Arab American University

Faculty of Graduate Studies

Lexicon-Based Sentiment Analysis for Arabic Slang Text

By

Khalil Edward Khalil Rantisi

Supervisor

Dr. Khalid Rabayah

This Thesis Was Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Computer Science.

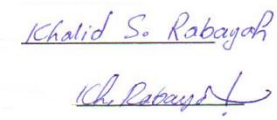
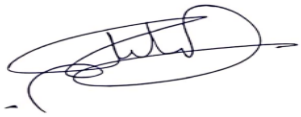
September / 2021

Lexicon-Based Sentiment Analysis for Arabic Slang Text

By

Khalil Edward Khalil Rantisi

This thesis was defended successfully on **11/09/2021** and approved by:

Committee Members	Signature
1. Supervisor: **Dr. Khalid Rabayah**	
2. Internal Examiner: **Dr. Mohammed Maree**	
3. External Examiner: **Dr. Ahmad Hasasneh**	

Declaration

This is to declare that the thesis entitled "**Lexicon-Based Sentiment Analysis for Arabic Slang Text** " under the supervision of " Dr. Khalid Rabayah " is my own work and does not contain any unacknowledged work or material previously published or written by another person, except where due reference is made in the text of the document.

Name: Khalil Edward Khalil Rantisi

Signature: 

Date: 11-9-2021

Dedication

The dedication of this work is split in Four ways:

To my father and mother,

To my supervisor,

To my colleagues,

And To my Wife and Children, who stuck

With me till the very end.

.

Acknowledgments

First and foremost, praise is to God, the Creator of all creations, who has given me strength and an iron will to pursue my research without a thought of weakness. He showered me with determination, his blessings and most of all, hope. Furthermore, *I* would like to extend my special appreciation to my Supervisor Dr. Khalid Rabayah for the continuous help and advice he has given me throughout my work. His constructive remarks and comments helped a great deal in advancing my work. *I* received a lot of guidance from him, he is always willing to share his ideas in every aspect of my work. *I* would like to thank him for all that she has done for the success of my work. *I am* thankful, because he has always brought a different perspective to my work through his varied experience and knowledge. His remarks enhanced the quality of my work. Special thanks for the thoughtful suggestions on many aspects that had helped me develop the work to what it is now. My thanks to my dr. Khalid for setting me in this way towards my goal by motivating me to take up this work. Thanks Dr. Mohammed Maree Internal Examiner and Dr.Ahmad Hasasneh External Examiner.

I am thankful to my second family, my boss and my colleagues in my work and for administrative and academic staff for their support and help.

Finally, *I am* thankful to my family including my father, mother, sister, wife, and my kids, for their endless patience, support and confidence in me. *I am* truly grateful for my family for making all the sacrifices while *I* was in my study.

Abstract

The rapid spread of social media generates a massive amount of data every day. Understanding and mining this data to determine the attitude of users towards products, services, events, and other topics are getting very beneficial for individuals as well as stakeholders. This kind of text mining requires a high-level language processing, a.k.a Natural Language Processing (NLP). This topic, s receiving increasing attention from all stakeholders including machine learning and artificial intelligence specialists, the business community, language specialists, etc. However, text mining and information extraction within the Arabic content still require extra efforts to arrive at the level of other languages like English. The Arabic language is one of the popular content-sharing Languages through social networks; the analysis of content written in Arabic faces various challenges, especially in the case of Colloquial/ Slang Arabic which is the widely used language in social media. This thesis is intended to place some efforts towards trying some approaches to enhance the sentiment analysis of social media content written in the Arabic language with the focus of Palestinian colloquial.

The proposed approach is based on the use of machine learning algorithms and lexicon tools to enhance outputs of sentiment analysis performed on Arabic social media content. The approach consists of two phases, the first phase is handled by the machine learning algorithm using three classifiers; Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Naïve Bayes (NB). The second phase is handled by a lexicon-based method using two classifiers: SVM, and NB. The second phase of the approach is implemented to enhance the result of the first stage. The output of the first stage was used to train the SVM and NB classifiers.

The proposed methods are tested using a customized dataset extracted from the Facebook pages of some public services provider in Palestine, the dataset consists thousands of comments and posts on various topics.

The results of the analysis revealed that lexicon-based approach improved the accuracy of comment polarity detection as the accuracy results increased from 90.57% to 90.68% using the proposed approach, while the F-measure results increased from 94.50% to 94.58%.the results also indicated that SVM was the best ML algorithm compared to NB and KNN for this research problem.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CEO	Chief Executive Officer
CTO	Chief Technical Officer
SIGL	Signal Advance Inc
FB	Facebook
NLP	Natural Language Processing
MVP	Minimum Viable Product
POS	Part of Speech
API	Application Programming Interface
BOW	Bag-Of-Words
RDF	Radial Distribution Function
ML	Machine Learning
MSA	Modern Standard Arabic
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
NB	Navie Bayes
TF-IDF	Term Frequency-Inverse Document Frequency
TF	Term Frequency
BTO	Binary Term Occurrences
ABSA	Aspect -Based Sentiment Analysis
RNN	Recurrent Neural Network
VSM	Vector Space Model

AOPI	Arabic opinions Polarity Identification
NN	Neural Network
DT	Decision Tree
NLTK	Natural Language Tool Kit
SSWIL	Slang Sentimental Words and Idioms Lexicon
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
P	Precision
R	Recall

Chapter 1

Introduction

1.1 Overview

With the great development of technology and communications, the world has become a small village. Social media which has become an integral part of our habits and our daily life, has led to its use in various aspects of life. It has an impact on the feelings and opinions of people and the spread of all information among people.

A short while ago, a tweet by US President Trump caused widespread chaos in the United States, especially in the corridors of the US Capitol building, which led to the declaration of a state of emergency in the country and disabling of the accounts of the US president on social media on charges of incitement, and so, he will be sued in court soon due to the incitement of violence [1].

Also, Elon Musk the founder, CEO, CTO, and chief designer of Space X, tweets and said simply, “Use Signal.” Musk was recommending the encrypted messaging service (Signal) that competes with Facebook's WhatsApp, but a frenzy ensued in Signal Advance (SIGL), nonetheless, has continued for days. [2]

The surge in Texas health care company "Signal Advance" continued as the stock rose by 438%, after previously gaining 1800% during one 24-hour. [2]

From here came the role of data analysis and mining, which includes sentiment analysis to find out and discover whether our feelings are negative or positive through comments and posting on social media platforms.

Sentiment analysis has recently become one of the growing areas of research related to text mining and natural language processing. The increasing availability of online resources and the popularity of rich and fast resources for opinion sharing like news online review sites and personal blogs caused several parties such as customers, companies, and governments to start analyzing and exploring these opinions.

Sentiment analysis is a particular case class for computational strategies that naturally extracts and summarizes those conclusions from opinions such a monstrous quantity of data that the average human reader is not able to process. In the area of advanced Internet technology, individuals from different locations worldwide are part of each other through social media Websites such as Facebook, google+, Twitter, Instagram, ...etc.

Facebook reports that 3.14 billion people use at least one of the company's core products (Facebook, WhatsApp, Instagram, or Messenger) each month [3].

Facebook has nearly 2.7 billion active users with feature optimization. It's been performed using millions of posts every day which makes it a main social media Internet site worldwide. As Facebook has this huge range of users and enormous data, it has usually been used as "an informative resource by various organizations to research public opinions and gather critical feedback" [3]. In a comment, users can write their views regarding any topic or general thoughts. People write in a concise manner by using slang which makes sentiment analysis a hard task. Sentiment analysis may be characterized similarly as the process of categorizing the opinions expressed through a comment to understand the user views about that topic. It is beneficial for the marketers as they can examine and analyze the opinion of the public towards their brand and existing/newly released products which would help them to evaluate their performance and improve it [3]. The findings can be used in many fields such as evaluating and tracking changes in feeling with a case, feelings about a specific brand or release of a specific item, analyzing the public opinion of government policies, ... etc. To classify the comments and evaluate the outcomes, a lot of studies have been performed on Facebook data. In this research, it is strived to review some research in this domain and study how to carry out Facebook data sentiment analysis.

This research is dedicated to the use of supervised machine learning for the purpose of enhancing lexicon-based sentiment analysis. Three supervised algorithms will be employed and in the course of the thesis we shall compare these algorithms with each other's.

General Background

Social media platforms like pages and groups, have played essential roles in several fields and specialties, like management, business, politics, sociology, etc. It gives the right to share his/her thoughts, opinions and sentiments in posts and comments. Recently, there has been renewed interest in the world of technology. People these days depend upon social websites to communicate with relatives and coworkers worldwide. Here sentiments come into play where opinions are shared by anyone in the time they feel and like to be shared. Sentiments are nothing but feelings with respect to events. Sentiment analysis is to determine the opinion of users related to some event or the statement describing the emotion of the user i.e., what he/she feels about it. The studies on sentiment analysis have been going for a long term. Today, the issue of sentiment analysis is hot and evolving in the fields of studies, research and technology.

Due to day by day significant rise in the number of users of social networking sites, a large amount of data are produced in the form of text, audio, video and images. This created a need to do sentiment analysis, such as texts analysis and mining to find out whether these sentiments are neutral, negative or positive.

1.2 Problem Definition and Objectives

The adoption of data analysis techniques has gained a huge boost and development in domains of use, in technologies and in size as well. The analysis of social network content is not an exception. Social media have brought birth to new communication techniques recognized as prompt (instance) messaging, where a high percentage of this communication is occurring in text

forms. In this way, there will be a considerable potential that this accessible online data could be utilized for analysis to extract information and knowledge that can be used to make better decisions. If implemented correctly, the analysis results will affect the daily decision-making processes and capabilities, in addition to affecting every sector of the community. The automatic extraction and detection of knowledge and ideas from social media text is turning into a hot topic, and quickly becoming amongst the highly researched issues.

Social media in general and Facebook in particular, are generating data that posed a huge challenge for text miners and natural language processing specialists. Among the most debated techniques that are continuously being implemented and tested is sentiment analysis, since social media is the platform that is adopted by users to express their feeling and opinions associated with users or entities, products, services, etc. Sentiment analysis is the primary tool that is used to identify and classify sentiments out of texts. Knowledge-based approaches in addition to machine learning tactics are used for sentiment classification.

With that in mind, this research aims to improve and enhance the accuracy of using lexicon-based approaches coupled with machine learning classification algorithms. The research aims to help in filling up the gap in the literature in regard to social media Arabic content analysis, mainly sentiment analysis. This area of research will have a significant impact and benefits on different sectors in a society like the marketers which enables them to explore the opinion of the public towards their brand and services. A matter that will assist them in conducting a performance evaluation of clients' satisfaction.

This research is also a contribution to the field of opinion mining and sentiment analysis. In particular, it aims at automatically identifying sentiments (e.g. positive or negative) of for example, Facebook posts or comments. Since in commercial and advertisements, social media is

heavily implemented, and since there is a lack of researches on supervised sentiment analysis for such a domain, this work is counted as the main contributor to the field especially in the Palestinian context.

Although the development of selling and buying online technologies is growing, customers do not see the product physically, this is why they need a kind of guarantee to ensure that the product is qualified and has the features they want. However, customers can trust each other, so they rely on each other's feedback (review) of a product they buy and use. The difficulty in this respect has to do with the huge number of reviews that they need to go over to help them make the right decision. An automatic technique can be of great help for those online customers and especially potential buyers. This technique is built to process and summarize those reviews in order to help customers in products evaluations. However, as the reviews are written as free text, unstructured and ungrammatical, the automatic systems may have some miss of accuracy.

Arab users use social network sites in their communications, and they express their views on these platforms. They post on Facebook and Twitter and they write comments on posts and tweets as opinions. To know the level of satisfaction and/or dissatisfaction of Arabic users comments, an approach is built to summarize and mine the Arabic comments written in an unstructured and ungrammatical way. People comment on topics based on their feeling towards those topics. So their feelings are translated via summarizing their comments.

1.3 Research Questions

The problem with the Arabic sentiment analysis has to do with the multi-meaning of the individual words, multi-accents and different writing styles that can be used by Arabs. This research will work to develop an approach that acts to overcome these issues. The approach will rely on machine learning algorithms that can be used to analyze content and text, with a high

level of accuracy, using the RapidMiner as a data mining instrument. RapidMiner is used in this thesis because it has a Text Processing package. This package includes many operators that support text mining. For example, there are operators for tokenization, stemming, filtering stopwords, and generating n-grams. The main reason for choosing RapidMiner is that the text Processing package can deal with the Arabic language.

This research will rely on a prepared Palestinian Arabic dataset that contains almost 6,600 comments. Using this data set, different supervised algorithms will be trained to build sentiment models that can be used to classify comments into positive, negative or neutral. The main research questions in this thesis are summarized below:

Main question: How can an effective framework for Arabic sentiment analysis be created?

The primary contribution of this thesis is to investigate the effectiveness of sentiment analysis of Arabic text using supervised ML algorithms.

Sub-question 1: Which machine learning algorithm produces a better Arabic Sentiment classifier?

Sub-question 2: Is there only one framework that covers all Arabic accents?

At this point, the research will try to test how features can be dealt with and cover different Arabic accents like Palestinian accent and others.

Sentiment analysis attempts to identify and analyze opinions and emotions. A common point in almost any work on sentiment analysis has to do with the need to identify which element of the language contribute more to express the subjectivity in text, with its general opinion-related (or properties (Contextual polarity) such as positivity, negativity or objectivity.

Contextual polarity means the polarity of the expression in which a word appears considering the context of the sentence and document. This approach first determines if the expression is polar (positive or negative), then taking the polar expressions for additional classification in order to determine the polarity for each polar expression. The main task in sentiment analysis is classifying a document or a sentence based on polarity. The polarity of a document or sentence can be positive or negative. An advanced task for sentiment analysis is the classifying of emotional states like anger, sadness, and happiness. However, this thesis scope does not include the contextual polarity detection.

1.4 Thesis Organization

This thesis is organized as follows; Chapter 1 presents a general introduction to the subject, where the problem statement is defined, the significance of the study is discussed, the objectives are determined. Chapter 2 introduces the theoretical basis of the study, where the basic concepts of machine learning algorithms used in content and text mining are discussed and explained. A review of existing studies is summarized in chapter 3. Chapter 4 discusses the methodology adopted in compiling the research project, where the data collection, data cleaning, preparation and analysis were described. In chapter 5, Data Analysis and Discussion of results are presented, and finally, in chapter 6, conclusions are presented, recommendations are provided, some directions for future researchers are given, and limitations to the study are discussed.

Chapter 2

Theoretical Basis of Sentiment Analysis

This chapter presents an overview of sentiment analysis. It goes through sentiment analysis history, definition, applications, and techniques. It also discusses Natural Language Processing (NLP), Text Analysis, and Computational Linguistics as sentiment analysis uses those fields to accomplish its task. Since this thesis focuses on Arabic sentiment analysis, this chapter also discusses sentiment analysis as applied in Arabic, and the characteristics of Arabic language content analysis.

2.1 Introduction of Sentiment Analysis

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [4]. This process identifies and categorizes People’s opinions to determine the writer's opinion about a specific topic. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, review mining, and so forth.

The main task in sentiment analysis is classifying a document or a sentence based on a polarity. The polarity of a document or sentence can be positive or negative. An advanced task for sentiment analysis is classifying the emotional states like anger, sadness, and happiness.

