node (leaf node) denotes the class prediction. At each node the algorithm chooses the best attribute to partition data into individual classes. The best attribute for partitioning is chosen by the attribute selection process with Information gain measure. The attribute with highest information gain is chosen for splitting the attribute. The information gain is of the attribute is found by

$$\mathrm{Info(D)} = \sum_{\mathrm{i}=١}^{m} \mathrm{p}i \ \log_{٢}(p) \qquad \ldots(٢،٧)$$

where p_i is the probability that a arbitrary vector in D belongs to class c_i. A log function to the base ٢ is used, because the information is encoded in bits. Info (D) is just the average amount of information needed to identify the class label in vector D. Before constructing the trees, base cases have to be taken into consideration with following points: If all the samples belong to the same class, it simply creates the leaf node for the decision tree. If no features provide any information gain, it creates a decision node higher up the tree using the expected value of the class.

The algorithm for decision tree induction is given as follows ١. Check for base cases. ٢. For each attribute a, find the information gain of each attribute for splitting ٣. Let a-best be the attribute with highest information gain ٤. Create a decision node that splits on a-best ٥. Recur on the sub lists obtained by splitting on a-best and add those nodes as children for the tree. The trees are constructed from top-down recursive approach which starts with training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is built. After the tree is built, for easy interpretation the rules are extracted using the leaf nodes of the tree, because rules give more comprehensibility than tree structure in case of big dataset.[٤٧].

٢,٤,٦. Relief

Kira and Rendall introduced relief as a traditional filtering feature selection strategy. Kononenko (١٩٩٤) describes it as an example-based learning strategy. Relief F determines which characteristics are relevant based on their feature weights. It begins by selecting a single R sample from the training set. We calculate the distance between each sample H and R for a sample set H in the same category as R, then choose the K samples that are the closest. Simultaneously, we must calculate each of the R separate categories of sample distance in the sample sets M and R, and then choose K closest samples as R not the same type of nearest neighbor sample. When calculating the feature weight of a certain feature A, there is a mismatch between two samples R and H in the training set. When the training set’s two samples R and Mj(c) disagree, feature A is given a low feature weight. When they are identical, feature A is given a high feature weight. A feature’ s feature weight is defined as [٤٨]:

$$W(A) = W(A) - \sum_{j=\text{١}}^{k} \frac{diff(A.R.Hj)}{m * k} +$$

$$\sum_{c \nexists class(R)} [\frac{P(c)}{\text{١} - P(class(R))} \sum_{j=\text{١}}^{k} diff(A.R.Mj(c))] /m * k \qquad \dots$$

(٢.٨)

٢،٤،٧. The Minimum Redundancy Maximum Relevance

The Minimum redundancy maximum relevance feature selection technique uses a forward selection filter. The technique works by adding one explanatory variable (feature) to the list at a time, enhancing relevance to the objective variable while reducing repetition among previously chosen variables. The purpose of the FS algorithm is to use mutual information to discover the subset S of explanatory variables on which Y has the largest reliance.

Assuming there are in total m features, and for a given feature Xi (i ∈ {١, ٢, ..., m}), its feature importance based on the mRMR criterion can be expressed as [٦]:

$$f^{mRMR}(X_i) = I(Y,X_i) - \frac{\text{١}}{|S|}\sum_{X_s \in S} I(X_s,X_i) \ldots(\text{٢،٩})$$

where Y is the response variable (class label), S is the set of selected features, |S| is the size of the feature set (number of features), X_s ∈ S is one feature out of the feature set S, X_i denotes a feature currently not selected: X_i ٦∈ S. The function I(·, ·) is the mutual information:

$$\mathrm{I(Y,X)} \ = \int_{\Omega\mathrm{Y}} \int_{\Omega\mathrm{Y}} \mathrm{p(x,y)}\ \log(\frac{p(x,y)}{p(x)p(y)}\ \mathrm{dxdy} \quad \ldots\ (\text{٢،١٠})$$

where Ω_Y and Ω_X are the sample spaces corresponding to Y and X, p(x, y) is the joint probability density, and p(˙) is the marginal density function.

For discrete variables Y and X, the mutual information formula takes the form:

$$\mathrm{I(X,Y)} = \sum_{y\in\Omega\mathrm{y}}\ \sum_{y\in\Omega x}\ \mathrm{p(x,y)}\ \log(\frac{p(x,y)}{p(x)p(y)}\ \ldots\ (\text{٢،١١})$$

In the mRMR feature selection process, at each step, the feature with the highest feature importance score max $x_{i\notin}f^{mRMR}(X_{i)}$ will be added to the selected feature set S.[٤٩].

٢,٤,٨. Light GBM

Light gradient boosting (LGB) model is an efficient implementation of the classic gradient boosting decision tree (GBDT) model. +e LGB model handles the classification, regression, and ranking problems in machine learning. GBDT obtains the final answer by combining multiple decision trees and by adding up the results of all the decision trees. +is process has been improved to obtain extreme gradient boosting (XGB). The difference between XGB and GBDT is in the way the tree is split and the way the value of the leaf node is determined. the core idea is to conduct a second-order Taylor expansion of the loss function to be fitted by GBDT and to introduce the regular term of the tree intelligently, so that the formula of the second-order Taylor expansion can be simplified and solved analytically. thus, a new tree splitting method and leaf node value determination method are derived. LGB is further optimized on the basis of XGB's improvement of GBDT formula. Figure ١ presents a flow chart of the transition from the gradient boosting method to the LGB model.

LGB model is further optimized on the basis of the XGB. These optimizations are performed to reduce the computational cost, but they can also play a role in preventing overfitting (because the original data are noisy, some rough processing may increase the generalization ability of the model). The computational cost of each leaf node split is

$$\text{Cost}_{\text{time}} = \text{feature}_{\text{num}} \times \text{sample}_{\text{num}} \times \text{point}_{\text{num}} \quad \ldots(٢,١٢)$$

where Cost $_{time}$ represents the time (s) consumed by calculation, feature $_{num}$ represents the number of features, sample $_{num}$ represents the number of samples, and point $_{num}$ indicates the number of candidate points.[٥٠].

٢,٥. Machine Learning (ML) Techniques

One of the most important and beneficial uses of artificial intelligence (AI) is machine learning, which enables computer systems to automatically learn and improve their functioning without explicit programming. The main goal of machine learning algorithms is to create automated systems that can access and utilize training data. (The training dataset, also known as the learning labeled data, is where the e-learning process begins. Real-world examples, reviews, examples, or user comments can all be used to identify trends in the data and inform decisions that will be made in the future. The primary goal of machine learning models is to automatically learn without human intervention. Fig. (٢,٤) depicts the three main categories of machine learning, which are employed for a variety of applications [٥١].

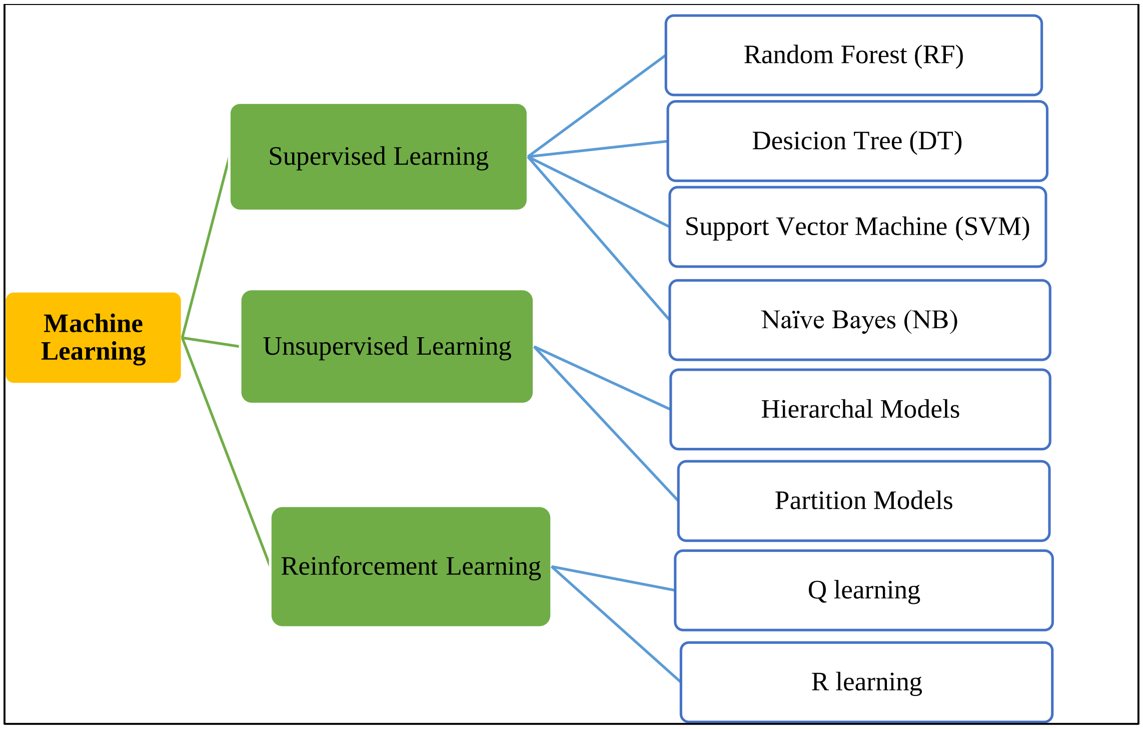


Figure (٢,٤). Types of machine learning [٥١].

The main types of machine learning are:

١) Supervised Learning

This learning strategy requires labeled data to better explain the relationship between the feature vector and the label. The Markov hidden model, decision tree, support vector machines, neural networks, and random forest were among the gait research techniques used. The prominence of the support vector machine in gait analysis stems from its ability to perform well even with minimal datasets. Kernels may solve both linear and nonlinear problems. In addition to the binary form, the classification efficiency was extended to numerous classes and is quite beneficial for gait analysis. Fig. (٢,٥) explains supervised learning method [٥٢].

٢) Unsupervised Learning

It is an unlabeled learning method. The algorithm creates the relationship between multiple inputs in order to evaluate output. The distance between all vector features is the focus of the majority of clustering approaches. These approaches were not investigated in gait research since accurately identifying the learning goals and managing a large number of feature vectors was a time-consuming task. Certain strategies can thus be used when the relationship between several outcomes is uncertain. Combining classification with some dimensional reduction approaches is important when working with large data sets. These unsupervised procedures can develop a variety of habits. An explanatory review can guarantee that the distance metrics for the challenge are correctly chosen [٥٣].

٣) Reinforcement Learning

In rehabilitation facilities, such as exoskeletons, reinforcement learning is required for systems to interact and walk in a dynamic context. Several recovery control mechanisms have been devised. Reinforcement learning and deep neural networks have been frequently employed in rehabilitation devices because of their ability to aid in recognizing participant heterogeneity and hence simplifying according to particular settings [٥٤].

٢,٥,١. Attacks Prediction in Online Social Networks Based on ML Techniques

This section briefly describes the commonly used supervised machine learning algorithms for social network attack prediction. These algorithms are:

٢,٥,١,١. Support Vector Machine (SVM) Algorithm

Both linear and non-linear data may be classified using the SVM method. Each piece of data is first mapped into an n-dimensional feature space, where n is the total number of features. The hyperplane that divides the data points into two classes is then identified, maximizing marginal distance for both classes and reducing classification errors. The distance between the decision hyperplane and the closest instance that belongs to the class is what is known as the marginal distance for that class. In a more formal setting, each data point is initially represented graphically as a point in an n-dimensional space (where n is the number of features), with the value of each feature being the value of a particular coordinate. Researchers must then locate the hyperplane that separates the two classes by the greatest margin in order to accomplish the classification. A simple representation of an SVM classifier is shown in Fig. (٢,٦) [٥٥].

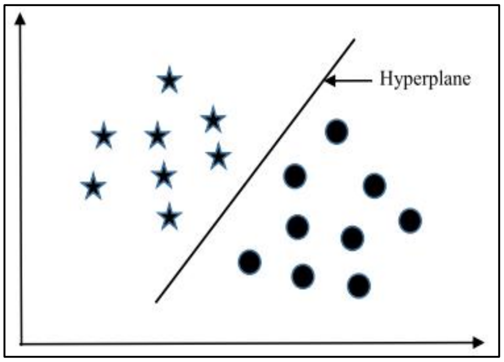


Figure (٢,٦). SVM classifier [٥٥].

If the dataset is linearly bounded, linear SVM may be used; if the dataset is non-linearly bounded, non-linear SVM can be used for classification tasks. Let us consider a dataset ($A_{١}$, $B_{١}$,…. A_n, B_n); where ($A_{١}$,… A_n) is the set of the input variable, ($B_{١}$,…, B_n) is the output variable, and ‘C’ is the intercept, then the SVM classifier is given as like Eq. (٢.٣).

$$SVM = \sum_{i=١}^{n} \beta_i - \frac{١}{٢} \sum_{i,j=١}^{n} b_i b_j C\left(a_i, a_j\right) \beta_i \beta_j \qquad \ldots (٢.٧)$$

Where, i=١،٢،٣….n; and C= $b_i\beta_i+b_j\beta_j$ [٥٦].

٢,٥,١,٢. Decision Tree (DT) Algorithm

One of the first and most well-known machine learning algorithms is DT. To classify data elements into a tree-like structure, a decision tree represents the decision logic, that is, the tests and outcomes that correspond to them. A DT tree's nodes typically have several layers, with the first or highest node is referred to as the root node. Tests on input variables or characteristics are represented by all internal nodes, which are nodes with at least one child. The classification algorithm branches in the direction of the suitable child node based on the results of the test, and then continue the test and branching procedure until it reaches the leaf node. The choice outcomes are represented by the leaf or terminal nodes. DTs are often used in various medical diagnostic regimens since they are simple to use and quick to learn [٥٧]. The results of all tests at each node along the path will be sufficient information to speculate on the classification of the sample when navigating the classification tree. Fig. (٢,٧) shows a representation of a DT together with its constituent parts and rules. The choice outcomes (Class A and Class B) are depicted by rectangles, and each variable (C١, C٢, and C٣) is represented by a circle. Each branch is given a label of “True” or “False” based on the result of the test of its ancestor node in order to correctly classify a sample to a class [٥٨].

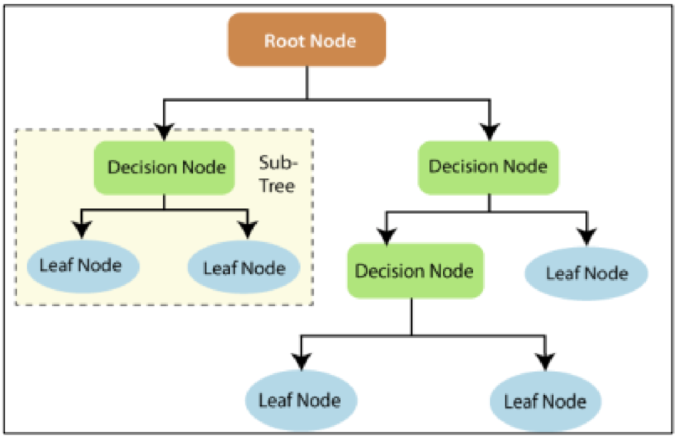


Figure (٢,٧). DT Classifier [٥٨].

٢,٥,١,٣. Random Forest (RF) Classifier

Like how a forest is made up of several trees, RF is an ensemble classifier made up of multiple DTs. Very deep DTs frequently lead to overfitting of the training data, which produces a large degree of classification variance. result for a little modification in the supplied data. They are prone to errors on the test dataset because they are particularly sensitive to their training data. The various DTs of an RF are trained using the various training dataset components. The input vector of a fresh sample must pass down with each DT of the forest to be classified. Each DT then considers a distinct aspect of the input vector and provides a classification result. The classification that receives the most “votes” (for a discrete classification outcome) or the average of all the trees in the forest is then chosen by the forest (for numeric classification outcome). The RF method can decrease the variance caused by the consideration of a single DT for the same dataset since it takes results from several distinct DTs into account. The RF method, which comprises three separate decision trees, is illustrated in Fig. (٢,٨). A random subset of the training data was used to train each of those three decision trees [٥٩].

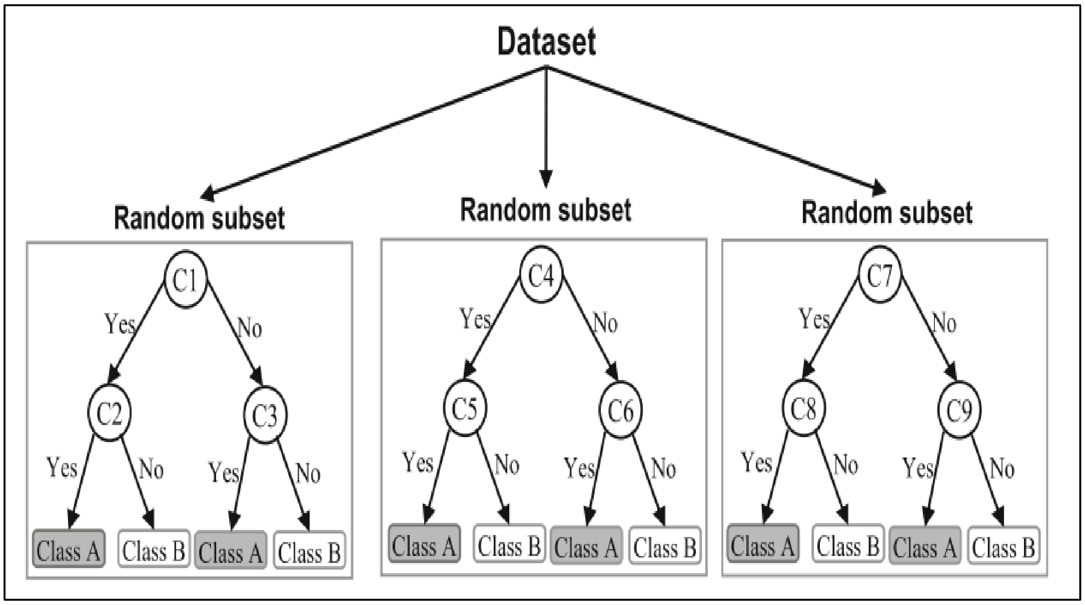


Figure (٢,٨). RF algorithm [٥٥].

More formally, for a p-dimensional random vector X = $(X_{١}.\,....X_p)^{T}$ represents the real-valued input or predictor variables and a random variable Y represents the real-valued response, it assumes an unknown joint distribution P_{XY} (X, Y). The goal is to find a prediction function f(X) for predicting Y. The prediction function is determined by a loss function L (Y, f(X)) and defined to minimize the expected value of the loss:

$$E_{xy} = \mathrm{L}\left(\mathrm{Y}, \mathrm{f(X)}\right) \qquad \ldots (٢.٨)$$

where the subscripts denote expectation with respect to the joint distribution of X and Y [٦٠].

