School of Computer Sciences

# CDS590 – Consultancy Project & Practicum

## Final Report

# [An Ensemble Machine Learning Model to Classify Literary Writers' Perspective on Anxiety, Stress and Trauma in War Narratives by Middle Eastern Women]

[Zhao Dan]

[P-COM0255/20]

Supervisor: [Ts. Dr. Pantea Keikhosrokiani]

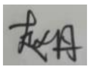
Mentor: [Dr. Moussa Pourya Asl]

Practicum place: School of Humanities, USM

SEM 1 2022/2023

# DECLARATION

"I declare that the following is my own work and does not contain any ***unacknowledged*** work from any other sources. This project was undertaken to fulfill the requirements of the Consultancy Project & Practicum for the Master of Science (Data Science and Analytics) program at Universiti Sains Malaysia".

Signature : 

Name : Zhao Dan

Date : 30th Jan 2023

# ACKNOWLEDGEMENTS

First and foremost, I would like to express my special thanks of gratitude to my supervisor, Dr. Pantea, who provided valuable advice and guidance on completing this project. She has taught me the methodology to carry out the research and helped me understand the project better. I am grateful to work and study under her supervision. I’d also like to this opportunity to express my gratitude toward her patience during discussion I had with her on completing the project.

I want to express my special appreciation and thanks to my mentor, Dr. Moussa, for giving me this golden opportunity to work on this project. He is always ready for help and shared with me useful information that was useful for the project. I would also like to express my gratitude to Dr. Nasuha Lee and Dr Chew Xin Ying for their guidance in Consultancy and Research skills.

I want to thank my parents and family, without their constant support and love, I would not complete the report so smoothly.

Finally, I would like to thanks all my friends for helping me throughout this project directly or indirectly.

# ABSTRACT

Stress, trauma, and anxiety are three typical negative sentiments. Stress refers to mental pressure or worry caused by problems in somebody's life or by having too much to do; trauma refers to a mental condition caused by severe shock, stress, or fear, especially when the harmful effects last for a long time; anxiety refers the state of feeling nervous or worried that something unwanted is going to happen. War narrative refers to literature that depicts wars. This project will study the three sentiments from selected novels by classifying the stress, trauma, and anxiety in war narratives from female middle-eastern writers. Before the experiment, the dataset is already on standby through data labeling and other data preparations (extracting features using the TF-IDF method and random over resampling method). During experiment. Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, and Decision Tree are adopted to classify the text based on the labeled dataset. The optimal individual models from each algorithm are selected and used as input to deploy four ensemble learning techniques. The performance of each model is evaluated using accuracy score, precision, recall, and F1-score. After that, interpretation has been conducted from technical and domain aspects. Finally, the project come to the conclusion that the optimal model is from the stacking technique, and ensemble learning techniques outperform the individual algorithms. Writers have an overall emphasis on the sentiment of trauma, and they believe that the sentiment of trauma is the most noticeable one of all.

# ABSTRAK

Tekanan, trauma, dan kebimbangan adalah tiga sentimen negatif yang tipikal. Tekanan merujuk kepada tekanan mental atau kebimbangan yang disebabkan oleh masalah dalam kehidupan seseorang atau kerana terlalu banyak kerja; trauma merujuk kepada keadaan mental yang disebabkan oleh kejutan, tekanan, atau ketakutan yang teruk, terutamanya apabila kesan berbahaya itu bertahan lama; kebimbangan merujuk keadaan berasa gementar atau bimbang sesuatu yang tidak diingini akan berlaku. Naratif perang merujuk kepada kesusasteraan yang menggambarkan peperangan. Projek ini akan mengkaji tiga sentimen daripada novel terpilih dengan mengklasifikasikan tekanan, trauma, dan kebimbangan dalam naratif perang daripada penulis wanita timur tengah. Sebelum percubaan, set data sudah bersedia melalui pelabelan data dan penyediaan data lain (mengekstrak ciri menggunakan kaedah TF-IDF dan kaedah rawak atas pensampelan semula). Semasa eksperimen. Regresi Logistik, Naïve Bayes, Mesin Vektor Sokongan, K-Nearest Neighbors, dan Decision Tree digunakan untuk mengklasifikasikan teks berdasarkan set data berlabel. Model individu yang optimum daripada setiap algoritma dipilih dan digunakan sebagai input untuk menggunakan empat teknik pembelajaran ensemble. Prestasi setiap model dinilai menggunakan skor ketepatan, ketepatan, ingat semula dan skor F1. Selepas itu, tafsiran telah dijalankan dari aspek teknikal dan domain. Akhirnya, projek ini membuat kesimpulan bahawa model optimum adalah dari teknik susun, dan teknik pembelajaran ensemble mengatasi algoritma individu. Penulis mempunyai penekanan keseluruhan pada sentimen trauma, dan mereka percaya bahawa sentimen trauma adalah yang paling ketara.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| APEX | - Accelerated Programme for Excellence |
| KNN | - K-Nearest Neighbour |
| LR | - Logistic Regression |
| NB | - Naïve Bayes |
| DT | - Decision Tress |
| SVM | - Support Vector Machine |
| NLP | - Natural Language Processing |
| MNB | - Multinomial Naïve Bayes |
| POS | - Part of Speech |
| TF-IDF | - Term Frequency-Inverse Document Frequency |

# 1 INTRODUCTION

## 1.1 Project Host Introduction

Universiti Sains Malaysia (USM) is a public research university situated in Penang Island in northern Malaysia. As a leading research institute founded in 1969, USM focuses on hotspot globally, collaborates across industries and multiple prestigious research institutes, and offers a wide range of courses to both undergraduate and postgraduate study programs. Forty years after its establishment, USM was entitled to implement the Accelerated Program for Excellence (APEX) in 2009, and such program has significantly improved the university's reputation and global influence. Like many high educational institutions, USM consists of multiple schools from three sectors, applied science and technological sector, pure science sector, and liberal arts sector, each of the school is expertise in its research fields respectively.

One school in the liberal arts sector is the school of humanities, one of the first batches of established schools with a history as old as USM itself. The school acts as a research center for the humanities, offering academic programs at undergraduate and postgraduate levels, creating innovation and new knowledge through research, and working towards the betterment of society through various community-industry initiatives. Its research area covers a wide spectrum of disciplines that include philosophy, civilisation, religion, language, literature, history, geography, and translation (Jasni Sulong). The project provided by the host is related to literary and text works.

## 1.2 Project Background Introduction

Middle East is a geographical term that mainly includes the Arab Peninsula, Egypt, and a part of West Asia. Due to the huge influence of Islam in this region, most of the dwellers here are muslins. Some Middle East countries suffer from invasions and civil wars since last century. Just in the passing 20 years, there were over 20 wars declared in this region. For example, Shia insurgency which lasts for over 10 years between Saudi Arabia and Yemen, and Syrian Civil War started from 2011. Some countries eventually earn peace after years of wars, others are currently trapped by wars after wars. Such situations catalyzed the presence of war narratives in the middle east. The war narrative is literary work, fiction or nonfiction, that tells stories that happened in the war, and it could be a great window for people to see the war from a variety of perspective. I selected 10 novels written by Middle East female writers which all tell stories about the wars in the Middle East, mostly in their home countries.

It is a common sense that wars have negative effect with mental health, and some negative sentiments are usually tied with the wars. For instance, fear, trauma, stress, anxiety, sorrow, depression, etc. The World Health Organization (WHO) has stated that in situations of armed conflict, "Around 10 percent of the people who experience traumatic events will have serious mental health problems, and another 10 percent will develop behavior that will hinder their ability to function effectively." Depression, anxiety, and psychosomatic problems such as insomnia are the most common effects (Huutsman Mental Health Institute). I chose three most commonly seen negative sentiments to study, and they are stress, trauma, and anxiety.

Overall, I would like to study stress, trauma, and anxiety in the 10 selected novels by applying machine learning models, and such a model should be expertise in classifying stress, trauma, and anxiety from the text data.

## 1.3 Problem Statement

A few problems act as motivation for me to take this opportunity to study the project, and persuade the project host to implement the solution I shall provide.

Firstly, the existing research on negative sentiments in middle east war is of scarcity, and I'd like to work on it to study stress, trauma, and anxiety in the middle eastern war narratives. Building a customed dictionary that contains vocabularies associated with each sentiment is even rarer. All of these motivate me to retrieve the existing keywords and build my own dictionary to classify the sentiments.

Secondly, manual classifying the sentiments involves human error and bias, and it is too time-consuming and a waste of human power. Besides, the lack of models to automatically classify stress, trauma, and anxiety from text data also calls for research. I'd like to develop a model to automatically classify stress, trauma, and anxiety. To automate the process, computerization is required and proper text processing steps should be followed to create a model. The machine learning algorithm will replace the role of human brain and learn from the dictionary created to classify the sentiments automatically.

Finally, studies in text classification for literary works is quite limited, which causes difficulties to choose algorithm and methods. Certain algorithm and learning method may outperform the others, testing multiple algorithms and tuning the hyperparameter is necessary for seeking the optimal model. Therefore, I’d like to tune multiple algorithms, parameters, and techniques to find the optimal model.

## 1.4 Research Questions

**RQ1: How to assign the correct sentiment to each sentence?**

Answering this question will help to identify the keywords that related to stress, trauma, and anxiety and create the dictionary respectively.

**RQ2: What preprocessing should be done before applying the machine learning techniques?**

Answering this question will help to get the data ready for machine learning model application.

**RQ3: How to apply the machine learning techniques?**

Answering this question will help to apply the machine learning techniques in python environment, and compare the models.

**RQ4: What is the optimal model and how to interpret the results?**

Answering this question will help to find the optimal model by the evaluation metrics chosen, visualize the results, and explain the insights gotten from the results.

## 1.5 Objectives of Project

The objectives and goals of this project are:

1. To identify the keywords that related to stress, trauma, and anxiety and create the dictionary respectively.

2. To get the data ready for machine learning model by following five steps, tokenization, normalization, punctuation removal, stopwords removal, and POS tag.

3. To apply the individual machine learning algorithm and ensemble learning techniques in python environment, and compare the models.

4. To identify the evaluation metrics used, select the optimal model based on the metrics chosen, visualize and interpret the results.

## 1.6 Expected Outcome

1. Solve the research problems mentioned.

2. Successfully implement the machine learning algorithms and techniques.

3. Generate an optimal model that is reliable and stable, which means at least enjoys a value of metrics over 0.9.

4. Discuss the modeling & model selection process, and interpret the modeling results.

## 1.7 Benefit of Project

The outcome of this project will help the client to have better insight for stress, trauma, and anxiety in the 10 selected novels written by Middle East women. Besides, the classification model expects at least 90% accuracy score, which means that the model is quite reliable and potentially can be further developed to implement in other war literary works. Finally, potential publication of the project may also improve the research reputation of the client.

# 2 RELATED WORKS

## 2.1 Introduction

This chapter will discuss related works from the domain perspective and machine learning technique perspective. Firstly, domain knowledge about stress, trauma, and anxiety and these sentiments in war narratives will be discussed. And then, the second part will mainly discuss text classification using individual machine learning algorithms and ensemble machine learning techniques.

## 2.2 Domain Perspective

### 2.2.1 Concept of stress, trauma, and anxiety

According to Oxford Advanced Learner's Dictionary (Oxford, n.d), stress is "mental pressure or worry caused by problems in somebody's life or by having too much to do"; trauma is "a mental condition caused by severe shock, stress or fear, especially when the harmful effects last for a long time"; anxiety is "the state of feeling nervous or worried that something bad is going to happen". From the definition, it is clear that they are all negative sentiments, and suffering from such negative in a long term can be quite harmful to mental health.

The term stress is used for a range from mild pressure in daily life to fierce situations during wartime. Existing research proposed that the term 'stress' should be restricted to conditions where an environmental demand exceeds the natural regulatory capacity of an organism, in particular situations that include unpredictability and uncontrollability. Physiologically, stress seems to be characterized by either the absence of an anticipatory response (unpredictable) or a reduced recovery (uncontrollable) of the neuroendocrine reaction. (Koolhaas, J. et al, 2011) The process of social stress can be seen as combining three major conceptual domains: the sources of stress, the mediators of stress, and the manifestations of stress. Each of these extended domains subsumes a variety of sub- parts that have been intensively studied in recent years. (Leonard I. et al. 1981)

Trauma is an emotional response to a terrible event like an accident, rape, or natural disaster. Immediately after the event, shock and denial are typical. Longer term reactions include unpredictable emotions, flashbacks, strained relationships, and even physical symptoms like headaches or nausea. (American Psychological Association, 2008)

Some people have difficulty moving on with their lives. Coping styles vary from action oriented to reflective and from emotionally expressive to reticent. Clinically, a response style is less important than the degree to which coping efforts successfully allow one to continue necessary activities, regulate emotions, sustain self-esteem, and maintain and enjoy interpersonal contacts. (National Library of Medicine, 2014)

According to the definition of anxiety, almost all kinds of anxiety have an element of future, e.g., from a signal of a shock to its application. Future anxiety is conceived of as a state of apprehension, uncertainty, fear, worry and concern of unfavorable changes in a more remote personal future. In an extreme case this would be a threat (panic) that something really catastrophic may happen to a person. (Zbigniew Zaleski, 1995)

Generalized anxiety disorder (GAD) usually involves a persistent feeling of anxiety or dread, which can interfere with daily life. It is not the same as occasionally worrying about things or experiencing anxiety due to stressful life events. People living with GAD experience frequent anxiety for months, if not years. (National Institute of Mental Health, 2022)

### 2.2.2 Stress, trauma, and anxiety in war narratives

As a form of literature which has a context of war, war narratives have a trajectory that is accompanied by an emotional progression. To be persuasive, a particular war narrative must also resonate with prototypical or archetypal themes. (Tony Vinson, Desmond McDonnell) Most war narratives revert to some common theme, such as call for peace, accuse of the ruthlessness of the war, compassion to civilians living in the war, tragic story of individual of family, etc. Therefore, there are plenty of resources in the war narratives that depicts the sentiments such as stress, trauma, and anxiety of human being in the war.

Among the most important catalogued in early studies of stress in the wars are: threat to life and limb; physical discomfort from exposure to the elements; disease and deprivation; loss of friends; exposure to the death and dying of others; the requirement to kill and harm others resulting in moral and value conflicts; and the feeling of helplessness in the face of forces beyond the individual's control. (Robert S.et al, 1984) Research of the effects of war stress, especially for children, can also been found in Europe. The war stress had a negative impact on the emotional functioning of all children, especially the refugees. (Ivanka, Zivcic, M.A. 2010)

The book, ‘Trauma in Contemporary Literature’, analyzes contemporary narrative texts in English in the light of trauma theory, including essays by scholars of different countries who approach trauma from a variety of perspectives. In this book, author analyzes and applies the most relevant concepts and themes discussed in trauma theory, such as the relationship between individual and collective trauma, historical trauma, absence vs. loss, the roles of perpetrator and victim. (Marita Nadal, Monica Calvo) In another book, Virginia Woolf, its characterization of Septimus Smith in Mrs. Dalloway illustrates not only the psychological injuries suffered by victims of severe trauma such as war but also the need for them to give meaning to their suffering in order to recover from the trauma. (Phdessay, 2021)

War anxiety can gradually sneak up on victims, or it can present suddenly in response to a trigger. Symptoms can be in one’s mind, in one’s body, or both. Physical symptoms of anxiety may include a racing heart, butterflies in the victim’s stomach, nausea, or dizziness. (Stephanie Collier, 2022) Many factors may contribute to the formation of anxiety, one factor, from previous research, is persecutory anxiety, which arises 'when a group is threatened in its prestige, income or its existence; i.e., when it declines and does not understand the historical process or is prevented from understanding it. (Neumann, F, 1972)

## 2.3 Machine Learning Perspective

### 2.3.1 Text Classification with individual algorithm

Text classification refers to applying machine learning technique to classify the given text into two or multiple categories. By doing this, useful information can be extracted from the text, and better insights can be generated. As text classification is a machine learning technique, it follows the typical machine learning process. From preprocessing to model fitting and interpret the results, etc.

Data preprocessing is the process of making the data eligible to be fitted into the model, text preprocessing distinguished from preprocessing of other data type in many ways due to the nature of the text data. Text data comprises noise in the form of emotions, punctuation, and text in different cases, among other things (Agrawal, 2021). Therefore, common text preprocessing steps include removing punctuation and special marks, removing stop words, and get all the data lowercase, etc.

Individual algorithm in classification problems refers to using a single algorithm to classify the input data to two or multiple categories. Due to the nature of the problem, only selected is eligible. For instance, K-Nearest Neigher (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Random Forest, and etc. A comparison table about popular classification algorithm can be found in table 2.1.

**Table 2.1 Comparison of Individual Classification Algorithm**

| Model | Advantage | Disadvantage |
|---|---|---|
| K-Nearest Neighbor (KNN) | 1) Straight forward and easy to interpret.<br>2) No strong assumptions | 1) Not applicable when weight or probability involved.<br>2) Dimensionality curse, hard to use with multiple features. |
| Random Forest (RF) | 1) Aggregated decision trees, less likely overfitting.<br>2) Flexible in classification and regression problem, and therefore cope well with both discrete and continuous data. | 1) Strong assumptions, may not be applicable when the condition doesn’t align with the assumption.<br>2) Relatively computationally expensive |

| | | |
|---|---|---|
| Support Vector Machine (SVM) | 1) High tolerance to noise, work well in noisy dataset.<br>2) Plenty of hyperparameters, and rich models can be generated.<br>3) Less constraints with dimensionalities. | 1) Hard to interpret the model and meaning of the hyperparameters.<br>2) Sensitive to noise, less effective on noisier dataset, especially when there are overlapping classes. |
| Decision Tree (DT) | 1) Rule based algorithm, easy to understand and interpret.<br>2) Have a clear indication that the importance hierarchy of each feature. | 1) Suffering from overfitting, and may generate too many classes.<br>2) Relatively computational expensive. |

Researchers have studied various classification problems using the algorithms above. The summary of related previous works can be found in table 2.2.

**Table 2.2 Summary of Previous Works on Classification Problem in Individual Algorithm**

| Paper Title | Description | Algorithm Used | Accuracy Obtained |
|---|---|---|---|
| Sentiment Analysis Framework of Twitter Data Using Classification (Khurana, M, 2018) | Classify the twitter post into positive, neutral, and negative using NB, SVM, and RF. | Naïve bayes (NB), Support Vector Machine (SVM), and Random Forest (RF) | **NB (88%),** SVM (81%), RF (63%) |
| Patent Text Classification based on Naive Bayes Method (Xiao, L, 2018) | Classify the patent to A, E, F, H using NB and SVM | Naïve Bayes (NB), and Support Vector Machine (SVM) | **NB (98.6%),** SVM (96%) *<br>*Take the first class as example. |

| | | | |
|---|---|---|---|
| Sentiment Analysis on Movie Review Data Using Machine Learning Approach (Rahman, A, 2019) | Classify the movie to four classes using NB, SVM, and DT | Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB), Support Vector Machine (SVM), and Decision Tree (DT) | MNB (88.5%), **BNB (87.5%)**, SVM (87.3%), and DT (80.17%) |
| Study on SVM Compared with the other Text Classification Methods (Liu, Z, 2019) | Classify the input data to four classes using KNN, NB, and SVM | K-Nearest Neighbor (KNN), Naïve Bayes (NB), Support Vector Machine (SVM) | **KNN (97.44%)**, NB (95.92%), SVM (97.31%). |

To better interpret the model and results from the model, it is usually recommended to do visualization work to refine the last mile of any project. It refers to a group of information analysis techniques and processes that use interactive graphical representations of textual data to facilitate knowledge discovery (Risch et al., 2008). Besides, visualization may also help the model builders to generate interesting insights and deliver the work he/she has done.

### 2.3.2 Text Classification with Ensemble Learning Techniques

Ensemble learning is also known as multiple classifier system or committee-based learning, and it refers to applying two or multiple models in ensemble learning technique usually for the purpose of improving model performance. The typical workflow of ensemble learning is training a set of individual learners first and then combining them via some strategies, where an individual learner is usually trained by an existing learning algorithm. An ensemble is said to be homogeneous if all individual learners are of the same type. (Zhou, ZH, 2021) There are four types of ensemble learning techniques, bagging, boosting, voting, and stacking, and each of them will be discussed later in this chapter.

Bagging is also known as bootstrap aggregation, and it builds model based on multiple models from same algorithm but with different training samples. After that, an average result will be given, and this result is expected to be more robust than individual algorithm. When it comes to classification problem in bagging technique, only one class will be the outcome with most votes. The working process of bagging technique is shown in figure 2.1.