The first research that matched sentiment analysis was published in 1940 [5]. This research relied on survey-based methods instead of opinion interviews. According to this work, personal interviews are slow and costly. The author's work was based on expert or public opinions not a specific group of customers or users.

Later on, three other papers were published from 1945 and 1947 that investigated public opinions in World War II countries (Japan, Italy, and Czechoslovakia) which suffered from the war [6] [7] [8]. Those papers also were based on survey methods.

In the 90s, a revolution in computing was started and new computer-based systems appeared. For example, a paper was published in 1995 [9], which used a computer-based system to analyze experts' opinions in the industrial safety domain in order to collect their opinions. The earliest work in sentiment analysis based on computerized systems was proposed in 1990 by Wiebe where a method for detecting the sentences subjectivity of a narrative was proposed [10]. Another paper was proposed in 1999 in order to classify sentences based on their subjectivity using a new standard [11].

Numerous persuasive papers were started in the turn of the 21st century that concentrates on sentiment analysis. In 1997, a paper was published to predict if the adjectives are the same orientation or different [12]. The authors built a list of positive and negative adjectives to do the prediction. The distribution of sentiment analysis published papers over years is illustrated in Figure 2.1 according to [13].

Sentiment Analysis uses Natural Language Processing [14], text analysis [15], and computational linguistics [16] to accomplish its task.

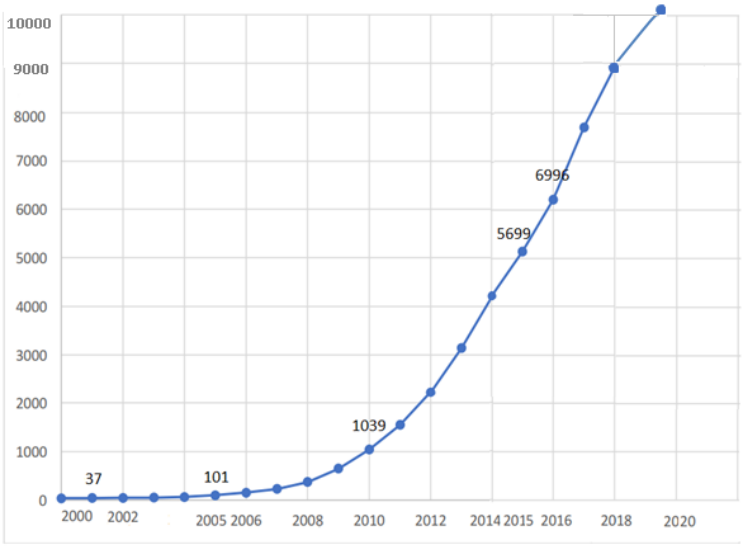


Figure 2. 1 Sentiment Analysis Paper Distribution Over Years [13]

2.1.1 Sentiment Analysis Applications

Sentiment analysis is very useful for everyone, at all levels, such as marketers, stakeholders, decision-makers, and managers. It is a vital task in business as it gives useful insights about customer's preferences, different points of view, and pros and cons in products or services. This valuable data helps organizations to improve their services and products. Sentiment analysis is so important also for individuals. It allows users to get relevant information about different products and services with a simple search on social media websites. For governments, sentiment analysis helps in investigating the impression of users' data and helps them to take a suitable decision.

There are many applications for sentiment analysis that reflect the need for sentiment analysis and its importance nowadays:

1. Application for computing customer satisfaction metrics: the firm can capture the customer satisfaction ratio by such applications from a number of negative and positive reviews on products and services.

2. Identifying disadvantages and problems: it helps customer care and customer services to focus on the problems and defects in services or products.
3. Promoting of services and products: good reviews and comments can be used to promote products and services. Customer good experiences are used as an advertisement for your products.
4. Measuring the public opinion: this kind of application is used to measure the public opinion towards a specific decision or public opinion case. It's so important for governments to watch citizens' manners.

Anticipating conflicts: it helps in anticipating the conflicts in unstable areas that may have revolutions or demonstrations. This expectation helps in taking precautions to prevent any unwanted behavior or violence.

2.2 Natural Language Processing

Natural Language Processing (NLP) [14], is a field of computer science that allows computers to analyze, communicate, study, and extract useful meaning from the human everyday language in an intelligent way. NLP helps in organizing and structuring knowledge from human language to perform some important tasks such as machine translation, automatic text summarization, speech recognition, text classification, text segmentation, and sentiment analysis.

Among the various applications of NLP are the following:

- **Machine translation**: It is the task of translating from the input language to the output language. It uses a bilingual data to build models for language and phrases which are used for translating text. It is called also automated translation [17].

- **Automatic text summarization:** It is the task of summarizing a document automatically by shortening the text. The idea is to extract or select the related words, sentences, and phrases from the original text for the purpose of creating a summary using internal semantic techniques [18].
- **Speech recognition:** It is the process of identifying human spoken language to convert them to a readable format. It's a speech-to-text conversion process [19].
- **Text classification (tagging**): It is the task of classifying text automatically by giving a label for every word. It uses mathematical models and algorithms to assign category to the input text [19].
- **Text segmentation:** It involves dividing written text into meaningful parts (words, sentences, and topics). It identifies the boundaries between words to extract word segments [20].
- **Sentiment analysis:** Defined as the process of extracting the polarity of the word or sentence or document. Usually, the polarity is positive or negative, and sometimes neutral, emotions. An advanced task of sentiment analysis is classifying emotion into anger, happiness, sadness, and so on. Sentiment analysis is the concern of this thesis [7].

2.3 Text Analysis

Text analysis is the process of converting unstructured data into structured and meaningful data. Its purpose is to create well-structured and meaningful text data in order to be managed and processed easily and professionally. Text analysis uses some machine learning techniques, like Naïve Bayes, K-nearest Neighbours, and support vector machine, statistical, and linguistics algorithms. In other words, it's the process of deriving high-quality information from text. This high-quality information is derived through the devising patterns through means such as

statistical pattern learning. Text mining usually posses three processes (structuring the input text, deriving patterns within the structured data, output interpretation and evaluation) [15].

Text analysis involves parsing texts in order to extract machine-readable facts from them. The purpose of text analysis is to create sets of structured data out of heaps of unstructured, heterogeneous documents. For example, through text analysis, the text in the sentence is divided into small chunks which are further classified. This is done by algorithms that were mentioned before that first parse the text, then extract facts about pre-defined types of events, people, things, entities or relationships.

Text analysis is sometimes called computational linguistics, which is an approach of understanding natural language (written and spoken language) to build patterns that can process and produce useful language. Computational linguistics is concerned with the statistical modeling of natural language. It can be considered an automatic processing of natural language [21].

2.4 Sentiment Analysis Techniques

The existing techniques for sentiment analysis can be classified into three groups, supervised machine learning techniques, unsupervised machine learning techniques, and hybrid technique [16]. The next three subsections describe the three techniques as follows:

2.4.1 Supervised Sentiment Analysis Technique

The sentiment analysis supervised technique uses machine learning methods to extract the polarity of the text [22]. It uses a training dataset (the corpus of reviews annotated with negative

or positive polarities) to extract dataset features (e.g. N-gram, POS (Part of Speech), Bag-of-words, …etc.).

- **Steps of sentiment analysis using supervised machine learning**

Supervised sentiment analysis technique needs several steps. The first step is dataset collection and preparation. The second step is extracting the features from the dataset. The third step is training the model using one of the supervised machine learning algorithms. The fourth and last step is testing the model using the test set. Figure 2.2 illustrates the supervised sentiment analysis phases. Supervised machine learning phases are discussed as follows:

Step1: Dataset collection and preparation in this step, a corpus of annotated reviews is collected from social media website (e.g. Facebook, Twitter, LinkedIn... etc.).

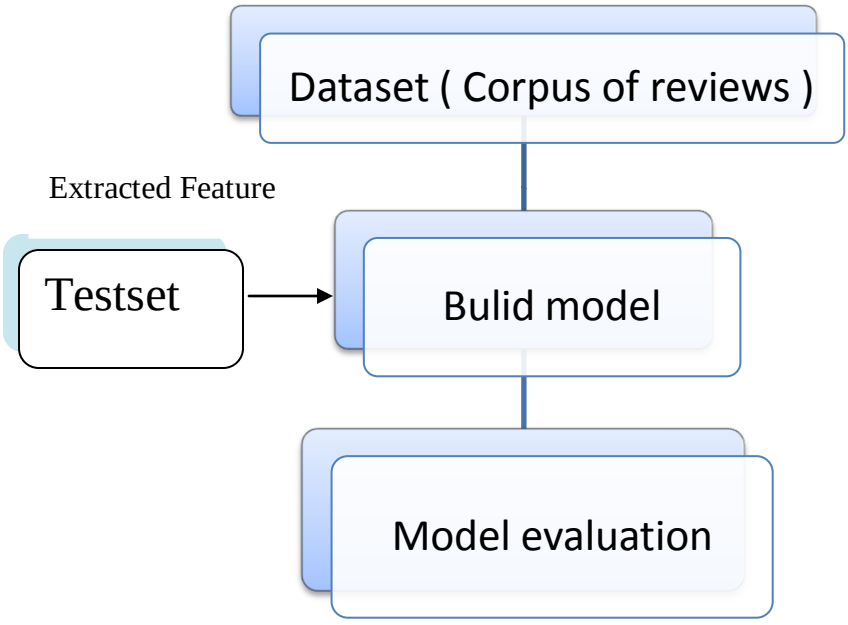


Figure 2. 2 Supervised Machine Learning Phases [23]

For the collection process, some Application Programming Interfaces (APIs) are used for crawling reviews from social media websites. In Twitter, Streaming API especially (GET/search)

command is used for crawling the tweets. Also, tweety API (API for accessing the entire Twitter RESTful API methods) is used for the same reason. In Facebook, Graph API (the primary way for apps to read and write to Facebook's platform) is used to get data out of Facebook's platform. The collected reviews are then prepared either manually or automatically. The data preparation process is done by eliminating any outliers like ads or special characters that are meaningless. The data is saved in a text file, excel sheet, or database file.

Step2: Feature extraction step: A feature is a property of characteristics in data that can be observed. Features extraction is the process of choosing the features that can be measured in data. Since, the input data is too large so it can be transformed into a set of features to be processed easily. These features reflect information about the input data to facilitate the machine learning algorithm to learn about data. Sentiment analysis deals with text so the features can be extracted using NLP models like bag-of-words [24], n-grams [25], and part-of-speech [26] model... etc.), each is discussed as follows [27]:

a) **Bag-of-words (BoW) model:** It is a model used in Natural Language Processing and Information Retrieval (IR). This model represents textual sentences or documents or even image pixels as a set of words. It can be called a unigram as the model review (*I* love this book) will be split using BOW (*I*, Love, this, book). Every word is processed on its own [24].

b) **N-grams model:** It is a model used in computational linguistics. N-gram is a continuous set of words (n items) of text. Unigram is n-gram of size one, bigram is n-gram of size two, trigram is n-gram of size three, and so on [25]. The example is illustrated in Table 2.1.

Table 2. 1 N-grams Example

I Like this TV. Show		
Unigrams	Bigrams	Trigrams
I	I Like	I Like this
Like	Like this	Like this TV.
This	This TV.	This TV. Show
TV.	TV. Show	
Show		

c) **Part-of-speech (POS):** It is the process of assigning words to categories based on their grammatical positions. This task is done using part-of-speech tagging which is an automatic assignment of words to specific tokens. Common parts or tags are nouns, verbs, adverbs, adjectives, pronouns, interjection, preposition and conjunctions [26]. An example is illustrated in Table 2.2.

Table 2. 2 Part of Speech Example

PUT your shoes in their place					
Put	Your	Shoes	In	Their	place
Verb	Pro-noun	Noun	Noun (letter)	Pro-noun	noun

Step3: Model training step: In this step, one of machine learning algorithms is selected for training using the collected dataset from step one to be the training set for the algorithm. The model is trained using the extracted features from the dataset.

Step4: Model evaluation step: In this step, the constructed model is tested using one of testing approaches. There are different approaches to test the model like cross validation [28].

Cross validation is a method used to evaluate the model which is better than the residuals [29]. In the residual assessment problem, it does not give an indicator that when the learner asks him to provide new predictive data, they have no notice. One way to solve this problem is to train learners without using the whole data set and to start training before some data is deleted. Then at the end of training, the deleted data can be used to test the performance of the new data model. This is the basic idea of a complete cross validation evaluation model. There are different cross validation methods (the hold-out, K-flod, Leave-one-out) which are discussed as follows:

• **The holdout method** is the simplest kind of cross validation. The data set is divided into two groups, the first is called training set and the second is test set. Function approximation is only applicable to the training set of functions. Then, function approximation requires that the output value of the data in the test set be predicted which is an unprecedented output value.

• **K-fold cross validation** is one way to improve over the holdout method. The data set is divided into k subsets and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed.

• **Leave-one-out cross validation** is K-fold cross validation taken to its logical extreme, with K equal to N, the number of data points in the set.

This means that *n* times discrete, the function approximates all data except one point and prediction point. The average error before calculation is used to evaluate the model.

2.4.1.1 Supervised Machine Learning Algorithms

Supervised machine learning algorithms [30] can be classified as classification algorithms [30] and regression algorithms [31].

Classification algorithms (Categorical response values) [30] is used to predict the output class. The data is separated into classes. The trained model either used to predict one of two classes and, in this case, it is called binary classification or it's used to predict one out of many classes and it's called mutli-classification. For example, predicting if the student will pass or fail is binary. Common classification algorithms are Support Vector Machine (SVM) [32], Naïve Bayes (NB) [33], Neural Networks [34], Decision Trees [35]... etc.

Regression algorithms [31] (continues response values) build models to predict continues data. It is used for predicting or learning numeric future data values. Regression is a field in both machine learning and statistics. Common regression algorithms are linear regression, nonlinear regression, and decision trees.

The coming three subsections define some of used supervised machine learning algorithms in sentiment analysis (Support vector Machine, Naive Bayes, and Support Vector Machine Light (SVM light)) [36].

2.4.1.1.1 Support Vector Machine

SVM [32] is a supervised machine learning algorithm. In SVM algorithm, each data item is plotted in (*n* features dimensional) space. After distributing all the points, a hyper-plane is found

in order to differentiate between classification classes. Figure 2.3 shows the basic concept of SVM.