٢,٥,١,٤. Stochastic Gradient Descent (SGD) Algorithm

The most often used approach for training data is Stochastic Gradient Descent (SGD). Due to the lengthier training times caused by larger networks and datasets, training on distributed systems is frequent, and distributed SGD variations, primarily asynchronous and synchronous SGD, are commonly employed. Asynchronous SGD is efficient in terms of communication, but its accuracy degrades because of delayed parameter updating. Regardless of its benefit, synchronous SGD becomes communication intensive as the number of nodes rises. This classifier is calculated by Eq. (٢.٥) as follows [٦١]:

$$e_{u.i} = r_{u.i} - \sum_{k=١}^{k} p_{uk} \cdot q_{ki} \qquad \ldots (٢.٩)$$

٢,٥,١,٥. Gaussian Naïve Bayes (GNB) Algorithm

One of the top ١٠ data mining methods is simply Naïve Bayes, a powerful classifier. The classifier Naïve Bayes is effective and frequently used in a variety of applications, including text categorization, document evaluation (such as spam filtering), and data stream classification. The generative model-based classifier Nave Bayes has a quick learning and testing cycle. The Bayesian rule and probability theorems are the foundation of Bayesian classifiers. An NP-hard issue exists when attempting to learn an ideal Bayesian classifier from training data [٦٧]. Naïve Bayes, a condensed form of the Bayesian classifier, makes two assumptions. The first is that attributes are conditionally independent given the class label, and the second is that no latent attribute effects on the label prediction process exist. A Bayesian classifier's benefit is that it can make classifications with a limited training dataset. This classifier performs multiclass prediction effectively and predicts the test data set's class quickly and accurately [٦٢].

$$p(c|x) = \frac{p(X|C)p(c)}{p(x)} \qquad \ldots (٢.١٠)$$

Where,

$p(c|x)$: given predictor, the posterior probability of class (c, target) (x, attributes).

$p(c)$: the likelihood of a certain class in the past.

$p(x|c)$: the likelihood, which is the likelihood of a predictor in a particular class.

$p(x)$: the predictor's prior probability.

Gaussian Naïve Bayes classification is a case of Naïve Bayes method with an assumption of having a Gaussian distribution on attribute values given the class label. For example, suppose that i^{th}attribute is continuous and its mean and variance are represented by $\mu_{c.i}$ and $\sigma^{٢}_{c.i}$ respectively, given the class label c. Hence, the probability of observing the value xi in i^{th} attribute given the class label c, is computed by Eq. (٢.٧) that is also called as normal distribution.

$$p(x_i|c) = \frac{١}{\sqrt{٢\pi\sigma^{٢}_{c.i}}}.\, exp^{\left(-\frac{x_i - \mu_{c.i}}{٢\sigma^{٢}_{c.i}}\right)} \qquad \ldots (٢.١١)$$

Even with Gaussian estimation, this method suffers the weakness of conditional independence of attributes [٦٣].

Table (٢.٢) shows the advantages and limitations of different supervised machine learning algorithms.

Table (٢.٢). Advantages and limitations of different supervised machine learning algorithms [٦٤].

Supervised algorithm	Advantages	Limitations
Support Vector Machine (SVM)	- More robust compared to LR - Can handle multiple feature spaces. - Performs well in classification of semi-structured or unstructured data, such as words, photos, etc. - Less possibility of overfitting	- Computationally expensive for large and complex datasets. - Performs poorly when the data are noisy. - It might be challenging to comprehend the model, weight, and influence of the final set of variables.

		- Unless expanded, a generic SVM can only classify up to two classes.
Decision Tree (DT)	- The classification tree that is produced is simpler to comprehend and interpret. - Preparing data is simpler. - Support for a variety of data formats, including numeric, nominal, and categorical. - Has the ability to produce reliable classifiers that can be tested using statistics.	- Demand that classes be exclusive of one another. - If any attribute or variable value for a non-leaf node is absent, the algorithm cannot branch. - The order of the qualities or variables affects the algorithm.
Random forest (RF)	- RF takes the average value from the results of its individual decision trees, which reduces variation and overfitting of training data compared to DT. - This ensemble-based classifier outperforms its individual base classifiers, or DTs, empirically. - It can offer estimates of what factors or characteristics are significant in the categorization and scales well for huge datasets.	- Costlier to compute and more complicated. - The quantity of base classifiers must be specified. - When assessing variable significance, it prioritizes variables or qualities that may take many alternative values. - Overfitting can happen quickly
Stochastic Gradient Descent (SGD)	- Because the network only processes one training sample,	- The steps taken to reach the minima are highly noisy because of frequent updates.

	it is simpler to fit in the memory. - Since only one sample is processed at a time, it is computationally quick. - For bigger datasets, it may converge more quickly since the parameters are updated more frequently. - Because of frequent updates, the steps necessary to reach the loss function minima feature oscillations that might assist you escape their local minimums (in case the computed position turns out to be the local minimum).	This frequently causes the gradient drop to tilt in another direction. - Additionally, it could take longer to get convergence to the loss function minima due to noisy steps. - Because one training sample is processed at a time, frequent updates are computationally costly. - As it only handles one sample at a time, it lacks the benefit of vectorized operations.
Gaussian Naïve Bayes (GNB)	- Easy to use and excellent for huge datasets. - Applicable to situations involving binary and many classes in classification. - Less training data are needed, according to this. - It is capable of handling both continuous and discrete data and can produce probabilistic predictions.	- Classes must be exclusive of one another. - Dependency between characteristics has a detrimental impact on categorization accuracy. - It presupposes that numeric properties are distributed normally.

٢,٦. Online Social Network Security Threats

OSN is an interaction-based tool that allows registered users to engage with individuals of various ages in the network. It may be further subdivided into three categories: classic threats, advanced persistent threats, and current threats [٦٥]:

٢,٦,١. Classic Threat

Because of the OSN structure, this danger spreads quickly and readily throughout the network among users. By using users' personal information, this malware impacts users' profiles and credentials. It spreads via clicking on malicious code or links. This hazard is further categorized into categories.

- **A. Man in the Middle Attack:** This form of attack is like spamming and is used to exploit OSNs on a wide scale.
- **B. Phishing attacks:** This sort of assault is increasing at an alarming rate, with more than ٨٥٪ of organizations experiencing this type of attack. The virus is best delivered via email attachments. To obtain all of the personal information, the attacker creates an identical social networking page.
- **C. Spamming Attack:** OSN users utilize third-party email providers to send unwanted messages and advertisements. Many researchers discovered that attackers construct a phony profile and engage individuals in harmful behavior without their awareness.
- **D. Malware:** It is a harmful code that includes Trojan horses, viruses, and worms. This has the potential to spread to many internet users. Most organizations are phasing out the problem of being unable to safeguard the material flow from users to servers.
- **E. Ransomware:** This type of assault is likewise malware, but the attackers prevent computer access until a quantity of money is paid.

٢,٦,٢. Advanced Threats

This sort of threat targets users' sensitive information by impersonating legitimate users in a deceptive manner. It is further subdivided into many assaults as follows:

A. **Whaling attack:** This attack gathers user information from many OSNs and poses as legitimate users to get personal information. This sort of assault gathers information about employees working for a certain firm.

B. **DDoS attack:** This assault originates from a variety of sources and uses users' computers without their knowledge.

C. **Speculation attack:** It is a geographical depiction of users under assault. When the number of users grows, so does the graph. This allows attackers to determine the location and use of users.

D. **Online chat attack:** This form of attack involves injecting harmful code into the conversation box. These chat rooms allow members to freely communicate any content and provide all personal information.

E. **Vicinity Attack:** The major hazard is information sharing via OSN. Users must safeguard their bank account numbers and other critical information.

F. **Sybill attack:** This assault focuses on peer-to-peer systems and dispersed environments.

٢,٦,٣. Modern Threats

These types of attack are basically related to OSN.

A. **Clickjacking:** It is a sort of attack that takes advantage of mouse clicks. When people click on some undesirable advertisements, they may be unaware of this form of assault.

B. **Social -Bots:** A social botnet is a collection of people that do harmful behaviors and time mimicking.

C. **SQL injection:** This attack primarily targets the backend by uploading malicious data into the system.

D. **Internet fraud:** It is the use of the internet to carry out illicit actions. This is done with malicious motives.

E. **Cross-site scripting:** This attack injects malicious code onto many websites.

٢,٧. Evaluation Metrics

In a binary classifier's confusion matrix Fig. (٢،٩). Actual values are denoted True (١) and False (٠), whereas anticipated values are marked Positive (١) and Negative (٠). Estimates of classification model possibilities are generated using the confusion matrix expressions TP, TN, FP, and FN [٧٠].

		Actual Condition	
		Positive	Negative
Predicted Condition	Positive	**TP**	**FP**
	Negative	**FG**	**TN**

Figure (٢.٩). Confusion matrix for the classification problem [٧٠].

For evaluating the performance of the system, certain parameters are used to determine its behavior [٧١]:

١. **Accuracy:** The percentage of instances properly classified out of all those presented. It's computed as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad \dots$$

(٢.١٢)

٢. **Precision:** For all those identified as class x, the proportion of true x-class occurrences. It is computed as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad \dots$$

(٢, ١٣)

٣. **Recall:** The proportion of instances classified as class x out of all examples classified as class x. It is computed as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad \dots$$

(٢. ١٤)

٤. **F- measure** or F-score is a measure of the test's accuracy. It is determined using the following formula in Eq. (٢,١١), which is based on accuracy and reminders:

$$\text{F}_{١} = ٢ * \frac{\text{precision}*\text{recall}}{\text{precision}+\text{recall}} \qquad \dots$$

(٢,١٥)

٥. **Specificity** is computed by dividing the total number of negatives (N) by the number of correct negative predictions (TN). The highest specificity is ١,٠, while the lowest is ٠,٠.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}}\ ١٠٠٪ \qquad \dots\ (٢,١٦)$$

Where,

TP = True positives are the number of cases that were anticipated to be positive but turned out to be true.

FP = False positives: the number of cases that were expected to be positive but turned out to be negative.

TN = True negatives are the number of cases that were expected to be negative but turned out to be true.

FN = False negatives: the number of cases that were expected to be negative but turned out to be positive.

٦. **Area under the ROC Curve:** Its area under the ROC curve is a performance metric for classification problems with varying thresholds. AUC is the degree or measure of specificity, whereas ROC is a probability curve. It indicates how well the model can discriminate between classes. The greater the AUC, the better the model predicts ٠s, ١s as classes.

Chapter Three
The Proposed System

Chapter Three

The Proposed System

٣,١. Introduction

This chapter discusses the suggested methods for achieving the objectives of the thesis. Also provided are the algorithms and techniques utilized to accomplish these goals. The suggested approach is made up of five classification models, which are trained on an ensemble of different models. It then predicts an output (class) based on the class that has the highest likelihood of being selected as the output. Based on the total majority of votes cast for each output type, the suggested technique forecasts outputs. The suggested technique includes the following models that cooperate to forecast results: Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Gaussian Naïve Bayes (GNB), Random Forest (RF), and Decision Tree (DT).

٣,٢. The Proposed System Architecture

The suggested approach is divided into numerous steps. Choosing a training approach is part of the initial stage. The data are preprocessed in the second stage to get them ready for classification. In the third stage, text features are retrieved using TF-IDF, and when detecting bogus accounts, additional features are analyzed, and rank feature selection methods are used to choose features with a high chance of making accurate predictions. The data will then be divided into training and testing sets and fed into a combination of models to forecast the results.

The proposed system has three use cases: the first case involves classifying social engineering text as an attack or a normal message using a dataset of tweeter-labeled messages, the second case involves classifying fake accounts using a dataset of fake and normal labeled tweeter accounts, and the third case involves classifying the text (tweet) as a sensitive or not sensitive tweet. The result determines whether the account is secure or whether more action is required to ensure the protection of personal information.

The block diagram in Fig. (٣،١) which contain three parts where the first part includes read one of the three dataset and doing data preprocessing (Character lower case, stop words removal, Non- English words removal, URL’s removal and merge other features if the chosen data set is for fake account detection that contain more features like time zone, geo enabled, followers count, like counts...), part two contain feature selection methods that’s have been used and third part contain the machine learning methods that were used to classify the dataset. The methods employed to accomplish these phases and actions are also made clear in Algorithm (٣،١).

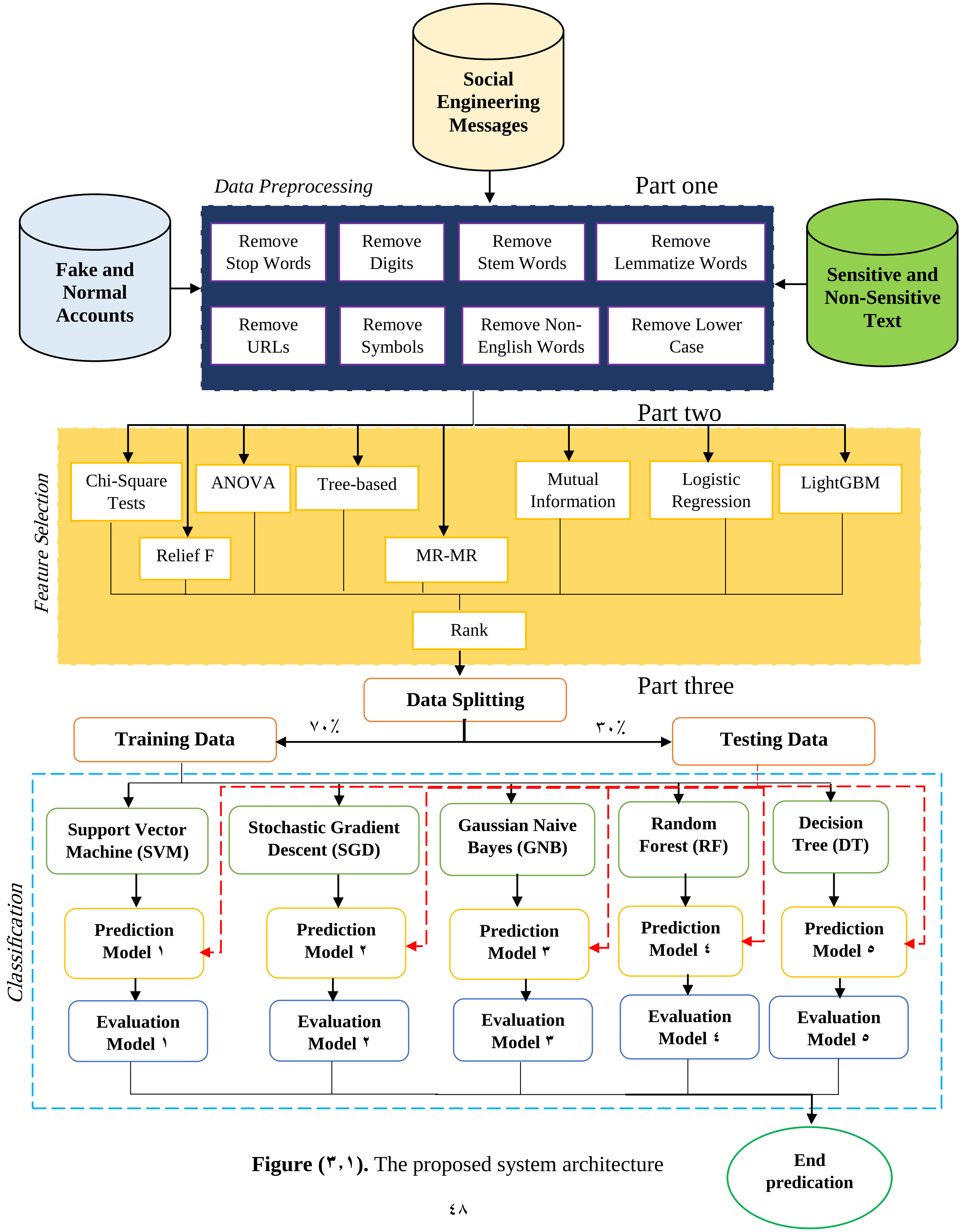


Figure (٣,١). The proposed system architecture

Algorithm (٣,١). The proposed system architecture

Input: Anyone of the three Datasets (Social engineering messages, Fake and normal accounts, and Sensitive and non-sensitive text).

Output: List of predictions

Begin

١: Training Mode Selection Stage

- Select Training mode (Social Engineering detection, Fake Account Detection, Sensitive Text Detection).

٢: Text Cleaning and Data Preprocessing Stage

- Remove stop words
- Removing digits(numbers)
- Remove links
- Stem words
- Lemmatize words
- Remove symbols
- Remove non-English words
- Lower case

٣: Text Feature Extraction Stage

- X = TF-IDF(Txt)
- If Training mode = fake account detection
- If (the chosen dataset is fake account)
- Concatenate (X, statuses_count, friends_count, listed_count, time_zone, profile_use_background_image, profile_background_tile…)

٤: Data Normalization Stage

- Min Max Scalar

٥: Splitting Stage

- Train=Scaler (Train)
- Test=Scaler (Test)

٦: **Best Feature Selection (Ranked Feature)**

- Return Features by Chi-Square Tests using equation (٢,٣)
- Return Features by ANOVA F-value using equation (٢,٤)
- Return Features by Mutual Information using equation (٢,٥)
- Return Features by selecting from model Logistic Regression using equation (٢,٦)
- Return Features by Tree-based using equation (٢,٧)
- Return Features by Relief F using equation (٢,٨)
- Return Features by (Minimum-Redundancy-Maximum-Relevance) using equation (٢,٩)
- Return Features by LightGBM using equation (٢,١٢)
- Return Features Similar in the results of the above methods

٧: **Classification Stage**

- Predication Result (١) = SVM (Features Set, Targets)
- Predication Result (٢) = SGD (Features Set, Targets)
- Predication Result (٣) = GNB (Features Set, Targets)
- Predication Result (٤) = DT (Features Set, Targets)
- Predication Result (٥) = RF (Features Set, Targets)

٨: Best Classifier = High (Accuracy)

End

Now each stage will be explained in more detail as follows:

٣,٢,١. Selecting Training Mode (First Stage)

Three datasets are used as input by the suggested system, as illustrated in Algorithm (٣,١). One of the benefits of this approach is the ability to train on numerous modes while using different datasets on the same code, which results in fewer code steps and faster processing. The user must select one of three alternatives from the proposed system to begin the treatment process in order to conduct one of these three modes. The first option is social engineering detection, the second is fake account detection, and the third is sensitive text detection. The proposed system reads each dataset and satisfies its features separately from one another, with the fake account dataset having both text and additional features while the social engineering and sensitive text datasets just have text features.

٣,٢,٢. Data Preprocessing (Second Stage)

The data is subjected to a number of operations in this step to get it ready for the classification process, including text cleaning procedures such as eliminating stop words, URLs, digits, lower case, stemming words, lemmatizing words, symbols, removing non-English terms, and normalizing data. The steps below summarize it and are also presented in Algorithm (٣,٢):

i. Text Cleaning

The suggested system's initial stage is data cleaning. It preprocessed the data content to remove unnecessary words from document text since they provide no use. There are various phases in this stage as follows: -

a) Remove Stop Words

To remove stop words from a phrase, first, use NLTK to obtain a list of stop words, then convert words to lowercase and split the text into words, and then remove the word if it appears in the list of stop words supplied by NLTK. The

NLTK library, which includes stopping word removal, is one of the oldest and most widely used Python libraries for Natural Language Processing.

b) Removing Digits (Numbers)

Texts typically contain random numbers that must be removed in order to find the correct English word and classify and identify them. Using an iterative loop that passes all the words in the text and is tested using the function of distinguishing numbers, nothing is placed in its place save a separator between the first and second words.

c) Stemming Lemmatizing Words

After removing stop words, use an iterative loop to pass all the words in the text, then find the root word for each one and return the base or dictionary form of a word.

d) Remove Links (URLs)

The links in the texts have been removed because they might be fake links intended to inflict harm on victims by tricking them into visiting phony websites or engaging in other fraudulent activities using these links. As a result, it was dropped and included in the project’s future development plans. Also, by using a "for loop" and an iterative loop that runs through all of the words in the text and checks its pattern (HTTPS: www. ... .com), it will be deleted.

e) Remove Symbols

Remove characters like (! @#$%^&*() _+=) by using the same for loop that’s used to remove links above.

f) Remove Any Non-English Words

Removing all symbols that represent non-English words.

g) Lower Case

Removing all symbols that represent the lower case.