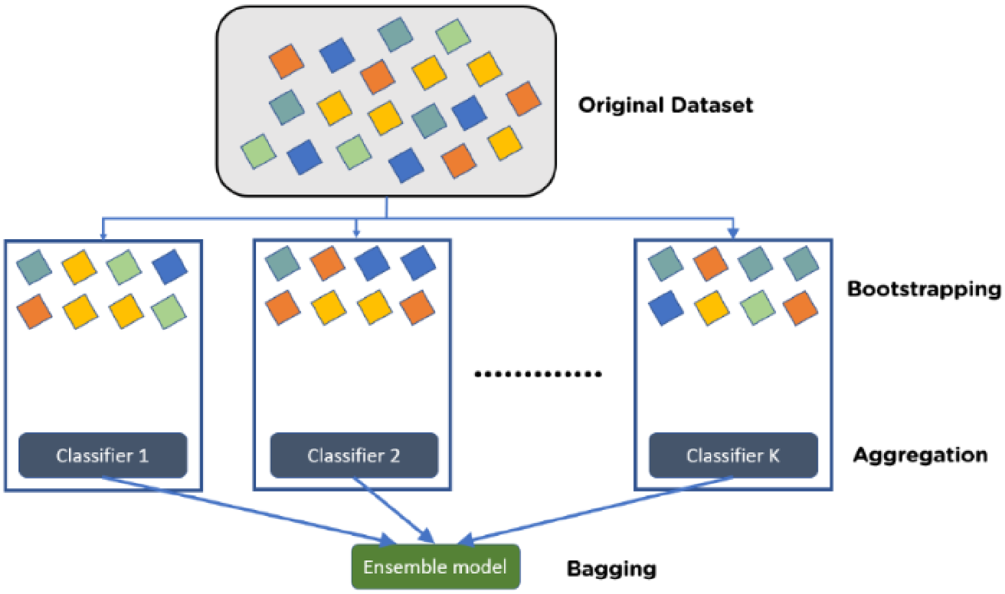



**Figure 2.1 Summary of Working Process of Bagging Process (Avijeet Biswal, 2021)**

Boosting is another ensemble learning technique that only utilizes multiple models from the same algorithm, and it builds stronger models from a bunch of weaker models by reweighing the training instance. After constructing each weak learner, modifications are made to the model weights. Low weights are assigned to simple examples that are successfully categorized by the weak learner, whereas large weights are applied to samples that are misclassified (Khairy et al.,2021). The process won't stop until minimum accuracy is achieved or no further improvements is possible. The working process of boosting technique is shown in figure 2.2

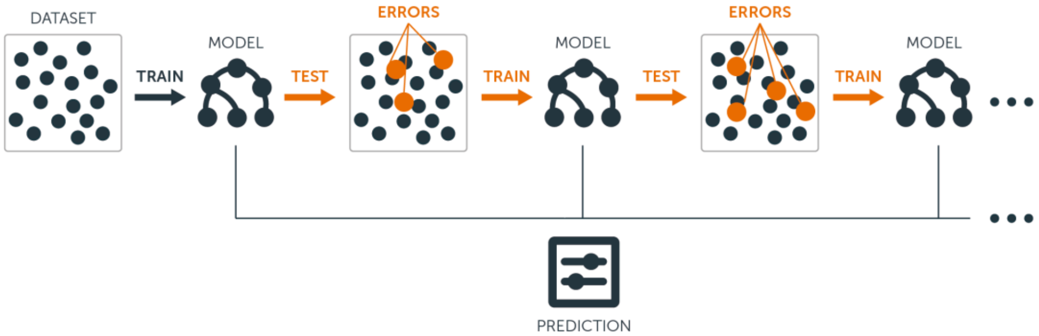



**Figure 2.2 Summary of Working Process of Boosting Process (Sadki, N, 2020)**

The voting technique can combine the predictions from multiple models from diversified algorithms, and then get a final prediction result by exploiting the voting rules (Kumar et al.,2017) The working process of boosting technique is shown in figure 2.3

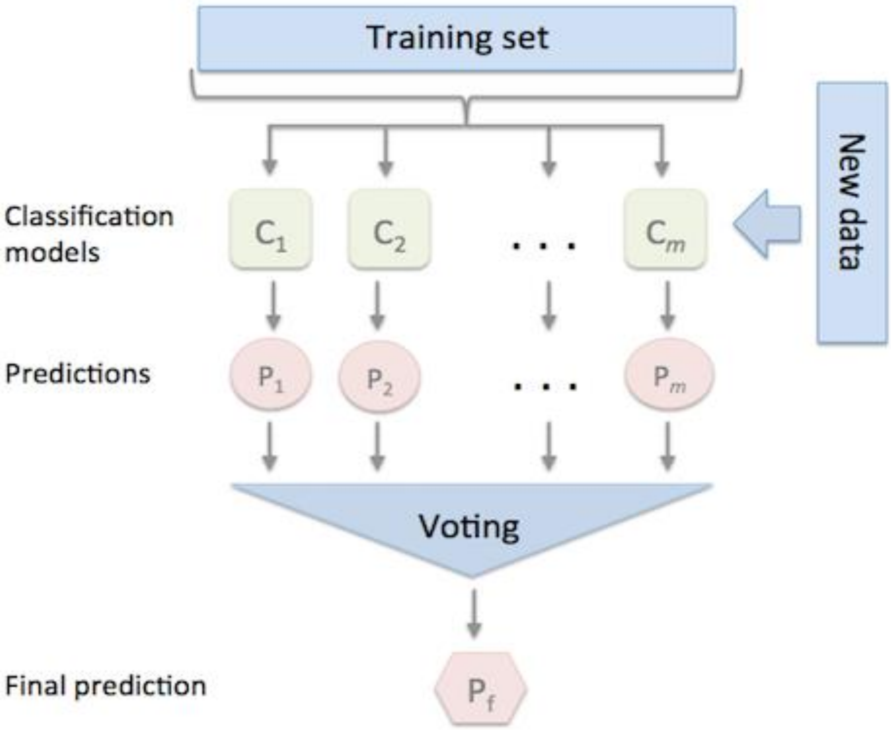



**Figure 2.3 Summary of Working Process of Voting Process (Raschka, 2022)**

There are two kinds of voting, soft voting, and hard voting. Soft voting refers to combining the prediction of each model and picking the prediction with the highest total probability. Therefore, only the probability-related algorithm is applicable in soft voting, and the sum of the probability should be 1. Hard voting refers to including or excluding the model, so only '1' and '0' are included, which indicates include or exclude. The comparison of hard and soft voting can be found in figure 2.4.

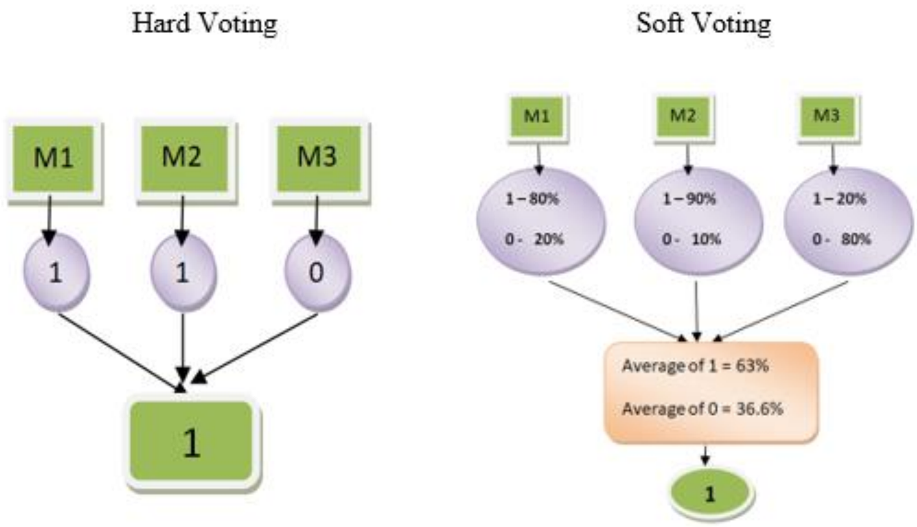



**Figure 2.4 Summary of Soft Voting vs Hard Voting (OpenGenus, 2020)**

Stacking is also known as Stacked Generalization, and it is an ensemble technique that combines multiple algorithms via a meta-algorithm or a final estimator. There are two kinds of estimators in stacking technique, base estimator and final estimator. Base estimator acts as the input generator for the final estimator, so the output of the base estimator is the input of the final estimator. The sequence of the base estimator won't affect the performance of the model, however, if there are any changes in the combination of base estimator or final estimator, the whole stacking model should be considered as a completely new model. The comparison of hard and soft voting can be found in figure 2.5.

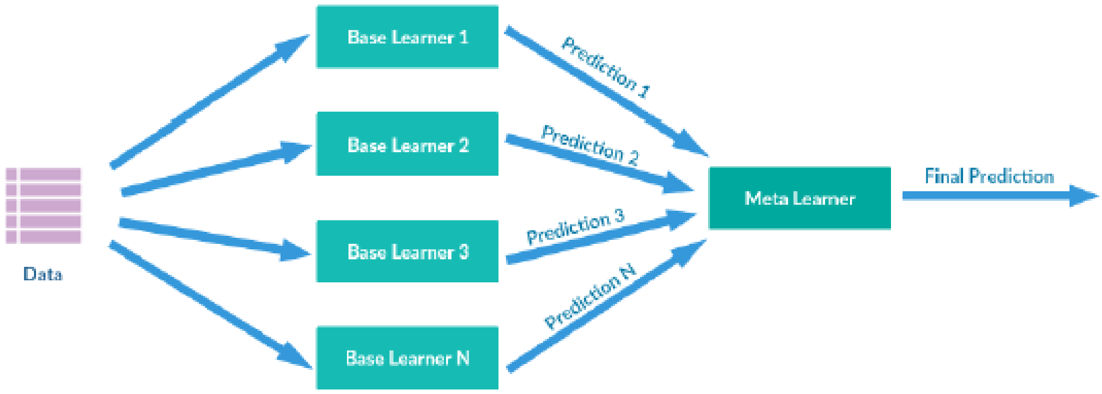



**Figure 2.5 Summary of Working Process of Stacking Process (Towardsai, 2022)**

After going through all the ensemble learning techniques, a comparison table of the techniques can be found in table 2.3.3.

**Table 2.3 Comparison of Ensembled Classification Algorithm**

| Ensemble Techniques | Advantages | Disadvantages |
|---|---|---|
| Bagging | 1. Reduce model variance<br>2. Suitable for complex models | 1. Hard to interpret the models<br>2. May suffer from model bias |
| Boosting | 1. Reduce model bias<br>2. Suitable for simple models | 1. Not suitable for noisy dataset<br>2. Hard to interpret the models |
| Voting | 1.Capability of overcoming defects of multiple algorithms by combining them.<br>2.Utilize the strengths of multiple algorithms | 1. Computationally expensive<br>2. Slow in training |
| Stacking | 1. Training is conducted with the entire dataset<br>2. Utilize multiple algorithms at a same time | 1. Computationally expensive<br>2. Slow in training |

### 2.3.3 Data Science and Analytical Tools

Python is an open-source programming language and is widely used in different domains, for instance, web development, machine learning, scripting, etc. Its rich packages are contributed by all the users around the world, and we the users call them libraries. Most of the time, libraries are predefined functions that can be called directly and build models by setting the parameters only. There are some famous libraries in data science domain. For example, Scikit is a library designed for machine learning; Tensorflow is for deep learning; Matplotlib is for visualization purposes. There are many other libraries out there both in and out of the data science domain. Besides, python also compatible with multiple operation systems such as Linux, Windows, MacOS, and the teams are working on it constantly to release the latest version of python. The lates version of it is Python 3.0 for now, and it is the main tools used in the project.

**Table 2.4 Python Comparison Summary(Prasanna, 2022)**

| Tools | Advantages | Disadvantages |
|---|---|---|
| Python | • Easy to learn and use<br>• Extensive libraries.<br>• Increases productivity.<br>• Flexible to integrate and deploy<br>• Supportive community | • Difficulties to work with other languages<br>• Low execution speed<br>• Many issues with the design of the language, which only gets displayed during runtime.<br>• Not suited for memory-intensive programs and mobile applications. |

# 3 RESEARCH METHODOLOGY

## 3.1 Business Understanding

As a research center, the School of Humanities wishes to know more about the stress, trauma, and anxiety in the war narratives by female writers from Middle East, but reading the books line by line and identifying the sentiment of each sentence is unrealistic when there is a huge number of narratives about wars out there. Besides, women in the war may have a unique perspective toward wars, and such perspective may lead people to understand the war from an interesting point of view. Therefore, the school wishes to have a model that performs well in automatically extracting and classifying the related expressions of stress, trauma, and anxiety from war narratives in the Middle East. In this way, the model will help researchers to uncover more insights about those negative sentiments in war narratives.

## 3.2 Project Framework

The framework of this project is illustrated in figure 3.1, and there are 6 steps in it.

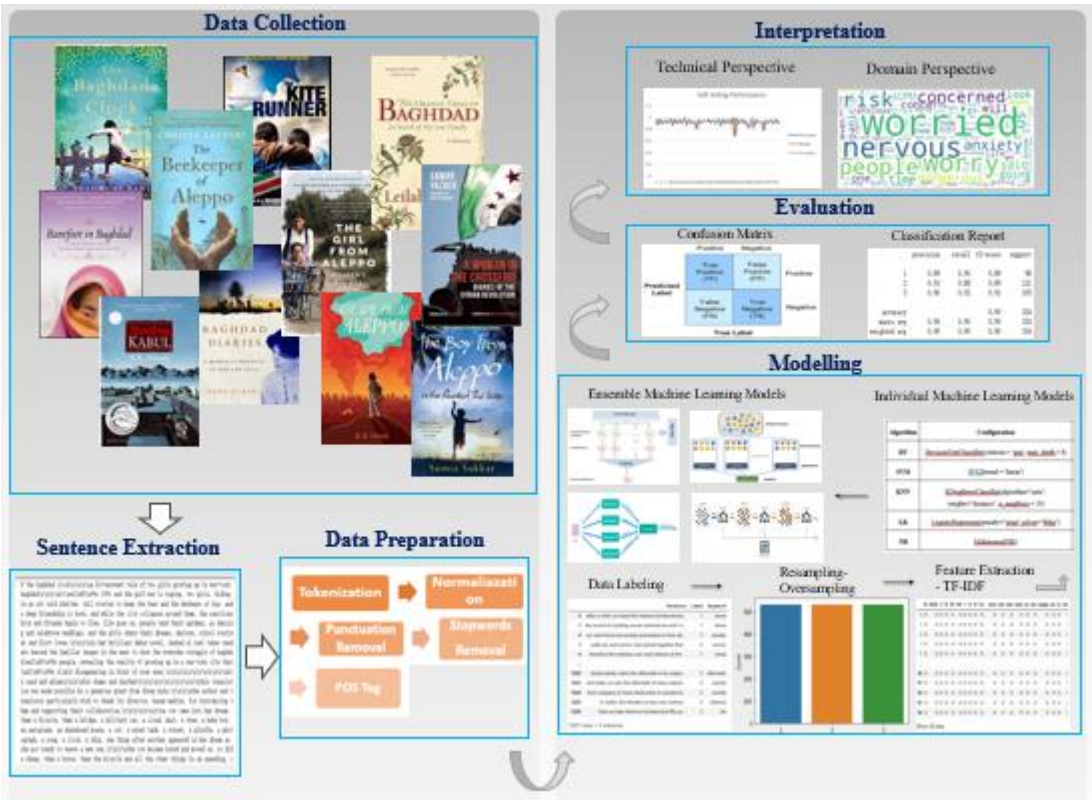



**Figure 3.1 Project Framework**

### 3.2.1 Data Collection and Data Understanding

From what Dr. Moussa has provided, ten typical novels about wars from eight famous female writers in Middle East act as the dataset. Due to the nature of the novel, all the data involved is text data. The summary of the dataset can be found in Table 3.1.

**Table 3.1 Dataset Description and Summary**

| No. | Name | Description | Writer |
|---|---|---|---|
| 1 | The Baghdad Clock | A tale of two friends growing up during the first and second Iraqi war | al Rawi, Shahad |
| 2 | Baghdad Diaries 1991-2002_ A Woman's Chronicle of War and Exile | Author's diary started from 1991 during the first Gulf war | al Rawi, Shahad, Nuha |
| 3 | The Beekeeper of Aleppo_ A Moving Testament to the Human Spirit | A beekeeper who was forced to flee away from Syria | Lefteri, Christy |
| 4 | The Girl from Aleppo_ Nujeen's Escape from War to Freedom | A disabled girl escape from Syria to Germany | Mustafa, Nujeen |
| 5 | The Orange Trees of Baghdad_ In Search of My Lost Family_ A Memoir | A story from the author about her immigration experience | Nadir, Leilah |
| 6 | Barefoot in Baghdad | A story about a lady moves to Iraq to help as many women as she can rebuild their lives | Omar, Manal M |
| 7 | Escape from Aleppo | A story about a family in Syria flees to safety | Senzai, N H |
| 8 | Shooting Kabul | A story about a family flees from Afghanistan to the U.S. | Senzai, N H |
| 9 | The Boy from Aleppo Who Painted the War | A story about a boy who attempts to understand the Syrian conflict and its effect on his life by painting his feelings | Sukkar, Sumia |
| 10 | A Woman in the Crossfire_ Diaries of the Syrian Revolution | A story about the author on how she struggles to protect herself and her young daughter in the war | Yazbek, Samar |

All the data mentioned in the table above mimic stories happened in the Middle East during war time, and they are all from female authors who are mainly from Syria and Iraq. It is also clear from the 'Description' that most of the novels focus on family or civilians' lives in the war.

### 3.2.2 Sentence Extraction

As the purpose of the study is to propose a model that is working well in the overall dataset, study the novels one by one is meaningless. Instead, I merged all the novels into one dataset and study it. Before preprocessing steps, a file of raw extracted sentence was generated, and the preprocessing actions would be taken based on this file. Figure 3.2 depicts the file.

```
: b'the baghdad clock\r\n\r\na bittersweet tale of two girls growing up in war-torn
  baghdad\r\n\r\nit\xe2\x80\x99s 1991 and the gulf war is raging. two girls, hiding
  in an air raid shelter, tell stories to keep the fear and the darkness at bay, and
  a deep friendship is born. and while the city collapses around them, the sanctions
  bite and friends begin to flee, life goes on. people tend their gardens, go dancin
  g and celebrate weddings, and the girls share their dreams, desires, school routin
  es and first loves.\r\n\r\nin her brilliant debut novel, shahad al rawi takes read
  ers beyond the familiar images in the news to show the everyday struggle of baghda
  d\xe2\x80\x99s people, revealing the reality of growing up in a war-torn city that
  \xe2\x80\x99s slowly disappearing in front of your eyes.\r\n\r\n\r\n\r\n\r\n\r\nfo
  r saad and ahlam\r\n\r\nfor shams and shather\r\n\r\n\r\n\r\n\r\n\r\nthis translat
  ion was made possible by a generous grant from diwan kufa.\r\n\r\nthe author and t
  ranslator particularly wish to thank its director, kanan makiya, for introducing t
  hem and supporting their collaboration.\r\n\r\n\r\n\r\na cow came into her dream.
  then a bicycle. then a bridge. a military car. a cloud. dust. a tree. a baby boy.
  an aeroplane. an abandoned house. a cat. a water tank. a street, a giraffe, a phot
  ograph. a song. a clock. a ship. one thing after another appeared in her dream as
  she got ready to weave a new one.\r\n\r\nthe cow became bored and moved on. so did
  a sheep. then a horse. then the bicycle and all the other things in an unending, c
```

**Figure 3.2 Merged Dataset Output**

### 3.2.3 Data Preparation

Data preparation is necessary in any project when the raw data is not ready to be fed to models, and different problems requires different preparation steps. In this project, data preparation follows typical preparation steps of Natural Language Processing (NLP) problem. The specific preparation steps can be found in figure 3.3.


**Figure 3.3 Text Preparation Steps**

#### 3.2.3.1 Tokenization

Tokenization is a common task in NLP, and it refers to process of breaking text data into sentences, words, or short phrases. The main purpose of tokenization is to create a structured dataset that is suitable to feed to the model. I conducted it by using the NLTK package in python.

There are four types of tokenization, which are word tokenizer, regular expression tokenizer, word punctuation tokenizer, and sentence tokenizer. The name of the tokenizer suggests the output of the tokenized results, and we need sentence specific output for the purpose of data labelling and modelling. Therefore, sentence tokenizer have been used in this study. Figure 3.4 illustrates the tokenized dataset.

| | 0 |
|---|---|
| 0 | the baghdad clock\r\n\r\na bittersweet tale of... |
| 1 | two girls, hiding in an air raid shelter, tell... |
| 2 | and while the city collapses around them, the ... |
| 3 | people tend their gardens, go dancing and cele... |
| 4 | in her brilliant debut novel, shahad al rawi t... |
| ... | ... |
| 46670 | index\r\n\r\n\r\na\r\n\r\nabbasid caliph mansu... |
| 46671 | her memoir the orange trees of baghdad won the... |
| 46672 | her fiction has appeared in descant and on cbc... |
| 46673 | she has written and broadcast commentaries for... |
| 46674 | www.leilahnadir.com |

46675 rows × 1 columns

**Figure 3.4 Tokenized Dataset**

#### 3.2.3.2 Normalization

Normalization in NLP can also be interpreted as standardization of tokens, and it usually consists of lemmatization, stemming, and lowercase. The purpose of this step is to convert the form of the word to its base form, and potentially the unique tokens can be decreased significantly by applying this step.