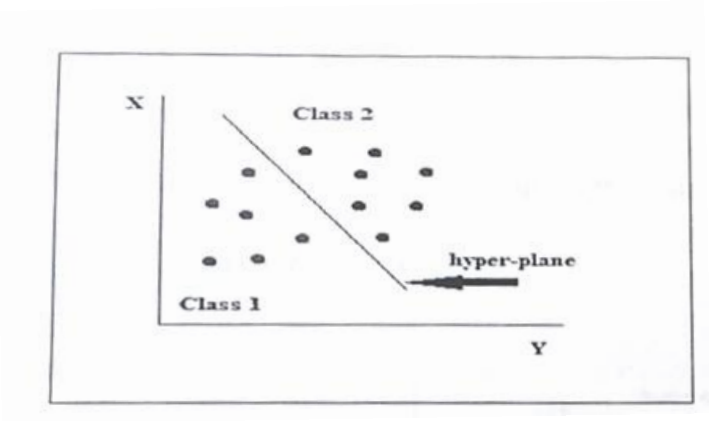


Figure 2. 3 Basic Concept of SVM

Support vectors are the data points nearest to the hyper-plane, the points of a data set that, if removed, would alter the position of the dividing hyper-plane. Because of this, they can be considered the critical elements of a dataset. SVM basic idea is to find a hyper-plane that best divides a dataset into two classes. In another word, it's goal is to find the optimal separating hyper-plane which maximizes the margin of the training data. While, the margin is the distance the hyper-plane and the nearest data point from either set. One of hyper-plane parameters is (C) parameter. C parameter controls the tradeoff between smooth decision boundary and classifying training points correctly. Another important parameter is the gamma parameter. The range of the gamma parameter is determined, reaching a training example, where a low value means "Far" and a high value means "close". Gamma parameter can be regarded as the inverse radius effect of sample selection as support vector.

The simplest way to separate the two types of data is a line in case of 2D data and a plane in case of 3D data. But it is not always possible to use lines or levels, and a nonlinear region is needed to

separate these categories. Support vector machines deal with locations like this, using kernel functions to allocate data to different spaces, where it can use linear hyper-plane to separate classes. This is called the technique of kernel function which transforms data into higher dimensional feature space making linear separation possible. Figure 2.4 shows the transformation using the kernel trick. There're many types of kernels, some of the popular ones are Linear Kernel, Polynomial Kernel, and Radial Base Function (RDF) kernel. However, it has been shown in literature that linear kernel has the best performance in the text classification [32].

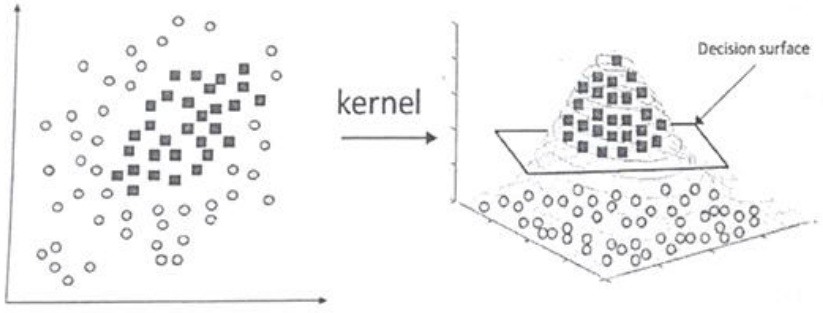


Figure 2. 4 Kernel Trick

2.4.1.1.2 Support Vector Machine Light

SVM light [36] is an implementation of Vapnik's Support Vector Machine [32] for the problem of pattern recognition, the problem of regression, and the problem of learning a ranking function. The algorithm has scalable memory requirements and can handle problems with many thousands of support vectors efficiently.

Training a SVM leads to a quadratic optimization problem with bound constraints and one linear equality constraint. Despite the fact that this type of problem is well understood, there are many

issues to be considered in designing an SVM learner. In particular, for large learning tasks with many training examples, off-the-shelf optimization techniques for general quadratic programs quickly become intractable in their memory and time requirements. SVM light is an implementation of an SVM learner which addresses the problem of large tasks.

The algorithm uses SVM light analysis strategy effectively to solve the problem of variable selection in working group. The algorithm also solves the shrinkage problem in the optimization process. This is particularly effective for discovering large learning tasks where SVS is a small part of sample size comparisons, or when many SVS is in the upper bound.

2.4.1.1.3 Naïve Bayes

Naïve Bayes (NB) [33] is a supervised machine learning algorithm. NB is based on calculating the probabilities using Bayes' theorem. The theorem calculates the probability by calculating the values frequency and values combinations in the data. Equation 2-1 illustrates the Bayes' theorem.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \text{Equation 2-1}$$

In Equation 2-1, The Probability that a given B belong to a Given A. The theory says to count the cases that A and B occur together, then divides it by the number of cases where A occurs alone.

Based on the Bayesian theorem, it is particularly suited when the dimensionality of the inputs is high. Parameter estimation for naive Bayes models uses the method of maximum likelihood. In

spite of over-simplified assumptions, it often performs better in many complex real-world situations.

There are many types of NB, the most common three types are Gaussian, Multinomial, and Bernoulli. **Gaussian** NB is used in classification and it assumes that features follow a normal distribution. **Multinomial** NB is used for discrete counts. For example, let's say, a text classification problem is available. Bernoulli trials can be considered which is one step further and instead of "word occurring in the document", "count how often word occurs in the document", you can think of it as "number of times outcome number x_ i is observed over the n trials". Finally, the binomial model is useful if feature vectors are binary (i.e. zeros and ones). One application would be text classification with 'bag of words' model where the 1s & 0s are "word occurs in the document" and "word does not occur in the document" respectively.

In sentiment analysis, the most commonly used classification techniques in sentiment analysis are **Support Vector Machine** and **Naïve Bayes.**

When talking about the text classification, it is referred to a supervised machine learning scenario. However, it is possible to create a simple classifier without machine learning and without the production line.

A set of comments are available where each comment (phrase) is rated with its corresponding emotion: negative or positive. For example, a comment that says, "هاي السيارة حلوة!"it's clearly expresses a positive feeling, while the comments reads, " مش رح اشترك بالخدمة بتاتا!!" has a negative feeling.

2.5.2 Unsupervised machine learning algorithms

Unsupervised machine learning algorithms [36] perform more complex tasks than supervised algorithms. Such algorithms are used when the data is neither classified nor labeled. The unsupervised algorithms don't predict the correct output values but explore the data and discover new hidden structures and relations in the input unlabeled data. Unsupervised algorithms task is to find groups in data where each group points are similar or close to each other.

The most common unsupervised learning method is clustering. Clustering finds a structure in a collection of unlabeled data. This process divides the data collection into "similar" clusters and another “dissimilar” cluster. A cluster is a group of data items which are similar between them and dissimilar to data in another group. The most popular clustering technique is K-means.

2.5.2.1 Unsupervised Sentiment Analysis Technique

The Unsupervised technique, also called the lexicon-based technique, usually builds a lexicon by containing roots or stems of words. Every root or stem is associated with polarity (e.g. - 1 for negative, +1 for positive). Basically, the lexicon-based technique consists of the following steps [22] dataset collection, tokenization, normalization, stop words removing, and stemming.

2.5.3Hybrid Sentiment Analysis Technique

The hybrid technique [23] combines both supervised and unsupervised techniques. The output of the lexicon-based technique is used as input for machine learning algorithms and the true classified reviews are used as a training set for the machine learning algorithm. After building the model using specific features. The unclassified reviews from the supervised technique are used to test the model in the unsupervised technique. Figure 2.5 illustrates the hybrid technique architecture.

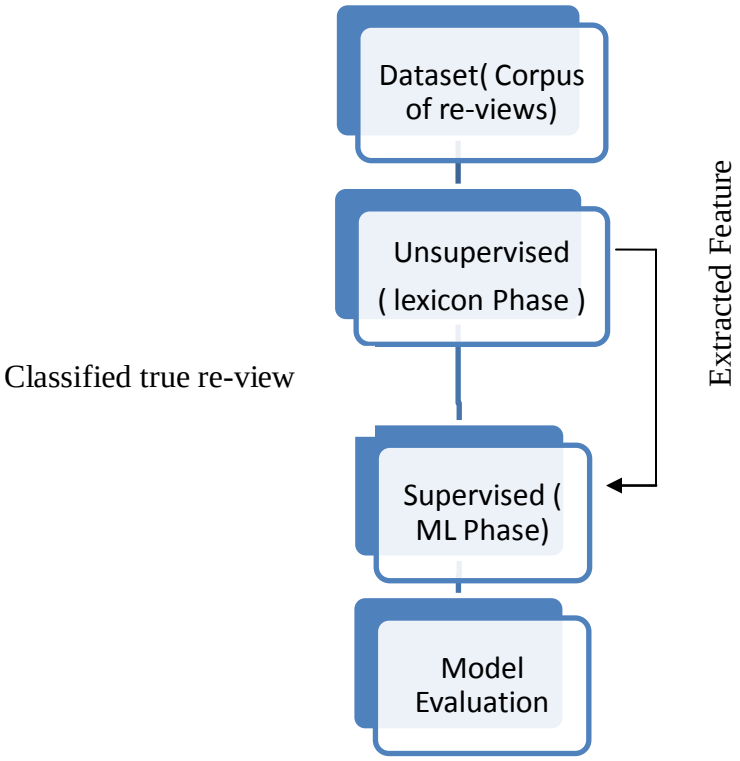


Figure 2. 5 Hybrid Sentiment Analysis Architecture

2.6 Arabic Language Characteristics and Challenges

The Arabic language is one of six official languages used in the United Nations [37] and it is the native spoken language for more than 300 million people worldwide [38]. It is one of the Semitic languages that is written from right to left and is consisted of 28 letters. The shapes of the written Arabic letters are changed according to their positions within the word. For instance, the letter "ج", pronounced as Jeem, is written as" ـجـ"if it comes at the beginning of a word, and it is written as "ـجـ " if it comes in between the letters, whereas it is written as " ـج " at the end of the word. In Addition, Arabic consists of long and short vowels. The long vowels are denoted by letters,

while the short vowels are not denoted as letters, but as diacritical marks instead. Each diacritical mark is written on either the top or the bottom of the letters. For example, the different diacritical of the letter "ن", pronounced as "noon" can be "نَ" or, "نُ"," نِ". In fact, the Arabic language is classified into three types; Classical, Modern Standard, and Colloquial Arabic language. The classical Arabic is the language of Holy Quran, whereas the Modern Standard Arabic (MSA) is the language of books, news, official statements and magazines. On the other hand, the colloquial Arabic is most widely spoken and written in daily live conversations with different dialects. For example, Palestinian Arabic is spoken in Palestine, Jordan. Egyptian Arabic is spoken in Egypt. In GULF area, people speak Khaliji (GULF Arabic). Levantine Arabic" Shami" is spoken by (Lebanon, Syria). Maghrebi Arabic is spoken in Libya, Tunisia, Algeria, and Morocco. In Sudan, people speak Sudanese Arabic. Omani Arabic is spoken in Oman. Yemeni Arabic is spoken in Yemen. And in Chad, people speak Chadian Arabic.

Usually, each Arabic word consists of prefixes, root, and suffixes. The automatic extraction of the root of the word using a stemming technique is not an easy task, particularly in Colloquial Arabic. Colloquial Arabic is written without diacritical marks, which adds ambiguity to understand the meaning of the words. Take for example, the word "حب" without diacritic can be understood alone as " حُب" love, or "حَب" a seed.

Chapter 3

Literature Review

3.1 Literature Review

The article of Badaro et .al (2019) [39] aims to give a detailed survey of the real advances in Arabic opinion mining. It doesn't cover previous work only, but it gives a comprehensive perspective by covering advances in various aspects including software tools, lexical sentiment and corpora resources. The article also presents directions for the future of opinion mining in Arabic. The study tries to help new researchers and industry developers.

The study of Al-Azani and El-Alfy (2018) [40] explores the impact of combining emojis-based features (pictographic symbols) which are commonly used in mass media on the sentiment classification of dialectical Arabic tweets. The features are extracted by using four methods: -(BOW), Latent Semantic Analysis, and two forms of word Embedding. The effect of using emojis is analyzed using a support vector classifier. It is found that simpler models can be constructed with better results when emojis are merged with word embedding. Nevertheless, we didn't apply this type of Sentiment analysis in our work. Because the most of the data is not having emojis.

A class-specific sentiment analysis framework is proposed by Hamdi and his team (2018) [41], It includes new annotation to build complex Arabic corpus and lexicon. This allows simultaneous annotation for different facets including domains, dialects, and polarity strengths. Each facet has multiple classes. The new annotations facilitate the development of new class-specific models and calculation. A hybrid model combining corpus-based and lexicon-based models are suggested for the new sentiment classification models. The corpus model has two phases to build: full corpus classification models and class -specific model. As a case study, 15274 reviews are collected and annotated from many sources including surveys, Facebook comments and

Twitter posts. In addition to that, a new web-based application is developed to apply the proposed framework on the case study. CIASENTI framework reaches up to 95% Accuracy and 93 % F1- score. Unfortunately, this data is not publicly available, and the researchers has communicated with the authors of this work requesting the data for experiments but there was no reply.

Real approaches based on supervised machine learning are offered by Al-Smadi and his team (2018) [42] in order to handle the challenges of aspect -based sentiment analysis (ABSA) of Arabic hotels' reviews. The researches in that work uses the: Recurrent Neural Network (RNN) and Support Vector Machine (SVM). To evaluate these approaches, a reference dataset of Arabic hotel's reviews is used. The results show that the SVM approach outperforms the other one in the research investigated tasks. and also, the RNN didn't applicable with thousands of features like text data.

To classify web-pages if containing reviews or not, khoja (2017) [43] and his team suggested a system and built a sentimental analyzer. They classified the reviews into three categories: positive, negative and neutral. That was based on the constructed lexicon of 600 positive words and more than 900 negative words as well as 100 neutral words, all of which being collected from frequent words over the internet pages. However, the unsupervised researches have some limitations: The technique based on lexicon is unable to deal with unseen words. Without the stem being found, a word is not treated, as well as the Accuracy which is not high enough. This is due to some Problems related to the stemmer while trying to find the root of words.

The study of Al-Kabi and his group (2017) [44] is based on a benchmark corpora having 3015 Arabic opinions taken from Facebook. They are distributed into three areas (food, sport and weather). For accomplishing the study, ten Arabic lexicons were formulated and a new tool is designed (called Arabic Opinions Polarity Identification AOPI) in order to identify this polarity. This study includes a comparison between the tool and two online sentiment analysis tools (Social Mention and Senti-Strength) supporting Arabic. Examining the effect of stemming on the Accuracy of the tools, it was found that AOPI is more effective than the other two online sentiment analysis tools.

Altyeb Altaher 2017 [45] proposed a hybrid approach for sentiment analysis of Arabic tweets is suggested, it is based on two stages: the first stage involves pre-processing methods like stop-word removal and two features weighting algorithms are used to give high weights to the most important feature of Arabic tweets. The second stage involves using the deep learning technique to classify Arabic tweets into positive and negative. Then, the performance of this approach is compared with some classifying methods such as Decision Tree (DT) and Neural Networks (NN) using the dataset collected from Arabic tweets. This approach surpasses the other approaches and the Accuracy reaches 90% and 93%.

However, a sentiment analysis system developed for one dialect might not work well for the others. The proposal model of Elouardighi et.al (2017) [46] is characterized by its novelty through its features. In this model, the sentiment is expressed by embeddings. The results show that the syntax ignorant embeddings could present different dialects efficiently.

The work on Arabic sentiment analysis as Aliane and his team work on (2016) [47] has gained more interest following the increasing interest in opinion mining by various research communities. With few polarities annotated datasets, many studies use these datasets to examine

the most famous supervised algorithms, the best reported algorithms in the Arabic sentiment analysis are the Naive Bayes and SVM. In this study, it is shown that using a genetic algorithm to select features and assist the quality of the training dataset improves the Accuracy of the learning algorithm.