Algorithm (٣,٢). Text Cleaning

```
Input: T is the set of documents
Output: Stop word Removing-List (T), Digits Removing-List (T), Stem-List (T), Links Removing-List (T), Symbols Removing-List (T), Non-English Removing-List (T), and Lower case (T).
Begin
Step ١:   Read the document from the dataset.
Step ٢:   While not (EOF) do:
        • Separate every word from others based on space for obtaining (tokens).
        • Eliminate all symbols token:
          If no. of token > ٠ then
              i= i+١
          End if
          Tokenization-List = Token[i]
        • Return (Tokenization-List T))
        • Put the word in a new list called (stop word Removing-List T)
        • For each word in (stop word Removing-List T) do If word ∈ stop word then
              Remove word
              Else
              Stop word Removing-List (T) = word
```

```
   End for
•  Return (Stop word Removing-List T))
•  Put the word in a new list called (Links Removing-List T)
•  For each text in (Links Removing-List T) do
        If text ∈ Links, Then
            Remove Links
            Else
           Links Removing-List (T) = text
         End if
   End for
 Return (Links Removing-List (T))
        Put the word in a new list called (Stem Removing-List T)
•  For each word in (Stem Removing-List T) do
        If word ∈ Stemming words, Then
            Remove word
            Else
            Stem Removing-List (T) = word
        End if
   End for
•  Return (Stem Removing-List (T))
    Put the word in a new list called (Lower Case Removing-List T)
•  For each word in (Lower Case Removing-List T) do
        If word ∈ Lower Case, Then
            Remove word
            Else
            Lower Case Removing-List (T) = word
            End if
        End for
    •   Return (Lower Case Removing-List (T))
            Put the word in a new list called (Non-English Removing-List T)
    •   For each word in (Non-English Removing-List T) do
            If word $\in$ Non-English, Then
                Remove word
                Else
                Non-English Removing-List (T) = word
            End if
        End for
    •   Return (Non-English Removing-List (T))
    •   Put the word in a new list called (preprocessed-List T)
End while
```

٣,٢,٣. Text Feature Extraction (Third Stage)

Create a feature vector for each preprocessed data set. The document should be converted from full text to a data vector before being represented as an array of features. The TF-IDF model was used to compute the frequency of each word or extract features from the data, and then the normalization procedure was employed.

i. Term Frequency Inverse Document Frequency (TF-IDF)

This process of finding the meaning of sentences made up of words helps the machine to read the words by numbers. The steps of this process are presented in Algorithm (٣,٣).

Algorithm (٣،٣). Calculating of TF-IDF

Input: preprocessed-List T

Output: F is the output file (table of ranking, TC, TF, M, IDF, and TF-IDF),

Function: String Matching (R, D), initialize F= ∅

```
Begin
  For each rule r in the prediction rule base R do
      For each example d ∈ D do
          If r refers to d (R, D)
              Add information to F
          End if
      End for
  End For
End
```

٣،٢،٤. Normalization Data (Fourth Stage)

Normalize Data using Min-Max Normalization preparing data to train and test set to ٠ and ١ in order to be under stainable by classifiers using min-max scale normalization. See the Algorithm (٣.٤) for more information on the min-max normalizing approach.

Algorithm (٣.٤). Min-Max Normalization

Input: ٢D array InP [n * m] in which n is the number of samples and m is the number of features.

Output: ٢D array OutP [n * m] after normalization process.

```
Begin
```

```
Let Min[i] and Max [i] two arrays holding minimum and maximum values
for each feature.
For i = ١ to m
    Find max and min values by setting them to the first value of feature i
    For j = ١ to n
        If  value < min then
            min = value
                else if value > max then
                max = value
                 End if
        End if
    End for
modify every value v in feature vector i using Eq. (٢،٢).
End For
End
```

٣,٢,٥. Feature Selection (Fifth Stage)

Finding the optimal features by integrating the Filter and Wrapper techniques for privacy-preserving feature selection. Eight methods of feature selection were used in this stage, Algorithm (٣،٦) explains the feature selection steps work.

Algorithm (٣.٥). Text Feature Selection

Input: Extracted Features

Output: Best Features

Begin

```
Calculate chi-square tests by using Eq. (٢,٣)
Calculate ANOVA F-value by using Eq. (٢,٤)
Calculate mutual information by using Eq. (٢,٥)
Calculate Logistic Regression by using Eq. (٢.٦)
Calculate Tree-based by using Eq. (٢.٧)
Calculate Relief F by using Eq. (٢.٨).
Calculate (Minimum-Redundancy-Maximum-Relevance) by using Eq. (٢.٩)
Calculate LightGBM. by using Eq. (٢.١٢)
Concatenate all returned features and put them in (all features array)
mode (all features array) (٠)
put final features in (Final Features array)
Return (Final Features)
End
```

٣,٢,٦. Data Splitting (sixth Stage)

It is the process of splitting supplied data into two halves for the purpose of cross-validation. The first set of data is used to build a prediction model, while the second is used to evaluate the model's performance. One way to assuring accurate generalization and prevent overtraining is to use cross-validation processes. The main idea is to divide data set T into two subsets: one for training (٧٠٪ of the dataset) and one for testing (٣٠٪ of the dataset). Cross-major validation seeks to generate a consistent and dependable approximation of the model's outputs. Algorithm (٣,٥) explain this stage in more detail.

Algorithm (٣.٦). Dataset Splitting

Input: Anyone of the three Datasets (Social engineering messages, Fake and

```
     normal accounts, and Sensitive and non-sensitive text) after applying the
     previous stages.
Output: Splitting dataset into a training set and testing set
Begin
  Define sets of model parameter values to evaluate.
      For each parameter set do
         For each resampling iteration do
             Hold-out specific samples
             Fit the model on the remainder
             Predict the hold-out samples
         End for
      Calculate the average performance across hold-out predictions.
      End for
      Determine the optimal parameter set.
      Fit the final model to all the training data utilizing the optimal parameter
      set.
End
```

٣,٢,٧. Classification Stage (Seventh Stage)

The most important phase in the proposed system is the machine learning data classification step. This section discusses the five classifiers employed in this work:

i. Classify data with Support Vector Machine (SVM) Classifier

Taking splatted sets that's result from algorithm (٣.٧) after splitting best features that's result from the eight feature selection methods in algorithm (٣،٦).

Algorithm (٣.٧) presents the working steps of this classifier.

Algorithm (٣،٧). Support Vector Machine (SVM) Classifier

```
Input: The split dataset
Output: Predication accuracy of SVM
Begin
    Select the optimal value of cost and gamma for SVM using Eq. (٢،٧).
    While (the stopping condition is not met) do:
        Implement the SVM train step for each data point.
        Implement SVM classification for testing data points.
    end while
    Return accuracy
End
```

ii. Classify data with Gradient Descent (SGD) Classifier

Taking splatted sets that's result from algorithm (٣،٧) after splitting best features that's result from the eight feature selection methods in algorithm (٣،٦).

Algorithm (٣.٨). Stochastic Gradient Descent Classifier (SGD)

```
Input: The split dataset, Learning Rate $\eta$, Reg. factor $\lambda$, n° of latent factors K,
Randomly initialize matrices P and Q ;
```

```
Output: Predication accuracy of SGD
Begin
    n = ٠
    While not(convergence) do:
    Randomly shuffle observed entries in S;
    For each (u, i) ∈ S then:
        Apply Eq. (٢,٩).
    End for
    n = n+١
    End while
End
```

iii. Classify data with Gaussian Naive Bayes (GNB)

Taking splatted sets that's result from algorithm (٣,٧) after splitting best features that's result from the eight feature selection methods in algorithm (٣,٦).

Algorithm (٣.٩). Gaussian Naïve Bayes (GNB) Classifier

```
Input: The split dataset
Output: Predication accuracy of GNB
Begin
    Splitting the data set training set and testing set
    P($C_i$) = Calculate the frequency and the probability for each class (fake or
real, sensitive or not  and attack or not) in the training dataset, utilizing Eq.
(٢,١٠).
    P($F_i$) = Calculate the frequency and the probability for each feature
        in all datasets depending on the class label, utilizing Eq. (٢,١٠).
```

```
For each $F_i$ in Feature Set
    $TF_i$= get the probability value from the training set for the feature of ($F_i$)
    $TF_i$ (C) = get the probability for (C )
    R ($C_i$) = R ($C_i$) * $TF_i$ ($C_i$)* P($C_i$)
Next $F_i$
Class = max R(C)
Enf for
End
```

iv. Classify data with Random Forest (RF) Classifier

v. Taking splatted sets that's result from algorithm (٣،٧) after splitting best features that's result from the eight feature selection methods in algorithm (٣،٦).

Algorithm (٣.١٠). Random Forest (RF) Classifier

```
Input: The split dataset
Output: Predication accuracy of RF
Begin
    Draw $n_{tree}$ bootstrap samples from the original data
    For each of the samples to generate an unpruned classification tree
     make the following changes:
            Rather than selecting the best split among all predictors, at each
            node, randomly sample n predictors and select the best split from
            among those variables.
    End for
```

```
٣:  Predict new data by aggregating the predictions of the $\boldsymbol{n_{tree}}$
End
```

vi. Classify data with Decision Tree Classifier (DT)

Taking splatted sets that’s result from algorithm (٣،٧) after splitting best features that’s result from the eight feature selection methods in algorithm (٣،٦).

Algorithm (٣.١١). Decision Tree Classifier (DT)

```
Input: The split dataset
Output: Predication accuracy of Decision Tree
Begin
١:  If stopping condition (S, F) = true then
        • Leaf = create Node
        • Leaf Label = classify(s)
        • Return Leaf
٢:  root = create Node
٣:  root test condition =find Best Spilt (S, F)
٤:  V= {v |v a possible outcomecfroot.test condition}
٥:  For each value v ∈V:
          • $\boldsymbol{S_v}$= {s | root test condition(s) = v and s ∈ S;
          • Child = Tree Growth ($\boldsymbol{S_v}$, F);
          • Add child as the descent of root and label the edge {root→child} as
            v.
        End For
٦:  Return (root)
End
```

Chapter Four

Experimental Results and Discussion

Chapter Four

Experimental Results and Discussion

٤,٧. Introduction

The research goals shown in chapter one were covered by the suggested system shown in chapter three. The first section includes a brief description of the three datasets used, along with the study needs. The experimental results are then presented and discussed in this chapter.

٤,٨. Software and Hardware Requirements

The proposed system requires the following hardware and software to achieve its aim:

٤,٨,١. Software Requirements

The Python ٣,٦ programming environment and PyCharm ٢٠٢٠ IDE were used to implement the suggested system. It runs on Microsoft Win١٠ Pro ٦٤-bits.

٤,٨,٢. Hardware Requirements

This work is implemented on a Lenovo computer with the following characteristics:

- CPU Intel Core i٧ ٩th gen ٦ Core
- GPU NVIDEA RTX ٢٠٧٠ ٦GB
- RAM ١٦ ٢٦٠٠ MHz.

- Hard Samsung m.٢ NVME

٤,٩. Datasets Description

This section discusses the datasets utilized by the proposed technique, which are three datasets: one for detecting social engineering words, another for detecting fake accounts, and a third for detecting sensitive content. In this study, the Institute of Informatics and Telematics of the Italian National Research Council (IIT-CNR) released the twitter dataset as described below with example.

٤,٩,١. Social Engineering detection

Dataset for detecting social engineering sentences: Here are some samples of data used in the proposed system’s operation.

Table (٤,١). Social engineering sentences dataset sample

Message	Class
Dear how is chechi. Did you talk to her	٠
This is Tim rose with rose industries. I am a security officer. I need you to write me your password so I can make sure it is secure. thanks bye.	١

٤,٩,٢. The Fake Accounts dataset

There are two types of Twitter accounts in the dataset: fake accounts, and real accounts. Were the dataset containing ٤١٥٥ Twitter accounts with ١٣٣٧ fake accounts and ٢٨١٨ real accounts. Table (٤.٢) shows attributes and their description.

Table (٤.٢). The dataset description. It contains ٤١٥٥ accounts, ٢٨٪ fake accounts, and ٣٠٪ trust accounts

Attributes	Description
Statuses count	The number of statuses posted from the account
Followers count	The number of followers for the account
Friends count	The number of friends for the account
Favorites count	the number of tweets that given user has marked as favorite
Listed count	The number of groups the account belongs to
Language	The language for the account
Time zone	The time zone of the account holder
Geo enabled	Twitter enables users to specify a location for individual Tweets
Profile use background image	Is this profile use background image
Profile text color	Color code of text that’s displayed to user
Profile sidebar border color	Color code of sidebar that’s displayed to user

Profile background tile	Is this profile use the background for Twitter profile
Profile sidebar fill color	Color code of sidebar fill that’s displayed to user
Profile background color	Color code of the background for that profile that’s displayed to user
Profile link color	Color code of profile link that’s displayed to user
UTC –offset	The UTC-offset given time zone
Description	The description that user used on his profile

٤,٩,٣. Sensitive or Not Sensitive Speech

Sensitive text detection: The text in this dataset, which comprises of ١٣٩٧ sentences, is classed as sensitive text or non-sensitive text (sensitive and non-sensitive). Many of these words are written by individuals and represent sensitive text that may be tweeted, and someone should be aware and think twice before publishing them, since this text represents sensitive information that may be used to damage the user, and it is critical to make the user aware of them. An example from this data shown below in table (٤.٣):

Table (٤.٣). Sensitive dataset sample

Text	Class
My aunt’s cell phone number is ٠٧٧٠٥٦٤٤٢٣	Sensitive
Might cry at the harry styles concert if he sings kiwi	Non-sensitive

٤,١٠. Implementation of the Proposed System

The system is implemented utilizing three datasets, and the implementation outcomes for each dataset depending on the metrics given in chapter two are shown in this section:

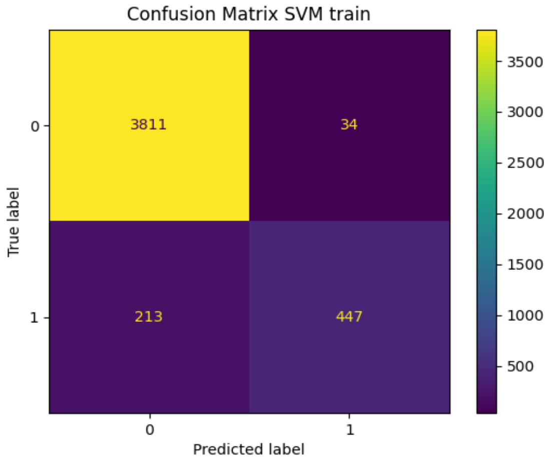
٤,١٠,١. Classification Results on Social Engineering Detection “Dataset ١”

In this section, the classification results for the first dataset are presented using the five machine learning classifiers in both the training and testing stage.

i. Training Phase “Dataset ١”

The results of the five classifiers in this stage are shown in figures from (٤,١) to (٤.١٠).

Support Vector Machine (SVM)



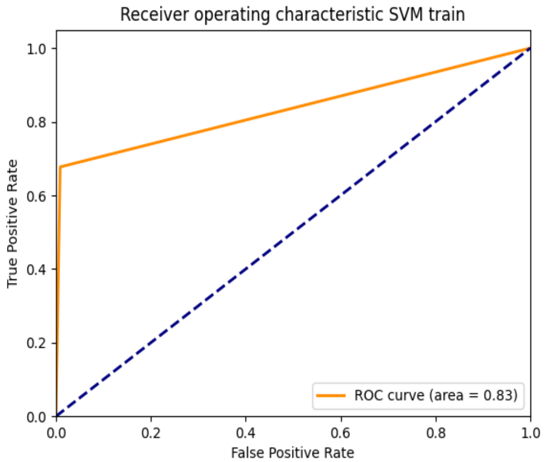


Figure (٤,١). Confusion matrix of SVM in training stage “dataset ١” where the square with the number ٣٨١١ represents the

Figure (٤.٢). ROC curve of SVM in training stage “dataset ١” and showing the performance of this classifier in classifying the correct number of correctly predicted class values and the square with the number ٣٤ represents the number of incorrectly predicted class values and the square with the number ٢١٣ represents the number of incorrectly predicted no-class values and the square with the number ٤٤٧ represents the number of correctly predicted no-class values.

predicted class and incorrect predicted class and the result is ٠,٨٣ which is above the threshold (٠,٥)

Stochastic Gradient Descent Classifier (SGD)

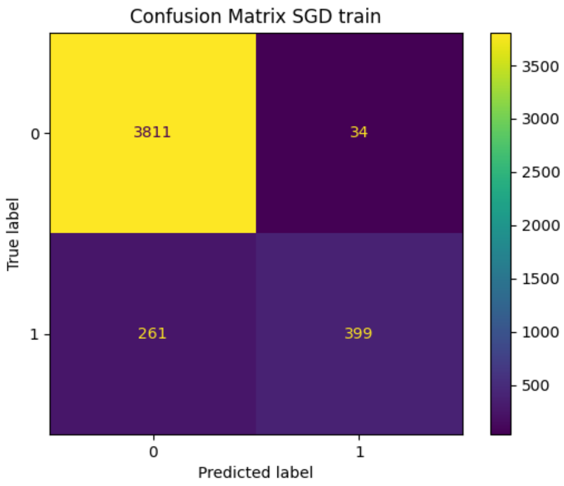


Figure (٤,٣). Confusion matrix of SGD in training stage “dataset ١”

where the square with the number ٣٨١١ represents the number of correctly predicted class values and the square with

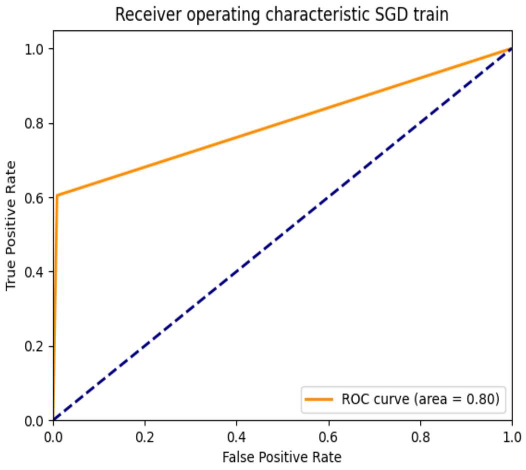
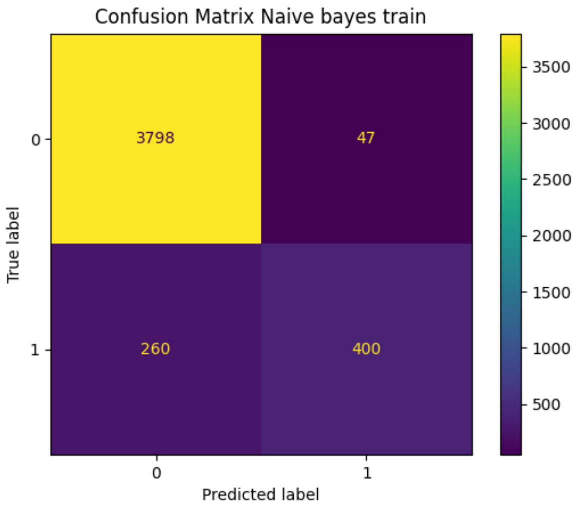


Figure (٤,٤). ROC curve of SGD in training phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class the number ٣٤ represents the number of incorrectly predicted class values and the square with the number ٢٦١ represents the number of incorrectly predicted no-class values and the square with the number ٣٩٩ represents the number of correctly predicted no-class values.

and the result is ٠,٨٠ which is above the threshold (٠,٥)

Gaussian Naïve Bayes (GNB)



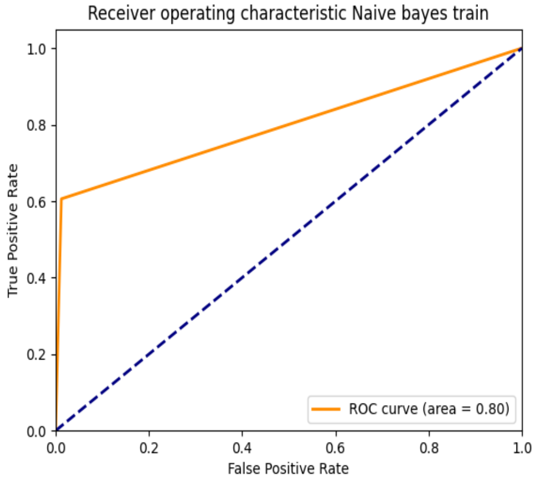


Figure (٤,٥). Confusion matrix of GNB in training phase “dataset ١” where the square with the number ٣٧٩٨ represents the number of correctly predicted class values and the square with the number ٤٧ represents the number of incorrectly predicted class values and the square with

Figure (٤,٦). ROC curve of GNB in training phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠.٨٠ which is above the threshold (٠,٥) the number ٢٦٠ represents the number of incorrectly predicted no-class values and the square with the number ٤٠٠ represents the number of correctly predicted no-class values.