In this study, only lemmatization has been applied because sometimes the results of stemming are meaningless when it removes the suffix or unable to convert the words to base form when the transformation of the words is irregular. A good example of this is that stemming will regard ‘see’, ‘saw’, ‘seen’ as three unique words. Unlike stemming, lemmatization convert the words to its root form ‘see’, which makes the research more meaningful and ease the process to establish the model and interpret the results. Figure 3.5 illustrates the normalized dataset.

| | Sentence |
|---|---|
| 0 | the baghdad clock a bittersweet tale of two gi... |
| 1 | two girls, hiding in an air raid shelter, tell... |
| 2 | and while the city collapse around them, the s... |
| 3 | people tend their gardens, go dancing and cele... |
| 4 | in her brilliant debut novel, shahad al rawi t... |
| ... | ... |
| 46670 | index a abbasid caliph mansur, 51 abraham, 57,... |
| 46671 | her memoir the orange tree of baghdad won the ... |
| 46672 | her fiction ha appeared in descant and on cbc ... |
| 46673 | she ha written and broadcast commentary for th... |
| 46674 | www.leilahnadir.com |

46675 rows × 1 columns

**Figure 3.5 Normalized Dataset**

3.2.3.3 Punctuation Removal

Removing punctuations is necessary because computer regards them as a unique symbol just as a word, but it is actually meaningless in human language. Besides, punctuation here doesn’t only mean the punctions when we use in English, it includes all the special symbol in the text. All the them shall be removed in the preparation stage. Figure 3.6 illustrates the punctuation removed dataset.

| | Sentence |
|---|---|
| 0 | the baghdad clock a bittersweet tale of two gi... |
| 1 | two girls hiding in an air raid shelter tell s... |
| 2 | and while the city collapse around them the sa... |
| 3 | people tend their gardens go dancing and celeb... |
| 4 | in her brilliant debut novel shahad al rawi ta... |
| ... | ... |
| 46670 | index a abbasid caliph mansur 51 abraham 57 11... |
| 46671 | her memoir the orange tree of baghdad won the ... |
| 46672 | her fiction ha appeared in descant and on cbc ... |
| 46673 | she ha written and broadcast commentary for th... |
| 46674 | wwwleilahnadircom |

46675 rows × 1 columns

**Figure 3.6 Punctuation Removed Dataset**

#### 3.2.3.4 Stopword Removal

Stopword are meaningless words, most of them are small words like preposition, pronoun, modals, and etc. They are in human language but carrying no meaning at all. Due to the grammatical function requirements, those words are usually quite high in frequency. Therefore, removal is required before modelling, otherwise it will cost too much program resources and bring noise to the modelling too. Luckily, NLTK provide a package of English stopwords, and I was allowed to add other stopwords to the list if needed. Figure 3.7 illustrates the stopwords removed dataset.

| | Sentence |
|---|---|
| 0 | baghdad clock bittersweet tale two girl gro... |
| 1 | two girls hiding air raid shelter tell story... |
| 2 | city collapse around sanction bite frien... |
| 3 | people tend gardens go dancing celebrate wed... |
| 4 | brilliant debut novel shahad al rawi take re... |
| ... | ... |
| 46670 | index abbasid caliph mansur 51 abraham 57 115... |
| 46671 | memoir orange tree baghdad 2008 george ry... |
| 46672 | fiction ha appeared descant cbc radio |
| 46673 | ha written broadcast commentary cbc globe... |
| 46674 | wwwleilahnadircom |

46675 rows × 1 columns

**Figure 3.7  Stopwords Removed Dataset**

### 3.2.3.5 Part-of-speech (POS) Tagging

Part-of-speech (POS) tagging refers to assigning categorical tags to tokens based on the syntactic functions of the tokens. This means that the same word may have different tags in different cases when the syntactic functions are different. Besides, other grammatical information can be generated too, such as tense, number, etc. Therefore, POS can help researchers to unwrap the secrets of the grammatical side of the text, the writing habit of the author, the style of the words, etc. In this study, POS was applied for data exploration purposes to let us know more about the narratives from a grammatical perspective. Table 3.2 is the list of Penn Treebank POS tags and its description, and Figure 3.8 illustrates the POS tagging output.

**Table 3.2 POS Tag List (**)**

| Tag | Description | Tag | Description |
|---|---|---|---|
| CC | coord, conjunction | RB | adverb |
| CD | cardinal number | RBR | adverb, comparative |

| | | | |
|---|---|---|---|
| DT | determiner | RBS | adverb, superlative |
| EX | existensial there | RP | particle |
| FW | foreign word | SYM | symbol |
| IN | preposition or subconjunction | TO | “to” |
| JJ | adjective | UH | interjection |
| JJR | adjective, comparative | VB | verb, base form |
| JJS | adjective, superlative | VBD | verb, past tense |
| LS | list item marker | VBG | verb, gerund |
| MD | modal | VBN | verb, past participle |
| NN | noun, singular or mass | VBP | verb, non-3sg pres |
| NNS | noun, plural | VBZ | verb, 3sg-pres |
| NNP | proper noun, sing. | WDT | wh-determiner |
| NNPS | proper noun, plural | WP | wh-pronoun |
| PDT | predeterminer | WPS | possessive wh- |
| POS | possessive ending | WRB | wh-adverb |
| PRP | personal pronoun | | |

** Source: https://ubc-cs.github.io/cpsc330/lectures/13_feature-engineering-selection.html

```
[('baghdad', 'NN'), ('clock', 'NN'), ('bittersweet', 'NN'), ('tale', 'JJ'), ('tw
o', 'CD'), ('girl', 'VBP'), ('growing', 'VBG'), ('wartorn', 'NNS'), ('baghdad', 'N
N'), ('1991', 'CD'), ('gulf', 'NN'), ('war', 'NN'), ('raging', 'VBG')]
[('two', 'CD'), ('girls', 'NNS'), ('hiding', 'VBG'), ('air', 'NN'), ('raid', 'VB
D'), ('shelter', 'NN'), ('tell', 'NN'), ('story', 'NN'), ('keep', 'VB'), ('fear',
'NN'), ('darkness', 'NN'), ('bay', 'NN'), ('deep', 'JJ'), ('friendship', 'NN'),
('born', 'VBN')]
[('city', 'NN'), ('collapse', 'NN'), ('around', 'IN'), ('sanction', 'NN'), ('bit
e', 'NN'), ('friend', 'NN'), ('begin', 'NN'), ('flee', 'JJ'), ('life', 'NN'), ('g
o', 'VB')]
[('people', 'NNS'), ('tend', 'VBP'), ('gardens', 'NNS'), ('go', 'VB'), ('dancing',
'VBG'), ('celebrate', 'JJ'), ('weddings', 'NNS'), ('girl', 'NNS'), ('share', 'N
N'), ('dreams', 'VBZ'), ('desires', 'VBZ'), ('school', 'NN'), ('routine', 'NN'),
('first', 'JJ'), ('loves', 'NNS')]
[('brilliant', 'JJ'), ('debut', 'NN'), ('novel', 'NN'), ('shahad', 'VBD'), ('al',
'JJ'), ('rawi', 'NNS'), ('take', 'VBP'), ('reader', 'NN'), ('beyond', 'IN'), ('fam
iliar', 'JJ'), ('image', 'NN'), ('news', 'NN'), ('show', 'NN'), ('everyday', 'J
J'), ('struggle', 'JJ'), ('baghdads', 'NNS'), ('people', 'NNS'), ('revealing', 'VB
G'), ('reality', 'NN'), ('growing', 'VBG'), ('wartorn', 'JJ'), ('city', 'NN'), ('t
```

**Figure 3.8 POS Tagging Output**

## 3.3 Text Exploration

Text exploration helps to reveal hidden patterns of the data, which could provide a rough idea for most of the topics in text data. To implement the exploration, I applied word cloud to show the word frequency, and bar chart to show the sentiments distribution. The exploration results will be discussed in Chapter 4.

## 3.4 Modeling

Model construction comes after the exploration procedure, and it mainly consists of three parts, data labeling, feature extraction, and modeling establishment.

### 3.4.1 Data Labeling

From Figure 3.7, it is clear that the dataset only one dimension, ‘Sentence’. However, the feature 'Sentence' alone is not enough to build the model, because we need to know the sentiment of the sentences. Therefore, assigning labels to each sentence is vital before constructing the model, and this process is also known as data labeling.

The labeling process started from determining the keywords of stress, trauma, and anxiety. From the existing literatures, a table of keywords which triggers the label has been generated, and the keywords dictionary can be found in table 3.3.

**Table 3.3 Table of List of Words for Keywords Dictionary of Stress, Trauma, and Anxiety (**)**

| Class | Words Related |
|---|---|
| stress (1) | stress, stressed, stresses, stressing, stressful, stressfully, tension, tense, tensed, tensing, pressure, pressures, pressuring, pressurize, pressurized, threat, threaten, threatened, threatening, threatens, threateningly, tightness, tight, tighten, tightens, tightening, tightened, intensity, tensity, forced, strain, strains, strained, straining, tenseness, |
| trauma (2) | agony, trauma, traumatized, traumatise, traumatised, traumatises, traumatising, traumatizing traumatize, traumatizes, traumatic, traumatically, offense, offend, offended, offends, abuse, abuses, abused, abusive, abusively, insult, insults, insulted, insulting, insultingly, affront, affronted, affronts, affronting, torture, tortures, tortured, torturing, massacre, massacred, massacres, massacring, tragedy, tragedies, tragic, tragical, tragically, beat, beaten, beating, beats, kill, kills, killing, distress, distressed, distressing, sorrow, sorrowful, sorrowfully, distressful anguish, anguishes, anguished, anguishing, torment, tormented, torments, tormenting, desperate, desperation, desperately, insecurity, insecure, gasp, gasps, gasped, arrest, arrests, arresting, arrested, tear, tears, teardrop, cry, crying, cried, die, died, dying, death, suicide, suicided, suicides, despair, despairing, despaired, despairs, doom, dead, shivering, shiver, shivered, humiliate, humiliation, humiliated, humiliating, scare, scared, scaring, violence, terrify, terrified, terrific, fear, fears, feared, fearful, aftermath, |
| anxiety (3) | anxiety, anxieties, anxious, anxiously, worry, worries, worried, worrying, concern, concerns, concerned, agitation, agitate, agitated, agitates, perturbation, perturbative, perturb, perturbed, perturbs, panic, panicked, panicking, panics, disquiet, disquiets, disquieting, disquieted, disquietingly, disquietude, risk, risks, risky, could be killed, would be killed, angst, agonize, agonizes, agonized, agonizing, unease, uneasy, uneasily, restless, troublous, unrestful, jittery, jitteriness, panicky, uptight, apprehensive, apprehensiveness, apprehensively, restlessness, nervous, nervously, nervousness, worse, worsen, worsening, worsened, worsens, hopeless, hopelessness, dangerous, danger, dangers, would have, could have, helpless |

*** There will be cases that one keyword is related to multiple sentiments; the final label is determined by manual selection. Such words are included in merely one kind of sentiment in table 3.3.*

According to the explanation of each sentiment in 2.2.1, the keywords come from the previous works and the synonyms. And then, searched every word in each sentence, and any sentence that contains the keyword would be labeled accordingly. Stress has been labelled as ‘1’, trauma has been labelled as ‘2’, and anxiety has been labelled as ‘3’. After searching every keyword in the sentences, a new dataset with keywords and label column has been generated, and it could be found in the figure 3.9. Due to the fact that only a small proportion of sentences contains elements related to ‘stress’, ‘trauma’, and ‘anxiety’, the dataset is expected to shrink after the labeling process.

| | Sentence | keyword | Label |
|---|---|---|---|
| **1** | two girls hiding air raid shelter tell story... | fear | 2 |
| **71** | heard intense bombardment followed sire... | tense | 1 |
| **89** | fear departed wa time sleep | fear | 2 |
| **105** | long week lived fear cold hunger sharing... | fear | 2 |
| **120** | im devil yes malaika started cry went si... | cry | 2 |
| **...** | ... | ... | ... |
| **46575** | emotional cried together | cried | 2 |
| **46596** | writes back house wa teardrop shape bend ... | tear | 2 |
| **46597** | river flow side part city anyone fam... | tear | 2 |
| **46641** | thank family canada england iraq heart ... | die | 2 |
| **46670** | index abbasid caliph mansur 51 abraham 57 115... | beat | 2 |

5073 rows × 3 columns

**Figure 3.9 Labeled Dataset Output**

However, keywords-oriented labeling alone cannot provide a satisfied degree of accuracy of label assignment due to the fact of complexity nature of human language. For example, the word ‘worry’ is surely attached with the sentiment anxiety, but detecting the word ‘worry’ doesn’t necessarily mean that the sentence should be labeled as ‘3’. Such sentence can be ‘Don’t worry.’ Therefore, manual processing and reassigning the labels where required is essential to improve the labeling accuracy. Only human can fully comprehend the sentiments behind the sentences. After doing so, the dataset shrank further, and the dataset after manual processing is shown in figure 3.10.

| | Sentence | Label | keyword |
|---|---|---|---|
| 0 | after a while we heard the intense bombardment... | 1 | tense |
| 1 | the sound of crashing waves reached me and i w... | 1 | tense |
| 2 | we went home but anxiety prevented us from sle... | 3 | anxiety |
| 3 | sadly joy and sorrow were joined together that... | 2 | sorrow |
| 4 | therefore the sadness was most intense at the ... | 1 | tense |
| ... | ... | ... | ... |
| 1222 | farahs family watch the aftermath of an explos... | 2 | aftermath |
| 1223 | and today we saw the aftermath of many explosi... | 2 | suicide |
| 1224 | from weapons of mass destruction to suicide bo... | 2 | suicide |
| 1225 | in reality, the situation in iraq was continui... | 2 | violence |
| 1226 | i find out later that two hundred and fifty pe... | 2 | die |

1227 rows × 3 columns

**Figure 3.10 Manual Labeled Dataset Output**

### 3.4.2 Data Resampling

To solve a classification problem, it is vital to check the balance status before training any models, because it will affect the model performance and the evaluation metrics applied. Figure 3.11 illustrate the status of the manual labeled dataset.

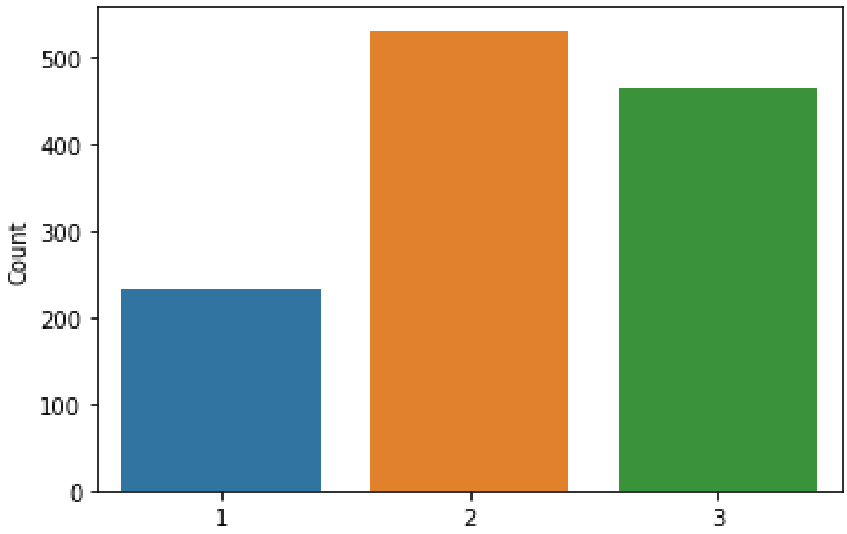



**Figure 3.11  Sentiments Distribution of Overall Dataset**

There are 565 samples in label ‘2’; 406 samples in label ‘3’; 256 samples in label ‘1’. It is obvious that the dataset is imbalanced, so I applied resampling technique to avoid any negative effect on performance it may have. There are three types resampling, oversampling, under-sampling, and hybrid. I have implemented the random oversampling technique to resample it, and it aligns the number of labeled data to the majority class. After resampling, the new distribution bar charts can be found in figure 3.12, and the number of each sentiment is 565.

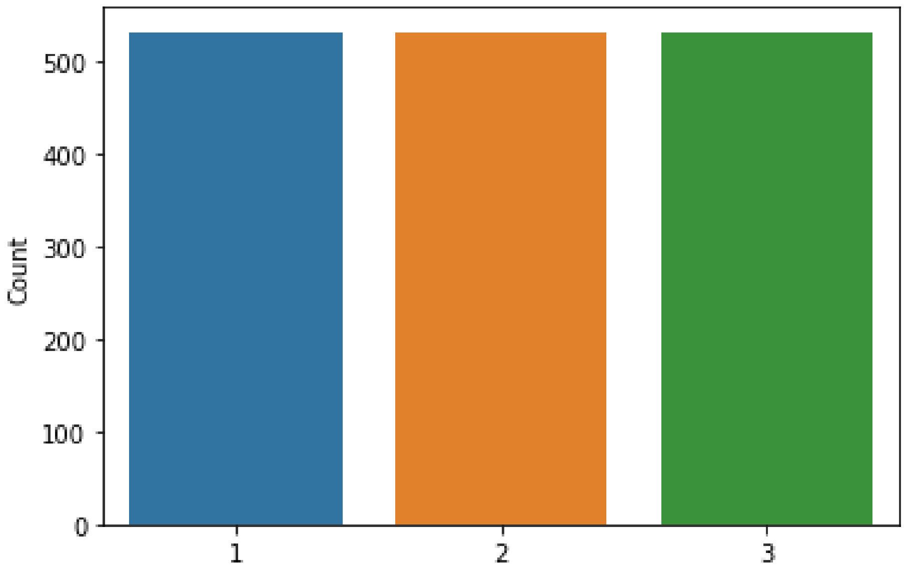



**Figure 3.12  Sentiments Distribution of Overall Dataset After Resampling**

After resampling, the final balanced labeled dataset generated is ready for next step, and there are around 1600 rows of data. It is shown in figure 3.13 as below.

| | Sentence | keyword | Label |
|---|---|---|---|
| 0 | after a while we heard the intense bombardment... | tense | 1 |
| 1 | the sound of crashing waves reached me and i w... | tense | 1 |
| 2 | we went home but anxiety prevented us from sle... | anxiety | 3 |
| 3 | sadly joy and sorrow were joined together that... | sorrow | 2 |
| 4 | therefore the sadness was most intense at the ... | tense | 1 |
| ... | ... | ... | ... |
| 1690 | i understood his anxiety | anxiety | 3 |
| 1691 | while living in the united states zafoona had ... | concern | 3 |
| 1692 | were worried how much endurance mrs merkel has | worried | 3 |
| 1693 | we heard they were running out of space for re... | worried | 3 |
| 1694 | the seconds passed with agonizing slowness but... | agonizing | 3 |

1695 rows × 3 columns

**Figure 3.13 Balanced Manual Labeled Dataset**

### 3.4.3 Feature Extraction

For now, the labeled dataset is ready, but computer cannot read the sentence, so we will need to transform the text data to numerical matrix in order for the algorithm to proceed. In this project, TF-IDF Vectorizer has been implemented. TF-IDF stands for Term Frequency - Inverse Dense Frequency, and it is a method for determining the meaning of sentences made out of words that overcomes the limitations of the Bag of Words technique, which is useful for text classification or assisting a machine in reading words in numbers (Madan, R., 2019). As a mature technique in text data preprocessing, TF-IDF Vectorizer is one of the most stable and widely used in extracting the features. The dataset with extracted features and labels can be found in figure 3.14.

| | 000 | 04022016 | 10 | 100 | 1000 | 10000 | 11 | 110 | 120 | 1230 | ... | zafoona | zainab | zalmay | zalmays | zayn | zeena | zigzagging | zone | zoo | Label |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1 |
| 1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1 |
| 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3 |
| 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2 |
| 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1690 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3 |
| 1691 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3 |
| 1692 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3 |
| 1693 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3 |
| 1694 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3 |

1695 rows × 4331 columns

**Figure 3.14 Dataset with Extracted Features**

### 3.4.4 Machine Learning Modeling Construction

The computer-readable dataset is ready, and the machine learning will be constructed based on the dataset shown in figure 3.14. As the dataset has already been labelled, only supervised machine learning algorithm would be considered. The modelling part consists of individual models and ensemble learning models. Individual models will be trained based on one model only among Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Naïve Bayes (NB), and K-Nearest Neighbor (KNN). The optimal individual models will be the input of the ensemble learning model. In ensemble learning section, ensemble learning techniques including bagging, boosting, voting, and stacking will be experimented. All the possible algorithm combination shall be included too.