Aldayel and Azmi (2016) [48] formulated a lexicon of MSA and Jordanian dialect made of 300, then to 3479 words (1262 positive words and 2217 negative ones). The approaches used were light stemming and root-based stemming. The Accuracy obtained was 85% for the latter and 78% for the former.

Semantria (2015) [49] allows the users of this service to analyze posts, surveys, reviews and tweets. Researchers, developers and students can use available application programming interfaces to analyze sentiment such as using the Python Natural Language Toolkit (NLTK). This is found to be the most commonly used platform for making programs that deal with human language data. Sentiment 1406 is another good API related to sentiment analysis in Twitter. That was developed by Stanford students and gave insights about useful features. Most of the research has been done on English texts.

Concuera and group (2015) [50] applied the SVM machine learning model which was used with features taken from the TASS dataset. The features were typical to systems for analyzing tweets such as has hashtags and emotions which were able to obtain an F-score of 60.6.

Rehab M. Duwairi et. all (2014) [51] applied an Arabic Data set of tweets and comments in three classification methods. They used machine learning approach SVM, Naive Bayes, and KNN particularly, SVM result gives the highest Precision while KNN gives the highest Recall.

This work has considered sentiment analysis in Arabic text. A dataset, which consists of 2591 tweets/comments, was collected and labeled using crowd sourcing. The

Tokenize, Stem (Arabic), Filter Stopwords (Arabic), and Generate-N-Grams (Terms) operators were used as text processing. The Naïve Bayes, SVM and KNN classifiers were used to detect the polarity of a given review. 10-fold cross validation was used to split the data into training and testing sets. The best precision was achieved by SVM and it equals to 75.25. The best recall was achieved in the case of KNN (K=10) and it equals 69.04.

Taysir H. A. Soliman and his team (2014) [52] constructed a Slang Sentimental Words and Idioms Lexicon (SSWIL) of opinion words. Also they proposed a gaussian kernel SVM classifier for Arabic slang language to classify Arabic news' comments on Facebook. They proposed a classifier to test performance and several Facebook news' comments are used. The result shows that 86.86% for Accuracy, 88.63 Precision and 78 Recall.

Chapter 4

Research

Methodology

4.1 The Proposed Method

This chapter presents the methodology applied to carry out this study. The methodology consists of three phases where each in turn contains many stages. Figure 4.1 summarizes these phases and included stages, and following is a detailed description of it.

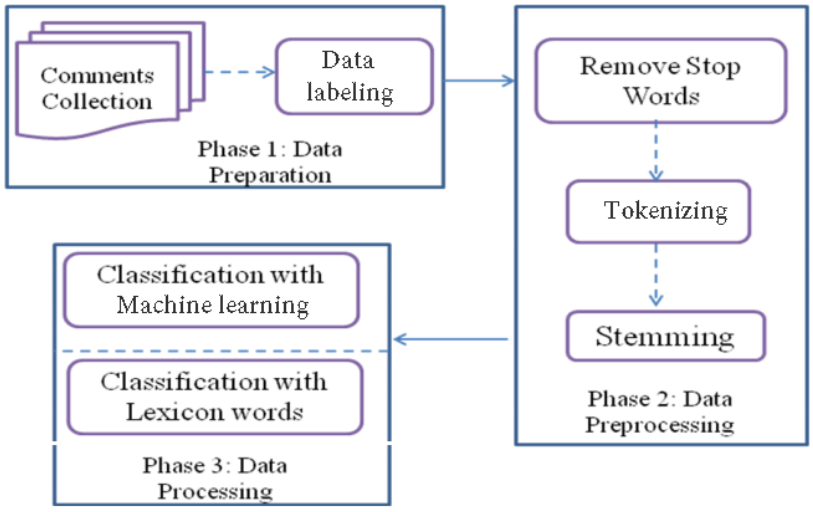


Figure 4. 1 Proposed System Phases

4.2 Data Preprocessing

Preprocessing tasks, such as cleaning and tokenization, are required to maintain a 'clean' dataset which is free of noisy data and which maintains textual emotional value. Next, It is focused on the tasks which have been adopted in this work due to their suitability for emotion analysis overall.

4.2.1 Cleaning

The first step in the preprocessing process is to remove the noise from the data instances. Initially all stop words are removed, i.e., words which carry no emotion information. These may

be pronouns, prepositions, or conjunctions (e.g., as "لماذا", "أين", "من","كيف"). Moreover, the URLs such as "www.jawwal.ps" are removed, the numbers, and the punctuation marks.

4.2.2 Filter Stop words

The quality of the dataset can be improved by removing stop words in the cleanup function. The stop words are frequently repeated words found in every language which are usually meaningless word such the conjunction, linking, and identification articles. For example, in Arabic, these words are "في", "الى", "على", "عن", "و"etc. Removing stop words is not always a wise choice and it depends a lot on the problem working on. After removing the stop words, a sentence like انا بحاجة الى سياره " will become "بحاجة سيارة", so who needs a new car is unknown. However, the removed words surely do not present any emotions that may affect the polarity of the instance.

4.2.3 Tokenization

The cleaned texts are then tokenized on all whitespace, i.e., a text is split into pieces (tokens) using whitespace as separator. Also, all characters of each word are converted to lower case. Tokenize (non letters) that removes any comments without letters that is empty is used. In tokenizing many methods are used, most common method which we apply in this research is the unigram method, in which each sentence is divided into single words, other methods such as bigram and trigram are not used in this study.

The Generate-n-grams operator is generating sequences of n-words and each sequence is considered one token. N, here, specifies the number of words or terms in a sequence. In this work, n was set to 1, i.e. we generated uni-grams.

4.2.4 Filtering by term length

Also Filtering Tokenize (By Length) this removes every word that is beyond 2 and 16 characters. This step will remove the words that are consist of less than 2 letters or more than 15 characters.

4.2.5 Stemming

Stemming is the process of generating the root word of any word. The root (stem) of each word is being kept by removing any morphological affixes. For instance, the stemmed word 'خدم' remains for all of this word's derivatives – استخدامها, يستخدم, …etc. it reduces a word to its smallest form. Stemming comes from linguistics and needs an in-depth knowledge of a particular language is needed if it is planned to create a mastering of that language. When Stemming, the smallest form of a word is given that may or may not be a word in the dictionary for the language to which the word belongs. In this study have used Stem (Arabic, Light). This thesis scope does not include the other morphological analysis steps.

4.3 Cross Validation

In this research, 5429 positive statements and 1200 negative statements, with a total of 6629 statements are available. The first step is to divide the above data into two groups, i.e. the training and testing groups. When an equal number of positive and negative categories in a binary rating scale is available, Precision, Accuracy, Recall, and F1 score in general are used. But here, due to the imbalance of the comments, the Fold 10 cross validation method is used.

- This method is used to estimate how accurately our predictive models would generalize to an independent data set. This method first splits the data set into 10 equal parts. It then trains on 9 partitions and validates on the 10th.

- To reduce variability, cross-validation was performed for 10 rounds, the validation set was rotated for each round, covering all possible combinations. Final evaluation of the model was created by averaging the results for all validation sets.

4.4 Data preparation

4.4.1 Lexicon adoption

Lexicon is a dictionary of terms and weights, weights come from processing very large words and terms inside a very big corpus, this weight of the words is refers to its importance in the whole corpus.

Based on what has been read and studied in the literature, the researcher found that there is a possibility of solving the study problem by introducing and integrating the dictionary, a lexicon-based sentiment analysis. A Lexicon containing 1367 words and their weight have been used. It has been provided from the National Research Council in Canada that provides English and Arabic sentiment lexicons and sentiment annotated corpora, [53].

when we did the train the classifiers , we will get the model, every term in the data comments output a weight , so we need to update the weight of terms we have that is also existed in the lexicon (We will give the widely spread word (term), greater and bigger weight)

Lexicon may contain a compound words and terms, but our model didn't exist compound words, because we use a unigram.

N-gram = 1

N=1

1 word = 1 feature

4.4.2 Sentiment dataset

Raw dataset was adopted from a previous work completed by Ahmed Qerem Supervised by Dr. Khalid Rabayah. The raw data were taken from user comments on Facebook on posts related to Jawwal Telecom Company and Car trade company "Citroen & others ". The extracted data were not processed by any processing tool by the researcher, therefore, we decided to use the data as our raw data for this project. To have the data ready for sentiment analysis the following pre-processing steps were performed;

1. Raw data is transformed from SQL to Excel.
2. All unwanted fields were removed from the data.
3. Annotating the data instances with the positive/negative labels, i.e. build the true labeling of the dataset. This procedure has been performed manually which took a long time. A set of 1200 negative comments and 5429 positive comments are manually marked, and revised. Following is a description of the generated Dataset:

Count of instances = 6629

Count of Positive instances (Comments) = 5429

Count of Negative instances (Comments) = 1200 comments

Table 4. 1 Table of Comments

All Data	**Positive Comments**	**Negative Comments**
6629	5429	1200

Total count of words in positive comments = 39375

Average of words in positive comments = 5.9 word

Maximum words in positive comment =671

Minimum words in positive comments =1

Table 4. 2 Table of Positive Comments

Total count of words in positive comments	Average of words in positive comments	Maximum words in positive comment	Minimum words in positive comments
39375	5.9 word	671	1

Table 4. 3 Table of Negative Comments

Total count of words in negative comments	Average of words in negative comments	Maximum words in negative comment	Minimum words in negative comments
13933	12.1 word	264	1

Total count of words in negative comments = 13933

Average of words in negative comments = 12.1 word

Maximum words in negative comment =264

Minimum words in negative comments =1

The data was imported into the RapidMiner Software for further processing.

4.5 Rapid Miner Software:

It is a software for machine learning and data mining. It is an open-source software and is based on Java. It contains a graphical interface (GUI) and it is possible to design the machine learning process and build models without the need to write code. One of the most valuable additions to the Rapid Miner Software is the word processing package which includes several operators that support text mining. For example, there are operators for Stemming, tokenizing, and filtering stop words. This is the main reason for choosing RapidMiner for its ease and ability to deal with the Arabic language. [54].

4.6 General Method Procedure

Algorithm 1: Genetic Algorithm-bio geographical based optimization neural network

input: Dataset, LexiconList, WordWeight

Output: Weighted Dataset

Main:

wordWeight=0;

for sentence **in** Dataset **do**

for word **in** sentence **do**

Weight= **find (**word, lexiconList); /* find function return if the word is found in the lexicon or not */

if not_zero (wordWeight) **then /*** the word doesn't have a weight in the lexicon***/**

weightedword ← word ∗ Weight;

else weightedword ← word;

// use the output as an input of the Rapid Miner

RapidMiner(WeightedDataset)

4.7 Evaluation metrics

In order to evaluate the performance, about 6600 comments divided to 5430 positive and 1200 negative have been taken and the N-fold cross validation was employed. Taking into account the device capabilities, 10-fold and 5-fold cross validation were used in order to evaluate the performances of the first four stages. The whole dataset was divided into ten sets with equal sized samples, where the classifier was trained on nine sets and the remaining set was used for testing. Regarding the fifth stage, a 10-fold cross validation was used in order to evaluate the performance by this stage the computational difficulty has been already reduced by selecting the most relevant subset of features and the best text representation. To measure the performance of the SVM classifier, the following evaluation metrics were chosen: Accuracy, Precision, Recall, and F1 for evaluating the SVM classifier; see equations 1, 2, 3 and 4 below [55].

(1) $$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

(2) $$Precision = \frac{TP}{TP+FP}$$

(3) $$Recall = \frac{TP}{TP+FN}$$

(4) $$F1 - Score = 2 * \frac{Precision*\text{Recall}}{Precision+\text{Recall}} = \frac{2TP}{2TP+FP+FN}$$

Where TP stands for True Positive, TN stands for True Negative, FP stands for False Positive, and FN stands False Negative.

Accuracy: is one of the simplest way used to evaluate mechanic learning algorithms. It determines how accurate your model is in reference to your search. If a model classified 90 of

phrases as correct, the accuracy will be 90% or 0.90. In this research, the data set is skewed, that is, the number of comments in one category outnumbers the number of comments in another. In these types of cases, it is not recommended to use accuracy as an assessment metric because it does not reflect the status of the data. So, high Accuracy might be achieved, but the model probably won't perform well when it comes to real-world data. In these cases, it is best to look at other metrics such as *precision*.

Precision: Before learning precision, definitions of some terms listed above are needed;

True Positive (or TP): Looking at the comments, if the model predicts the comment to be positive, and the actual target of that positive comment is, it is considered a true positive.

True Negative (or TN): Looking at the comments, if the model predicts the comment to be negative, and the actual objective of that comment is negative, it is considered a true negative.

In short, if the model correctly predicts a positive category, this is a true positive, if your model correctly predicts a negative category, then this is a true negative.

False Positive (or FP): Looking at the comments, if the model predicts the Comment to be positive and the actual target for that comment is negative, then it is a false positive result.

False Negative (or **FN**): Looking at the comments, if the model predicts the comment to be negative, and the actual target of that comment is positive, it is a false negative result.

In short, if the model incorrectly (or wrongly) predicts a positive category, that is a false positive. If the model incorrectly (or erroneously) predicts a negative category, it will be a false negative.

Recall: Recall is defined as:

$$Recall = TP/(TP + FN)$$

For a "good" model, the Precision and/or Recall values must be high.

F1-Measure: The F1 score is a measure that combines Precision and Recall [56]. It is defined as a weighted average of Precision and Recall. If Precision is referred to P and Recall to R, the degree of F1 is represented as follows;

$$F1 = 2PR/(P + R)$$

A little math will lead us to the following equation for F1 based on TP, FP and FN

$$F1 = 2TP/(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$$

Instead of looking at Precision and Recall individually, it is better to look at the F1 score. As for Precision, Recall, and Accuracy, the F1 score ranges from 0 to 1, and the ideal prediction model has F1value equal to 1.

When dealing with datasets that contain skewed targets, F1 (Precision, Recall) should be considered rather than Accuracy.

Confusion Matrix: In binary or multiclass classification, it is also common to look at the confusion matrix. It is very easy. The confusion matrix is nothing but a table of True positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Using the confusion matrix, it could be quickly perceived how many comments were misclassified and how many were correctly classified. If you understand the TP, FP, TN, FN, the Accuracy and Recall, it will

be easy to understand and explain the confusion matrix. Sometimes people also prefer to label FP as a Type-I error and FN as a Type-II error.

the binary confusion matrix could be expanded into a multiclass confusion matrix. If there are classes of magnitude N, it will be an array of size NxN.

- **Term Frequency–Inverse Document Frequency (TF-IDF)**

It is commonly used in some known approaches, all words in the text are treated as equally important - there's no idea that some words in a document are more important than others. Term Frequency–Inverse Document Frequency (TF-IDF) addresses this issue. It aims to determine the importance of a particular word in relation to other words in the document and in the group. It is a commonly used representation scheme for information retrieval systems, to extract related documents from a collection for a given text query.

The idea behind TF-IDF is as follows: if the word *w* appears many times in the *di* document but doesn't appear often in the rest of the *dj* documents in the set, then the *w* word should be of great importance to the *di* document. The importance of *w* should increase in proportion to its frequency in *di*, but at the same time, its importance should decrease in proportion to the frequency of the word in the other *dj* documents in the set. Mathematically, this is captured using two quantities: TF and IDF. The two are then combined to reach the TF-IDF score.