Random Forest (RF)

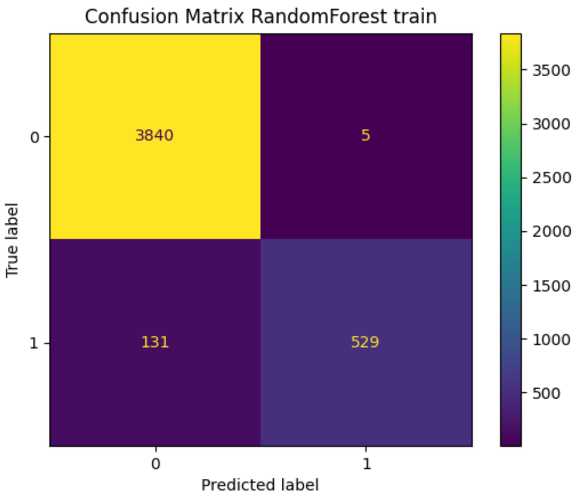


Figure (٤,٧). Confusion matrix of RF in training phase “dataset ١”

where the square with the number ٣٨٤٠ represents the number of correctly predicted class values and the square with the number ٥ represents the number of incorrectly predicted class values and the

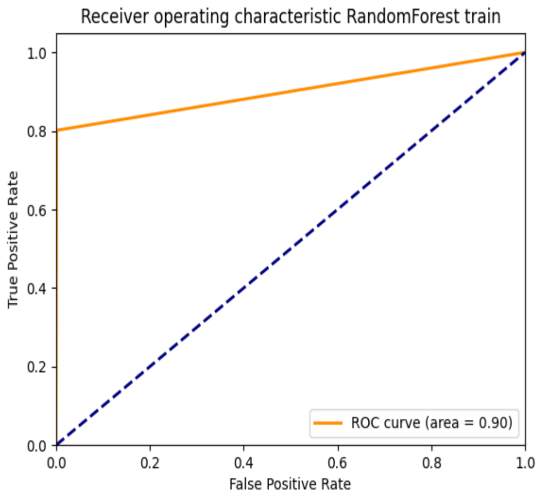


Figure (٤,٨). ROC curve of RF in training phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠.٩٠ which is above the threshold (٠,٥) square with the number ١٣١ represents the number of incorrectly predicted no-class values and the square with the number ٥٢٩ represents the number of correctly predicted no-class values.

Decision Tree Classifier (DT)

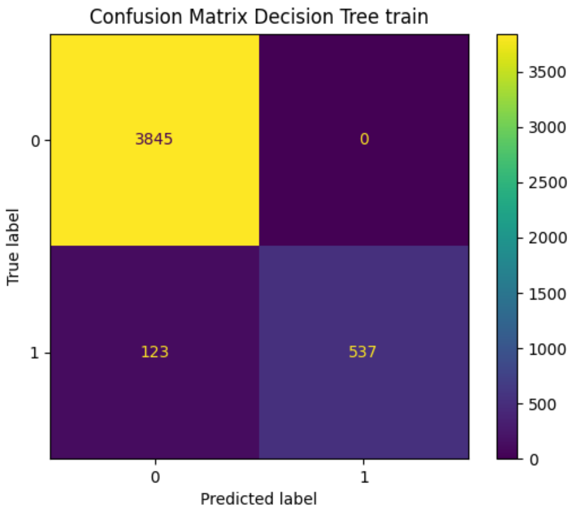


Figure (٤,٩). Confusion matrix of DT in training phase "dataset ١"

where the square with the number ٣٨٤٥ represents the number of correctly predicted class values and the square with the number ٠ represents the number of incorrectly predicted class values and the

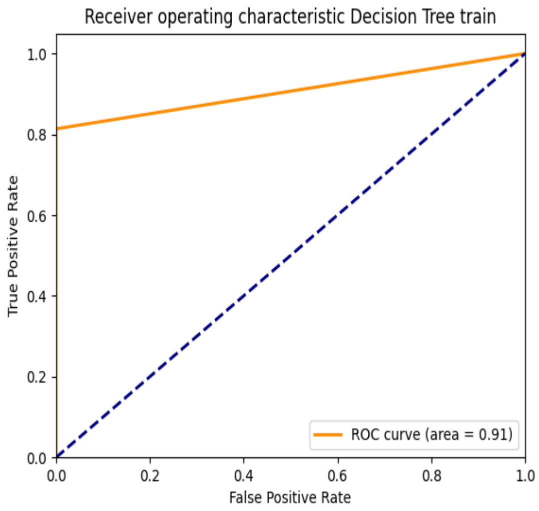


Figure (٤,١٠). ROC curve of DT in training phase "dataset ١" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩١ which is above the threshold (٠,٥) square with the number ١٢٣ represents the number of incorrectly predicted no-class values and the square with the number ٥٣٧ represents the number of correctly predicted no-class values.

Table (٤.٤) explain the results of the measurements for the training phase of the five classifiers on “dataset ١”.

Table (٤.٤). Results of training phase on “dataset ١”

	Accuracy	**Precision**	**Recall**	**F-score**	**Specificity**
SVM	٩٤%	٩٩%	٩٤%	٩٦%	٩٢%
SGD	٩٣%	٩٩%	٩٣%	٩٥%	٩٢%
GNB	٩٣%	٩٨%	٩٣%	٩٥%	٨٩%
RF	٩٦%	٩٩%	٩٦%	٩٨%	٩٩%
DT	٩٧%	١٠٠%	٩٧%	٩٩%	١٠٠%

ii. Testing phase “dataset ١”

The results of the five classifiers in this stage are shown in figures from (٤.١١) to (٤.٢٠).

Support Vector Machine (SVM)

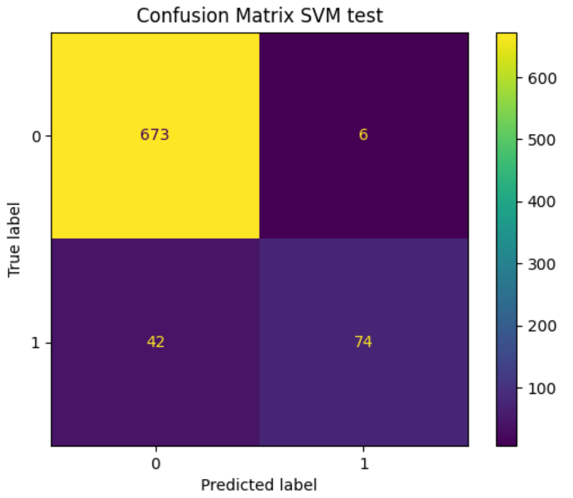


Figure (٤.١١). Confusion matrix of SVM in testing phase “dataset ١”

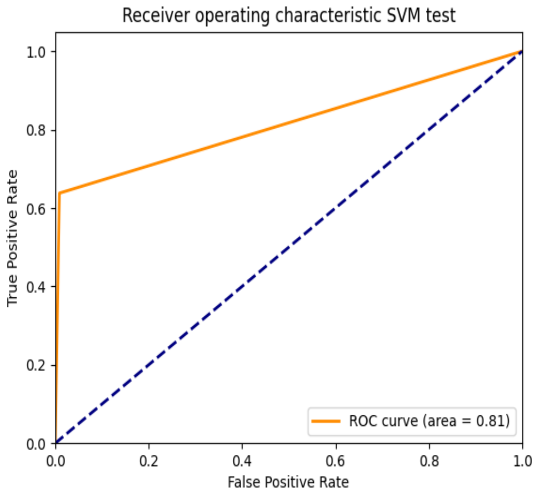


Figure (٤.١٢). ROC curve of SVM in testing phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٨١ which is above the threshold (٠,٥)

where the square with the number ٦٧٣ represents the number of correctly predicted class values and the square with the number ٦ represents the number of incorrectly predicted class values and the square with the number ٤٢ represents the number of incorrectly predicted no-class values and the square with the number ٧٤ represents the number of correctly predicted no-class values.

Stochastic Gradient Descent Classifier (SGD)

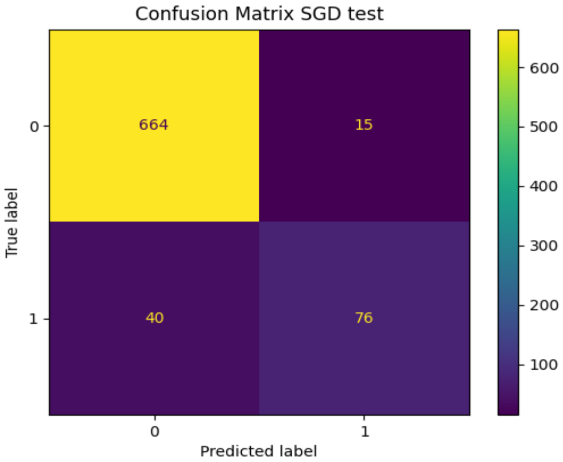


Figure (٤.١٣). Confusion matrix of SGD in testing phase "dataset ١"

where the square with the number ٦٦٤ represents the number of correctly predicted class values and the square with the number ١٥ represents the number of incorrectly predicted class values and the square with the number ٤٠ represents the number of incorrectly predicted no-class values and the square with the number ٧٦ represents the number of correctly predicted no-class values.

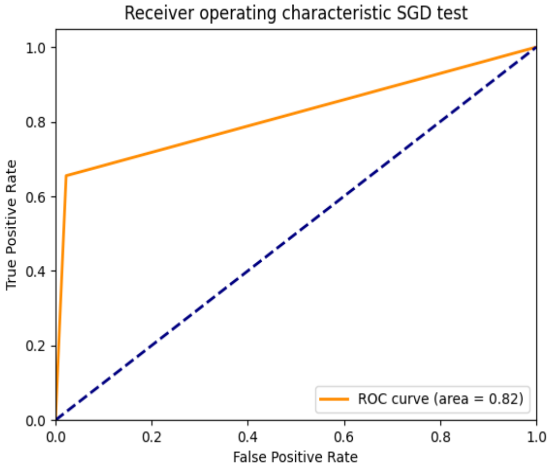


Figure (٤.١٤). ROC curve of SGD in testing phase "dataset ١" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٨٢ which is above the threshold (٠,٥)

Gaussian Naïve Bayes (GNB)

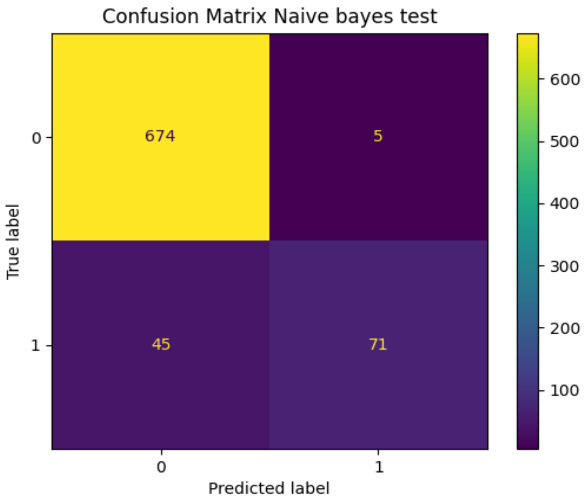


Figure (٤.١٥). Confusion matrix of GNB in testing phase “dataset ١” where the square with the number ٦٧٤ represents the number of correctly predicted class values and the square with the number ٥ represents the number of incorrectly predicted class values and the square with the number ٤٥ represents the number of incorrectly predicted no-class values and the square with the number ٧١ represents the number of correctly predicted no-class values.

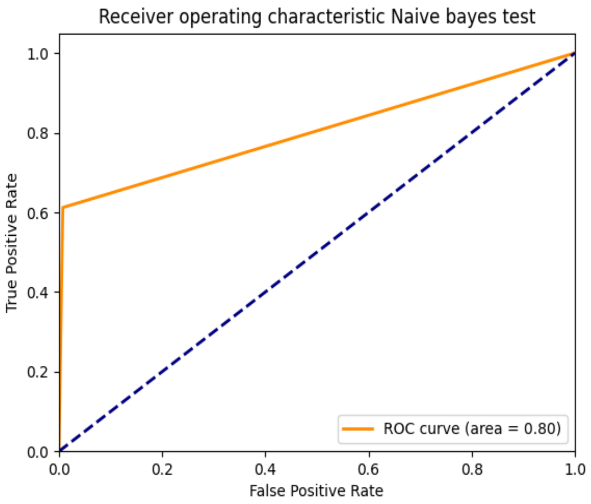


Figure (٤.١٦). ROC curve of GNB in testing phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٨٠ which is above the threshold (٠,٥)

Random Forest (RF)

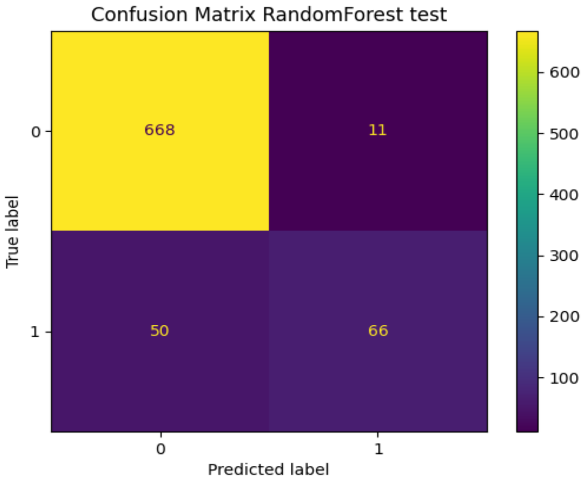


Figure (٤.١٧). Confusion matrix of RF in testing phase “dataset ١” where the square with the number ٦٦٨ represents the number of correctly predicted class values and the square with the number ١١ represents the number of incorrectly predicted class values and the square with the number ٥٠ represents the number of incorrectly predicted no-class values and the square with the number ٦٦ represents the number of correctly predicted no-class values.

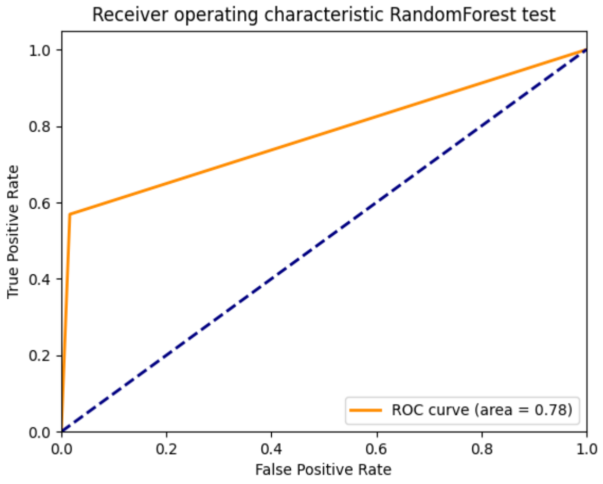


Figure (٤.١٨). ROC curve of RF in testing phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٧٨ which is above the threshold (٠,٥)

Decision Tree Classifier (DT)

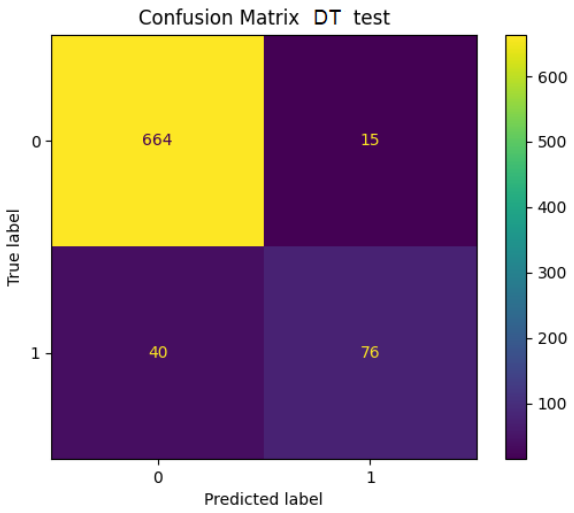


Figure (٤.١٩). Confusion matrix of DT in testing phase “dataset ١”

where the square with the number ٦٦٤ represents the number of correctly predicted class values and the square with the number ١٥ represents the number of incorrectly predicted class values and the square with the number ٤٠ represents the number of incorrectly predicted no-class values and the square with the number ٧٦ represents the number of correctly predicted no-class values.