Bagging technique prepares the dataset to multiple subsets. Each subset is exactly the same size, and each subset is applied to train the model. There are five individual algorithms in the project, but only one algorithm can be applied in a single experiment. This indicates that there will be five models generated from bagging technique, and they are models adopting KNN/LR/SVM/DT/NB respectively. After training, the test set is applied to the model to generate the prediction. The final bagging model will be formed finally after aggregation. The architecture of bagging technique is shown in figure 4.15.

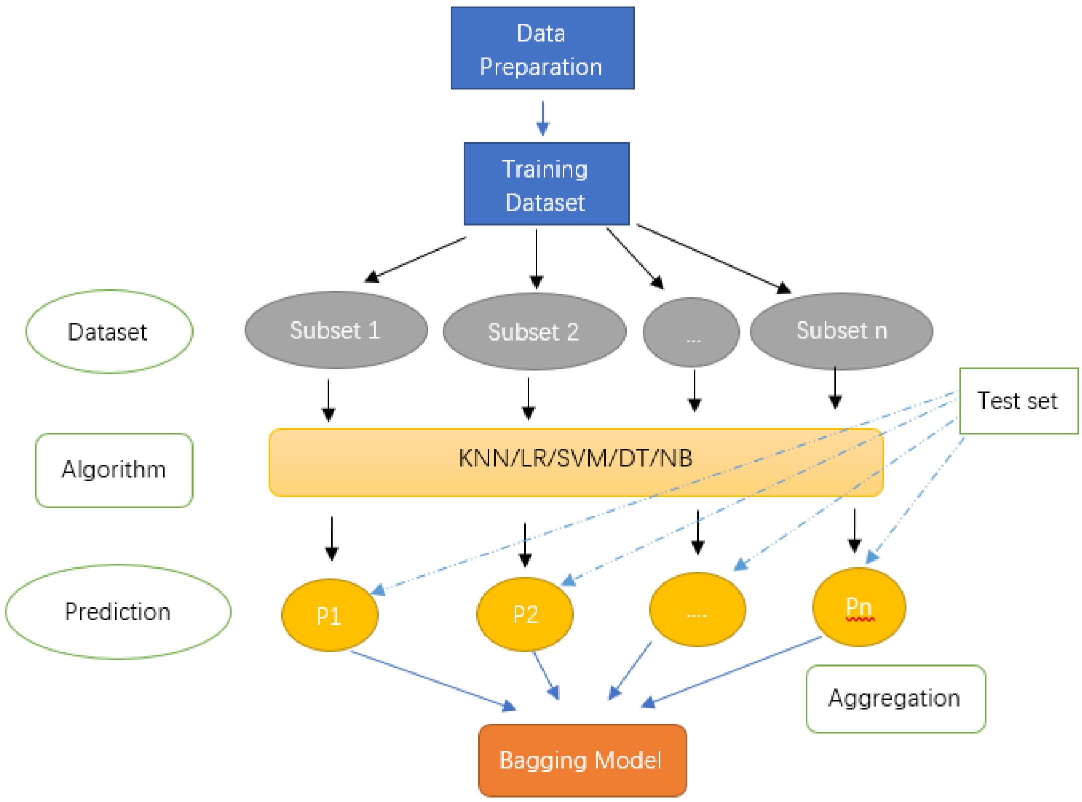



**Figure 3.15 Bagging Technique Architecture**

Boosting technique is quite similar with bagging technique. It also merely applies one algorithm a time, and follow the same process from subset generation, algorithm application, prediction, and aggregation. The different point between the two is the way to generate the subset. In Boosting, the subset is generated in a row by assigning different weight to all samples, and the weight is determined by the previous experiment. Finally, five models applying KNN/LR/SVM/DT/NB will be generated. The architecture of boosting technique is shown in figure 4.16.

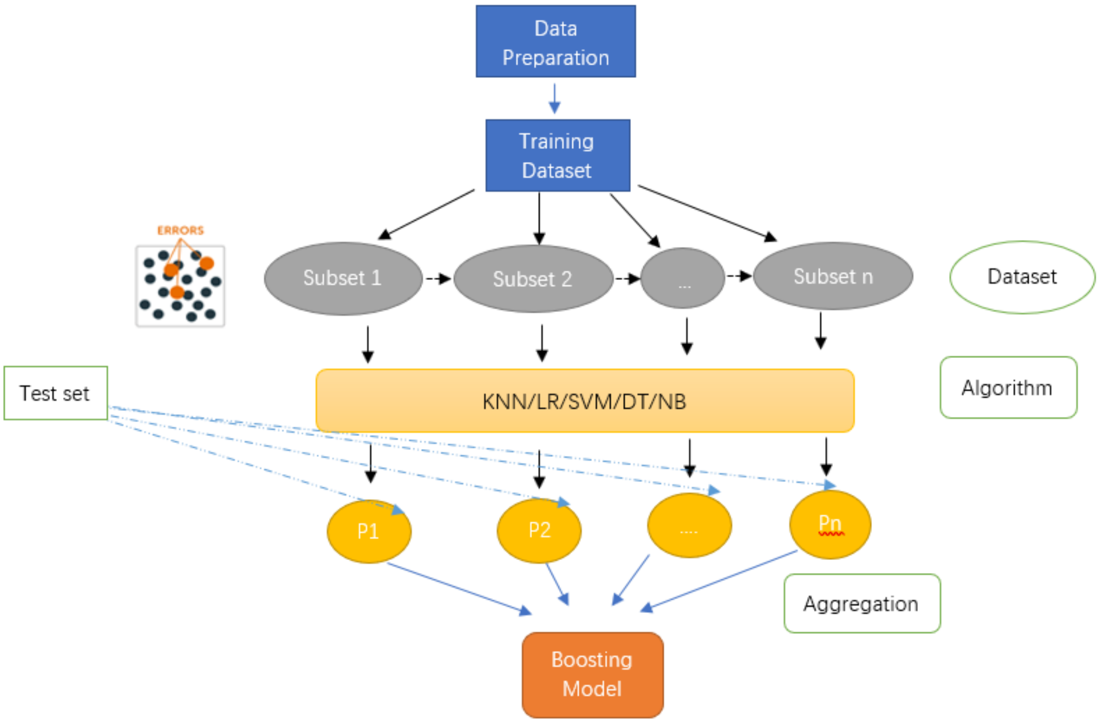



**Figure 3.16 Boosting Technique Architecture**

Voting technique differs from bagging technique and boosting technique mainly by the dataset preparation and algorithm arrangement. There will be no subset generation, and only the original training set will be applied in voting technique. As for algorithms arrangement, more than one algorithm must be chosen from the algorithm pool because if there is only one algorithm exists in the technique, the results it gets is the results of that individual algorithm. Though it follows the same architecture from algorithm to aggregation, the subset generation part disappears and the combination of multiple algorithms appear. The number of algorithms in the combination pool ranges from two from five, and each of the combination will be deployed in later experiments. There will be 26 models generated by applying the voting technique. The architecture of voting technique is shown in figure 4.17.

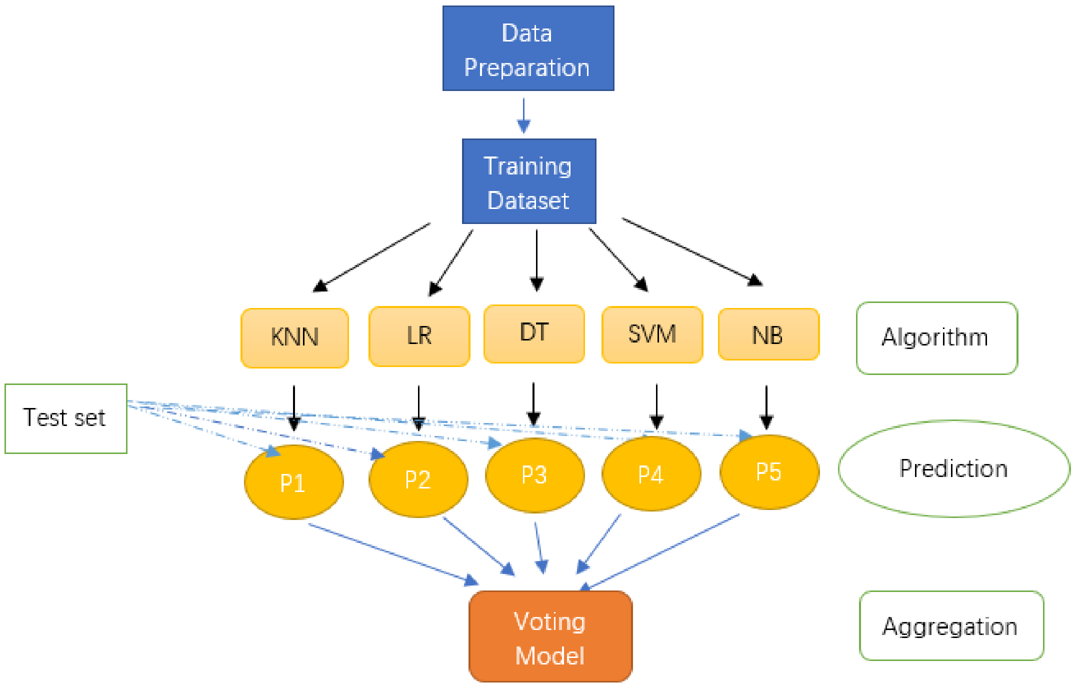



**Figure 3.17 Voting Technique Architecture**

Stacking technique differs from the first three techniques in many ways. Like voting technique, there is no step for subset generation, and the original dataset will be the only dataset exists in the experiments. Stacking technique aggregates the prediction by using a final estimator, the output of the base estimators will be the inputs of final estimator to generate the final results. As the function of the base estimator and final estimator is quite different, the complexity of combination increased. For example, for the combination of KNN/SVM/DT/LR/NB, it no longer stands for one model. Instead, it represents five models, each individual algorithm acts as final estimator. There will be 60 models generated from the stacking technique. The architecture of stacking technique is shown in figure 4.18.

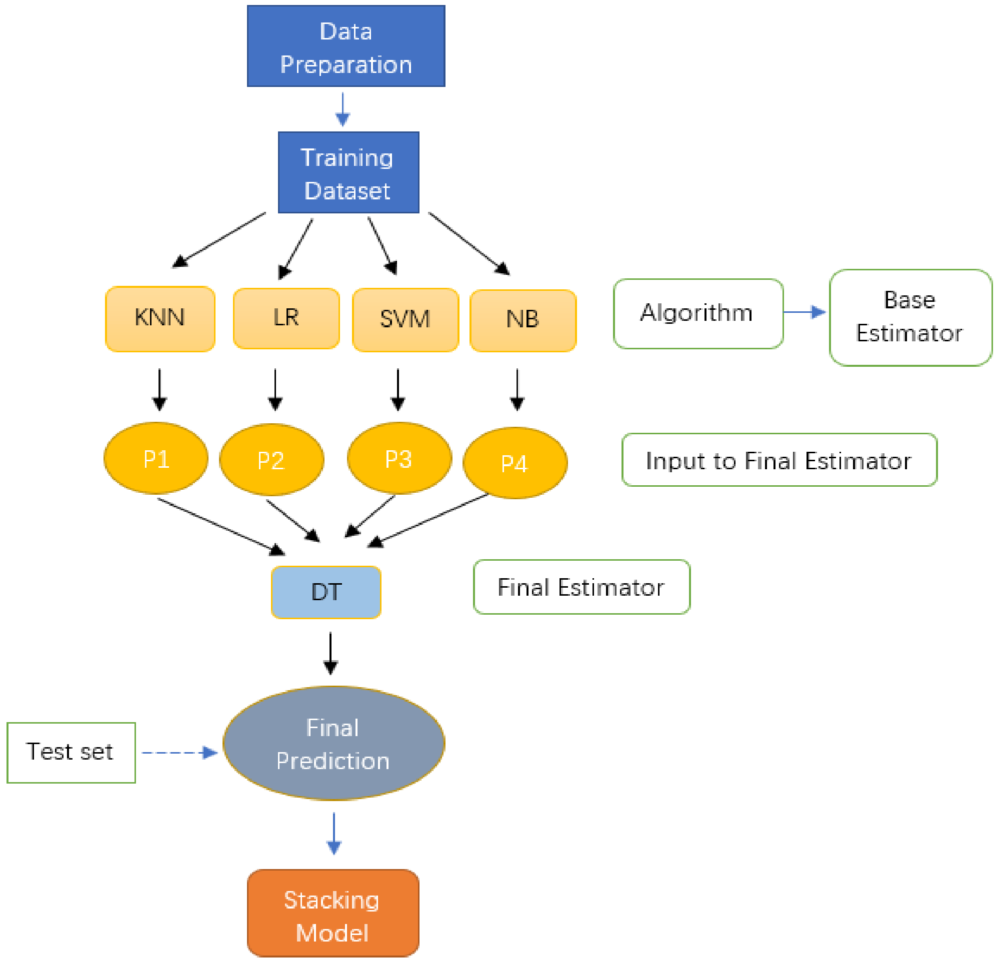



**Figure 3.18 Stacking Technique Architecture**

## 3.5 Evaluation

There are multiple metrics available to evaluate the models, but only some of them are applicable because of the nature of the model. In our project, only supervised classification metrics will be considered. Different metrics may give different conclusion, so evaluation is a crucial step after modelling.

The main evaluation method used in in this project is confusion metrics, and there are multiple metrics can be generated from the confusion matrix. I used classification report to reveal most of them such as precision, recall, f1-score, accuracy score. The final recommendation will be given based on all the four metrics above. Accuracy score is calculated based on the overall prediction capability, which means that if the data is imbalanced, the metrics can be twisted. Instead, precision, recall, and f1-score are focused more on the performance of each class. Though the performance of each class is less crucial compared to the overall model performance, it still conveys rich information about prediction capability. Precision predicts how many cases are truly positive among the stated positive cases. Recall predicts the fraction of predicted positive cases among all the truly positive cases. F1-score combines both precision and recall, and it is a harmonical mean of precision and recall. All of the metrics range from 0 to 1, and the higher the score, the better the performance of the model. The relationship of confusion matrix, accuracy score, precision, recall, and F1-score are shown in figure 3.19 and table 3.4.

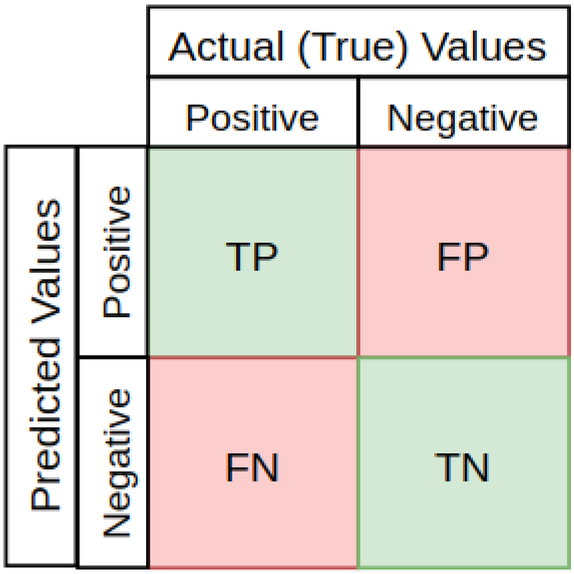



**Figure 3.19 Confusion Metrics (Teemu Kanstrén, 2021)**

**Table 3.4 Table of Accuracy, Precision, Recall, and F1-score (Teemu Kanstrén, 2021)**

| Metrics | Formula |
|---|---|
| Accuracy Score | $\frac{\text{True Positives + True Negatives}}{\text{True Positives + True Negatives + False Positives + False Negatives}} = \frac{\text{N. of Correct Predictions}}{\text{N. of all Predictions}} = \frac{\text{N. of Correct Predictions}}{\text{Size of Dataset}}$ |
| Precision | $\frac{\text{True Positives}}{\text{True Positives + False Positives}} = \frac{\text{N. of Correctly Predicted Positive Instances}}{\text{N. of Total Positive Predictions you Made}} = \frac{\text{N. of Correctly Predicted People with Cancer}}{\text{N. of People you Predicted to have Cancer}}$ |

| | |
|---|---|
| Recall | $\frac{\text{True Positives}}{\text{True Positives + False Negatives}} = \frac{\text{N. of Correctly Predicted Positive Instances}}{\text{N. of Total Positive Instances in the Dataset}} = \frac{\text{N. of Correctly Predicted People with Cancer}}{\text{N. of People with Cancer in the Dataset}}$ |
| F1-score | $2 * \frac{\text{Precision * Recall}}{\text{Precision + Recall}}$ |

# 4 RESULTS AND DISCUSSION

## 4.1 Technical Modelling Results and Discussion

The experiment was designed into five sections, the first part is individual algorithms, and there are five individual algorithms in this study, which are KNN, SVM, DT, LR, and NB. After generating the optimal individual models from each algorithm, ensemble learning techniques have been applied to the individual optimal algorithms. There are four ensemble learning techniques, bagging, boosting, voting, and stacking, and each technique has been discussed in a separate part. All the results of the experiments mentioned above can be found in table 7.1 to table 7.6 in Appendix A. The optimal individual models have been summarized in table 4.1.

**Table 4.1 Individual Optimal Models Summary**

| Algorithm | Configuration | Accuracy Score |
|---|---|---|
| **DT** | DecisionTreeClassifier(criterion = 'gini', max_depth = 8) | 0.636 |
| **SVM** | SVC(kernel = 'linear') | 0.935 |
| **KNN** | KNeighborsClassifier(algorithm="auto", weights="distance", n_neighbors = 20) | 0.820 |
| **LR** | LogisticRegression(penalty="none",solver="lbfgs") | 0.923 |
| **NB** | MultinomialNB() | 0.897 |

The optimal individual models were generated after setting up the experiments in table 7.1, and each individual algorithm provided an optimal model. The accuracy score of the optimal model provided by each algorithm varies. It ranges from 0.636 to 0.935, and it is safe to claim that the performance of individual algorithms is unstable with a gap of around 0.3. The accuracy score of the SVM model is quite satisfactory, but refining the model and improving the score is still necessary. And the five models in table 4.1 would be the input in the ensemble models.

Based on the five optimal algorithms, the top five ensemble learning models were generated, and their configuration and summaries are shown in table 4.2 and 4.3 respectively.

**Table 4.2 Top 5 Ensemble Learning Models Summary**

| Technique | Configuration | Accuracy Score |
|---|---|---|
| **Stacking** | model_stacking = StackingClassifier(estimators = [('dt', clf1),('svm', clf2), ('mnb',clf5)], \final_estimator = linear_model.LogisticRegression(penalty="none",solver="lbfgs")) | 0.947 |
| **Voting** | model_voting = VotingClassifier(estimators = [('svm', clf2), ('knn', clf3),('mnb', clf5)], \voting = 'soft') | 0.944 |
| **Voting** | model_voting = VotingClassifier(estimators = [('svm', clf2),('mnb', clf5)], \voting = 'soft') | 0.941 |
| **Stacking** | model_stacking = StackingClassifier(estimators = [('dt', clf1),('svm', clf2),('knn', clf3), ('mnb',clf5)], \ final_estimator = linear_model.LogisticRegression(penalty="none",solver="lbfgs")) | 0.941 |
| **Stacking** | model_stacking = StackingClassifier(estimators = [('dt', clf1),('svm', clf2)], \final_estimator = linear_model.LogisticRegression(penalty="none",solver="lbfgs")) | 0.935 |

**Table 4.3 Top 5 Ensemble Learning Models Performance Summary**

**** Used to indicate the final estimator in the combination*

| Technique | Combination | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|---|
| Stacking | SVM, DT MNB, LR*** | 1 | 1.00 | 0.95 | 0.97 | 0. 947 |
| | | 2 | 0.89 | 0.96 | 0.93 | |
| | | 3 | 0.95 | 0.93 | 0.94 | |
| | | Avg. | 0.947 | 0.947 | 0.947 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Voting | KNN, SVM MNB | 1 | 0.96 | 0.95 | 0.96 | 0. 944 |
| | | 2 | 0.92 | 0.94 | 0.93 | |
| | | 3 | 0.95 | 0.94 | 0.95 | |
| | | Avg. | 0.943 | 0.943 | 0.947 | |
| Voting | MNB, SVM | 1 | 0.98 | 0.94 | 0.96 | 0.941 |
| | | 2 | 0.89 | 0.94 | 0.92 | |
| | | 3 | 0.95 | 0.94 | 0.95 | |
| | | Avg. | 0.940 | 0.940 | 0.943 | |
| Stacking | DT, SVM KNM, MNB LR*** | 1 | 1.00 | 0.95 | 0.97 | 0.941 |
| | | 2 | 0.87 | 0.97 | 0.92 | |
| | | 3 | 0.96 | 0.91 | 0.93 | |
| | | Avg. | 0.943 | 0.943 | 0.940 | |
| Stacking | SVM, DT LR*** | 1 | 0.97 | 0.95 | 0.96 | 0.935 |
| | | 2 | 0.89 | 0.93 | 0.91 | |
| | | 3 | 0.94 | 0.92 | 0.93 | |
| | | Avg. | 0.922 | 0.922 | 0.922 | |

** The row 'Avg.' in table 4.9 is the average value of precision, recall, and F1-score of each class in a model respectively.*

The main metric used to select the recommended model is the accuracy score, because, unlike precision, recall, or F1-score, it reflects the overall performance of the model, rather than just a class in the model. From table 4.2, it is clear that the optimal model is the combination of SVM, DT, LR, and MNB using the stacking technique as highlighted with an accuracy score as high as 0.947. Besides, a deeper dive is necessary for the top five models because other evaluation metrics can also reveal the secret of model performance. In a class-specified way, precision, recall, and F1-score were listed in table 4.3, and a row 'Avg' was generated for each metric. The average value of each metric represents the overall performance of the model using the metrics. Among all the five models, the optimal model is the combination of SVM, DT, LR, and MNB using the stacking technique as highlighted in table 4.3. Table 4.2 and Table 4.3 lead to the same consequence.