TF (Term Frequency) measures the number of times a term or word appears in a given document. Since different documents in the set of documents may be of different lengths, a term

may be repeated more often in a longer document than in a shorter document. To flatten these numbers, we divide the number of iterations over the length of the document:

IDF (Inverse Document Frequency) measures the significance of a term across a group. In TF calculation, all terms are given equal importance (weight). However, it is well known that stop words like الى, من, في ,عن …. etc.., is not important, although it is repeated. To account for such cases, the IDF reduces the weight of very common terms across a group and increases the weight of rare terms.

The TF-IDF score is the product of these two terms. Hence, the TF-IDF score = TF * IDF. Some terms appear in only one document, some appear in two, and others appear in three documents. TF-IDF is a representation commonly used in application scenarios such as information retrieval and text classification. Until today, TF-IDF is still a popular scheme for many NLP tasks, especially the initial versions of the solution.

TF-IDF (Term Frequency – Inverse Document Frequency)

That TF is the term frequencies, and IDF is the inverse document frequency. It may seem difficult with these terms, but things will become clear their equations.

$$TF(t) = \frac{\textit{Number of times a term t appears in a document}}{\textit{Total number of terms in the document}}$$

$$IDF(t) = LOG\left(\frac{\textit{Total number of documents}}{\textit{Number of documents with term t in it}}\right)$$

In another way, Term Frequency – Inverse Document Frequency (TFIDF) is a numerical statistical value that reflects how important a word is to a document in a collection or corpus. The value of TFIDF increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus. TFIDF often used as a weighting factor in information retrieval and text mining such as search engines. TFIDF for the term *w* can by computed as follow:

TF: Term Frequency, which refers to how frequently a term occurs in a document

TF (w) = Number of Times *w* appears in a document / Total number of terms in the document

IDF: Inverse Document Frequency which measures how important a term is.

IDF (w) = log Total number of documents / Number of documents with w in it

TFIDF: The multiplication of the TF and IDF terms.

TFIDF (w) = TF $*$ IDF

Chapter 5

Results and Discussion

5.1 Experiments and Results

In this section, the experimental results of assessing the performance of the classifier are reported. The goal is to investigate and compare different structures and representations when an SVM classifier is used.

Extraction techniques fail to extract the opinion words at the first classification type but it performs well at the second classification type after adding the Lexicon. Applying the proposed mining in all comments, the percent of classified comments in the first type not classified (without lexicon and pre-processing) produces 88.68%, accuracy rate while the second classification type (using pre-processing) produces accuracy rate 89.21%. The Third classification type (using lexicon) produces 90.68%, as illustrated in the tables below. In addition, applying the system using pre-processing (stop words removal, tokenizing and stemming) and cross validation gives 90.57% as a percent of comments classified correctly.

The results got enhanced after applying the lexicon. In addition, the F1 score and the Recall give the highest rate in case of classifying comments, with 90.68 % Accuracy, 99.23% Recall, 90.34% Precision and an F-measure of 94.58%.

5.1.1 Support Vector Machine SVM

Table 5.1 shows the results of SVM investigating the three weighting schemes TF-IDF, TF, and BTO (Binary Term Occurrences when a 10-Fold cross validation is used and for the TF-IDF when a 5-Fold cross validation is used. The results show that the classifier performed better with TF-IDF in terms of accuracy and perception compared to TF and BTO. It can be seen that the Recall based on BTO was the highest, which means a high number of positive documents were

correctly classified as positive. The TF-IDF with accuracy of 90.57 % was selected as a vector representation.

Table 5.2 shows the Results of SVM investigating the weighting schemes of TF-IDF, it compares the results when Number of Folds are 10 and 5.

Table 5. 1: Uni-gram(Support Vector Machine) SVM weighting schemes cross validation results

Weighting Schemes	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score**
TF-IDF	90.57	90.38	99.02	94.50
TF	90.49	90.19	99.19	94.48
BTO	89.63	89.28	99.26	94.01

Table 5. 2: Term Frequency — Inverse Document Frequency (TF-IDF) weighting schemes results

# of folds	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score**
5	90.51	90.32	99.02	94.47
10	90.57	90.38	99.02	94.50

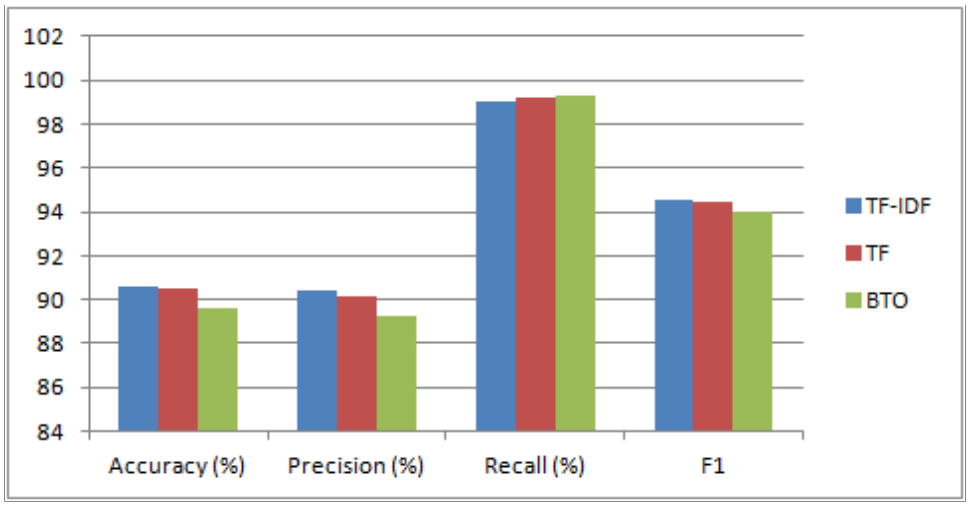


Figure 5.1: Uni-gram SVM weighting schemes cross validation results

Figure 5.1 shows the Accuracy, Precision, Recall, And F1-score Results of NB classifier weighting schemes cross validation TF-IDF, TF, BTO.

Table 5. 3: The cross validation results after Applying different pre-processing techniques for uni-gram features SVM.

Pre- Process	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1**
Stop words removal + Tokenizing	88.68	87.87	99.98	93.53
Light stemming+ Tokenizing	88.71	87.90	99.98	93.55
Tokenizing	89.21	88.55	99.72	93.80

Table 5.3 shows the results of the cross validation results after applying pre-processing techniques for SVM classifier, stop words removal, light stemming and tokenizing. The results of the comparison showed that the classification performed better with Accuracy and Precision in Tokenizing compared to Recall in stop words removals and light Stemming. It can be seen that the Recall in light Stemming and stop word removal was the highest.

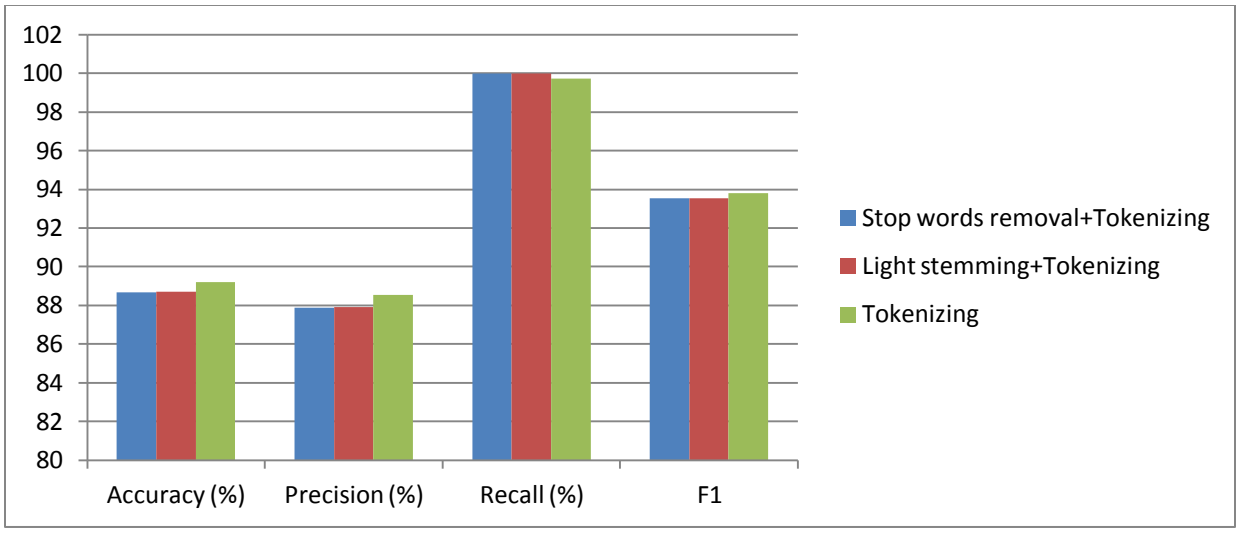


Figure 5. 2 :The cross validation results after applying different pre-processing techniques for uni-gram features SVM

Figure 5.2 Shows the performance measurements of SVM classifier results after Applying different pre-processing techniques like Stop words removal, Light Stemming, and Tokenizing.

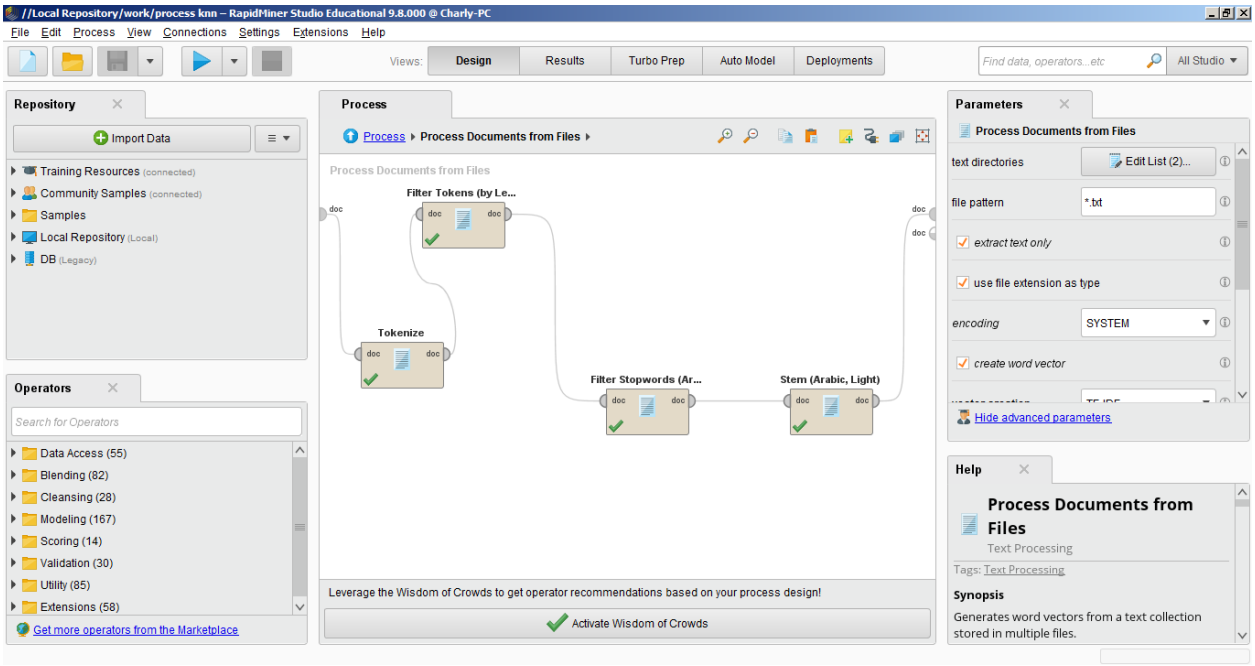


Figure 5.3: Rapid Miner tool after Applying different pre-processing techniques like Stop words removal, Light Stemming, and Tokenizing for uni-gram features SVM classifier

Table 5.4: Accuracy Result Differences for Support Vector Machine (SVM) classifier features and Techniques.

	Weighting Schemes	***Accuracy (%)***
Classifying before applying pre-processing	SVM	88.68
SVM -after applying different pre-processing techniques	Stop words removal+ Tokenizing	88.68
	Light stemming+ Tokenizing	88.71
	Tokenizing	89.21
SVM	TF-IDF	90.57
	TF	90.49
	BTO	89.63
After applying Lexicon (Addition Weights)	SVM	89.8
After applying Lexicon (Multiply Weights)	SVM	90.68
After Applying Lexicon (5 fold)	SVM	89.86

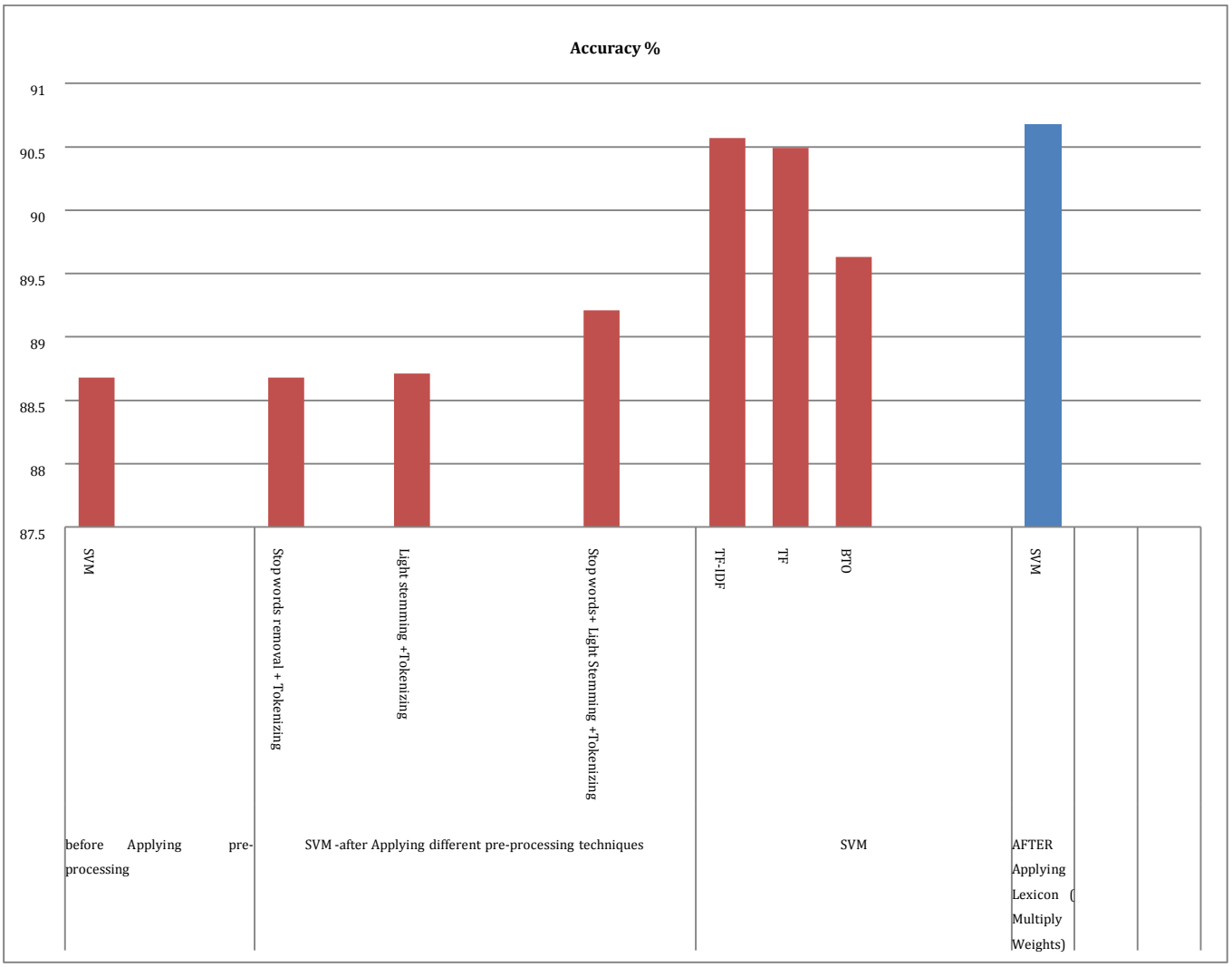


Figure 5.4: Accuracy results for all uni-gram features and techniques

The results of the comparison showed that results are enhanced after applying Lexicon and the classifier performed better with TF-IDF in and 10 folds cross validation. The TF-IDF with Accuracy 90.68 % was selected as a vector representation. In addition, F1 score and Recall give the highest rates in case of classifying comments, Recall which are 99.23%, and Precision of 90.34 %, and having F-measure of 94.58%.