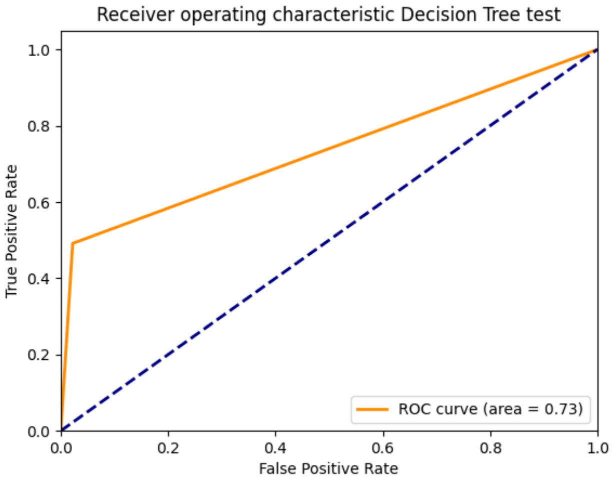


Figure (٤.٢٠). ROC curve of DT in testing phase “dataset ١” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٧٣ which is above the threshold (٠,٥)

Table (٤.٥) explain the results of the measurements for the testing phase of the five classifiers on “dataset ١”.

Table (٤.٥). Results of testing phase on “dataset ١”

	Accuracy	Precision	Recall	F-score	Specificity
SVM	٩٣%	٩٩%	٩٣%	٩٥%	٩٢%
SGD	٩٣%	٩٧%	٩٣%	٩٦%	٨٣%
GNB	٩٣%	٩٩%	٩٣%	٩٥%	٩٣%
RF	٩٢%	٩٨%	٩٢%	٩٤%	٨٥%
DT	٩٣%	٩٧%	٩٣%	٩٥%	٨٣%

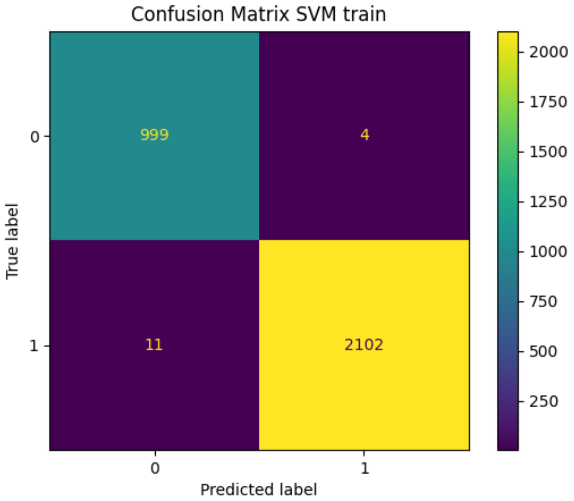
٤,١٠,٢.Classification Results on Fake Account “Dataset ٢”

In this section, the classification results for the second dataset are presented using the five machine learning classifiers in both the training and testing stage.

i. Training phase “dataset ٢”

The results of the five classifiers in this stage are shown in figures from (٤.٢١) to (٤.٣٠).

Support Vector Machine (SVM)



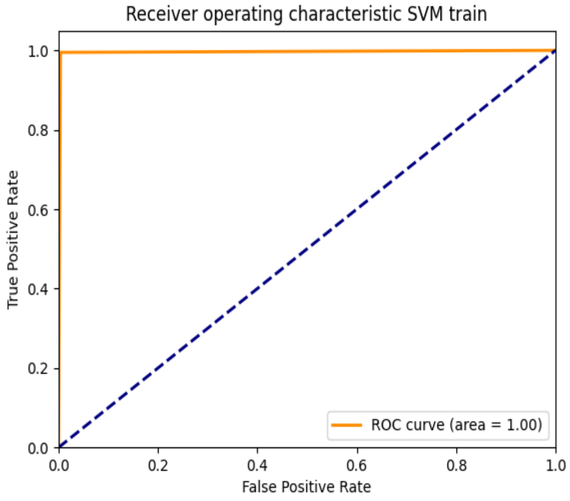


Figure (٤.٢١). Confusion matrix of SVM in training stage “dataset ٢”

where the square with the number ٩٩٩ represents the number of correctly predicted class values and the square with the number ٤ represents the number of incorrectly predicted class values and the square with the number ١١ represents the number of incorrectly predicted no-class values and the square with the number ٢١٠٢ represents the number of correctly predicted no-class values.

Figure (٤.٢٢). ROC curve of SVM in training stage “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Stochastic Gradient Descent Classifier (SGD)

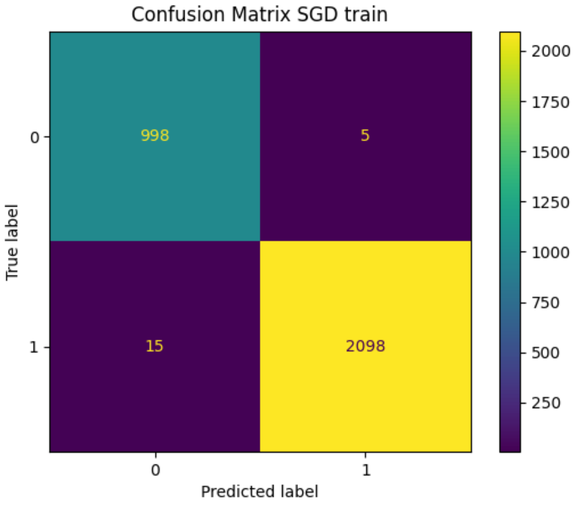


Figure (٤.٢٣). Confusion matrix of SGD in training stage “dataset ٢”

where the square with the number ٩٩٨ represents the number of correctly predicted class values and the square with the number ٥ represents the number of incorrectly predicted class values and the square with the number ١٥ represents the number of incorrectly predicted no-class values and the square with the number ٢٠٩٨ represents the number of correctly predicted no-class values.

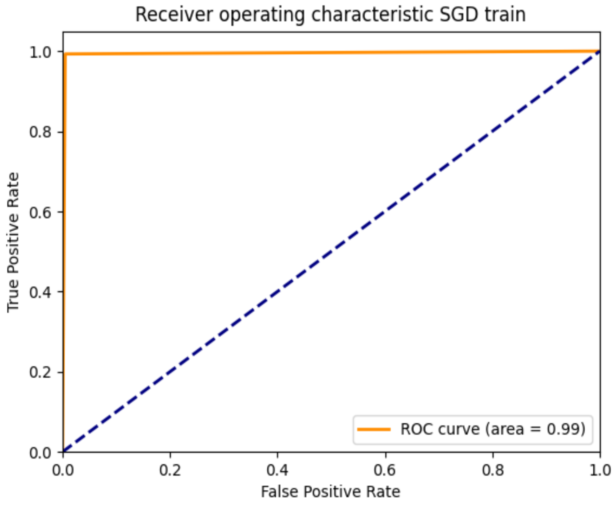


Figure (٤.٢٤). ROC curve of SGD in training phase “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥)

Gaussian Naïve Bayes (GNB)

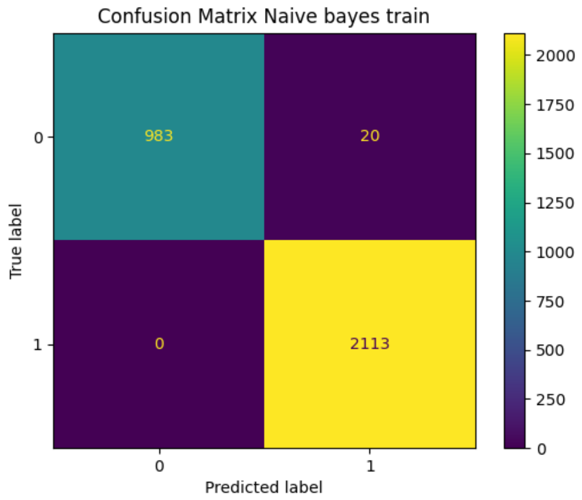


Figure (٤.٢٥). Confusion matrix of GNB in training phase “dataset ٢”

where the square with the number ٩٨٣ represents the number of correctly predicted class values and the square with the number ٢٠ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٢١١٣ represents the number of correctly predicted no-class values.

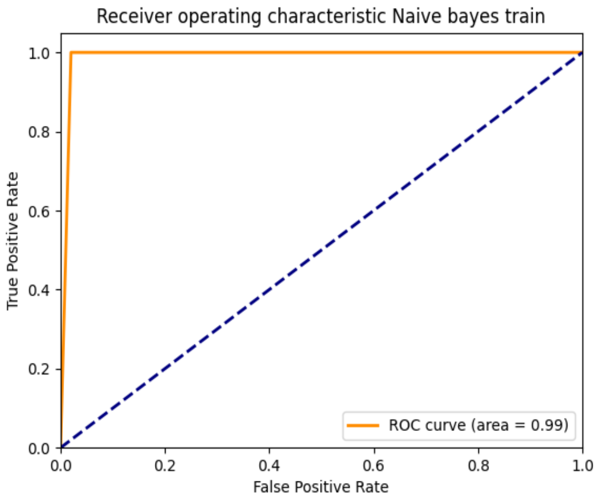


Figure (٤.٢٦). ROC curve of GNB in training phase “dataset٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥)

Random Forest (RF)

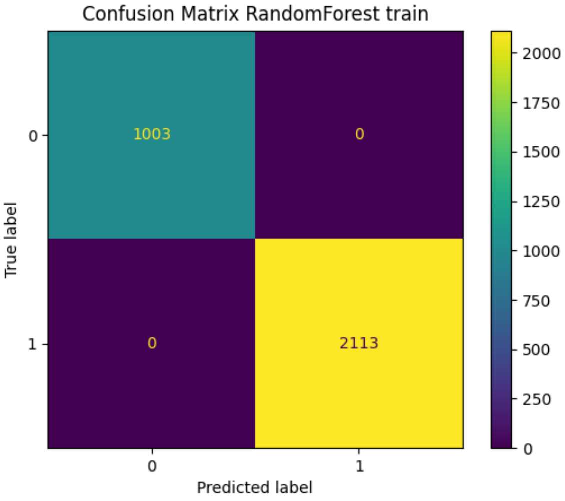


Figure (٤.٢٧). Confusion matrix of RF in training phase “dataset ٢”

where the square with the number ١٠٠٣ represents the number of correctly predicted class values and the square with the number ٠ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٢١١٣ represents the number of correctly predicted no-class values.

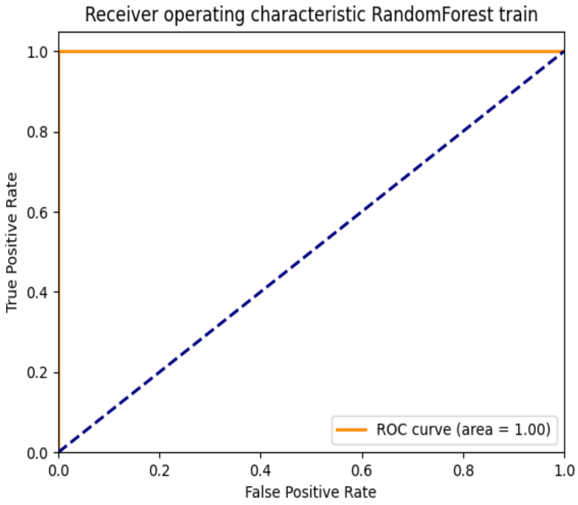


Figure (٤.٢٨). ROC curve of RF in training phase “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Decision Tree Classifier (DT)

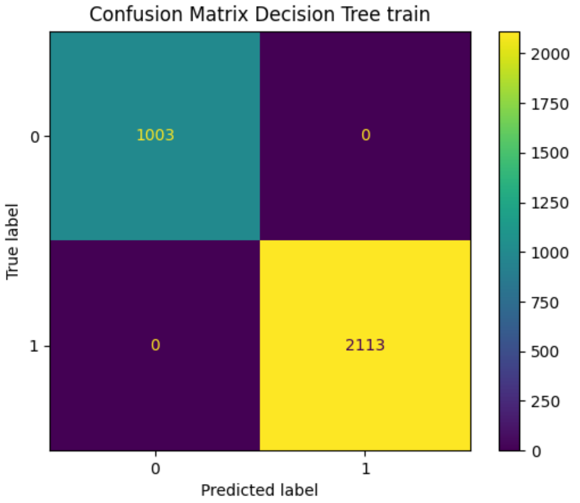


Figure (٤.٢٩). Confusion matrix of DT in training phase "dataset ٢" where the square with the number ١٠٠٣ represents the number of correctly predicted class values and the square with the number ٠ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٢١١٣ represents the number of correctly predicted no-class values.

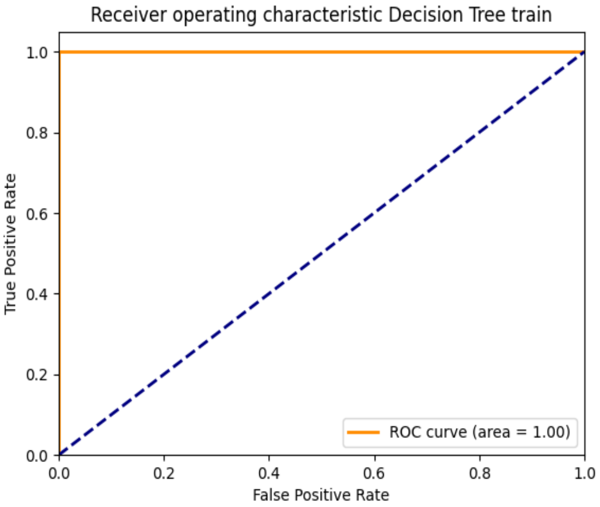


Figure (٤.٣٠). ROC curve of DT in training phase "dataset ٢" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Table (٤.٦) explains the results of the measurements for the training phase of the five classifiers on "dataset ٢".

Table (٤.٦). Results of training phase on “dataset ٢”

	Accuracy	Precision	Recall	F-score	Specificity
SVM	٩٩,٥%	٩٩,٨%	٩٩,٥%	٩٩,٨%	٩٩,٨%
SGD	٩٩,٣%	٩٩,٥%	٩٨,٥%	٩٨,٩%	٩٩,٧%
GNB	٩٩,٣%	٩٨%	١٠٠%	٩٩,٢%	٩٩%
RF	١٠٠٪	١٠٠٪	١٠٠٪	١٠٠٪	١٠٠٪
DT	١٠٠٪	١٠٠٪	١٠٠٪	١٠٠٪	١٠٠٪

ii. Testing phase “dataset ٢”

The results of the five classifiers in this stage are shown in figures from (٤.٣١) to (٤.٤٠).

Support Vector Machine (SVM)

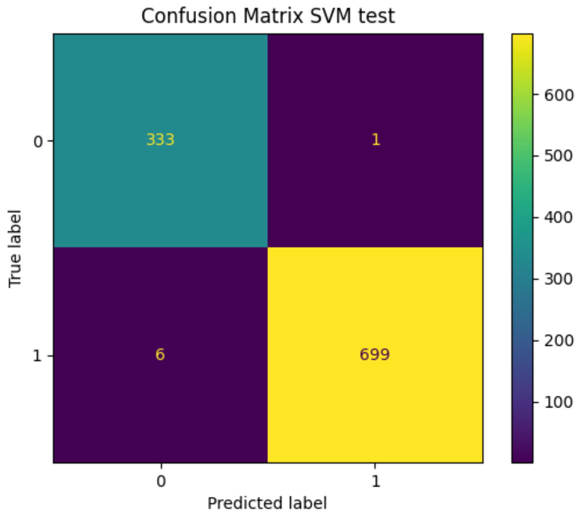


Figure (٤.٣١). Confusion matrix of SVM in testing phase “dataset ٢”

where the square with the number ٣٣٣ represents the number of correctly predicted class values and the square with

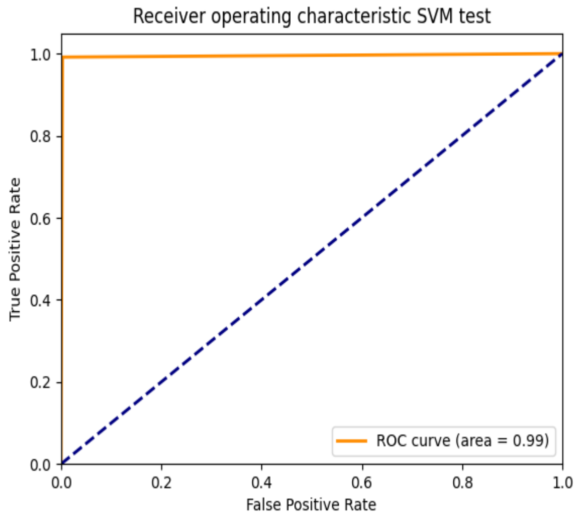


Figure (٤.٣٢). ROC curve of SVM in testing phase “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥) the number ١ represents the number of incorrectly predicted class values and the square with the number ٦ represents the number of incorrectly predicted no-class values and the square with the number ٦٩٩ represents the number of correctly predicted no-class values.

Stochastic Gradient Descent Classifier (SGD)

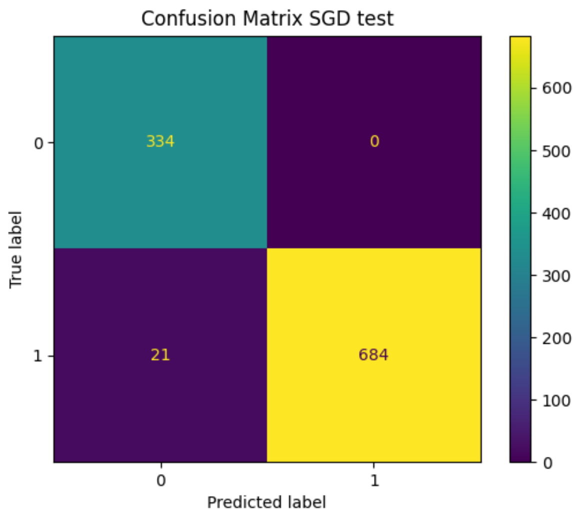


Figure (٤.٣٣). Confusion matrix of SGD in testing phase "dataset ٢"

where the square with the number ٣٣٤ represents the number of correctly predicted class values and the square with the number ٠ represents the number of incorrectly predicted class values and the square with the number ٢١ represents the number of incorrectly predicted no-class values and the square with the

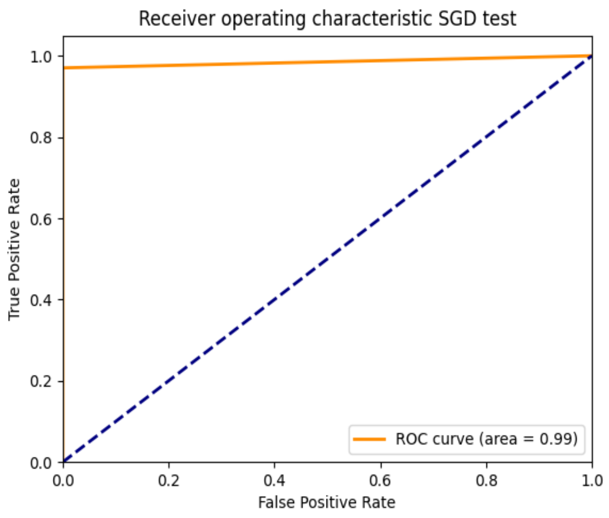


Figure (٤.٣٤). ROC curve of SGD in testing phase "dataset ٢" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥) number ٦٨٤ represents the number of correctly predicted no-class values.