This project involves both individual algorithms and ensemble learning techniques, understanding their performance and comparing them are essential either. Accuracy score is the main metric used in the project, the comparison of individual algorithms, and ensemble learning techniques are shown in Table 4.4, and Table 4.5.

**Table 4.4 Individual Algorithms Comparison in Accuracy Score**

| Algorithm | Average Score | Highest Score | Hyperparameter List |
|---|---|---|---|
| SVM | 0.919 | 0.935 | kernel, degree, gamma, C |
| LR | 0.923 | 0.923 | penalty, solver |
| NB | 0.832 | 0.897 | gaussian, bernoulli, multinomial |
| KNN | 0.782 | 0.820 | n_neighbors, algorithm, weights, |
| DT | 0.611 | 0.636 | criterion, max_depth |

SVM has the highest score of 0.935, and DT has the lowest score of 0.636. The average score is able to reflect the classification stability of the model because it represents the overall performance. LR has the highest average score of 0.923, and DT has the lowest average score of 0.636. It is easy to notice that the most stable algorithm is LR, while the optimal individual algorithm is SVM. Hyperparameter is important when it comes to tuning the models and searching the optimal model. During the experiment, I discovered that DT has a problem with overfitting, especially when the max_depth is not set properly. N_neighbor is the most important hyperparameter of KNN, so I applied 'grid search' to search the optimal number of neighbors.

**Table 4.5 Ensemble Learning Techniques Comparison in Accuracy Score**

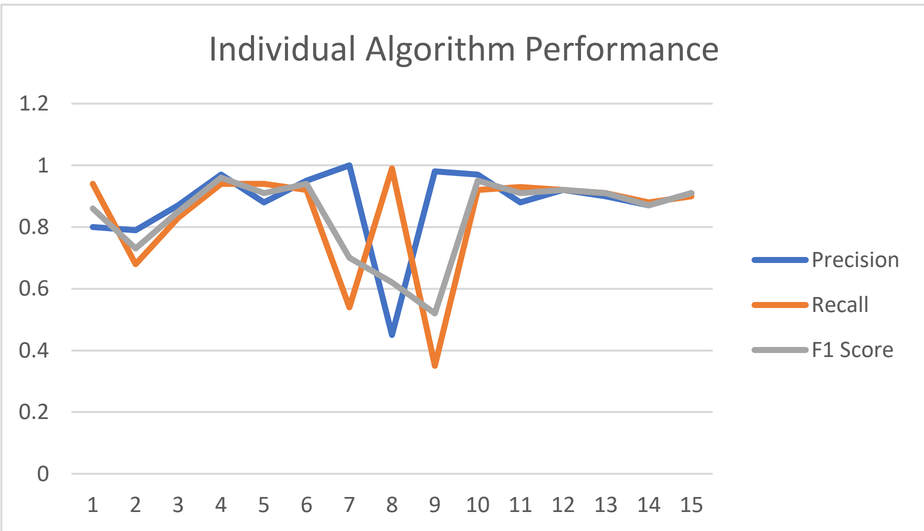
| Technique | Average Score | Highest Score | Hyperparameter List |
|---|---|---|---|
| Bagging | 0.850 | 0.929 | base_estimator, n_estimators |

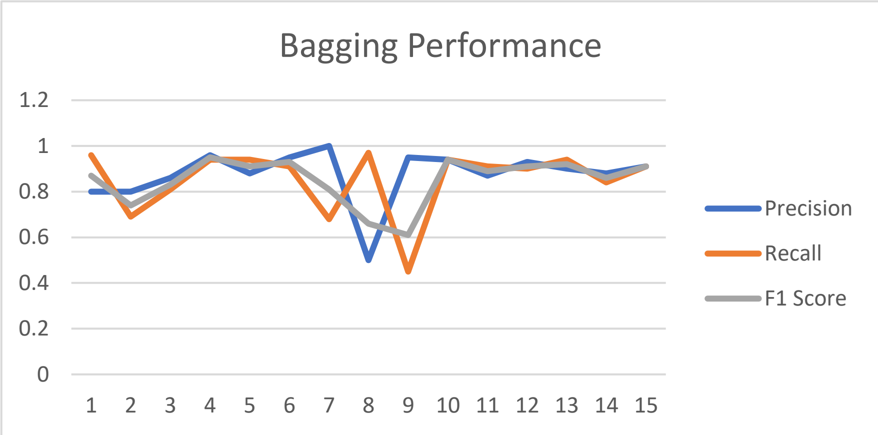
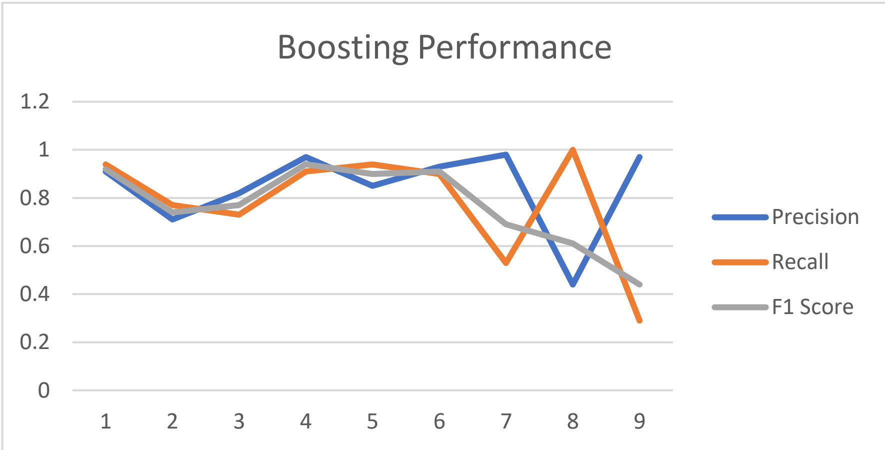
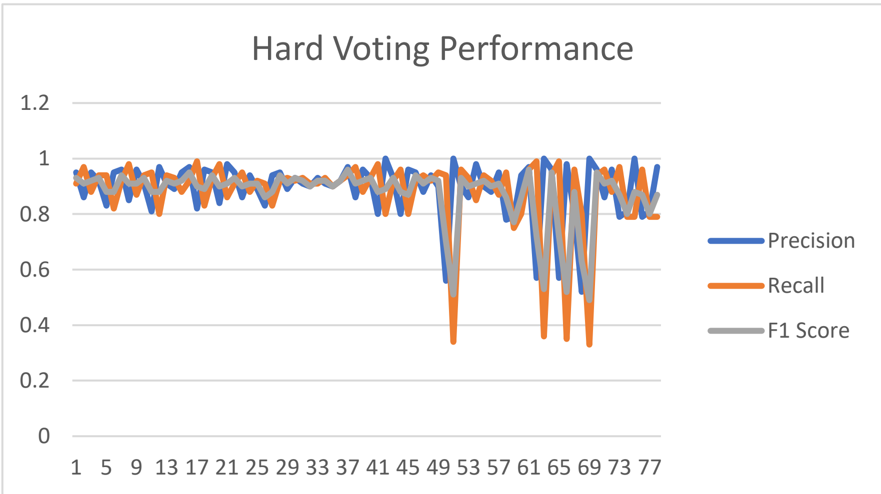
| | | | |
|---|---|---|---|
| Boosting | 0.774 | 0.917 | base_estimator, n_estimators |
| Hard Voting | 0.877 | 0.929 | estimators, voting |
| Soft Voting | 0.919 | 0.944 | estimators, voting |
| Stacking | 0.889 | 0.947 | estimators, final_estimator |

The most common hyperparameters in all techniques are estimators and the number of estimators, while different techniques may have their unique hyperparameters. The optimal ensemble learning model is from the stacking technique with an accuracy score of 0.947, while the most stable ensemble learning technique is soft voting with an accuracy score of 0.919. Therefore, the most reliable techniques are soft voting and stacking.

Though precision, recall, and F1-score are not the main metrics, they are worth analyzing too. The performance of each metric of the individual algorithm and ensemble learning techniques were summarized into line charts, and they can be found in table 4.6.

**Table 4.6 Performance of individual and ensembled techniques**

| Technique | Performance |
|---|---|
| Individual Algorithm |   **Figure 4.1 Individual Algorithm Performance in three metrics** |

| | |
|---|---|
| Bagging | <br>**Figure 4.2 Bagging Technique Performance in three metrics** |
| Boosting | <br>**Figure 4.3 Boosting Technique Performance in three metrics** |
| Hard Voting | <br>**Figure 4.4 Hard Voting Technique Performance in three metrics** |

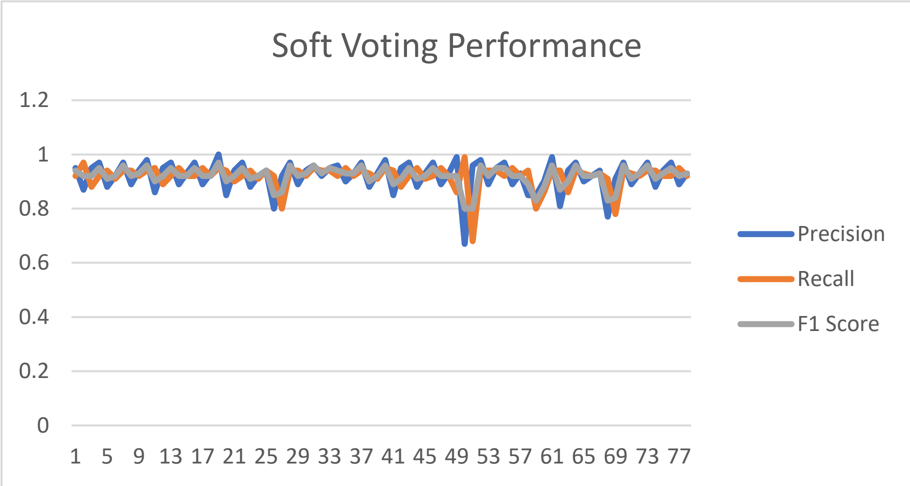
<table>
<tr>
<td>Soft Voting</td>
<td>


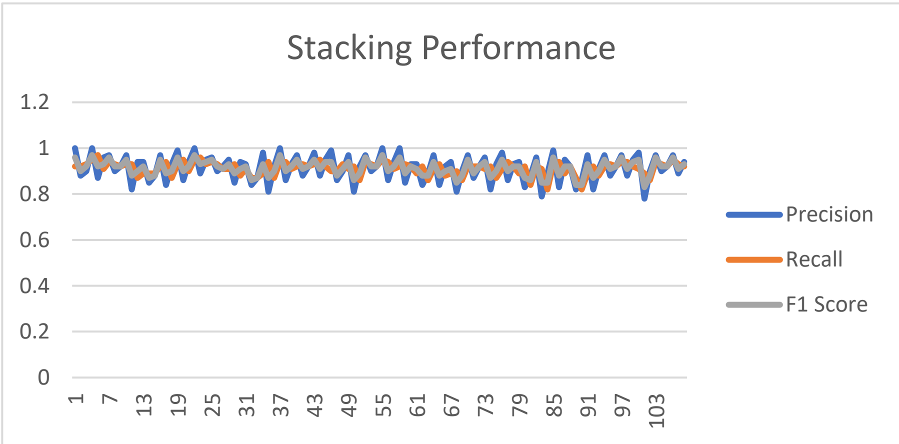

**Figure 4.5 Soft Voting Technique Performance in three metrics**
</td>
</tr>
<tr>
<td>Stacking</td>
<td>



**Figure 4.6 Stacking Technique Performance in three metrics**
</td>
</tr>
</table>

From table 4.6, it is clear that the individual algorithm, bagging, and boosting technique all suffer from the drastic fluctuation of the evaluation metrics, which makes the overall performance of it doubtful, and less reliable. In the voting technique, the performance of hard voting ranges from around 0.4 to 1.0, while that of soft voting ranges from about 0.6 to 1.0. This indicates that the probability strategy in the soft voting technique outperforms the highest number of voting strategies in the hard voting technique due to the fact that it is more stable and trustworthy. When it comes to the stacking technique, the performance range of it even narrows, and it is roughly from 0.8 to 1.0.

From the optimal model generated and the analysis above, the ensemble learning model outperforms the individual algorithms. However, this is not necessarily the case. There are times that the ensemble learning model underperforms the individual. It is more likely to happen when the performance of individual algorithms is already good enough, or the number of trials is not enough. There are some examples that the ensemble models underperform the individual algorithms listed in table 4.7.

**Table 4.7 Underperformed Ensemble Learning Models Samples**

| Technique | Combination | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|---|
| Boosting | MNB | 1 | 0.98 | 0.53 | 0.69 | 0.590 |
| | | 2 | 0.44 | 1.00 | 0.61 | |
| | | 3 | 0.97 | 0.29 | 0.44 | |
| Stacking | DT, SVM*** | 1 | 1.00 | 0.54 | 0.70 | 0.616 |
| | | 2 | 0.45 | 0.98 | 0.62 | |
| | | 3 | 0.96 | 0.36 | 0.72 | |
| Stacking | DT, LR*** | 1 | 0.99 | 0.61 | 0.75 | 0.631 |
| | | 2 | 0.46 | 0.98 | 0.63 | |
| | | 3 | 0.96 | 0.34 | 0.51 | |
| Stacking | DT, KNN*** | 1 | 0.98 | 0.54 | 0.70 | 0.602 |
| | | 2 | 0.44 | 0.98 | 0.61 | |
| | | 3 | 0.95 | 0.32 | 0.48 | |

There are three models in Table 4.7 involve algorithm 'DT', the main reason for this is that the performance of DT algorithm is not satisfied enough to be a trustworthy input to the ensemble learning models. Besides, Table 4.7 has a large proportion that overlaps with Table 4.8, which indicates that the over proportional false positive and false negative samples have a great negative impact on the model performance.

Apart from the optimal model selection, some interesting insights can be generated too.

The nature of the algorithms and techniques matters in building the models. In table 7.3, there are only three models listed. The reason why the SVM model is not included is that the base classifier is too strong, and Adaboost refuse to boost the models, the error received is 'AdaBoostClassifier with algorithm='SAMME.R' requires that the weak learner supports the calculation of class'. For KNN, the reason why it is not included is that Adaboost works from assigning different proportion of each individual models to get a better model, while KNN doesn't allow probability feature in the algorithm. The error message received from KNN is 'probabilities with a predict_proba method'. Therefore, it is clear due to the nature of the algorithm and ensemble techniques, not all algorithm can apply boosting technique. Speaking of the nature of the techniques, in soft voting techniques, as it chooses the model by assigning the probability to the different models, some adjustments are required when the algorithm is not naturally probability involved. Set the parameter probability to 'True' could solve the issue. In stacking technique, there are two kinds of estimators, and they are base estimators and final estimators. When the MNB model acts as a final estimator, the results generated from the base estimator is invalid as the input of MNB cannot be negative. An error message was received as 'Negative values in data passed to MultinomialNB (input X)'. Therefore, not all algorithms can be used as final estimator in stacking technique due to the nature of the algorithm and ensemble techniques.

Apart from algorithm nature, in appendix A, there are some cases when the value of precision and recall has a negative correlation. Table 4.8 contains all such cases in the project. All of them shall be filtered out during the selection process.

**Table 4.8 Table of the over proportional false negative and false positive cases**

| Technique | Combination | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|---|
| Individual Algorithm | DT | 1 | 1.00 | 0.46 | 0.63 | 0.569 |
| | | 2 | 0.95 | 0.20 | 0.33 | |
| | | 3 | 0.45 | 1.00 | 0.62 | |
| Bagging | DT | 1 | 1.00 | 0.68 | 0.81 | 0.690 |
| | | 2 | 0.50 | 0.97 | 0.66 | |
| | | 3 | 0.95 | 0.45 | 0.61 | |
| Boosting | MNB | 1 | 0.98 | 0.53 | 0.69 | 0.590 |
| | | 2 | 0.44 | 1.00 | 0.61 | |
| | | 3 | 0.97 | 0.29 | 0.44 | |
| Hard Voting | MNB, DT | 1 | 0.90 | 0.95 | 0.92 | 0.735 |
| | | 2 | 0.56 | 0.94 | 0.70 | |
| | | 3 | 1.00 | 0.34 | 0.51 | |
| Hard Voting | DT, SVM | 1 | 0.97 | 0.96 | 0.96 | 0. 758 |
| | | 2 | 0.57 | 0.99 | 0.72 | |
| | | 3 | 1.00 | 0.36 | 0.53 | |
| | | 1 | 0.80 | 0.96 | 0.88 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Hard Voting | KNN, DT | 2 | 0.52 | 0.80 | 0.63 | 0.690 |
| | | 3 | 1.00 | 0.33 | 0.49 | |
| Stacking | DT, SVM*** | 1 | 1.00 | 0.54 | 0.70 | 0.617 |
| | | 2 | 0.45 | 0.98 | 0.62 | |
| | | 3 | 0.96 | 0.36 | 0.52 | |

When the precision or recall equals to 1, the other one can be as low as 0.2. The reason of this is that the false samples highly overweigh the true samples, which means that the model predicts too many positive cases as negative or too many negative cases as positive, and the prediction capability of that class is weak. Therefore, no matter how good the F1-score and accuracy score are, the model is not recommended because it is too strict in predicting the true cases.

## 4.2 Sentiments Results and Discussion

Understanding the classification results from a domain perspective and exploring the sentiments revealed from the selected novels is equally essential as modeling. The exploration results of each sentiment from the attribute 'Sentence' can be found in table 4.9, and exploration results in an novel-specific manner from the attribute 'keyword' can be found in table 4.10.

**Table 4.9 Exploration Results of Each Sentiments**

| **Name** | **Sentiment Results** |
|---|---|
| **Stress** |   **Figure 4.7 Word Cloud of Stress** |

<table>
<tr><td>Trauma</td><td>

Figure 4.8 Word Cloud of Trauma</td></tr>
<tr><td>Anxiety</td><td>

Figure 4.9  Word Cloud of Anxiety</td></tr>
</table>

WordCloud calculates all the words in the sentences, and the bigger font of the word, the high frequent of the word. Table 4.9 shows the most frequent words for each sentiment. It is clear that the most frequent words in different sentiments are different. The most frequent words in stress are ‘forced’, and ‘threat; the most frequent words in trauma are ‘massacre, ‘beat’, and ‘people’; the most frequent words in anxiety are ‘worried, ‘worry, and ‘risk’. Some of them are from the keywords dictionary while some are not, but all of them are a good indicator to show the nature of respective sentiment.

Figure 4.7 illustrates the overall perspective of writers toward stress, and they believe that people are especially stressful when they are forced or threatened to do something unwilling. Figure 4.8 illustrates the overall perspective of writers toward trauma, and they believe that people are especially traumatized when people had some horrible experience. Figure 4.9 illustrate the overall perspective of writers toward anxiety, and they believe that people are especially anxious when people are worried, nervous, or exposed to risks.

As stress, trauma, and anxiety are all negative sentiments, it seems inevitable that there is an overlapping proportion in the wordcloud. The expression of English language enriches the way for writers to depict stress, trauma, and anxiety sentiments. It is easy to notice that the writers use the word ‘security’ and ‘death’ in stress and trauma. However, the sentiments carried in the sentence should be based on the meaning of the sentence, so the same word can express different sentiment under the writers’ pen.