5.1.2 Navie Bayes

Table 5.5 shows the results of Navie Bayes investigating the three weighting schemes TF-IDF, TF, and BTO when a 10-Fold cross validation is used and TF-IDF when a 5-Fold cross validation is used. Similar to SVM, the results of the comparison showed that the classifier performed better with TF-IDF in terms of accuracy and perception compared to TF and BTO. It can be seen that the precision based on TF-IDF was the highest. This means that our model is 95% correct when it attempts to identify positive comments. The TF-IDF is with Accuracy of 77.58 % was selected as a vector representation.

Table 5.5: Uni-gram Navie Bayes weighting schemes cross validation results.

Weighting Schemes	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1**
TF-IDF	77.58	95.0	76.66	84.85
TF	71.79	94.74	69.41	80.12
BTO	59.5	93.86	54.10	68.64

Table 5.6: NB TF-IDF weighting schemes results

# of folds	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	**F-score**
5	78.0	94.91	77.29	85.20
10	77.58	95.0	76.66	84.85

Table 5.6 shows the Results of NB (Navie Bayes) investigating the weighting schemes TF-IDF, it compares the results when Number of Folds is 10 folds and 5 folds.

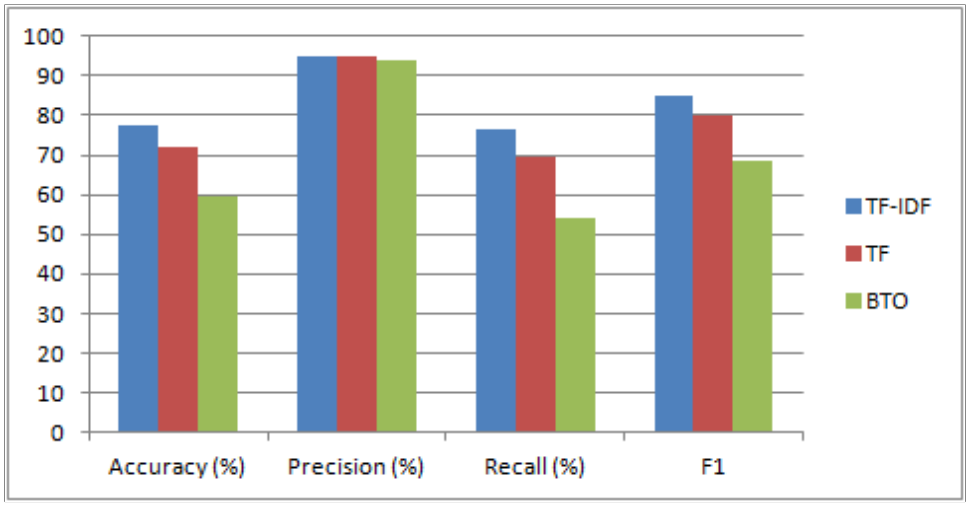


Figure 5.5: Uni-gram Navie Bayes weighting schemes cross validation results

Figure 5.5 Shows the Accuracy, Precision, Recall, And F1 Results of NB classifier weighting schemes cross validation TF-IDF, TF, BTO.

5.1.3 K-nearest Neighbors (KNN)

Table 5.7: Uni-gram KNN weighting schemes cross validation results.

Weighting Schemes	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	F1
TF-IDF	86.69	86.09	99.89	92.48
TF	87.43	86.76	99.91	92.87
BTO	87.85	87.09	100	93.10

Table 5.8: KNN TF-IDF weighting schemes results

# of folds	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	F-score
5	85.98	85.44	99.91	92.11
10	86.69	86.09	99.89	92.48

Table 5.7 shows the results of KNN (K-nearest neighbors) investigating the three weighting schemes TF-IDF, TF, and BTO when a 10-Fold cross validation is used and TF-IDF when a 5-Fold cross validation is used. The results of the comparison showed that the classifier performed better with BTO (Binary Term Occurrences) in terms of Accuracy and Recall compared to TF-IDF and TF. It can be seen that the Recall based on BTO was the highest. That means a high number of positive documents were correctly classified as positive. The BTO with Accuracy 87.85 % and the Recall 100 % was selected as a vector representation.

Table 5.8 shows the Results of KNN investigating the weighting schemes TF-IDF, it compares the results when Number of Folds is 10 folds and 5 folds.

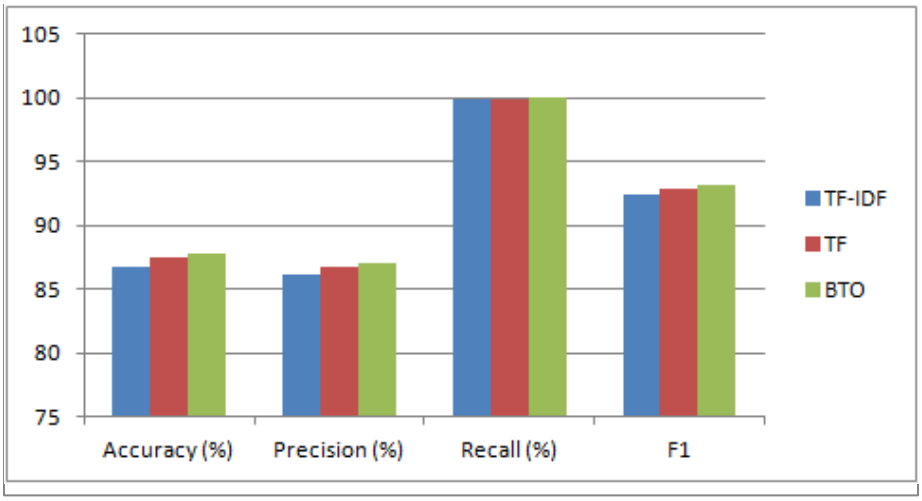


Figure 5.6: Uni-gram KNN weighting schemes cross validation results

Figure 5.6 shows the Accuracy, Precision, Recall, And F1 Results of KNN classifier weighting schemes cross validation TF-IDF, TF, BTO.

5.1.4 Comparison of results SVM VS NB VS KNN

Table 5.9 shows the results of the cross validation results before applying pre-processing techniques for uni-gram features. SVM (Support Vector Machine), KNN, and Navie Bayes when a 10-Fold cross validation is used and TF-IDF. The results of the comparison showed that the classification with SVM performed better with SVM in terms of Accuracy and Perception compared to KNN and Navie Bayes. It can be seen that the Recall based on KNN classifier was the highest.

Table 5.9: The cross validation results before Applying pre-processing techniques for uni-gram features.

Weighting Schemes	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	F1
SVM	88.68	87.87	99.98	93.53
KNN	83.74	83.43	100	90.97
Navie Bayes	41.75	99.94	28.90	44.83

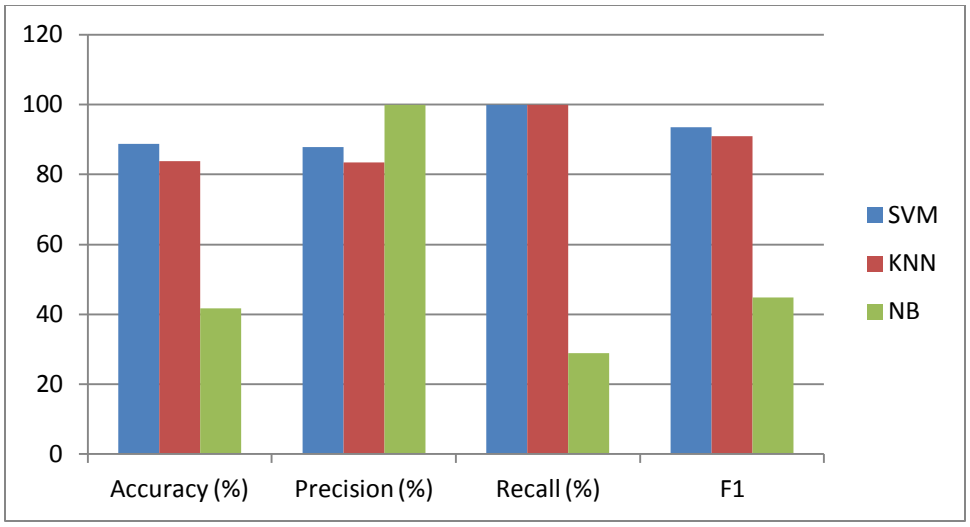


Figure 5. 7 :-The cross validation results before applying pre-processing techniques for uni-gram features.

Figure 5.7 shows the Accuracy, Precision, Recall, And F1 results of SVM, KNN, and NB classifiers results before Applying pre-processing techniques.

5.1.5 Comparison of results SVM VS NB

Table 5. 10 : The cross-validation results AFTER Applying Lexicon for uni-gram features. (Addition Weights)

Weighting Schemes	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	F1
SVM	89.80	89.46	99.24	94.10
Navie Bayes	70.78	94.43	68.36	79.31

Table 5.10 shows the results of the cross validation results AFTER applying Lexicon for uni-gram features. (Addition Weights). The weights of the Lexicon are taken and added them to the main model. The results of classifiers show that the SVM classification performed better than Navie Bayes in Accuracy and Recall, compared to Precision in Navie Bayes which are higher.

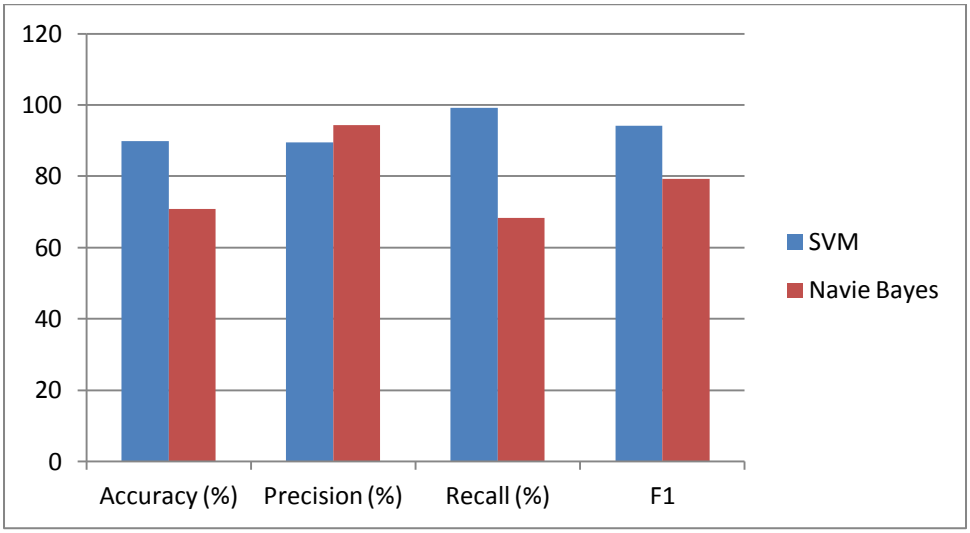


Figure 5. 8 :The cross validation results after applying lexicon for uni-gram features. (Addition Weights)

Figure 5.8 Shows the Accuracy, Precision, Recall, And F1 Results of SVM classifier results after applying Lexicon for uni-gram features by Addition Weights.

Table 5. 11 : The cross validation results after applying Lexicon for uni-gram features (5 fold)

(Addition Weights) ***Weighting Schemes***	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	F1
SVM	89.56	89.29	99.15	93.96
Navie Bayes	71.11	94.48	68.74	79.58

Table 5.11 shows the results of the cross validation results after applying Lexicon for uni-gram features. (Divided data to 5 folds and use stratified sampling type). The weights of the Lexicon are taken and added them to the main model. The results of classifiers show that the SVM classification performed better than Navie Bayes in Accuracy and Recall, compared to Precision in Navie Bayes which are higher.

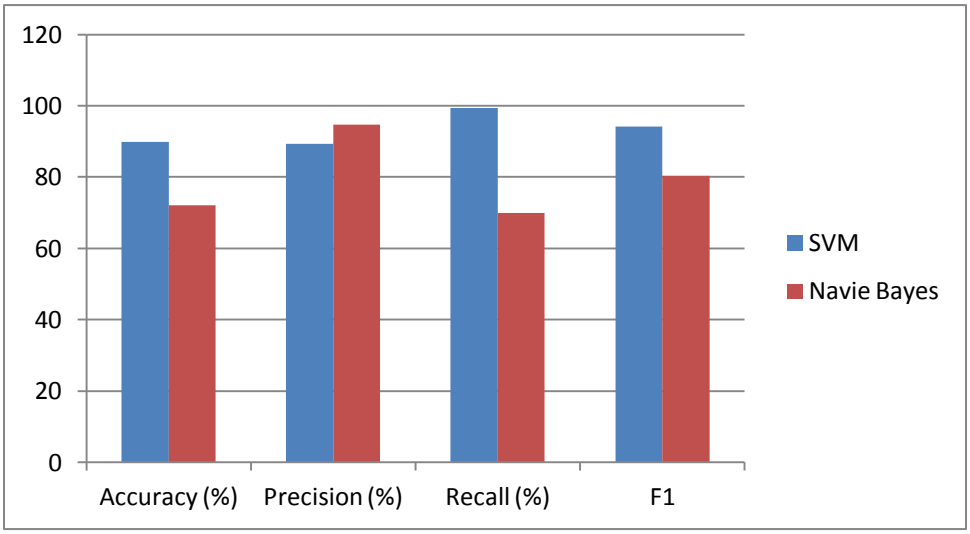


Figure 5. 9 : The cross validation results after applying Lexicon for uni-gram features (5 fold)

Figure:5.9 Shows the Accuracy, Precision, Recall, And F1 results of SVM classifier results after applying lexicon for uni-gram features, investigating the weighting schemes TF-IDF for 5 Fold.