Gaussian Naïve Bayes (GNB)

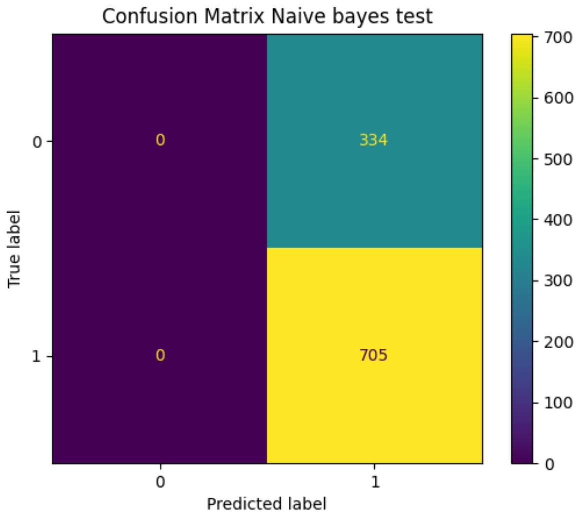


Figure (٤.٣٥). Confusion matrix of GNB in testing phase “dataset ٢”

where the square with the number ٠ represents the number of correctly predicted class values and the square with the number ٣٣٤ represents the number of incorrectly predicted class values and

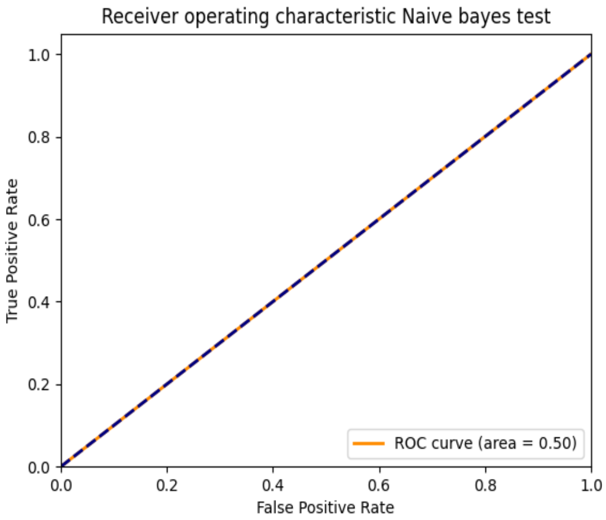


Figure (٤.٣٦). ROC curve of GNB in testing phase “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٥٠ which is equal to the threshold (٠,٥) the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٧٠٥ represents the number of correctly predicted no-class values.

Random Forest (RF)

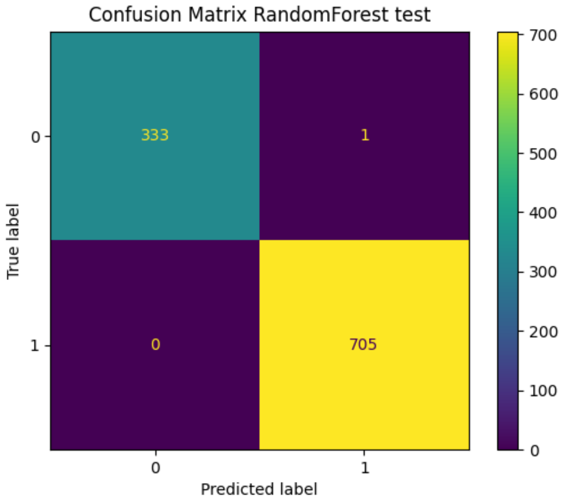


Figure (٤.٣٧). Confusion matrix of RF in testing phase “dataset ٢”

where the square with the number ٣٣٣ represents the number of correctly predicted class values and the square with the number ١ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٧٠٥ represents the number of correctly predicted no-class values.

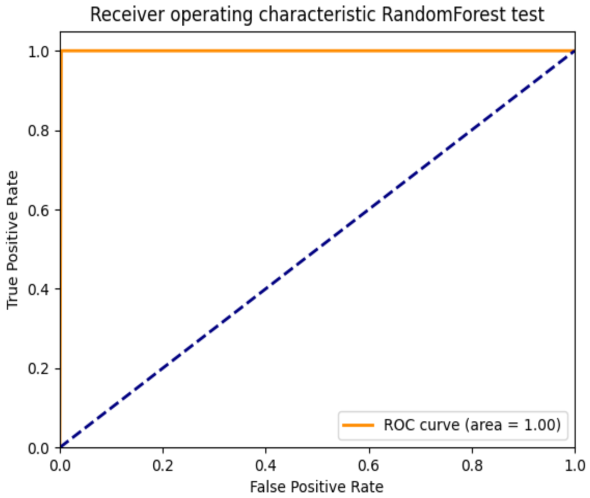


Figure (٤.٣٨). ROC curve of RF in testing phase “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Decision Tree Classifier (DT)

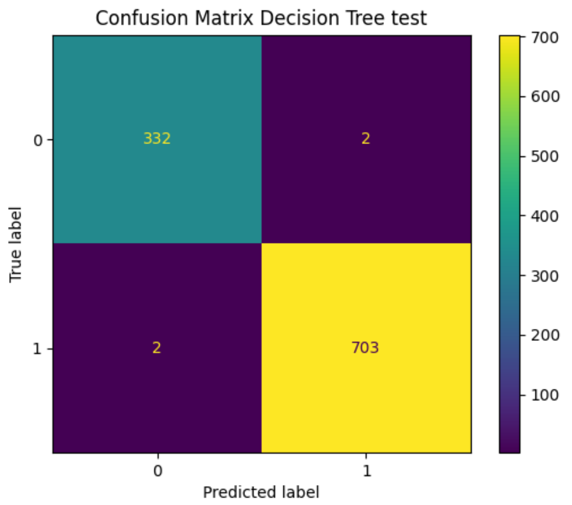


Figure (٤.٣٩). Confusion matrix of DT in testing phase “dataset ٢”

where the square with the number ٣٣٢ represents the number of correctly predicted class values and the square with the number ٢ represents the number of incorrectly predicted class values and the square with the number ٢ represents the number of incorrectly predicted no-class values and the square with the number ٧٠٣ represents the number of correctly predicted no-class values.

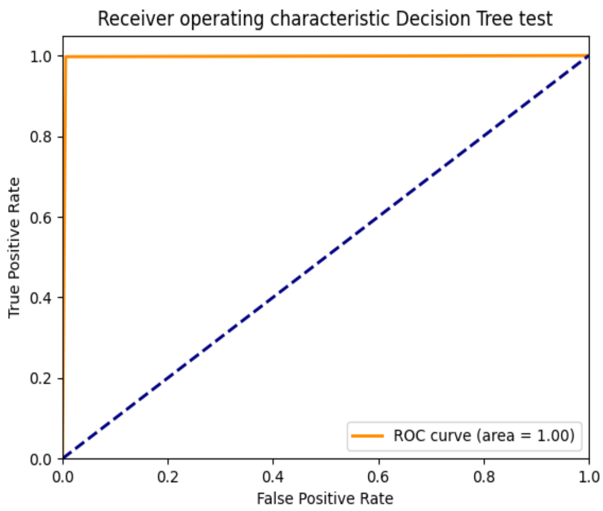


Figure (٤.٤٠). ROC curve of DT in testing phase “dataset ٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Table (٤.٧) explains the results of the measurements for the testing phase of the five classifiers on “dataset ٢”.

Table (٤.٧). Results of testing phase on “dataset ٢”

	Accuracy	Precision	Recall	F-score	Specificity
SVM	٩٩,٣%	٩٩,٧%	٩٨,٢%	٩٩,٣%	٩٩,٨%
SGD	٩٧,٩%	١٠٠%	٩٤%	٩٧,٩%	١٠٠%
GNB	٦٧,٨%	٦٧,٩%	٦٧,٩%	٦٧,٧%	٦٧,٨%
RF	٩٩,٩%	٩٩,٧%	١٠٠%	٩٩,٨%	١٠٠%
DT	٩٩,٦%	٩٩,٤%	٩٩,٤%	٩٩%	٩٩,٧%

٤,١٠,٣. Classification Results on Sensitive “Dataset ٣”

In this section, the classification results for the third dataset are presented using the five machine learning classifiers in both the training and testing stage.

i. Training phase “dataset ٣”

The results of the five classifiers in this stage are shown in figures from (٤.٤١) to (٤.٥٠).

Support Vector Machine (SVM)

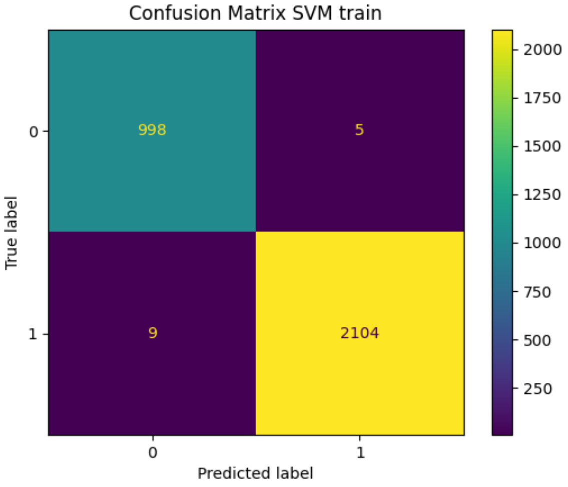


Figure (٤.٤١). Confusion matrix of SVM in training stage “dataset ٣”

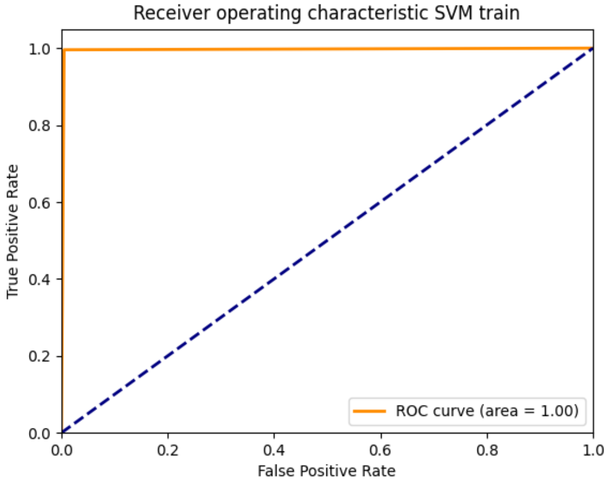


Figure (٤.٤٢). ROC curve of SVM in training stage “dataset ٣” and showing the performance where the square with the number ٩٩٨ represents the number of correctly predicted class values and the square with the number ٥ represents the number of incorrectly predicted class values and the square with the number ٩ represents the number of incorrectly predicted no-class values and the square with the number ٢١٠٤ represents the number of correctly predicted no-class values.

of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Stochastic Gradient Descent Classifier (SGD)

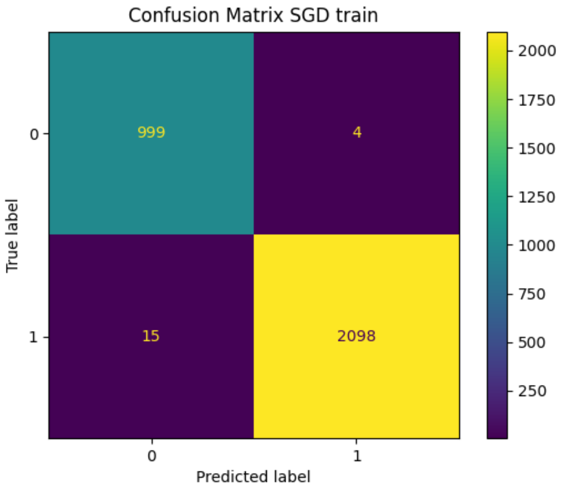


Figure (٤.٤٣). Confusion matrix of SGD in training stage “dataset ٣”

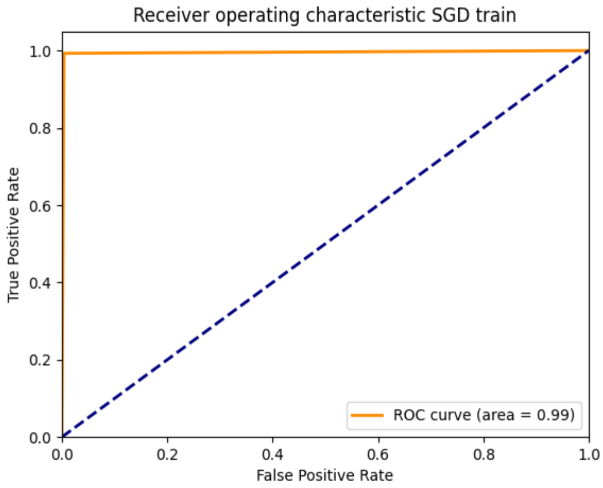


Figure (٤.٤٤). ROC curve of SGD in training phase “dataset ٣” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥)

where the square with the number ٩٩٩ represents the number of correctly predicted class values and the square with the number ٤ represents the number of incorrectly predicted class values and the square with the number ١٥ represents the number of incorrectly predicted no-class values and the square with the number ٢٠٩٨ represents the number of correctly predicted no-class values.

Gaussian Naïve Bayes (GNB)

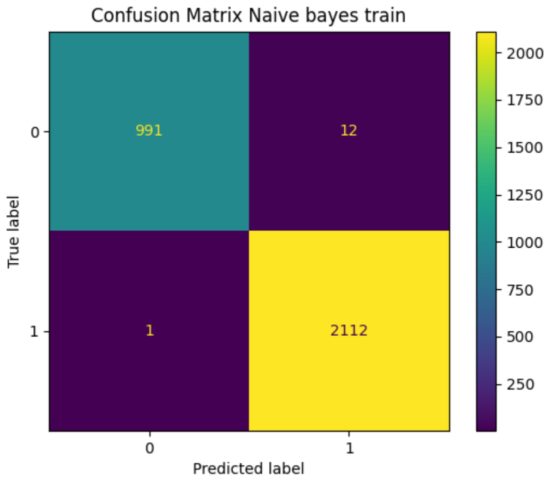


Figure (٤.٤٥). Confusion matrix of GNB in training phase “dataset ٢”

where the square with the number ٩٩١ represents the number of correctly predicted class values and the square with the number ١٢ represents the number of incorrectly predicted class

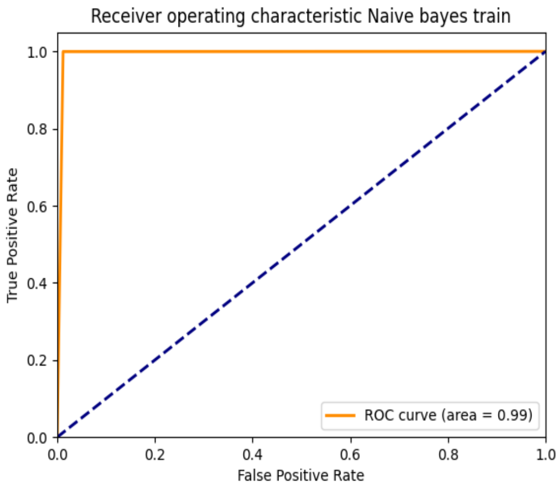


Figure (٤.٤٦). ROC curve of GNB in training phase “dataset٢” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥) values and the square with the number ١ represents the number of incorrectly predicted no-class values and the square with the number ٢١١٢ represents the number of correctly predicted no-class values.

Random Forest (RF)

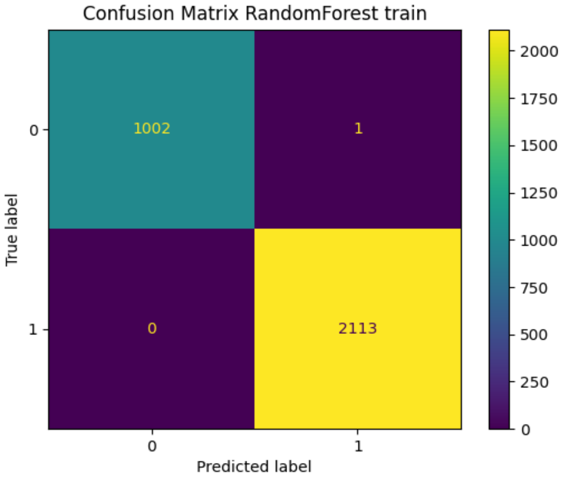


Figure (٤.٤٧). Confusion matrix of RF in training phase "dataset ٣"

where the square with the number ١٠٠٢ represents the number of correctly predicted class values and the square with the number ١ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٢١١٣ represents the

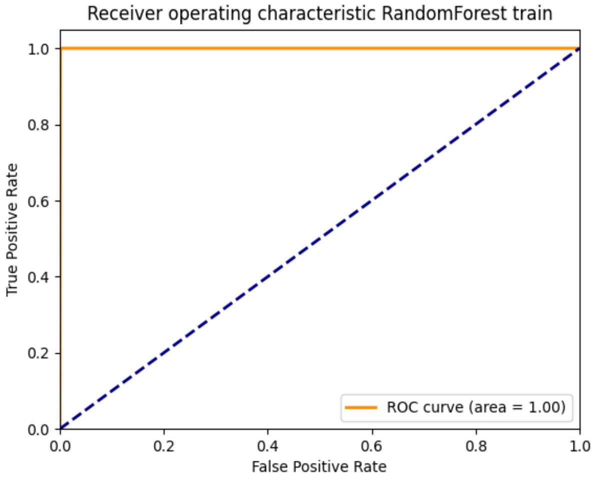
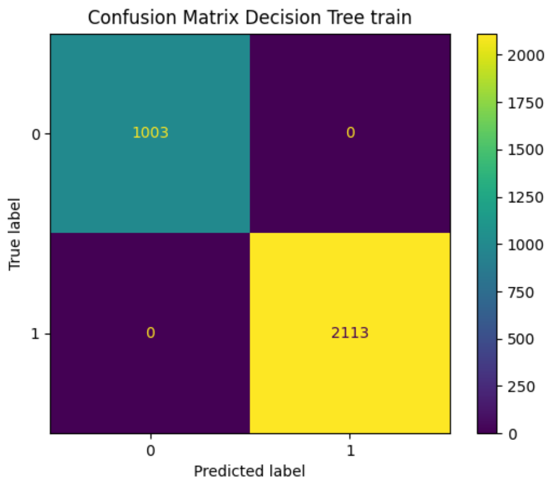


Figure (٤.٤٨). ROC curve of RF in training phase "dataset ٣" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥) number of correctly predicted no-class values.

Decision Tree Classifier (DT)



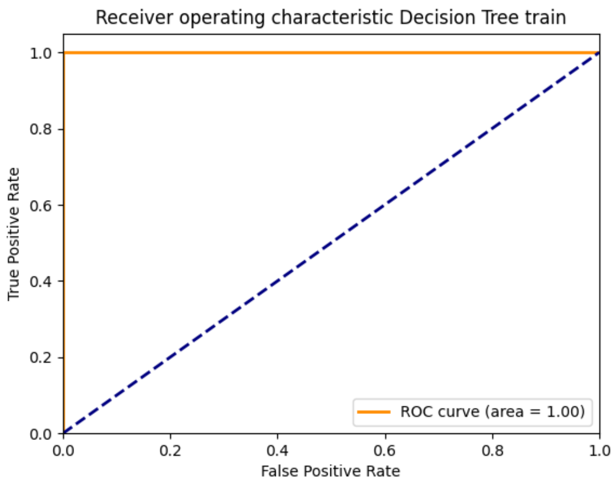


Figure (٤.٤٩). Confusion matrix of DT in training phase “dataset ٣” where the square with the number ١٠٠٣ represents the number of correctly predicted class values and the square with the number ٠ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٢١١٣ represents the number of correctly predicted no-class values.