Besides, according to Figure 3.11, the writers have an overall emphasis of the sentiment trauma, and the sentiment stress has the least emphasis from the writers. Therefore, as a whole, the writers believe that the sentiment trauma is the most important sentiment of the three. However, sentiment-specific exploration can only reveal the insights of all the novels as a whole, and studying the sentiments from a novel-specific perspective is also crucial. Exploration based on each novel has also been conducted too. I used word cloud, and bar chart to illustrate the word frequency, and sentiment distribution in Table 4.10.

**Table 4.10 Exploration Results of Each Writer**

| **Title** | **Sentiment Results** |
|---|---|
| **The Baghdad Clock** | <br>**Figure 4.10 Frequent Word of The Baghdad Clock** |

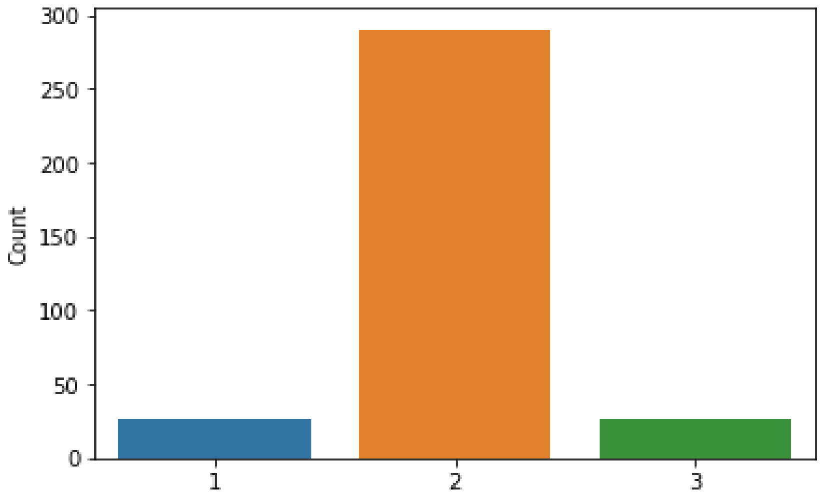
<table>
<tr>
<td></td>
<td>


**Figure 4.11 Sentiments Distribution of The Baghdad Clock**</td>
</tr>
<tr>
<td>**Baghdad Diaries 1991-2002_ A Woman's Chronicle of War and Exile**</td>
<td>


**Figure 4.12 Sentiments Distribution of The Baghdad Clock**

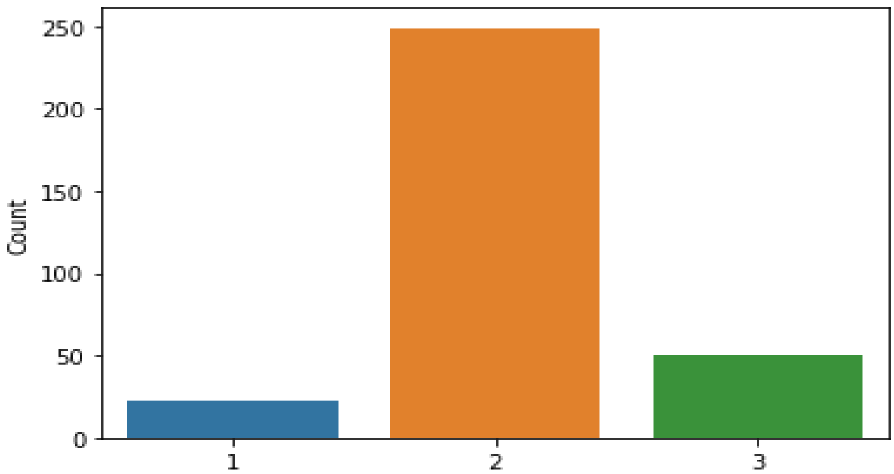



**Figure 4.13 Sentiments Distribution of Baghdad Diaries**</td>
</tr>
</table>

<table>
<tr>
<td>The Beekeeper of Aleppo_ A Moving Testament to the Human Spirit</td>
<td>



Figure 4.14  Frequent Word of The Beekeeper of Aleppo

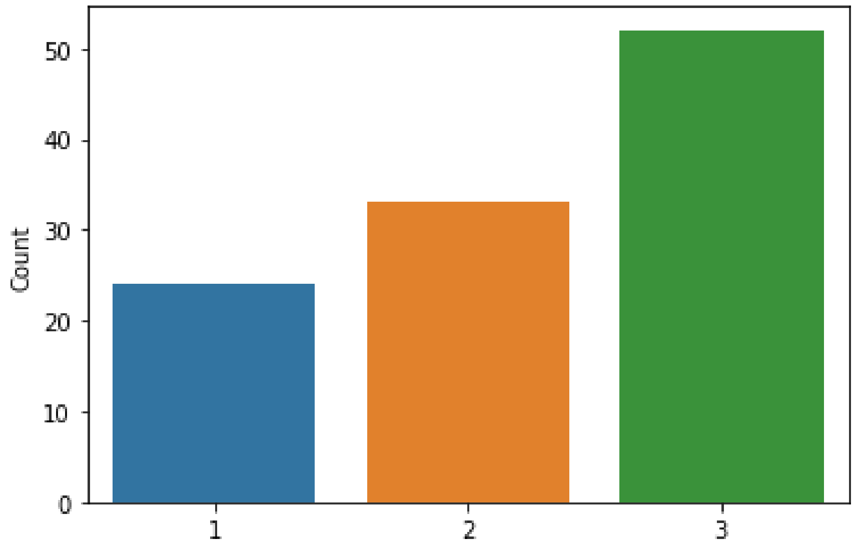
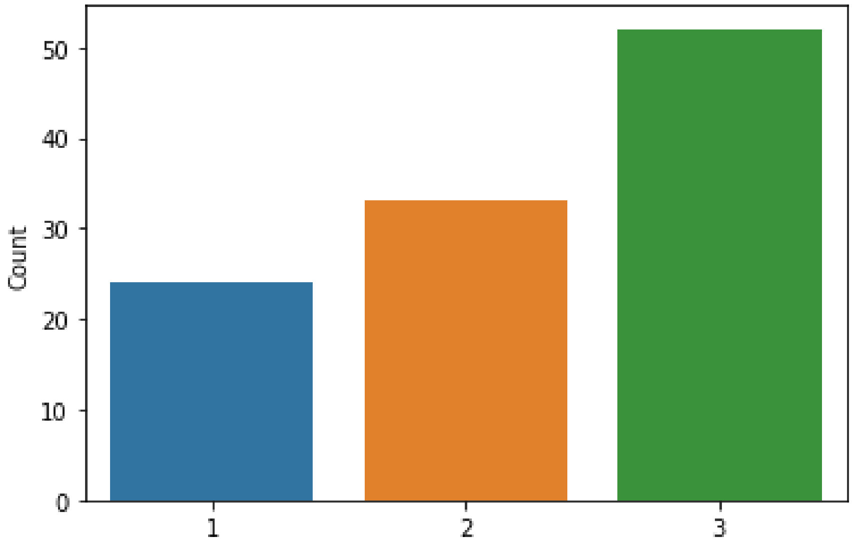
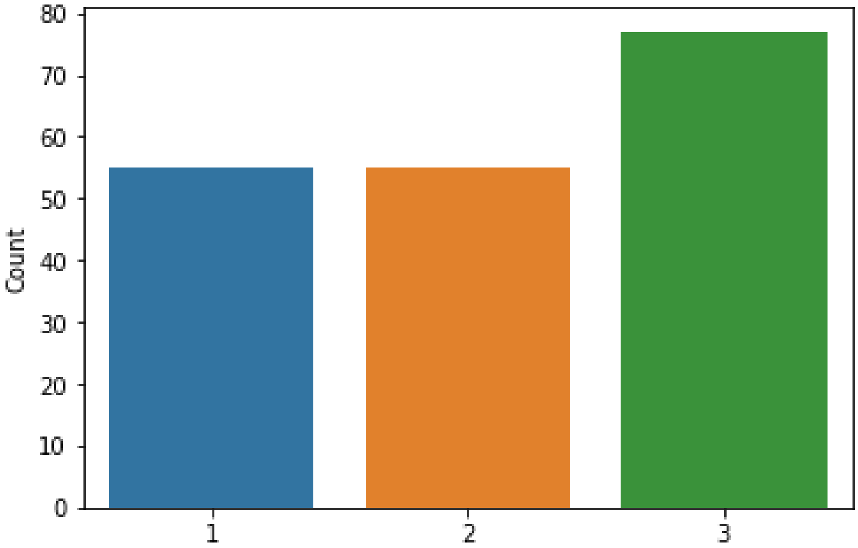


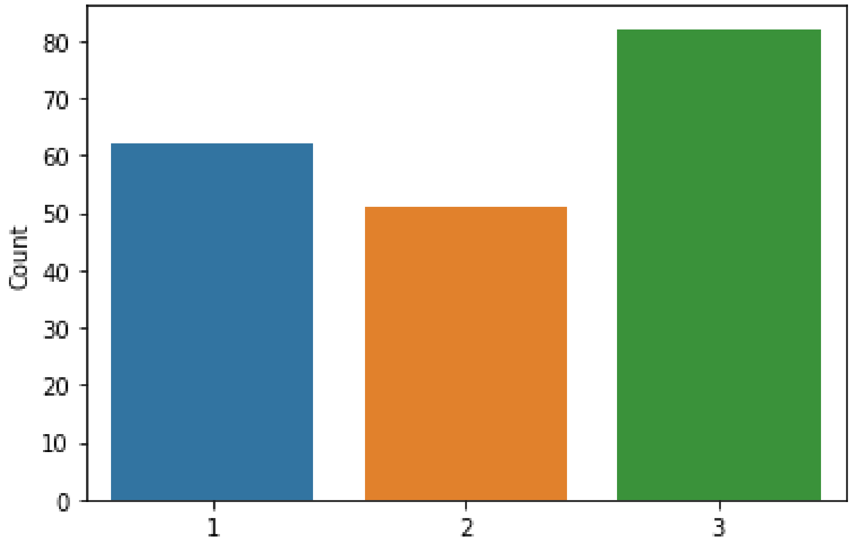
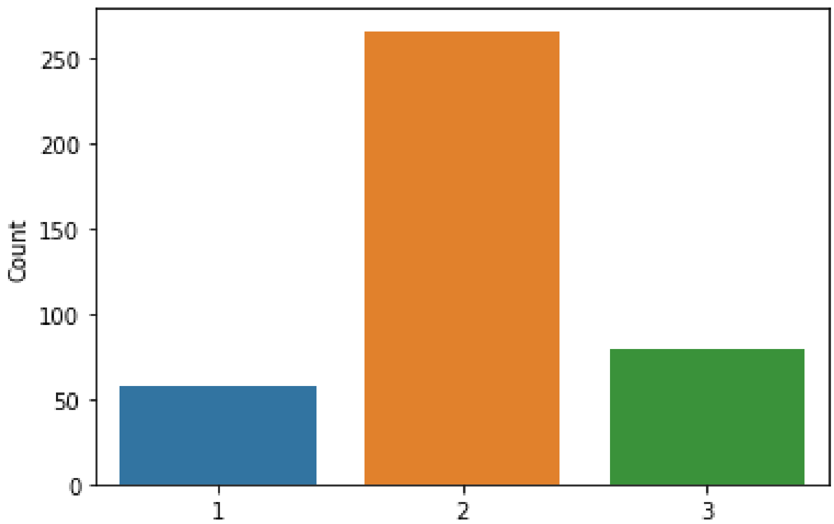

Figure 4.15  Sentiments Distribution of The Beekeeper of Aleppo
</td>
</tr>
<tr>
<td>The Girl from Aleppo_ Nujeen's Escape from War to Freedom</td>
<td>



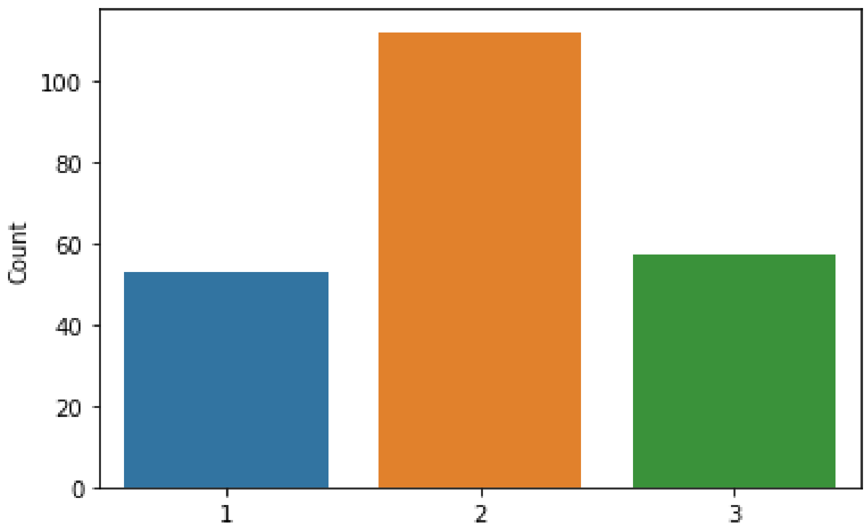
Figure 4.16  Frequent Word of The Girl from Aleppo
</td>
</tr>
</table>

| | |
|---|---|
| | <br><br>**Figure 4.17 Sentiments Distribution of The Girl from Aleppo** |
| **The Orange Trees of Baghdad_ In Search of My Lost Family_ A Memoir** | <br><br>**Figure 4.18 Frequent Word of The Orange Trees of Baghdad** |

| | |
|---|---|
| | <br>**Figure 4.19 Sentiments Distribution of The Orange Trees of Baghdad** |
| **Barefoot in Baghdad** | <br>**Figure 4.20 Frequent Word of Barefoot in Baghdad** |

| | |
|---|---|
| |  <br>**Figure 4.21 Sentiments Distribution of Barefoot in Baghdad** |
| **Escape from Aleppo** |  <br>**Figure 4.22 Frequent Word of Escape from Aleppo**<br>  |

<table>
<tr><td></td><td>Figure 4.23 Sentiments Distribution of Escape from Aleppo</td></tr>
<tr><td>Shooting Kabul</td><td>

<br>Figure 4.24 Frequent Word of Shooting Kabul<br>


<br>Figure 4.25 Sentiments Distribution of Shooting Kabul</td></tr>
<tr><td>The Boy from Aleppo Who Painted the War</td><td>
</td></tr>
</table>

<table>
<tr>
<td></td>
<td>
Figure 4.26 Frequent Word of The Boy from Aleppo

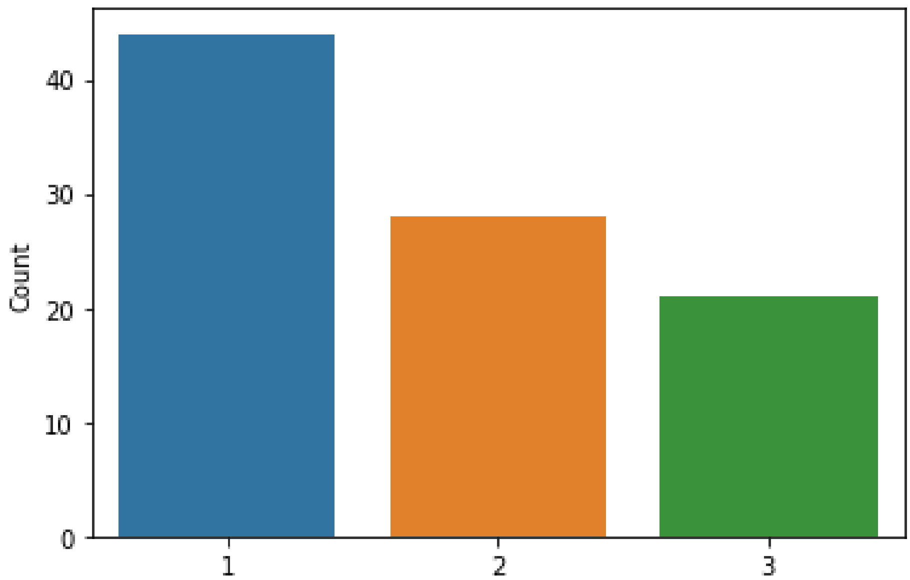
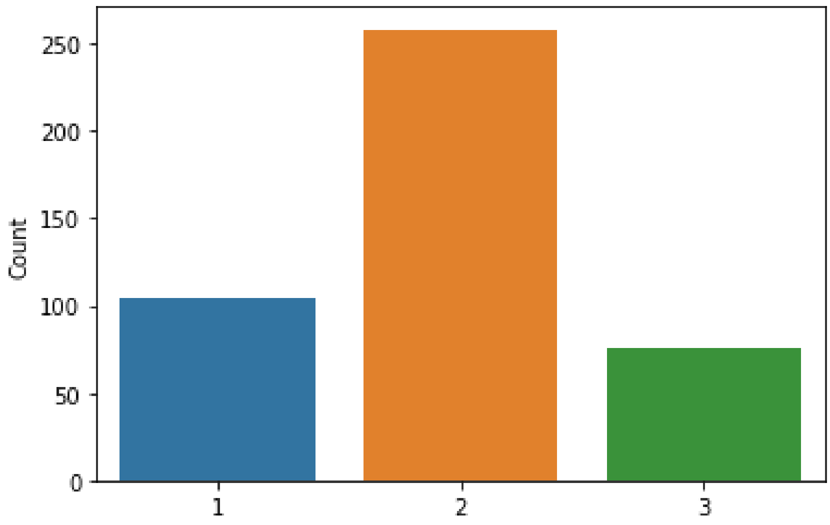



Figure 4.27 Sentiments Distribution of The Boy from Aleppo
</td>
</tr>
<tr>
<td>A Woman in the Crossfire_ Diaries of the Syrian Revolution</td>
<td>



Figure 4.28 Frequent Word of Diaries of the Syrian Revolution
</td>
</tr>
</table>

| | |
|---|---|
| |   **Figure 4.29 Sentiments Distribution of Diaries of the Syrian Revolution** |

There are ten novels from eight writers in this study. The top frequent words and labeled sentences vary in multiple aspects. Writers emphasize different words and elements in different novels in Table 10. Besides, the sentiments in the books are not balanced either, which means that the novels from different writers emphasize different sentiment. Table 4.11 is the sentiment summary of each novel.

**Table 4.11 Sentiment Summary of Each Novel**

| No. | Writer | Novels | Emphasized Word | Emphasized Sentiment |
|---|---|---|---|---|
| 1 | al Rawi, Shahad | The Baghdad Clock | Tear<br>Fear<br>Die | Trauma |
| | | Baghdad Diaries 1991-2002_ A Woman's Chronicle of War and Exile | Kill<br>Die<br>Dead | Trauma |
| 2 | Lefteri, Christy | The Beekeeper of Aleppo_ A Moving Testament to the Human Spirit | Worried<br>Beat<br>Forced | Anxiety |
| 3 | Mustafa, Nujeen | The Girl from Aleppo_ Nujeen's Escape from War to Freedom | Worried<br>Worry<br>Beat | Anxiety |
| 4 | Nadir, Leilah | The Orange Trees of Baghdad_ In Search of My Lost Family_ A Memoir | Worried<br>Risk | Anxiety |

| | | | | |
|---|---|---|---|---|
| | | | Beat | |
| 5 | Omar, Manal M | Barefoot in Baghdad | Concern<br>Forced<br>Risk | Anxiety |
| 6 | Senzai, N H | Escape from Aleppo | Kill<br>Die<br>Tight | Trauma |
| | | Shooting Kabul | Tight<br>Beat<br>Cried | Trauma |
| 7 | Sukkar, Sumia | The Boy from Aleppo Who Painted the War | Tight<br>Worry<br>Torture | Stress |
| 8 | Yazbek, Samar | A Woman in the Crossfire_ Diaries of the Syrian Revolution | Massacre<br>Beat<br>Torture | Trauma |

It is clear that most of the novels emphasize sentiment trauma, only four novels emphasize sentiment anxiety, and one emphasizes sentiment stress.

The majority of the novels which emphasize trauma use the words such as ‘kill’, ‘die’, and ‘massacre’. Four novels which emphasize the sentiment of anxiety all use the word ‘worried’, ‘risk’, and ‘forced’. The novel, ‘The Boy from Aleppo Who Painted the War’, emphasizes the sentiment of stress, and the most frequent word used is ‘tight’.

All writers include the sentiment of stress, trauma, and anxiety in a single novel, which means that the writers believe that all the three sentiments studied in the project are all typical negative sentiments in war literatures. Though it is natural to have emphasis on different sentiment, writers use similar words to depict a sentiment. Besides, there are two authors write more than one novel in Table 4.11. The emphasized sentiment of two novels written by al Rawi, Shahad is trauma. However, the emphasized words are different, although the two novels both emphasize death, writer use the word ‘fear’ and ‘tear’ more frequently in ‘The Bagdad Clock’ while use the word ‘kill’ more frequently in ‘Baghdad Diaries 1991-2002’. The similar case also exists in ‘Escape from Aleppo’ and ‘Shooting Kabul’.