Table 5. 12 : The cross validation results AFTER Applying Lexicon for uni-gram features (Multiply Weights)

Weighting Schemes	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	F1
SVM	90.68	90.34	99.23	94.58
Navie Bayes	71.60	94.71	69.20	79.97

Table 5.12 shows the results of the cross validation results after applying lexicon for uni-gram features. (Multiply Weights). The weights of the Lexicon are being taken and Multiplying them to the main model. The results of classifiers with Lexicon show that Lexicon based classification performed better than baseline classification without lexicon.

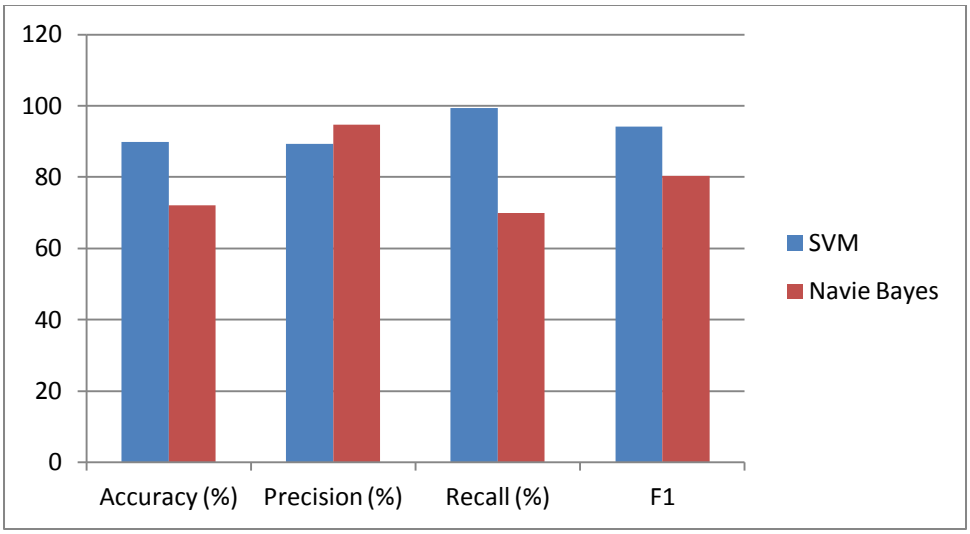


Figure 5. 10 : The cross validation results after applying lexicon for uni-gram features (Multiply Weights)

Figure 5.10 Shows the Accuracy, Precision, Recall, And F1 results of SVM and NB classifiers results after applying Lexicon for uni-gram features by multiplying weights.

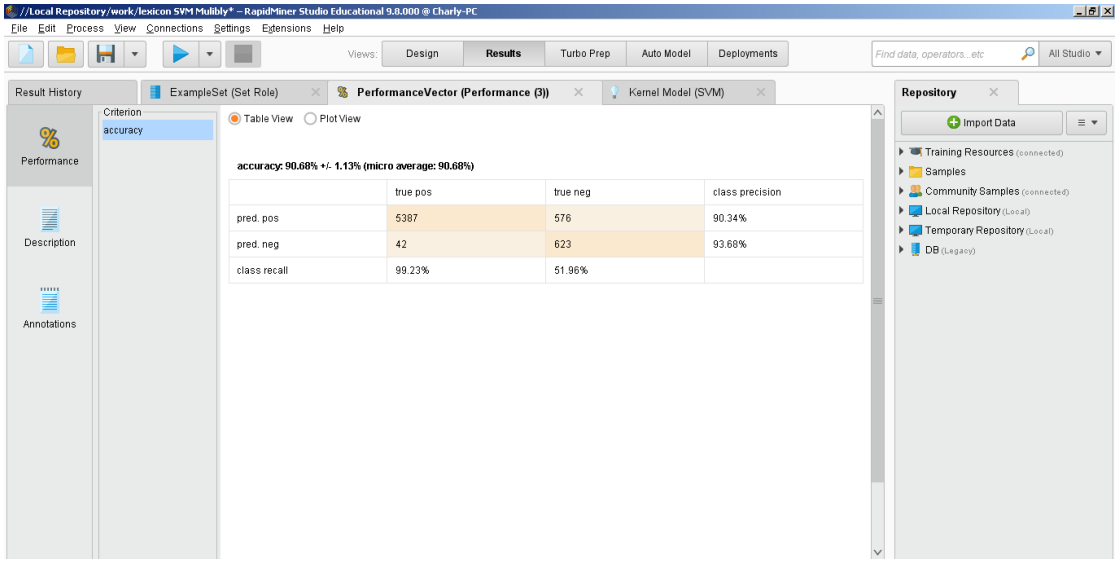


Figure 5. 11: Confusion Matrix from RapidMiner

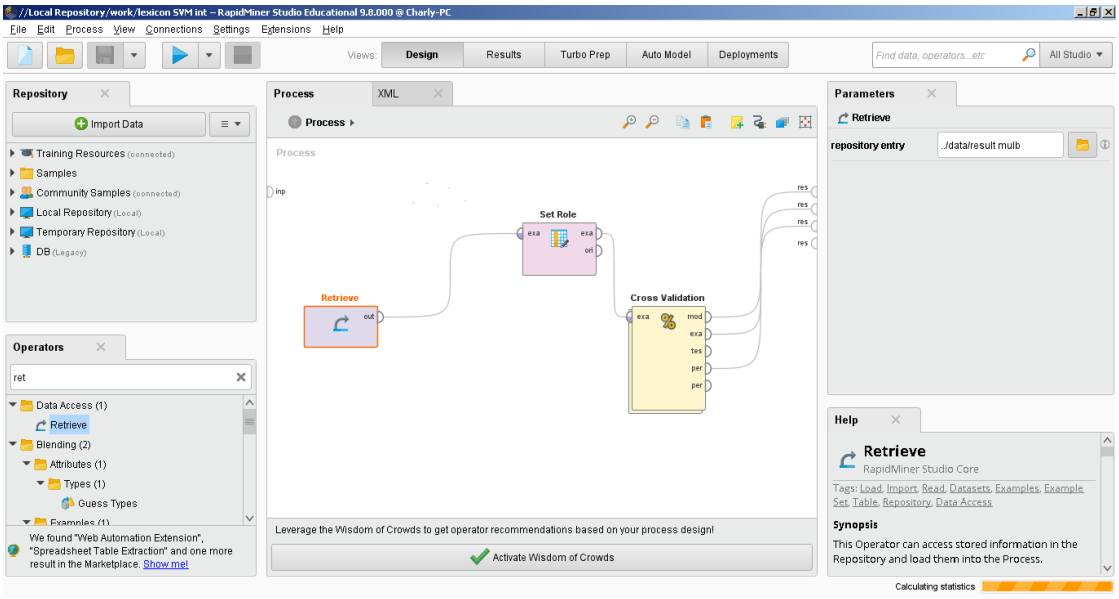


Figure 5. 12: Rapid Miner importing and processing data

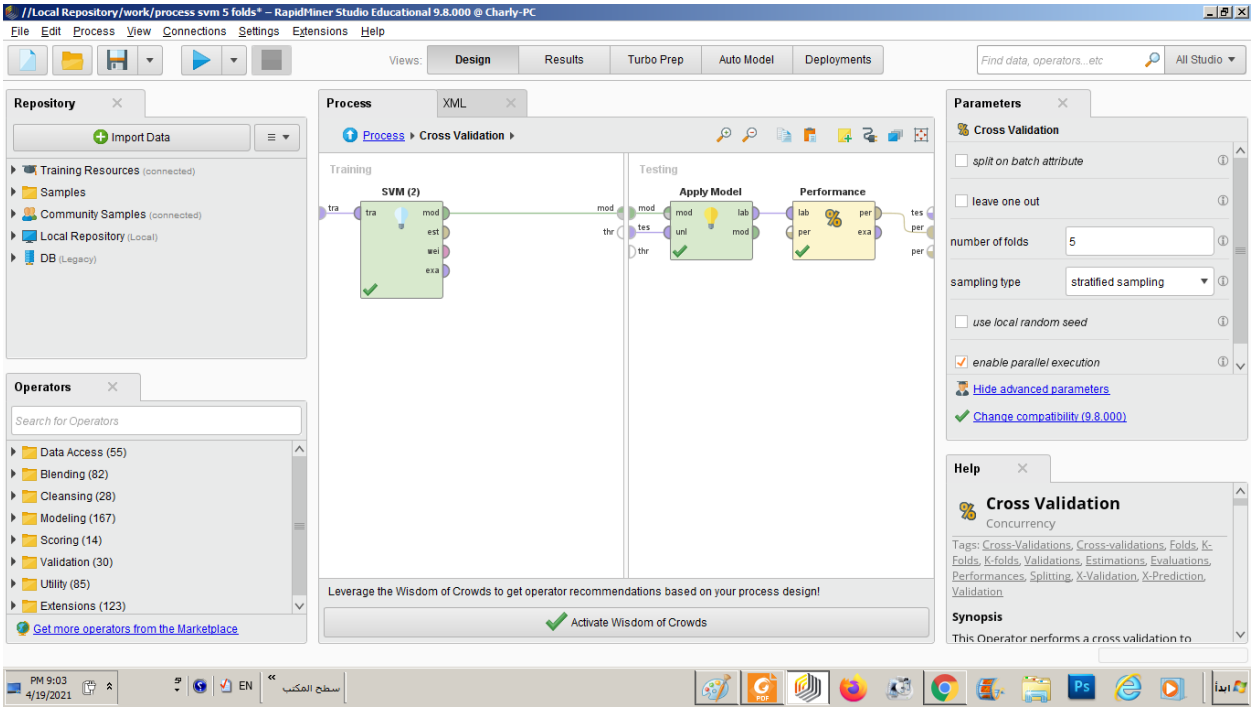


Figure 5. 13: Cross Validation Process in Rapid Miner Tool

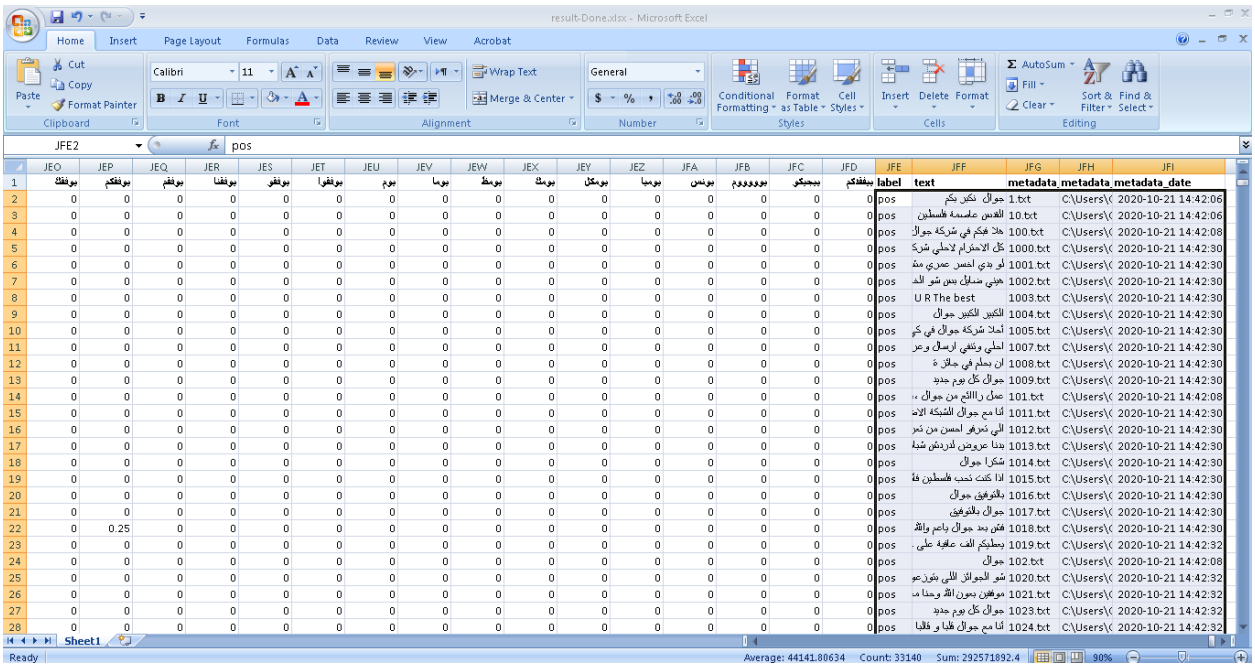


Figure 5. 14 : Excel Sheet Generating and updating Weights from the first phase SVM results methodology

Table 5. 13: The cross validation results after applying Lexicon for uni-gram features (5 fold) (Multiply Weights)

Weighting Schemes	***Accuracy (%)***	***Precision (%)***	***Recall (%)***	**F1**
SVM	89.86	89.37	99.45	94.14
Navie Bayes	72.10	94.71	69.85	80.40

Table 5.13 shows the results of the cross validation results after applying lexicon for uni-gram features, after divide data to 5 folds and use stratified sampling type, the weights of the Lexicon are being taken and Multiplying it to the main model. The results of classifiers with Lexicon shows that Lexicon based classification performed better than baseline classification without lexicon.

Table 5. 14 : Comparison Between different Classifiers Accuracy results for All uni-gram features and Techniques.