Figure (٤.٥٠). ROC curve of DT in training phase “dataset ٣” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Table (٤.٨) explains the results of the measurements for the training phase of the five classifiers on “dataset ٣”.

Table (٤.٨). Results of training phase on “dataset ٣”

	Accuracy	Precision	Recall	F-score	Specificity
SVM	٩٩,٥%	٩٩,٥%	٩٩,١%	٩٩,٥%	٩٩,٧%
SGD	٩٩,٣%	٩٩,٦%	٩٨,٥%	٩٩%	٩٩,٨%
GNB	٩٩,٥%	٩٨,٨%	٩٨,٨%	٩٩,٥%	٩٩,٤%
RF	٩٩,٩%	٩٩,٩%	١٠٠%	١٠٠%	٩٩,٩%
DT	١٠٠%	١٠٠%	١٠٠%	١٠٠%	١٠٠%

ii. Testing phase “dataset ٣”

The results of the five classifiers in this stage are shown in figures from (٤.٥١) to (٤.٦٠).

Support Vector Machine (SVM)

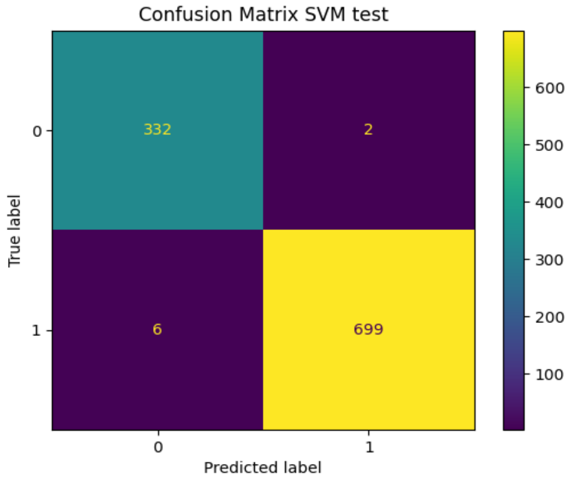


Figure (٤.٥١). Confusion matrix of SVM in testing phase “dataset ٣”

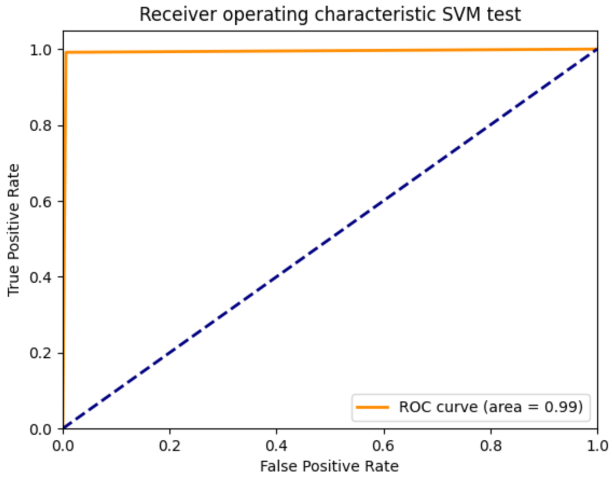


Figure (٤.٥٢). ROC curve of SVM in training phase “dataset ٣” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and where the square with the number ٣٣٢ represents the number of correctly predicted class values and the square with the number ٢ represents the number of incorrectly predicted class values and the square with the number ٦ represents the number of incorrectly predicted no-class values and the square with the number ٦٩٩ represents the number of correctly predicted no-class values.

the result is ٠,٩٩ which is above the threshold (٠,٥)

Stochastic Gradient Descent Classifier (SGD)

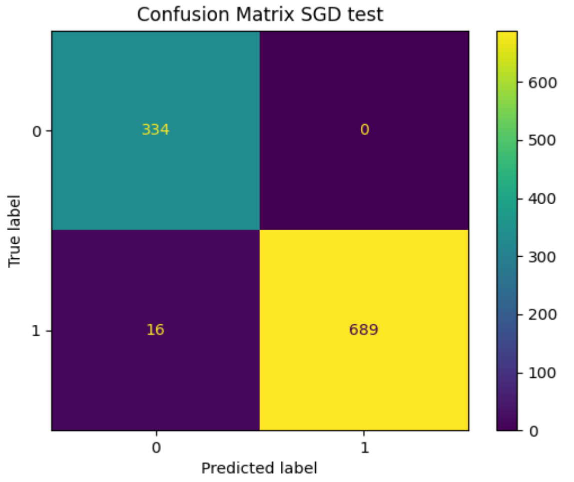


Figure (٤.٥٣). Confusion matrix of SGD in testing phase “dataset ٣”

where the square with the number ٣٣٤ represents the number of correctly predicted class values and the square with the number ٠ represents the

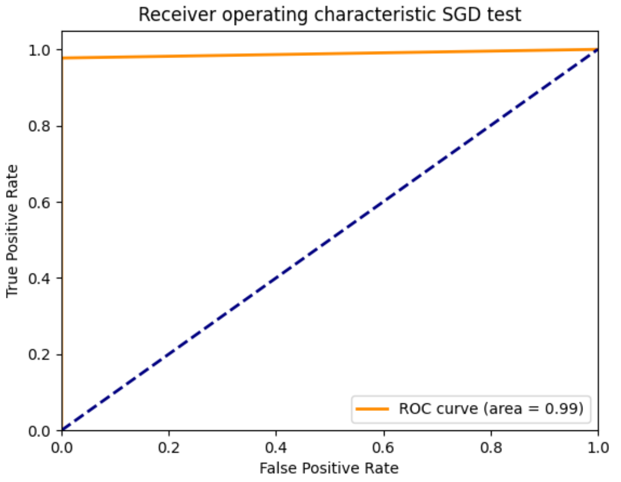


Figure (٤.٥٤). ROC curve of SGD in testing phase “dataset ٣” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩٩ which is above the threshold (٠,٥) number of incorrectly predicted class values and the square with the number ١٦ represents the number of incorrectly predicted no-class values and the square with the number ٦٨٩ represents the number of correctly predicted no-class values.

Gaussian Naïve Bayes (GNB)

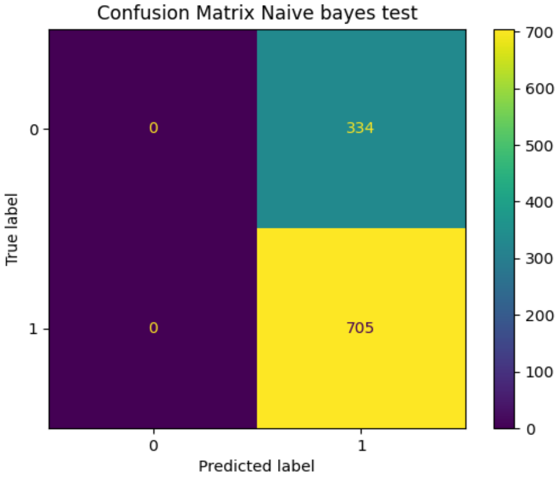


Figure (٤.٥٥). Confusion matrix of GNB in testing phase “dataset ٣”

where the square with the number ٠ represents the number of correctly predicted class values and the square with the number ٣٣٤ represents the number of incorrectly predicted class

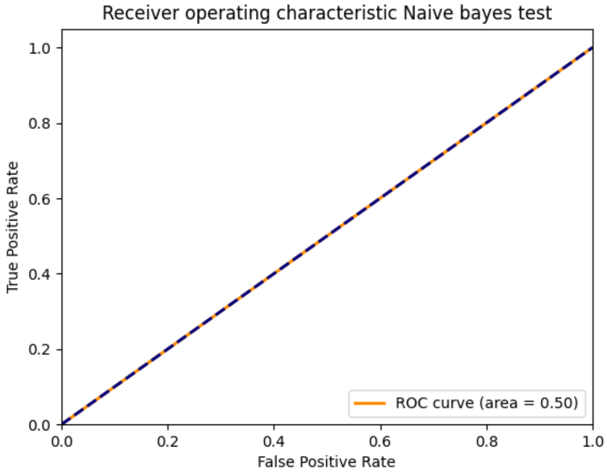


Figure (٤.٥٦). ROC curve of GNB in testing phase “dataset ٣” and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٥٠ which is equal to the threshold (٠,٥) values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٧٠٥ represents the number of correctly predicted no-class values.

Random Forest (RF)

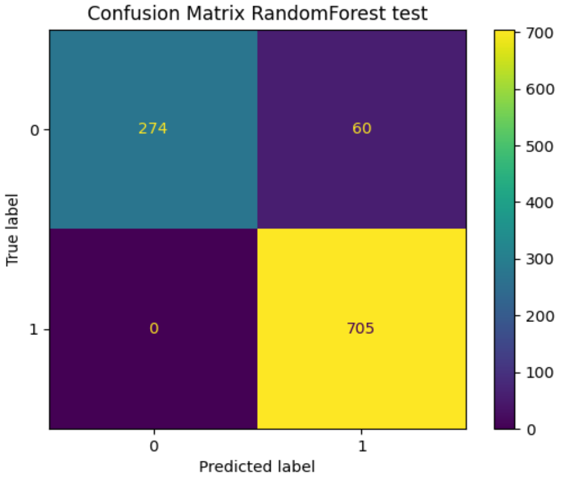


Figure (٤.٥٧). Confusion matrix of RF in testing phase "dataset ٣"

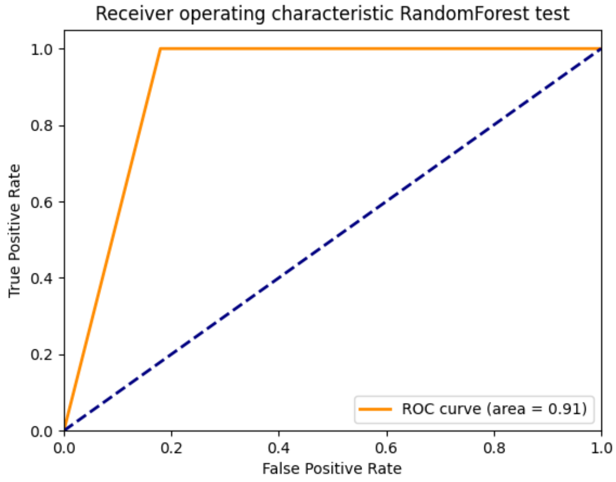


Figure (٤.٥٨). ROC curve of RF in testing phase "dataset ٣" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ٠,٩١ which is above the threshold (٠,٥)

where the square with the number ٢٧٤ represents the number of correctly predicted class values and the square with the number ٦٠ represents the number of incorrectly predicted class values and the square with the number ٠ represents the number of incorrectly predicted no-class values and the square with the number ٧٠٥ represents the number of correctly predicted no-class values.

Decision Tree Classifier (DT)

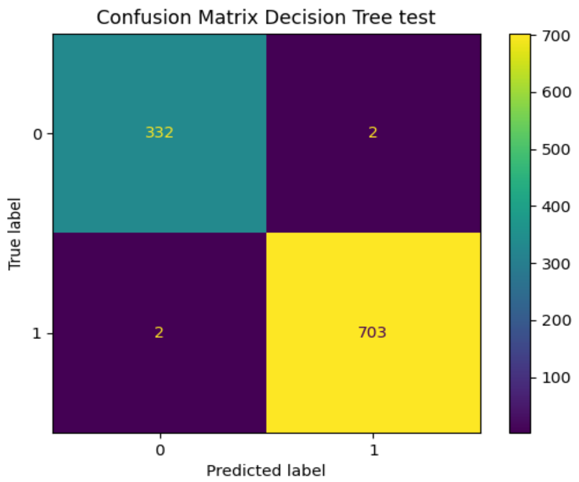


Figure (٤.٥٩). Confusion matrix of DT in testing phase "dataset ٣"

where the square with the number ٣٣٢ represents the number of correctly predicted class values and the square with the number ٢ represents the number of incorrectly predicted class values and the square with the number ٢ represents the number of incorrectly predicted no-class values and the square with the number ٧٠٣ represents the number of correctly predicted no-class values.

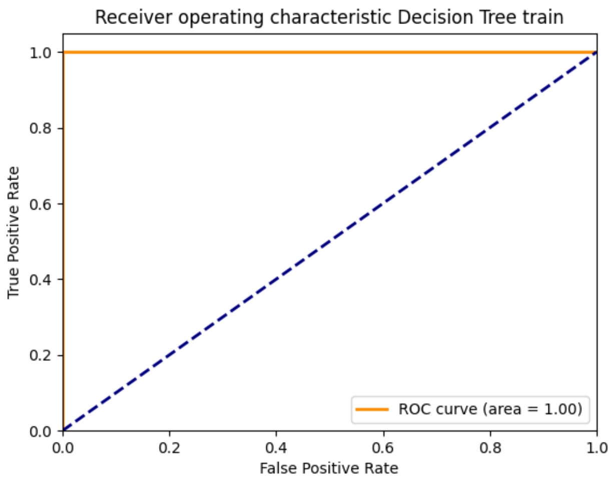


Figure (٤.٦٠). ROC curve of DT in testing phase "dataset ٣" and showing the performance of this classifier in classifying the correct predicted class and incorrect predicted class and the result is ١,٠٠ which is above the threshold (٠,٥)

Table (٤.٩) explains the results of the measurements for the testing phase of the five classifiers on "dataset ٣".

Table (٤.٩). Results of testing phase on “dataset ٣”

	Accuracy	Precision	Recall	F-score	Specificity
SVM	٩٩,٢%	٩٩,٤%	٩٨,٢%	٩٩%	٩٩,٧%
SGD	٩٨,٤%	١٠٠%	٩٥,٤%	٩٧,٨%	١٠٠%
GNB	٦٧,٨%	٠%	٠%	٠%	٦٧,٨%
RF	٩٤,٢%	٨٢%	١٠٠%	٩٠,١%	٩٢,١%
DT	٩٩,٦%	٩٩,٤%	٩٩,٤%	٩٩,٧%	٩٩,٧%

٤,١١. Discussion of Experimental Results

In this section, the performance of each classifier will be discussed with each data set in the training and testing phases, and the best and worst classifiers will be mentioned as follows:

٤,١١,١. Classifiers performance with “dataset ١”

In the training phase, the DT classifier was the best, with a classification accuracy of ٩٧%, followed by an RF classifier with an accuracy rate of ٩٦%, then SVM, and finally, SGD and GNB shared a classification accuracy equivalent to ٩٣%. In the testing phase, all classifiers showed an accuracy of ٩٣% except for the RF which classified the data with an accuracy of ٩٢%. It can be said that the decision tree classifier is best with this dataset.

٤,١١,٢. Classifiers performance with “dataset ٢”

In the training phase, the DT, and RF classifiers shared an ideal accuracy rate of ١٠٠%, followed by the SVM classifier with a high accuracy rate also reached ٩٩,٥%, while the results of both SGD and GNB were equal to ٩٩,٣٪. In the testing phase, the RF classifier was the best in terms of the possibility of classification with an accuracy rate equal to ٩٩,٩%, followed by the DT classifier with an accuracy rate of ٩٩,٦%. The accuracy value of the SVM classifier was also close to the mentioned classifiers with an accuracy rate of ٩٩,٣%, and then the SGD classifier came with an accuracy rate of ٩٧. ٩% As for the GNB classifier, it was far worse than the rest of the used classifiers, with an accuracy of ٦٧.٨٪.

٤,١١,٣. Classifiers performance with “dataset ٣”

In the training phase, the best classifier was DT with a ١٠٠٪ accuracy rate, followed by the RF classifier with ٩٩,٩%, and both SVM and GNB shared the same value equal to ٩٩,٥٪ while the SGD had the lowest accuracy rate with ٩٩,٣٪. In the testing phase, DT also outperformed the rest of the classifiers in accuracy, which reached ٩٩,٦%, followed by the SVM classifier with an accuracy rate of ٩٩,٢%, and the accuracy of the SGD and RF classifiers was ٩٨.٤٪ and ٩٤,٢%, respectively, while GNB was also worse with this data set with an accuracy rate of ٦٧,٨٪.

From the above results, it can be confirmed that the DT classifier is the best among the five classifiers with the three datasets.

٤,١٢. Comparison with Previous Studies

In this section, the results of the accuracy of the proposed system are compared with the studies mentioned in chapter one. Fig. (٤.٦١) show the accurate results of the proposed system and these studies. Results are shown in Table (٤.١٠).

By comparing the proposed system results with the previous work, the proposed system using Decision Tree gives an accuracy of ٩٧٪ for social engineering attack Which depends only on the texts sent via messages, which was supposed to be ٥٪ higher than the method of detecting social engineering attacks that use texts to phish victims, but as its result in the previous study was ٩٢% as shown in table (٤,١٠) reference number ١٧, that is, ٥%less than the proposed method. And for Fake Account Detection the proposed system using Random force gives ٩٩,٩٪ of accuracy which is higher then previous study on detection of spam bots in reference number ١٦ that used DT-SVMNB which gives ٩٨% and for Sensitive Text Detection the proposed system using Decision tree gives ١٠٠٪ of accuracy which is higher than previously studies reference number [١٢] using SVM, NB and NN where the NN had ٩٠٪ of accuracy less than the proposed system results. **See Table (٤,١٠).**

Table (٤.١٠). Accuracy Comparison with related studies

Ref. No.	Techniques	Accuracy
[١١]	KNN+ RF+ DT+MLP	٩٧,١٣٪
[١٢]	SVM, NB, NN	٩٠٪
[١٣]	SWRL+ OWL	٩٤,٩٪
[١٤]	SVM+KNN	٩٧,٧٪
[١٥]	Bag of bots' word model	٩٩,٢٪
[١٦]	DT-SVMNB	٩٨٪
[١٧]	DT	٩٢٪
[١٨]	SVM	٨٨٪
[١٩]	RF	٨٨٪
[٢٠]	SVM+RF+ANN	٩٧,٤٥٪
Social Engineering Detection "dataset ١"	DT	٩٧٪
Fake Account Detection "dataset ٢"	RF	٩٩,٩٪
Sensitive Text Detection "dataset ٣"	DT	١٠٠٪

Chapter Five

Conclusions and Recommendations

Chapter Five

Conclusions and Recommendations

٥,٣. Conclusions

From the comparison of the proposed system results with the previous studies it can bee seen the effect of choosing the best features among the results of different methods improving the classifier’s performance and accuracy in classifying texts as sensitive or not and attack or not and fake profile or real. Focusing on selecting features is better than incorporating more than one Machine learning algorithm in one system which consume less time in training.

The training database for detecting sensitive texts or social engineering attack texts can be modified depending on the vision of the organization benefiting from the system to protect its employees and business policy according to its needs.