# 5 CONCLUSION & LESSON LEARNED

The problems I got before the project have been well answered.

***RQ1: How to assign the correct sentiment to each sentence?***

By building the keywords dictionary and searching the keywords in the sentences using python as a scripting tool, there more than 5000 sentences that contains the elements of stress, trauma, and anxiety in the extracted data. Based on the automatically labeled dataset, I also conducted manual labelling, and there are over 1200 labeled sentences. The detailed process and interpretation can be found in 3.4.1.

***RQ2: What preprocessing should be done before applying the machine learning techniques?***

The preprocessing part followed the steps shown in figure 3.2, and there are 5 steps during the preprocessing, tokenization, normalization, punctuation removal, stopwords removal, and POS tag. The detail of each step is discussed from 3.2.2.1 to 3.2.2.5.

***RQ3: How to apply the machine learning techniques?***

The machine learning techniques were separate as individual algorithm, and ensemble techniques. Five algorithms (KNN, LR, NB, DT, SVM) were experimented, and the optimal individual algorithm acts as the input of the ensemble learning technique. Ensemble techniques included bagging, boosting, voting, and stacking, and all the experiments were held in python 3.0 environment.

***RQ4: What is the optimal model and how to interpret the results?***

The optimal model is the combination of SVM, MNB, LR, and DT, while LR acts as final estimator under stacking technique. Four metrics (Accuracy score, Precision, Recall, and F1-score) were applied to select the optimal model from all the experiments. The detailed process of selecting the optimal model and discussing the models can be found in 4.1. The results were interpreted from both technical and domain aspects too. The detailed interpretation about the sentiments can be found in 4.2.

## 5.1 Conclusion – Technical Perspective

This study generated a bunch of models in a python environment and analyzed the performances of both individual and ensembled machine learning models. The optimal model generated has an accuracy score as high as 0.947, and the model was from the stacking technique. Overall, the ensemble learning technique has outperformed individual algorithms in terms of accuracy score, precision, recall, and F1-value, and the most stable and reliable models are from soft voting and stacking. Besides, there are two interesting insights I discovered during the modeling process.

## 5.2 Conclusion – Domain Perspective

Writers' perspective toward stress, trauma, and anxiety based on the classification results has been revealed too. Writers have an overall emphasis on the sentiment of trauma. Although novels tell different stories, their writers use similar words to depict stress, trauma, and anxiety. Different novels have different emphasized sentiments, while all three sentiments are included. Overall, writers believe wars impose stress, trauma, and anxiety on people, and the sentiment of trauma is the most noticeable one of all.

## 5.3 Lesson Learned

Undergoing the practicum project, I learned how to perform text classification and analysis. Sloving a text classification problem from scratch is not an easy task for me, because I had fairly limited experience in NLP (Natural Language Processing). Reading and studying the existing surely helped me, but not as much as discussing with my supervisor and having trials on my own. Doing so taught me how to label the data, refine the model, and interpret the results. Besides, I learned the importance of domain knowledge. Finishing the coding and technical part is far from distinct, data scientists like us should understand the business of our clients, as well as the consequences the clients are looking for. Finally, I also learned the brutal nature of wars, and how people suffer from stress, anxiety, and trauma in a war. Innocent civilians are the most vulnerable group of people in a war, and I witnessed their hopelessness and sorrow through the novels.

## 5.4 Challenges and Solution

The main three challenges I met in the project are data labelling, data preparation, and model & results interpretation.

Firstly, labelling the data manually is quite challenging and time-consuming. The most reliable method to conduct the labelling process should be manually labelling the data by domain experts and proof reading it by another group of people. However, due to the limited time available for this project and restricted manpower, it is unrealistic to do it all manually. Instead, I chose a hybrid method. After generating the keywords dictionary and label the data with the dictionary, I manually went through the labelled dataset, and came out the final labelled dataset.

Secondly, unique preparation procedure for text data is unfamiliar to me. Due to the nature of the data of the project, I had to prepare the data to feed them to the models. Learning and accepting new concepts and knowledge is of great importance, and by learning from previous works, online channels and resources, and asking for guidance from supervisor, I finally got the data ready for modelling.

Finally, selecting the recommended model and interpreting the results can be confusing for this project. According to the four chosen evaluation metrics, the optimal model is quite straightforward. However, generating more interesting insights from the experiments are equally important. Chapter 4 consists of results and interpretation from technical and domain prospects.

## 5.5 Future Work

Despite this project having been successfully implemented. There is still more that awaits researchers to explore. Firstly, there are many negative sentiments can be detected from the war narratives. Apart from stress, trauma, and anxiety, other sentiments such as fear, and anger are also worth exploring. Secondly, manually labeling all the data is more recommended due to the complex nature of human nature. Finally, the performance of the ensemble learning models varies from one algorithm combination to another. More algorithms can be added, tuned, and combined to potentially find a model with a better performance.

# 6 REFERENCES


Jasni, Sulong. Welcome to the School of Humanities, Universiti Sains Malaysia. https://humanities.usm.my/#:~:text=The%20School%20of%20Humanities%20prides,history%2C%20geography%2C%20and%20translation

Huutsman Mental Health Institute. (2022). The Mental Health Effects of War: Backed by Science. https://healthcare.utah.edu/hmhi/news/2022/war-mental-health.php#:~:text=The%20World%20Health%20Organization%20(WHO,.%22%20Depression%2C%20anxiety%2C%20and

*Oxford Learner's Dictionaries | Find definitions, translations, and grammar explanations at Oxford Learner's Dictionaries*. (n.d.). Oxford Learner's Dictionaries. https://www.oxfordlearnersdictionaries.com

Koolhaas, J., Bartolomucci, A., Buwalda, B., De Boer, S., Flügge, G., Korte, S., . . . Fuchs, E. (2011). Stress revisited: A critical evaluation of the stress concept. Neuroscience & Biobehavioral Reviews, 35(5), 1291-1301. doi:10.1016/j.neubiorev.2011.02.003

Leonard I. Pearlin, Elizabeth G. Menaghan, Morton A. Lieberman and Joseph T. Mullan. (1981). The Stress Process. Journal of Health and Social Behavior , Dec., 1981, Vol. 22, No. 4 (Dec., 1981), pp. 337-356

American Psychological Association. (2008). Trauma. https://www.apa.org/topics/trauma#:~:text=Trauma%20is%20an%20emotional%20response,symptoms%20like%20headaches%20or%20nausea

National Library of Medicine. (2014). Trauma-Informed Care in Behavioral Health Services. https://www.ncbi.nlm.nih.gov/books/NBK207191/

Zbigniew Zaleski. (1995). Future Anxiety: Concept, Measurement, and Preliminary Research. Vol. 21. No. 2. pp. 165-174.

National Institute of Mental Health. (2022). Anxiety Disorders. https://www.nimh.nih.gov/health/topics/anxiety-disorders#part_2223

Tony Vinson, Desmond McDonnell. War Narratives.

Robert S. Laufer, M. S. Gallops and Ellen Frey-Wouters. (1984). War Stress and Trauma: The Vietnam Veteran Experience. Journal of Health and Social Behavior, Mar., 1984, Vol. 25, No. 1 (Mar., 1984), pp. 65-85

Ivanka, Zivcic, M.A. (2010). Emotional Reactions of Children to War Stress in Croatia. Volume 32, Issue 4, July 1993, Pages 709-713.

Marita Nadal, Monica Calvo. Trauma in Contemporary Literature, Narrative and Representation.

Phdessay. (2021). Trauma and recovery in Virginia woolf's Mrs Dalloway, https://phdessay.com/trauma-and-recovery-in-virginia-woolfs-mrs-dalloway/

Stephanie Collier, MD, MPH. (2022), War anxiety: How to cope. https://www.health.harvard.edu/blog/war-anxiety-how-to-cope-202205232748#:~:text=Symptoms%20of%20war%20anxiety&text=Physical%20symptoms%20of%20anxiety%20may,Others%20may%20feel%20numb

Neumann, F. (1972). Anxiety in Politics. In Conspiracy: the fear of subversion in American history, (eds.) Richard O. Curry & Thomas M. Brown. New York: Holt, Rinehart & Winston

Agrawal, R. (2021, June 14). *Must Known Techniques for text preprocessing in NLP*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/must-known-techniques-for-text-preprocessing-in-nlp/#:%7E:text=Text%20preprocessing%20is%20a%20method,text%20in%20a%20different%20case

Khurana, M., Gulati, A., & Singh, S. (2018). Sentiment analysis framework of Twitter data using classification. 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC). doi:10.1109/pdgc.2018.8745748

Xiao, L., Wang, G., & Liu, Y. (2018). Patent text classification based on naive Bayesian method. 2018 11th International Symposium on Computational Intelligence and Design (ISCID). doi:10.1109/iscid.2018.00020

Rahman, A., & Hossen, M. S. (2019). Sentiment analysis on movie review data using Machine Learning Approach. 2019 International Conference on Bangla Speech and Language Processing (ICBSLP). doi:10.1109/icbslp47725.2019.201470

Liu, Z., Lv, X., Liu, K., & Shi, S. (2010). Study on SVM compared with the other text classification methods. 2010 Second International Workshop on Education Technology and Computer Science. doi:10.1109/etcs.2010.248

Risch, J., Kao, A., Poteet, S. R., & Wu, Y. J. J. (2008). Text visualization for visual text analytics. In *Visual data mining* (pp. 154-171). Springer, Berlin, Heidelberg.

Zhou, ZH. (2021). Ensemble Learning. In: Machine Learning. Springer, Singapore. https://doi.org/10.1007/978-981-15-1967-3_8

Biswal, A. (2021, September 14). What is bagging in machine learning and how to perform bagging. Retrieved November 26, 2022, from https://www.simplilearn.com/tutorials/machine-learning-tutorial/bagging-in-machine-learning#:~:text=Bagging%2C%20also%20known%20as%20Bootstrap,variance%20of%20a%20prediction%20model

Khairy, R. S., Hussein, A., & ALRikabi, H. T. H. S. (2021). The detection of counterfeit banknotes using ensemble learning techniques of AdaBoost and voting. International Journal of Intelligent Engineering and Systems, 14(1), 326-339. https://doi.org/10.22266/ijies2021.0228.31

Sadki, N. (2020, February 13). Understand different types of boosting algorithms. Retrieved November 26, 2022, from https://iq.opengenus.org/types-of-boosting-algorithms/

Kumar, U. K., Nikhil, M. S., & Sumangali, K. (2017). Prediction of breast cancer using voting classifier technique. In 2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM) (pp. 108-114). IEEE. https://doi.org/10.1109/ICSTM.2017.8089135

Raschka, S. (n.d.). Ensemblevoteclassifier: A majority voting classifier. Retrieved November 26, (2022), from http://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/

OpenGenus. (2020). Hard vs soft voting classifier python example. Retrieved November 27, 2022, from https://vitalflux.com/hard-vs-soft-voting-classifier-python-example/

Towardsai. (2022). Machine learning model stacking in python. Retrieved November 27, 2022, from https://towardsai.net/p/l/machine-learning-model-stacking-in-python

Madan, R. (2019). TF-IDF/term frequency technique: Easiest explanation for text classification in NLP with python. Retrieved November 27, 2022, from https://medium.com/analytics-vidhya/tf-idf-term-frequency-technique-easiest-explanation-for-text-classification-in-nlp-with-code-8ca3912e58c3

Prasanna. (2022). Advantages and disadvantages of python: Python language advantages, disadvantages and its applications. Retrieved November 27, 2022, from https://www.aplustopper.com/advantages-and-disadvantages-of-python/

Teemu Kanstrén. (2021). A look at precision, recall, and F1-score. Retrieved November 27, 2022, from https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec

# 7 APPENDIX A

**Table 7.1 Performance of Optimal Individual Algorithm**

| Models | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| KNN | 1 | 0.80 | 0.94 | 0.86 | 0.820 |
| | 2 | 0.79 | 0.68 | 0.73 | |
| | 3 | 0.87 | 0.83 | 0.85 | |
| SVM | 1 | 0.97 | 0.94 | 0.96 | 0.935 |
| | 2 | 0.88 | 0.94 | 0.91 | |
| | 3 | 0.95 | 0.92 | 0.94 | |
| DT | 1 | 1 | 0.54 | 0.70 | 0.617 |
| | 2 | 0.45 | 0.99 | 0.62 | |
| | 3 | 0.98 | 0.35 | 0.52 | |
| LR | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| | 2 | 0.88 | 0.93 | 0.91 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| NB | 1 | 0.90 | 0.91 | 0.91 | 0.897 |
| | 2 | 0.87 | 0.88 | 0.87 | |
| | 3 | 0.91 | 0.90 | 0.91 | |

**Table 7.2 Performance of Bagging Technique**

| Models | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| KNN | 1 | 0.80 | 0.96 | 0.87 | 0.820 |
| | 2 | 0.80 | 0.69 | 0.74 | |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | 0.86 | 0.81 | 0.83 | |
| SVM | 1 | 0.96 | 0.94 | 0.95 | 0.930 |
| | 2 | 0.88 | 0.94 | 0.91 | |
| | 3 | 0.95 | 0.91 | 0.93 | |
| DT | 1 | 1.00 | 0.68 | 0.81 | 0.690 |
| | 2 | 0.50 | 0.97 | 0.66 | |
| | 3 | 0.95 | 0.45 | 0.61 | |
| LR | 1 | 0.94 | 0.94 | 0.94 | 0.914 |
| | 2 | 0.87 | 0.91 | 0.89 | |
| | 3 | 0.93 | 0.90 | 0.91 | |
| NB | 1 | 0.90 | 0.94 | 0.92 | 0.897 |
| | 2 | 0.88 | 0.84 | 0.86 | |
| | 3 | 0.91 | 0.91 | 0.91 | |

**Table 7.3 Performance of Boosting Technique**

| **Models** | **Class** | **Precision** | **Recall** | **F1-Score** | **Accuracy** |
|---|---|---|---|---|---|
| DT | 1 | 0.91 | 0.94 | 0.92 | 0.814 |
| | 2 | 0.71 | 0.77 | 0.74 | |
| | 3 | 0.82 | 0.73 | 0.77 | |
| LR | 1 | 0.97 | 0.91 | 0.94 | 0.917 |
| | 2 | 0.85 | 0.94 | 0.90 | |
| | 3 | 0.93 | 0.90 | 0.91 | |
| NB | 1 | 0.98 | 0.53 | 0.69 | 0.590 |
| | 2 | 0.44 | 1.00 | 0.61 | |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | 0.97 | 0.29 | 0.44 | |

**Table 7.4 Performance of Voting Technique Under Hard Voting**

| Models | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| MNB, DT | 1 | 0.95 | 0.91 | 0.93 | |
| SVM, LR | 2 | 0.86 | 0.97 | 0.91 | 0.920 |
| KNN | 3 | 0.95 | 0.88 | 0.92 | |
| MNB, DT | 1 | 0.92 | 0.94 | 0.93 | |
| LR, KNN | 2 | 0.83 | 0.94 | 0.88 | 0.900 |
| | 3 | 0.95 | 0.82 | 0.88 | |
| MNB, DT | 1 | 0.96 | 0.92 | 0.94 | |
| SVM, LR | 2 | 0.85 | 0.98 | 0.91 | 0.920 |
| | 3 | 0.96 | 0.87 | 0.91 | |
| MNB, DT | 1 | 0.91 | 0.94 | 0.93 | |
| SVM, KNN | 2 | 0.81 | 0.95 | 0.88 | 0.894 |
| | 3 | 0.97 | 0.80 | 0.88 | |
| MNB, SVM | 1 | 0.91 | 0.94 | 0.92 | |
| LR, KNN | 2 | 0.89 | 0.93 | 0.91 | 0.917 |
| | 3 | 0.95 | 0.88 | 0.92 | |
| DT, SVM | 1 | 0.97 | 0.92 | 0.95 | |
| LR, KNN | 2 | 0.82 | 0.99 | 0.90 | 0.912 |
| | 3 | 0.96 | 0.83 | 0.89 | |
| MNB, DT | 1 | 0.95 | 0.93 | 0.94 | |
| SVM | 2 | 0.84 | 0.98 | 0.90 | 0.920 |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | 0.98 | 0.86 | 0.91 | |
| MNB, DT LR | 1 | 0.95 | 0.91 | 0.93 | 0.914 |
| | 2 | 0.86 | 0.95 | 0.90 | |
| | 3 | 0.94 | 0.88 | 0.91 | |
| MNB, DT KNN | 1 | 0.89 | 0.92 | 0.91 | 0.885 |
| | 2 | 0.83 | 0.91 | 0.86 | |
| | 3 | 0.94 | 0.83 | 0.88 | |
| MNB, SVM LR | 1 | 0.95 | 0.92 | 0.94 | 0.926 |
| | 2 | 0.89 | 0.93 | 0.91 | |
| | 3 | 0.93 | 0.92 | 0.93 | |
| MNB, SVM KNN | 1 | 0.91 | 0.93 | 0.92 | 0.914 |
| | 2 | 0.90 | 0.91 | 0.90 | |
| | 3 | 0.93 | 0.91 | 0.92 | |
| MNB, LR KNN | 1 | 0.91 | 0.93 | 0.92 | 0.914 |
| | 2 | 0.90 | 0.90 | 0.90 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| DT, SVM LR | 1 | 0.97 | 0.94 | 0.96 | 0.930 |
| | 2 | 0.86 | 0.97 | 0.91 | |
| | 3 | 0.96 | 0.88 | 0.92 | |
| DT, SVM KNN | 1 | 0.93 | 0.93 | 0.93 | 0.900 |
| | 2 | 0.80 | 0.98 | 0.88 | |
| | 3 | 1.00 | 0.80 | 0.89 | |
| DT, LR KNN | 1 | 0.93 | 0.92 | 0.93 | 0.891 |
| | 2 | 0.80 | 0.96 | 0.88 | |
| | 3 | 0.96 | 0.80 | 0.87 | |
| LR, SVM | 1 | 0.95 | 0.92 | 0.94 | |

| | | | | | |
|---|---|---|---|---|---|
| KNN | 2 | 0.88 | 0.94 | 0.91 | 0.926 |
| | 3 | 0.94 | 0.92 | 0.93 | |
| MNB, DT | 1 | 0.90 | 0.95 | 0.92 | 0.735 |
| | 2 | 0.56 | 0.94 | 0.70 | |
| | 3 | 1.00 | 0.34 | 0.51 | |
| MNB, SVM | 1 | 0.90 | 0.96 | 0.93 | 0.912 |
| | 2 | 0.86 | 0.93 | 0.90 | |
| | 3 | 0.98 | 0.85 | 0.91 | |
| LR, MNB | 1 | 0.90 | 0.94 | 0.92 | 0.909 |
| | 2 | 0.88 | 0.92 | 0.90 | |
| | 3 | 0.95 | 0.87 | 0.91 | |
| KNN, MNB | 1 | 0.78 | 0.95 | 0.86 | 0.832 |
| | 2 | 0.79 | 0.75 | 0.77 | |
| | 3 | 0.94 | 0.80 | 0.86 | |
| DT, SVM | 1 | 0.97 | 0.96 | 0.96 | 0.758 |
| | 2 | 0.57 | 0.99 | 0.72 | |
| | 3 | 1.00 | 0.36 | 0.53 | |
| LR, DT | 1 | 0.96 | 0.94 | 0.95 | 0.749 |
| | 2 | 0.57 | 0.99 | 0.72 | |
| | 3 | 0.98 | 0.35 | 0.52 | |
| DT, KNN | 1 | 0.80 | 0.96 | 0.88 | 0.690 |
| | 2 | 0.52 | 0.80 | 0.63 | |
| | 3 | 1.00 | 0.33 | 0.49 | |
| LR, SVM | 1 | 0.96 | 0.94 | 0.95 | 0.926 |
| | 2 | 0.86 | 0.96 | 0.91 | |
| | 3 | 0.96 | 0.88 | 0.92 | |

| | | | | | |
|---|---|---|---|---|---|
| KNN, SVM | 1 | 0.79 | 0.97 | 0.87 | 0.853 |
| | 2 | 0.81 | 0.79 | 0.80 | |
| | 3 | 1.00 | 0.79 | 0.88 | |
| LR, KNN | 1 | 0.79 | 0.96 | 0.87 | 0.847 |
| | 2 | 0.81 | 0.79 | 0.80 | |
| | 3 | 0.97 | 0.79 | 0.87 | |

**Table 7.5 Performance of Voting Technique Under Soft Voting**

| Models | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| MNB, DT | 1 | 0.95 | 0.92 | 0.94 | 0.923 |
| SVM, LR | 2 | 0.87 | 0.97 | 0.92 | |
| KNN | 3 | 0.95 | 0.88 | 0.92 | |
| MNB, DT | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| LR, KNN | 2 | 0.88 | 0.94 | 0.91 | |
| | 3 | 0.92 | 0.91 | 0.92 | |
| MNB, DT | 1 | 0.97 | 0.94 | 0.96 | 0.935 |
| SVM, LR | 2 | 0.89 | 0.94 | 0.92 | |
| | 3 | 0.94 | 0.92 | 0.93 | |
| MNB, DT | 1 | 0.98 | 0.94 | 0.96 | 0.926 |
| SVM, KNN | 2 | 0.86 | 0.95 | 0.90 | |
| | 3 | 0.95 | 0.89 | 0.92 | |
| MNB, SVM | 1 | 0.97 | 0.92 | 0.95 | 0.929 |
| LR, KNN | 2 | 0.89 | 0.95 | 0.92 | |
| | 3 | 0.93 | 0.92 | 0.92 | |

| | | | | | |
|---|---|---|---|---|---|
| DT, SVM LR, KNN | 1 | 0.97 | 0.92 | 0.95 | 0.929 |
| | 2 | 0.89 | 0.95 | 0.92 | |
| | 3 | 0.93 | 0.92 | 0.92 | |
| MNB, DT SVM | 1 | 1.00 | 0.95 | 0.97 | 0.929 |
| | 2 | 0.85 | 0.94 | 0.90 | |
| | 3 | 0.94 | 0.90 | 0.92 | |
| MNB, DT LR | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| | 2 | 0.88 | 0.94 | 0.91 | |
| | 3 | 0.92 | 0.91 | 0.92 | |
| MNB, DT KNN | 1 | 0.94 | 0.94 | 0.94 | 0.882 |
| | 2 | 0.80 | 0.92 | 0.85 | |
| | 3 | 0.92 | 0.80 | 0.86 | |
| MNB, SVM LR | 1 | 0.97 | 0.94 | 0.96 | 0.935 |
| | 2 | 0.89 | 0.94 | 0.92 | |
| | 3 | 0.94 | 0.92 | 0.93 | |
| MNB, SVM KNN | 1 | 0.96 | 0.95 | 0.96 | 0.944 |
| | 2 | 0.92 | 0.94 | 0.93 | |
| | 3 | 0.95 | 0.94 | 0.95 | |
| MNB, LR KNN | 1 | 0.96 | 0.92 | 0.94 | 0.932 |
| | 2 | 0.90 | 0.95 | 0.93 | |
| | 3 | 0.93 | 0.92 | 0.93 | |
| DT, SVM LR | 1 | 0.97 | 0.94 | 0.96 | 0.926 |
| | 2 | 0.88 | 0.93 | 0.90 | |
| | 3 | 0.93 | 0.91 | 0.92 | |
| DT, SVM KNN | 1 | 0.98 | 0.95 | 0.96 | 0.923 |
| | 2 | 0.85 | 0.94 | 0.89 | |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | 0.95 | 0.88 | 0.91 | |
| DT, LR KNN | 1 | 0.97 | 0.92 | 0.95 | 0.926 |
| | 2 | 0.88 | 0.95 | 0.91 | |
| | 3 | 0.93 | 0.91 | 0.92 | |
| LR, SVM KNN | 1 | 0.97 | 0.92 | 0.95 | 0.929 |
| | 2 | 0.89 | 0.95 | 0.92 | |
| | 3 | 0.93 | 0.92 | 0.92 | |
| MNB, DT | 1 | 0.99 | 0.86 | 0.92 | 0.838 |
| | 2 | 0.67 | 0.99 | 0.80 | |
| | 3 | 0.96 | 0.68 | 0.80 | |
| MNB, SVM | 1 | 0.98 | 0.94 | 0.96 | 0.941 |
| | 2 | 0.89 | 0.94 | 0.92 | |
| | 3 | 0.95 | 0.94 | 0.95 | |
| LR, MNB | 1 | 0.97 | 0.92 | 0.95 | 0.929 |
| | 2 | 0.89 | 0.95 | 0.92 | |
| | 3 | 0.93 | 0.92 | 0.92 | |
| KNN, MNB | 1 | 0.85 | 0.94 | 0.89 | 0.867 |
| | 2 | 0.85 | 0.80 | 0.83 | |
| | 3 | 0.90 | 0.86 | 0.88 | |
| DT, SVM | 1 | 0.99 | 0.94 | 0.96 | 0.912 |
| | 2 | 0.81 | 0.94 | 0.87 | |
| | 3 | 0.94 | 0.86 | 0.90 | |
| LR, DT | 1 | 0.97 | 0.94 | 0.96 | 0.929 |
| | 2 | 0.90 | 0.93 | 0.92 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| | 1 | 0.94 | 0.93 | 0.93 | |

| | | | | | |
|---|---|---|---|---|---|
| DT, KNN | 2 | 0.77 | 0.91 | 0.83 | 0.870 |
| | 3 | 0.91 | 0.78 | 0.84 | |
| LR, SVM | 1 | 0.97 | 0.94 | 0.96 | 0.932 |
| | 2 | 0.89 | 0.93 | 0.91 | |
| | 3 | 0.93 | 0.92 | 0.93 | |
| KNN, SVM | 1 | 0.97 | 0.94 | 0.96 | 0.932 |
| | 2 | 0.88 | 0.94 | 0.91 | |
| | 3 | 0.94 | 0.92 | 0.93 | |
| LR, KNN | 1 | 0.97 | 0.92 | 0.95 | 0.932 |
| | 2 | 0.89 | 0.95 | 0.92 | |
| | 3 | 0.93 | 0.92 | 0.93 | |

**Table 7.6 Performance of Stacking Technique**

(*** Final Estimator in the Model)

| Models | Class | Precision | Recall | F1-Score | Accuracy |
|---|---|---|---|---|---|
| MNB, DT | 1 | 1.00 | 0.92 | 0.96 | 0.926 |
| SVM, LR | 2 | 0.88 | 0.92 | 0.90 | |
| KNN*** | 3 | 0.90 | 0.93 | 0.92 | |
| MNB, DT | 1 | 1.00 | 0.95 | 0.97 | 0.941 |
| SVM, LR*** | 2 | 0.87 | 0.97 | 0.92 | |
| KNN | 3 | 0.96 | 0.91 | 0.93 | |
| MNB, DT | 1 | 0.97 | 0.94 | 0.96 | 0.929 |
| SVM***, LR | 2 | 0.90 | 0.93 | 0.92 | |
| KNN | 3 | 0.92 | 0.92 | 0.92 | |
| MNB, DT*** | 1 | 0.97 | 0.93 | 0.95 | |

| | | | | | |
|---|---|---|---|---|---|
| SVM, LR | 2 | 0.82 | 0.93 | 0.88 | 0.909 |
| KNN | 3 | 0.94 | 0.87 | 0.90 | |
| MNB***, DT | 1 | NA | NA | NA | NA |
| SVM, LR | 2 | NA | NA | NA | |
| KNN | 3 | NA | NA | NA | |
| DT, LR | 1 | NA | NA | NA | NA |
| MNB***, KNN | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| DT***, LR | 1 | 0.94 | 0.89 | 0.92 | 0.891 |
| MNB, KNN | 2 | 0.85 | 0.89 | 0.87 | |
| | 3 | 0.88 | 0.89 | 0.88 | |
| DT, LR*** | 1 | 0.97 | 0.92 | 0.95 | 0.912 |
| MNB, KNN | 2 | 0.84 | 0.94 | 0.89 | |
| | 3 | 0.93 | 0.87 | 0.90 | |
| DT, LR | 1 | 0.99 | 0.93 | 0.96 | 0.926 |
| MNB, KNN*** | 2 | 0.86 | 0.95 | 0.90 | |
| | 3 | 0.94 | 0.90 | 0.92 | |
| MNB, SVM | 1 | 1.00 | 0.95 | 0.97 | 0.947 |
| LR***, DT | 2 | 0.89 | 0.96 | 0.93 | |
| | 3 | 0.95 | 0.93 | 0.94 | |
| MNB, SVM*** | 1 | 0.96 | 0.94 | 0.95 | 0.926 |
| LR, DT | 2 | 0.90 | 0.93 | 0.92 | |
| | 3 | 0.92 | 0.91 | 0.91 | |
| MNB, SVM | 1 | 0.95 | 0.91 | 0.93 | 0.909 |
| LR, DT*** | 2 | 0.85 | 0.93 | 0.89 | |
| | 3 | 0.94 | 0.88 | 0.91 | |

| | | | | | |
|---|---|---|---|---|---|
| MNB***, SVM<br>LR, DT | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| DT, SVM<br>KNN, MNB*** | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| DT***, SVM<br>KNN, MNB | 1 | 0.93 | 0.91 | 0.92 | 0.882 |
| | 2 | 0.84 | 0.87 | 0.86 | |
| | 3 | 0.87 | 0.87 | 0.87 | |
| DT, SVM***<br>KNN, MNB | 1 | 0.98 | 0.89 | 0.93 | 0.900 |
| | 2 | 0.81 | 0.94 | 0.87 | |
| | 3 | 0.92 | 0.87 | 0.90 | |
| DT, SVM<br>KNN***, MNB | 1 | 1.00 | 0.94 | 0.97 | 0.929 |
| | 2 | 0.86 | 0.94 | 0.90 | |
| | 3 | 0.93 | 0.91 | 0.92 | |
| LR, KNN<br>MNB***, SVM | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| LR, KNN<br>MNB, SVM*** | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| | 2 | 0.88 | 0.93 | 0.91 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| LR***, KNN<br>MNB, SVM | 1 | 0.98 | 0.93 | 0.95 | 0.935 |
| | 2 | 0.88 | 0.95 | 0.91 | |
| | 3 | 0.95 | 0.92 | 0.94 | |
| LR, KNN***<br>MNB, SVM | 1 | 0.99 | 0.90 | 0.94 | 0.912 |
| | 2 | 0.86 | 0.90 | 0.88 | |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | 0.90 | 0.93 | 0.91 | |
| SVM, KNN | 1 | 0.97 | 0.91 | 0.94 | |
| LR, DT*** | 2 | 0.81 | 0.92 | 0.86 | 0.897 |
| | 3 | 0.92 | 0.86 | 0.89 | |
| SVM***, KNN | 1 | 0.97 | 0.94 | 0.96 | |
| LR, DT | 2 | 0.90 | 0.93 | 0.92 | 0.929 |
| | 3 | 0.92 | 0.92 | 0.92 | |
| SVM, KNN | 1 | 1.00 | 0.95 | 0.97 | |
| LR***, DT | 2 | 0.86 | 0.94 | 0.90 | 0.932 |
| | 3 | 0.94 | 0.91 | 0.92 | |
| SVM, KNN*** | 1 | 1.00 | 0.92 | 0.96 | |
| LR, DT | 2 | 0.85 | 0.93 | 0.89 | 0.923 |
| | 3 | 0.93 | 0.92 | 0.92 | |
| LR, DT | 1 | NA | NA | NA | |
| MNB*** | 2 | NA | NA | NA | NA |
| | 3 | NA | NA | NA | |
| LR, DT*** | 1 | 0.93 | 0.89 | 0.91 | |
| MNB | 2 | 0.84 | 0.92 | 0.88 | 0.888 |
| | 3 | 0.89 | 0.86 | 0.88 | |
| LR***, DT | 1 | 0.97 | 0.91 | 0.94 | |
| MNB | 2 | 0.84 | 0.93 | 0.88 | 0.909 |
| | 3 | 0.92 | 0.88 | 0.90 | |
| MNB***, DT | 1 | NA | NA | NA | |
| KNN | 2 | NA | NA | NA | NA |
| | 3 | NA | NA | NA | |
| MNB, DT*** | 1 | 0.94 | 0.89 | 0.91 | |

| | | | | | |
|---|---|---|---|---|---|
| KNN | 2 | 0.81 | 0.90 | 0.85 | 0.879 |
| | 3 | 0.90 | 0.86 | 0.88 | |
| MNB, DT KNN*** | 1 | 0.97 | 0.92 | 0.95 | 0.920 |
| | 2 | 0.87 | 0.92 | 0.89 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| LR MNB***, KNN | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| LR*** MNB, KNN | 1 | 0.96 | 0.91 | 0.94 | 0.903 |
| | 2 | 0.82 | 0.92 | 0.87 | |
| | 3 | 0.93 | 0.87 | 0.90 | |
| LR MNB, KNN*** | 1 | 0.98 | 0.91 | 0.95 | 0.923 |
| | 2 | 0.86 | 0.94 | 0.90 | |
| | 3 | 0.93 | 0.92 | 0.92 | |
| KNN LR, DT*** | 1 | 0.94 | 0.89 | 0.92 | 0.882 |
| | 2 | 0.83 | 0.92 | 0.87 | |
| | 3 | 0.88 | 0.84 | 0.86 | |
| KNN LR***, DT | 1 | 0.96 | 0.92 | 0.94 | 0.879 |
| | 2 | 0.79 | 0.91 | 0.85 | |
| | 3 | 0.89 | 0.82 | 0.85 | |
| KNN*** LR, DT | 1 | 0.99 | 0.93 | 0.96 | 0.920 |
| | 2 | 0.83 | 0.94 | 0.88 | |
| | 3 | 0.95 | 0.89 | 0.92 | |
| MNB***, SVM DT | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |

| | | | | | |
|---|---|---|---|---|---|
| MNB, SVM DT*** | 1 | 0.92 | 0.92 | 0.92 | 0.867 |
| | 2 | 0.82 | 0.87 | 0.84 | |
| | 3 | 0.86 | 0.82 | 0.84 | |
| MNB, SVM*** DT | 1 | 0.97 | 0.89 | 0.93 | 0.900 |
| | 2 | 0.82 | 0.92 | 0.87 | |
| | 3 | 0.91 | 0.88 | 0.90 | |
| LR, SVM MNB*** | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| LR, SVM*** MNB | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| | 2 | 0.88 | 0.93 | 0.91 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| LR***, SVM MNB | 1 | 0.97 | 0.94 | 0.96 | 0.935 |
| | 2 | 0.88 | 0.94 | 0.91 | |
| | 3 | 0.95 | 0.92 | 0.94 | |
| LR, DT*** SVM | 1 | 0.98 | 0.91 | 0.95 | 0.885 |
| | 2 | 0.78 | 0.89 | 0.83 | |
| | 3 | 0.90 | 0.86 | 0.88 | |
| LR, DT SVM*** | 1 | 0.97 | 0.94 | 0.96 | 0.929 |
| | 2 | 0.90 | 0.93 | 0.92 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| LR***, DT SVM | 1 | 0.97 | 0.95 | 0.96 | 0.935 |
| | 2 | 0.89 | 0.93 | 0.91 | |
| | 3 | 0.94 | 0.92 | 0.93 | |
| SVM, KNN MNB*** | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | NA | NA | NA | |
| SVM***, KNN MNB | 1 | 0.96 | 0.89 | 0.93 | 0.897 |
| | 2 | 0.82 | 0.93 | 0.87 | |
| | 3 | 0.92 | 0.87 | 0.89 | |
| SVM, KNN*** MNB | 1 | 0.98 | 0.93 | 0.95 | 0.932 |
| | 2 | 0.90 | 0.93 | 0.92 | |
| | 3 | 0.92 | 0.93 | 0.93 | |
| SVM, KNN DT*** | 1 | 0.98 | 0.92 | 0.95 | 0.882 |
| | 2 | 0.81 | 0.88 | 0.84 | |
| | 3 | 0.86 | 0.85 | 0.86 | |
| SVM***, KNN DT | 1 | 0.99 | 0.91 | 0.95 | 0.876 |
| | 2 | 0.77 | 0.92 | 0.84 | |
| | 3 | 0.89 | 0.81 | 0.85 | |
| SVM, KNN*** DT | 1 | 1.00 | 0.94 | 0.97 | 0.912 |
| | 2 | 0.84 | 0.91 | 0.87 | |
| | 3 | 0.90 | 0.89 | 0.89 | |
| LR, KNN SVM*** | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| | 2 | 0.88 | 0.93 | 0.91 | |
| | 3 | 0.92 | 0.92 | 0.92 | |
| LR***, KNN SVM | 1 | 0.98 | 0.93 | 0.95 | 0.926 |
| | 2 | 0.86 | 0.94 | 0.90 | |
| | 3 | 0.94 | 0.91 | 0.92 | |
| LR, KNN*** SVM | 1 | 0.99 | 0.89 | 0.94 | 0.914 |
| | 2 | 0.86 | 0.92 | 0.89 | |
| | 3 | 0.90 | 0.93 | 0.92 | |
| | 1 | NA | NA | NA | |

| | | | | | |
|---|---|---|---|---|---|
| DT, MNB*** | 2 | NA | NA | NA | NA |
| | 3 | NA | NA | NA | |
| DT***, MNB | 1 | 0.92 | 0.90 | 0.91 | 0.867 |
| | 2 | 0.76 | 0.89 | 0.82 | |
| | 3 | 0.93 | 0.82 | 0.87 | |
| MNB, LR*** | 1 | 0.94 | 0.89 | 0.91 | 0.894 |
| | 2 | 0.82 | 0.92 | 0.87 | |
| | 3 | 0.92 | 0.87 | 0.90 | |
| MNB***, LR | 1 | SVM | SVM | SVM | NA |
| | 2 | SVM | SVM | SVM | |
| | 3 | SVM | SVM | SVM | |
| MNB***, KNN | 1 | SVM | SVM | SVM | NA |
| | 2 | SVM | SVM | SVM | |
| | 3 | SVM | SVM | SVM | |
| MNB, KNN*** | 1 | 0.96 | 0.89 | 0.92 | 0.903 |
| | 2 | 0.83 | 0.92 | 0.88 | |
| | 3 | 0.92 | 0.90 | 0.91 | |
| MNB***, SVM | 1 | NA | NA | NA | NA |
| | 2 | NA | NA | NA | |
| | 3 | NA | NA | NA | |
| MNB, SVM*** | 1 | 0.95 | 0.89 | 0.92 | 0.894 |
| | 2 | 0.82 | 0.92 | 0.87 | |
| | 3 | 0.91 | 0.88 | 0.90 | |
| DT, LR*** | 1 | 0.99 | 0.61 | 0.75 | 0.631 |
| | 2 | 0.46 | 0.98 | 0.63 | |
| | 3 | 0.95 | 0.34 | 0.51 | |

| | | | | | |
|---|---|---|---|---|---|
| DT***, LR | 1 | 0.94 | 0.89 | 0.92 | 0.885 |
| | 2 | 0.82 | 0.90 | 0.86 | |
| | 3 | 0.90 | 0.87 | 0.88 | |
| DT, KNN*** | 1 | 0.98 | 0.54 | 0.70 | 0.602 |
| | 2 | 0.44 | 0.98 | 0.61 | |
| | 3 | 0.95 | 0.32 | 0.48 | |
| DT***, KNN | 1 | 0.95 | 0.83 | 0.89 | 0.805 |
| | 2 | 0.64 | 0.91 | 0.75 | |
| | 3 | 0.92 | 0.69 | 0.79 | |
| DT***, SVM | 1 | 0.98 | 0.93 | 0.95 | 0.879 |
| | 2 | 0.78 | 0.89 | 0.83 | |
| | 3 | 0.88 | 0.82 | 0.85 | |
| DT, SVM*** | 1 | 1.00 | 0.54 | 0.70 | 0.617 |
| | 2 | 0.45 | 0.98 | 0.62 | |
| | 3 | 0.96 | 0.36 | 0.52 | |
| LR, KNN*** | 1 | 0.95 | 0.89 | 0.92 | 0.891 |
| | 2 | 0.80 | 0.89 | 0.84 | |
| | 3 | 0.93 | 0.89 | 091 | |
| LR***, KNN | 1 | 0.94 | 0.89 | 0.91 | 0.847 |
| | 2 | 0.74 | 0.88 | 0.81 | |
| | 3 | 0.88 | 0.78 | 0.83 | |
| LR***, SVM | 1 | 0.96 | 0.93 | 0.95 | 0.926 |
| | 2 | 0.87 | 0.94 | 0.90 | |
| | 3 | 0.95 | 0.91 | 0.93 | |
| LR, SVM*** | 1 | 0.97 | 0.92 | 0.95 | 0.923 |
| | 2 | 0.88 | 0.93 | 0.91 | |

| | | | | | |
|---|---|---|---|---|---|
| | 3 | 0.92 | 0.92 | 0.92 | |
| SVM, KNN*** | 1 | 0.95 | 0.92 | 0.94 | 0.909 |
| | 2 | 0.87 | 0.90 | 0.88 | |
| | 3 | 0.90 | 0.91 | 0.90 | |
| SVM***, KNN | 1 | 0.94 | 0.87 | 0.90 | 0.844 |
| | 2 | 0.73 | 0.89 | 0.80 | |
| | 3 | 0.88 | 0.78 | 0.83 | |