Social engineering attacks are one of the most common cyber-attacks. Phishing attack is a type of social engineering attack, which can be performed over several ways. Text phishing attack is a prevalent kind of phishing attack. This thesis addressed the text attack using machine learning and natural language processing. Trained five classification models to distinguish text from normal text. Support Vector Machine (SVM), Gaussian Naïve Bayes (GNB), Stochastic Gradient Descent (SGD), Random Forest (RF), and Decision Tree (DT) algorithms were used with best features of preprocessed data for training the classifiers. Found that the most effective model against text attack is the Decision Tree model.

٥,٤. Suggestions for Future Works

The following issues have been noticed and should be resolved in the future to improve the suggested systems:

١. Using other classifiers from machine learning classifiers such as K-Nearest Neighbors (KNN), Logistic Regression (LR), and so on.

٢. Deep learning techniques can be used instead of machine learning techniques to discover vulnerabilities and problems with the Twitter application.

٣. Work on datasets for other applications such as Facebook and Instagram.

References

References

[١] B. Kavin et al., “Machine Learning-Based Secure Data Acquisition for Fake Accounts Detection in Future Mobile Communication Networks,” Hindawi, Wireless Communications and Mobile Computing, ٢٠٢٢.

[٢] Kondeti, Priyanka, Lakshmi Pranathi Yerramreddy, Anita Pradhan, and Gandharba Swain. "Fake account detection using machine learning." In *Evolutionary computing and mobile sustainable networks*, pp. ٧٩١-٨٠٢. Springer, Singapore, ٢٠٢١.

[٣] Pathak, Ajeet Ram, Aditee Mahajan, Keshav Singh, Aishwarya Patil, and Anusha Nair. "Analysis of techniques for rumor detection in social media." *Procedia Computer Science* ١٦٧ (٢٠٢٠): ٢٢٨٦-٢٢٩٦.

[٤] R. Pugliese, “Machine Learning-Based Approach: Global Trends, Research Directions, And Regulatory Standpoints,” Data Science and Management, vol. ٤, pp. ١٩-٢٩, December ٢٠٢١.

[٥] A. el Azab et al., “Fraud News Detection for Online Social Networks Web Usage Mining Techniques and Application across Industries,” igi global, ٢٠١٧.

[٦] J. Jiang et al., “Understanding latent interactions in online social networks,” in Proceedings of the the١٠th ACM SIGCOMM Conference on Internet Measurement, ACM, pp. ٣٦٩–٣٨٢, Melbourne, Australia, ٢٠١٩.

[٧] C. Buntain and J. Golbeck, “Automatically identifying fake news in popular Twitter threads,” in Proceedings of the IEEE International Conference on Smart Cloud, pp. ٢٠٨–٢١٥, New York, ٢٠١٧.

[٨] S. Bickley et al., “Artificial Intelligence in The Field of Economics,” Scientometrics, vol. ١٢٧, pp. ٢٠٥٥–٢٠٨٤, ٢٠٢٢.

[٩] A. Jobin et al., “The Global Landscape of AI Ethics Guidelines,” Nature Machine Intelligence, vol. ١, no. ٩, pp. ٣٨٩–٣٩٩, ٢٠١٩.

[١٠] Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN Computer Science* ٢, no. ٣ (٢٠٢١): ١-٢١.

[١١] A. Al-Zoubi et al., “Spam Profiles Detection on Social Networks Using Computational Intelligence Methods: The Effect of the Lingual Context”, Journal of Information Science, vol. ٤٧, no.١, pp. ٥٨–٨١, ٢٠١٩.

[١٢] Basarslan, Muhammet Sinan, and Fatih Kayaalp. "Sentiment analysis with machine learning methods on social media." (٢٠٢٠).

[١٣] M. Jabardi and A. Hadi, “Twitter Fake Account Detection and Classification using Ontological Engineering and Semantic Web Rule Language,” Karbala International Journal of Modern Science, vol. ٦, no, ٤, ٢٠٢٠.

[١٤] Sarker, Aditi, Partha Chakraborty, SM Shaheen Sha, Mahmuda Khatun, Md Rakib Hasan, and Kawshik Banerjee. "Improvised technique for analyzing data and detecting terrorist attack using machine learning approach based on twitter data." *Journal of Computer and Communications* ٨, no. ٧ (٢٠٢٠): ٥٠-٦٢.

[١٥] Ramalingaiah, A., S. Hussaini, and S. Chaudhari. "Twitter bot detection using supervised machine learning." In *Journal of Physics: Conference Series*, vol. ١٩٥٠, no. ١, p. ٠١٢٠٠٦. IOP Publishing, ٢٠٢١.

[١٦] S. Rahman et al., “An Efficient Hybrid System for Anomaly Detection in Social Networks,” Springer, vol. ٤, no.١٠, ٢٠٢١.

[١٧] Alsufyani, Asma A., and S. M. Alzahrani. "Social Engineering Attack Detection Using Machine Learning: Text Phishing Attack." (٢٠٢١): ٧٤٣-٧٥١.

[١٨] Luo, Xiaoyu. "Efficient english text classification using selected machine learning techniques." *Alexandria Engineering Journal* ٦٠, no. ٣ (٢٠٢١): ٣٤٠١-٣٤٠٩.

[١٩] Heidari, Maryam, H. James Jr, and Ozlem Uzuner. "An empirical study of machine learning algorithms for social media bot detection." In ٢٠٢١ *IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, pp. ١-٥. IEEE, ٢٠٢١.

[٢٠] B. Kavin et al., “Machine Learning-Based Secure Data Acquisition for FakeAccounts Detection in Future Mobile Communication Networks,” Hindawi, Wireless Communications and Mobile Computing, ٢٠٢٢.

[٢١] Y. Chawla and G. Chodak, “Social Media Marketing for Businesses: Organic Promotions of Web Links on Facebook,” Journal of Business Research, vol. ١٣٥, pp. ٤٩-٦٥, October ٢٠٢١.

[٢٢] A. Alalwan et al., “Social Media in Marketing: A Review and Analysis of the Existing Literature,” Telematics and Informatics, vol. ٣٤, pp. ١١٧٧-١١٩٠, ٢٠١٧.

[٢٣] S. Farivar et al., “Followers’ problematic engagement with influencers on social media: An attachment theory perspective,” Computers in Human Behavior, vol. ١٣٣, August ٢٠٢٢.

[٢٤] A. Jain et al., “Online social networks security and privacy: comprehensive review and analysis,” Complex & Intelligent Systems, vol. ٧, pp.١٥٧–٢١٧٧, ٢٠٢١.

[٢٥] S. Sahoo and B. Gupta, “Fake profile detection in multimedia big data on online social networks,” International Journal of Information Computer Security, vol. ١٢, pp. ٣٠٣–٣٣١, ٢٠٢٠.

[٢٦] W. Akram, "A Study on Positive and Negative Effects of Social Media on Society," International Journal of Computer Sciences and Engineering, vol. ٥, no.١٠, ٢٠١٨.

[٢٧] S. Siddiqui and T. Singh, "Social Media its Impact with Positive and Negative Aspects," International Journal of Computer Applications Technology and Research, vol. ٥, no. ٢, pp. ٧١ - ٧٥, ٢٠١٦.

[٢٨] Y. Zhang et al., "Twitter-aided decision making: a review of recent developments," Applied Intelligence, ٢٠٢٢.

[٢٩] A. Helmond, "Facebook's Evolution: Development of a Platform-As-Infrastructure," Internet Histories, vol. ٣, no. ٢, pp. ١٢٣–١٤٦, ٢٠١٩.

[٣٠] S. Yang et al., "The science of YouTube: What factors influence user engagement with online science videos?", PLOS ONE, vol.١٧, no. ٥, ٢٠٢٢.

[٣١] K. Johnson, "Using LinkedIn to Teach Students How to Build Their Professional Network and Enhance their Personal Brand," Global Research in Higher Education, vol. ٤, no. ٢, ٢٠٢١.

[٣٢] K. Miltner and Y. Gerrard, "Tom had us all doing front-end web development: a nostalgic (re)imagining of Myspace," Internet Histories, vol. ٦, pp. ٤٨-٦٧, ٢٠٢٢.

[٣٣] M. Beye et al., "Literature Overview - Privacy in Online Social Networks," Centre for Telematics and Information Technology (CTIT), ٢٠١٠.

[٣٤] S. Zhang et al., "Applications of Social Network Analysis to Obesity: A Systematic Review," Obesity Reviews, ٢٠١٨.

[٣٥] S. Zeebaree et al., "Social Media Networks Security Threats, Risks and Recommendation: A Case Study in the Kurdistan Region," International Journal of Innovation, Creativity, and Change, vol. ١٣, no. ٧, ٢٠٢٠.

[٣٦] A. Kadhim, “An Evaluation of Preprocessing Techniques for Text Classification,” International Journal of Computer Science and Information Security, vol. ١٦, no. ٦, pp. ٢٢-٣٢, ٢٠١٨.

[٣٧] H. Alshuraiqi, “Improved Term Frequency Inverse Document Frequency (TF-IDF) Method for Arabic Text Classification,” International Journal of Advanced Trends in Computer Science and Engineering, vol. ٩, no.٥, pp. ٦٩٣٩-٦٩٤٦, ٢٠٢٠.

[٣٨] H. Ali, “Machine Learning Methods to Predict the Genes Expression that Affect Stages of Alzheimer’s Disease”. Master Thesis, University of Babylon, College of Information Technology, ٢٠٢١.

[٣٩] M. Little et al., “Using and Understanding Cross-Validation Strategies,” Giga Science, ٢٠١٧.

[٤٠] Q. Ren et al., “Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives,” Journal of Big Earth Data, ٢٠١٩.

[٤١] J. Villarreal et al., “Secure Learning for Android Malware Detection,” Faculty of Engineering Systems Engineering Program ITCS Department ICESI University Project, ٢٠١٩.

[٤٢] D. Sondakh, “A Comparative Study of Classification Algorithms: k-Folds and Holdout as Accuracy Estimation Methods,” International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), vol. ٦, no. ١, January ٢٠١٦.

[٤٣] C. Rayat, ٢٠١٨. “Statistical Methods in Medical Research || Chi-Square Test (χ٢ – Test)”, Chapter ٩, ©Springer, pp. ٦٩–٧٩.

[٤٤] M. Kumar et al., “Feature Selection and Classification of Microarray Data using MapReduce based ANOVA and K-Nearest Neighbor,” Procedia Computer Science, vol. ٥٤, pp. ٣٠١–٣١٠, ٢٠١٥.

[٤٥] F. Zhao, “A Filter Feature Selection Algorithm Based on Mutual Information for Intrusion Detection,” Applied Sciences, vol. ٨, no. ٩, ٢٠١٨.

[٤٦] Zakharov, Roman, and Pierre Dupont. "Ensemble logistic regression for feature selection." In IAPR International Conference on Pattern Recognition in Bioinformatics, pp. ١٣٣-١٤٤. Springer, Berlin, Heidelberg, ٢٠١١.

[٤٧] Gayatri, N., S. Nickolas, A. V. Reddy, S. Reddy, and A. V. Nickolas. "Feature selection using decision tree induction in class level metrics dataset for software defect predictions." In Proceedings of the world congress on engineering and computer science, vol. ١, pp. ١٢٤-١٢٩. ٢٠١٠.

[٤٨] T. Le et al., “Statistical Inference Relief (STIR) Feature Selection,” Bioinformatics, ٢٠١٨.

[٤٩] Zhao, Zhenyu, Radhika Anand, and Mallory Wang. "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform." In ٢٠١٩ *IEEE international conference on data science and advanced analytics (DSAA)*, pp. ٤٤٢-٤٥٢. IEEE, ٢٠١٩.

[٥٠] S. Duan et al. “LightGBM Low-Temperature Prediction Model Based on LassoCV Feature Selection,” Mathematical Problems in Engineering, ٢٠٢١.

[٥١] N. Dutta et al., “Centrifugal Pump Cavitation Detection Using Machine Learning Algorithm Technique,” IEEE Access, ٢٠١٨.

[٥٢] A. Kadhim, “Survey on Supervised Machine Learning Techniques for Automatic Text Classification,” AI Review, Vol.٥٢, No.١, pp.٢٧٣-٢٩٢, ٢٠١٩.

[٥٣] A. Viloria et al., “Unsupervised Learning Algorithms Applied to Grouping Problems”, Procedia Computer Science, vol. ١٧٥, pp. ٦٧٧–٦٨٢, ٢٠٢٠.

[٥٤] A. Muzio et al., “Deep Reinforcement Learning for Humanoid Robot Behaviors”, Journal of Intelligent Robot System, vol. ١٠٥, no. ١٢, ٢٠٢٢.

[٥٥] S. Uddin et al., “Comparing Different Supervised Machine Learning Algorithms for Disease Prediction,” BMC Medical Informatics and Decision Making, vol. ١٩, no. ٢٨١, ٢٠١٩.

[٥٦] P. Saigal and V. Khanna, “Multi‑ category News Classification Using Support Vector Machine Based Classifiers,” SN Applied Sciences vol. ٢, no. ٤٥٨, ٢٠٢٠.

[٥٧] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning”, Journal of Applied Science and Technology Trends, vol. ٢, no. ١, pp. ٢٠-٢٨, ٢٠٢١.

[٥٨] B. Jijo and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” Journal of Applied Science and Technology Trends, vol. ٠٢, no. ٠١, pp. ٢٠ – ٢٨, ٢٠٢١.

[٥٩] H. Fei et al., “Cotton Classification Method at the County Scale Based on Multi-Features and Random Forest Feature Selection Algorithm and Classifier,” Remote Sensor, vol. ١٤, no. ٨٢٩, ٢٠٢٢.

[٦٠] P. Divya et al., “Application of Random Forest Algorithm in Bio Informatics”, International Journal of Information Technology Insights & Transformations, vol. ٥, no. ١, ٢٠٢١.

[٦١] S. Hyland and S. Tople, “An Empirical Study on The Intrinsic Privacy of Stochastic Gradient Descent,” arXiv:١٩١٢,٠٢٩١٩v٤ [cs. LG] ٢٨ Feb ٢٠٢٢.

[٦٢] M. Anand et al., “Gaussian Naıve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer,” Hindawi, Mobile Information Systems, ٢٠٢٢.

[٦٣] D. Berrar, “Bayes’ theorem and Naïve Bayes classifier,” Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, pp. ٤٠٣–٤١٢, Elsevier Science Publisher, Amsterdam, Netherlands, ٢٠١٨.

[٦٤] C. Hassan et al., “Comparison of Machine Learning Algorithms in Data classification,” ٢٤th International Conference on Automation and Computing (ICAC), ٢٠١٨.

[٦٥] A. Sharma, “A survey on: Online Social Networking Attacks Detection Techniques,” International Journal of Scientific Research in Computer Science Engineering and Information Technology, vol. ٧, no. ٣, pp. ٤٤-٥٠, ٢٠٢١.

[٦٦] M. Fire et al., “Online Social Networks: Threats and Solutions,” IEEE Communications Surveys & Tutorials, vol. ١٦, no. ٤, pp. ٢٠١٩–٢٠٣٦, ٢٠١٤.

[٦٧] I. Markoulidakis et al., “Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem,” Technologies, vol. ٩, no. ٨١, ٢٠٢١.

[٦٨] Ž. Vujović, “Classification Model Evaluation Metrics,” International Journal of Advanced Computer Science and Applications (IJACSA), vol.١٢, no. ٦, ٢٠٢١.

الخلاصة

في هذه الايام ، تعد شبكات التواصل الاجتماعي إحدى أهم المدونات الصغيرة وتبادل البيانات. لديها ملايين المستخدمين في جميع أنحاء العالم ، ويتواصل مستخدموها معًا عبر الرسائل والمنشورات. بسبب البنية المفتوحة وسلوكها ، فهي عرضة للهجمات من حسابات وهمية وعدد هائل من البرامج الآلية ، أو "الروبوتات". تعتبر الروبوتات ضارة لأنها تستخدم الإنترنت لإرسال رسائل غير مرغوب فيها إلى رواد الشبكة الاجتماعية. يضع مستخدمو الشبكات الاجتماعية قيمة عالية لأمن البيانات والخصوصية حيث يجب تلبية هذه المطالب إذا كان للشبكة أن تحتفظ باهتمام المستخدم وتحفظ خصوصيته. لمعالجة هذه الصعوبات ، يلزم اتباع نهج فعال للكشف عن الحسابات المزيفة وتصنيفها.

في هذه الأطروحة ، يتم استخدام مجموعة متنوعة من تقنيات اختيار الميزات بالإضافة إلى مجموعة متنوعة من المنهجيات القائمة على التعلم الآلي لتحديد الحسابات المزيفة التي يمكن أن تخدع المستخدمين والنصوص الخطرة المرسلة عبر الرسائل والنصوص ذات المعلومات الحساسة. تم الجمع بين ثماني استراتيجيات لاختيار الميزات وخمس طرق للتعلم الآلي لمعالجة مجموعة البيانات مسبقًا. تتضمن عملية اختيار الميزة (التي تستخدم طرقًا متعددة لتصنيف الميزات) ثماني عمليات اختيار ميزة(Chi-Square Test Feature Selection, ANOVA Feature Selection, Mutual Information Feature Selection, Logistic Regression Feature Selection, Additional Tree Feature Selection, Embossed Feature Selection, mrMR features Selection and, Light GBM Trait Selection) فقط لتحديد الميزات المثالية التي يتم تسجيلها بواسطة المصنف للحصول على أفضل أداء. تم استخدام واختبار اكثر من طريقة تصنيف من بين خمس خوارزميات للتعلم الآلي Support Vector Machine (SVM)و Gaussian Naïve Bayes (GNB) و SGD و Random Forest (RF) و Decision Tree (DT)، وتم اختيار افضلها اداء من حيث الدقة واقل معدل خطا.

تم استخدام خمسة مصنفات لغرض تصنيف حسابات مواقع التواصل الاجتماعي والنصوص الخطرة والنصوص ذات المعلومات الحساسة. تم تطبيق مجموعة واسعة من عمليات المعالجة المسبقة لتنقية البيانات وتسهيل معالجة البيانات ، بالإضافة إلى تجربة عدد كبير من تقنيات استخراج الميزات واختيار افضل الميزات من نتائجها. النظام المقترح تم اختباره باستخدام ثلاث مجموعات بيانات ، بعض هذه البيانات كانت متاحة على الإنترنت ، والبعض الآخر تم جمعه. تشير النتائج التي تم الحصول عليها إلى أن المصنف DT كان الأفضل لمجموعة البيانات الأولى ، بينما كان RF هو المصنف الأفضل بمعدل دقة يساوي ٩٩,٩٪ لمجموعة البيانات الثانية. بالنسبة لمجموعة البيانات الثالثة ، يتمتع المصنف DT بأعلى دقة تصنيف مثالية وهي ١٠٠٪.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جـامعـــة بـابـــل
كلية تكنولوجيا المعلومات
قسم البرمجيات

منع وكشف هجمات مواقع التواصل الاجتماعي باستخدام طرق التعلم الآلي

رسالة مقدمة

إلى مجلس كلية تكنولوجيا المعلومات - جامعة بابل كجزء من متطلبات

نيل درجة الماجستير في تكنولوجيا المعلومات / البرمجيات

من قبل :

كرار صادق محسن جواد

بأشراف:

أ .د. وسام سمير عبد علي بهيه

١٤٤٤ هـ **٢٠٢٢ م**