	Weighting Schemes	*Accuracy (%)*
SVM	**TF-IDF**	**90.57**
	TF	**90.49**
	BTO	**89.63**
Navie Bayes	**TF-IDF**	**77.58**
	TF	**71.79**
	BTO	**59.5**
KNN	**TF-IDF**	**86.69**
	TF	**87.43**
	BTO	**87.85**
before Applying pre-processing	**SVM**	**88.68**
	KNN	**83.74**
	Navie Bayes	**41.75**
SVM -after applying different pre-processing techniques	**Stop words Removal + Tokenizing**	**88.68**
	Light Stemming + Tokenizing	**88.71**
	Tokenizing	**89.21**
After applying lexicon (Addition Weights)	**SVM**	**89.80**
	Navie Bayes	**70.78**
After applying lexicon (Multiply Weights)	**SVM**	**90.68**
	Navie Bayes	**71.60**
After applying lexicon (5 fold)	**SVM**	**89.86**
	Navie Bayes	**72.10**

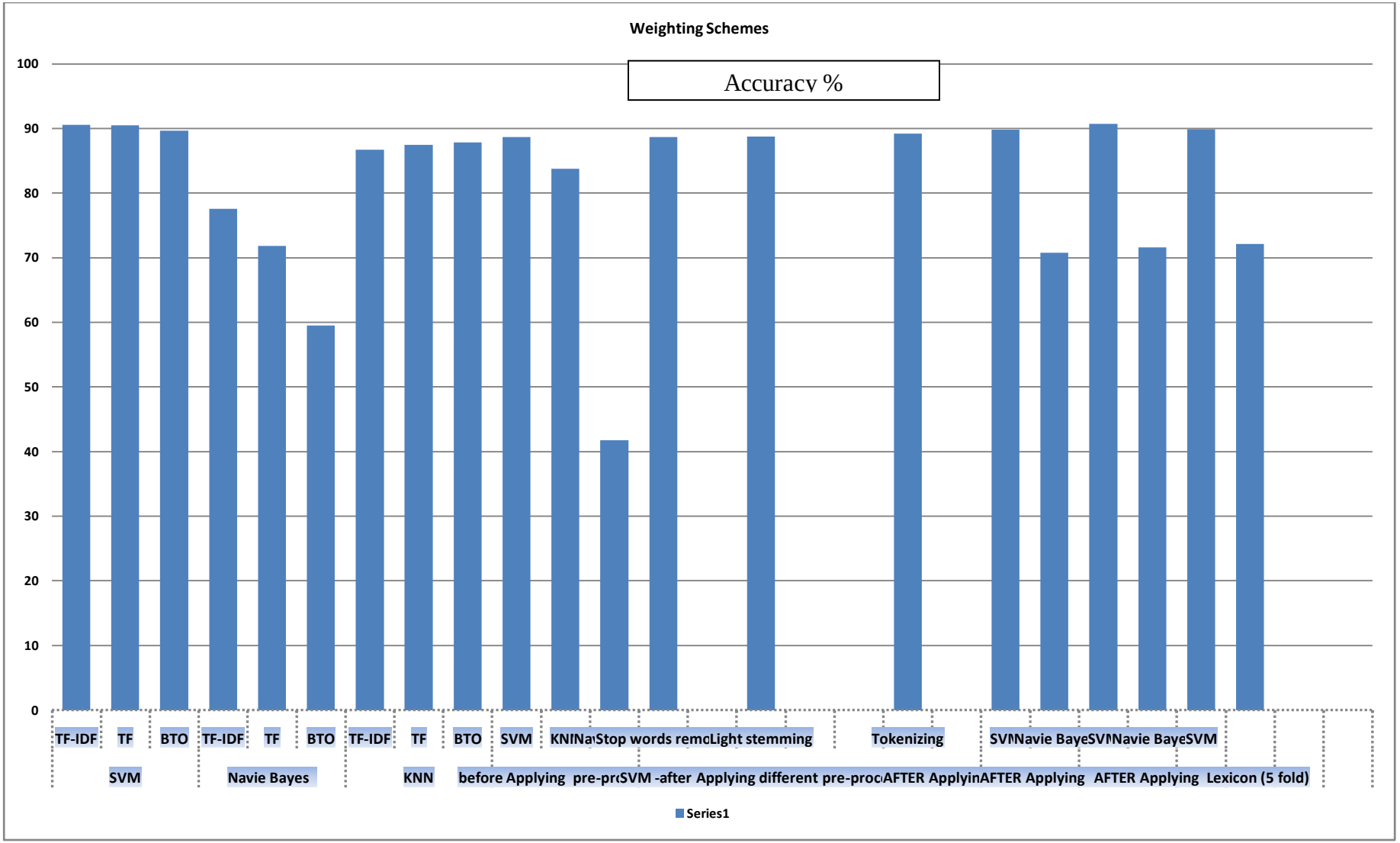


Figure 5. 15 **:** Accuracy results for All uni-gram features and Techniques.

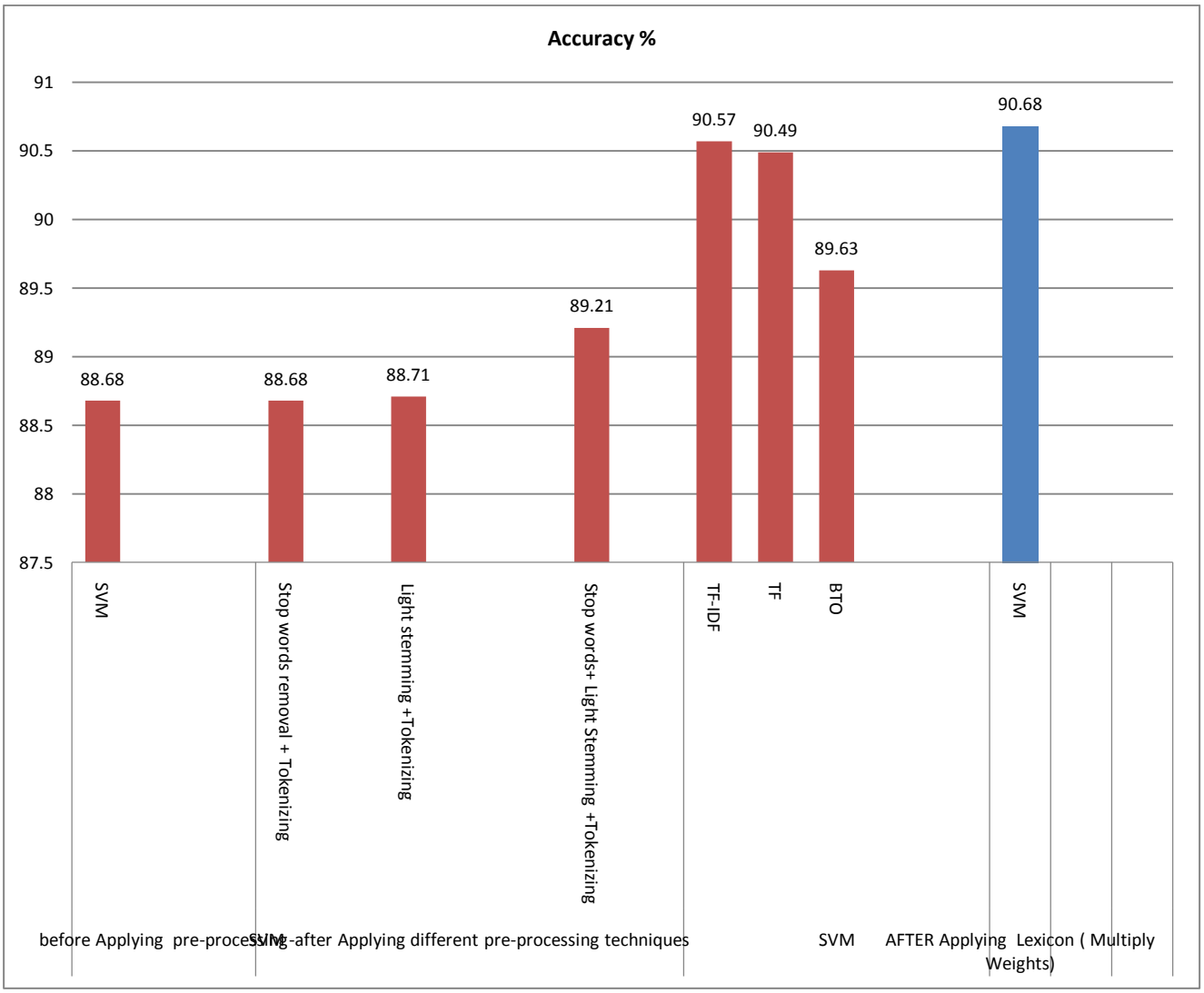


Figure 5. 16 :Accuracy results for All uni-gram features and Techniques

The results of the comparison showed that results are enhanced after applying Lexicon and the classifier performed better with TF-IDF in and 10 folds cross validation. The TF-IDF with Accuracy 90.68 % was selected as a vector representation. In addition, F1 score and Recall give the highest rates in case of classifying comments, Recall which are 99.23%, and Precision of 90.34 %, and having F-measure of 94.58%.

Chapter 6

Conclusions and Future Work

6.1 Conclusion and Future Works

The pervasiveness of social media generates a massive amount of data every day. Understanding and mining this data to determine the attitude of users towards products, services, events, and topics is very beneficial for individuals as well as stakeholders. The Arabic language is one of the popular content sharing languages through social networks. The Arabic language faces many challenges in analysis, especially in the case of colloquial Arabic which is the most widely spoken and written Arabic language. This thesis proposed a new hybrid approach to analyzing Palestinian colloquial Arabic sentiment. The proposed approach used two methods of sentiment analysis in Arabic: machine learning and a lexicon-based approach. The proposed approach consists of two stages. The first stage of sentiment analysis was handled by machine learning approach using the three classifiers SVM, KNN and Naive Bayes, and the result showed that the SVM classifier is the best an Accuracy of 90.57% and Precision of 90.38 %, and having F -measure of 94.50% and Recall of 99.02 %

A second stage is proposed to enhance the result of the first stage. The output of the first stage was used to train the SVM and NB classifiers. The lexicon was used to update the SVM weights. The second stage of sentiment analysis was handled by lexicon based approach using the classifiers of SVM and Naive Bayes, and the result showed that the SVM classifier exhibited better results with Accuracy of 90.68 %, Precision of 90.34 %, F-measure of 94.58 %, and Recall of 99.23 %

The results of classification using SVM were better than classification using Navie Bayes, as in previous studies and the literature review showed the superiority of SVM over the rest of the other classification results, especially as it is known in the classification of the text. And when the

dictionary (lexicon) is used, it is improved and the results are developed for the better, and this is confirmed by recent studies.

In the future work, both of data set and lexicon will be extended to cover more Arabic colloquial dialects. In addition, an effective solution will be searched, for solving sarcastic reviews and the ambiguous reviews problems. Semantics and ontology may help to solve such problems.

6.2 Challenges and Difficulties

Dealing with computers and analysis software with the Arabic language is much more difficult than dealing with the English language and takes longer time in issuing the results. Also, analysis with a large number of data requires a computer and high specifications to perform calculations and processing in record time. What forced me to divide the processing and calculation process and make equations through Microsoft Excel, which took great time and effort. Also, the data set that I have, which is homemade and unbalanced, is unbalanced in terms of negative and positive statements, and this may affect the results.

Also, some of the challenges in Sentiment Analysis are: people express opinions in complex ways and tend to express a lot of remarks in the form of sarcasm, irony, implication, …etc. which are very difficult to interpret. For Example," How can someone buy this cmera" is extremely negative sentiment yet contains no negative lexographic word. Even if an opinion word is presented in the text, there can be cases where an opinion word that is considered to be positive in one situation may be considered negative in another situation. In informal medium like Facebook or blogs (Social media), more likely people combine different opinions in the same sentence which is easy for a human to understand, but more difficult for a computer to parse. In addition, as there is constrain on

the number of characters used for each message, users follow informal grammar. They use misspellings, creative spellings, slang, URLs, emoticons, new words, and genre-specific terminology and special abbreviations. Working with such informal text opens a new trend in natural language processing.

Bibliography

[1] *The 65 days that led to chaos at the Capitol—BBC News*. (n.d.). Retrieved September 20, 2021, from https://www.bbc.com/news/world-us-canada-55592332

[2] *Elon Musk tweet about Signal messaging app sent stock with same name flying—Axios*. (n.d.). Retrieved September 20, 2021, from https://www.axios.com/stock-signal-advance-bubble-sign-d3c0f33b-7e7f-475b-b341-02377e1ed75d.html

[3] *Facebook MAU worldwide 2021*. (n.d.). Statista. Retrieved September 20, 2021, from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

[4] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5, no. 1 (2012): 1-167.

[5] R. Stagner, "The cross-out technique as a method in public opinion analysis," The Journal of Social Psychology, vol. 11, no. 1, pp. 79-90, 1940.

[6] A. L. Knutson, "Japanese opinion surveys: the special need and the special difficulties," Public Opinion Quarterly, vol. 9, no. 3, p. 313-319, 1945.

[7] P. L. FEGIZ, "Italian Public Opinion," Public Opinion Quarterly, vol. 11, no. 1, P. 92-96, 1947.

[8] A. CENEK and V. IVAN, "Polls come to Czechoslovakia," Public Opinion Quarterly, vol. 11, no. 4, p. 548-552, 1947.

[9] S. Sandri, D. Dubois and H. Kalfsbeek, "Elicitation, assessment, and pooling of expert judgments using possibility theory," IEEE Transactions on Fuzzy Systems , vol. 3, no. 3, p. 313-335, 1995.

[10] Wiebe, Janyce Marbury. "Recognizing subjective sentences: A computational investigation of narrative text." PhD diss., State University of New York at Buffalo, 1990.

[11] Wiebe, Janyce, Rebecca Bruce, and Thomas P. O'Hara. "Development and use of a gold-standard data set for subjectivity classifications." In Proceedings of the 37th annual meeting of the Association for Computational Linguistics, pp. 246-253. 1999.

[12] Hatzivassiloglou, Vasileios, and Kathleen McKeown. "Predicting the semantic orientation of adjectives." In 35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics, pp. 174-181. 1997.

[13] Viking Mäntylä, Mika, Daniel Graziotin, and Miikka Kuutila. "The Evolution of Sentiment Analysis-A Review of Research Topics, Venues, and Top Cited Papers." arXiv e-prints (2016): arXiv-1612.

[14] Olivie, J., Christianson, C. and McCarry, J., 2011. Handbook of natural language processing and machine translation.

[15] Bauer, Martin W., Aude Bicquelet, and Ahmet K. Suerdem. Text analysis: An introductory manifesto. (pp. xxi-xlvii). Sage, 2014.

[16] M. Dholariyq A. Ganatra and D. Bhoi, "Survey on Sentiment Analysis: Tools and Technique," International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, no. 3, pp. 6046-6050, 2017.

[17] D.J. Arnold, Lorna Balkan, Siety Meijer, R.Lee Humphreys and Louisa Sadler *Machine Translation: an Introductory Guide*, Blackwells-NCC, London, 1994, ISBN: 1855542-17x.

[18] A. Nenkova and K. McKeown, 'Automatic Summarization," Foundations and Trends in Information Retrieval, vol. 5, no. 2, p. 103-233, 2011.

[19] Clark, A., C. Fox, and S. Lappin. "Handbook of Computational Linguistics and Natural Language Processing." (2010).

[20] John, Adebayo Kolawole, Luigi Di Caro, and Guido Boella. "Text segmentation with topic modeling and entity coherence." In International conference on hybrid intelligent systems, pp. 175-185. Springer, Cham, 2016.

[21] Bolshakov, I.A., Gelbukh, A.: Computational linguistics: Models, resources, applications. IPN–UNAM–FCE, (2004) 187 pp.

[22] W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications, A survey; Ain Shams Engineering Journal, vol. 5, no. 4. pp. 1093-1113,2014.

[23] A Review on Classification Techniques in Machine Learning R. Vijaya Kumar Reddy1 , Dr. U. Ravi Babu2 , Volume No7, Special issue No.3 2018, *International Journal of Advance Research in Science and Engineering* (*IJARSE*).

[24] Y. Mang, IL lin and Z.-H. Zhou, "Underfinding bag-of-words model: a statistical framework," Intemational loumal of Machine Learning and Cybentetim. vol. 1, no. 1, pp. 43-52,2010.

[25] D. Jurafsky and J. H. Marvin, "N.grams," in Speech and Language Processing, New Jersey, Alan Apt, 1999, pp. 189,33.

[26] Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments." In ACL (Short Papers). 2011. Jan 1.

[27] Guyon, Isabelle, and André Elisseeff. "An introduction to feature extraction." In Feature extraction, pp. 1-25. Springer, Berlin, Heidelberg, 2006.

[28] Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." In Ijcai, vol. 14, no. 2, pp. 1137-1145. 1995.

[29] N. Kargard, "Estimation criterion, residuals and prediction evaluation," Computational Statistics & Data Analysis, vol. 5, no. 4. pp. 443-450, 2002.

[30] S. B. Kolsientis, Supervised Machine Learning: A Review of Classification Technique; Informatics, vol. 31, no. 3, pp. 249-268.2007.

[31] I.L.H.A.N.U.Y.S.A.L and H.A.L.T.A.Y.G.U.VENIR , “An overview of regression techniques for knowledge discovery," The Knowledge Engineering Review Cambridge University Press, vol. 14, no. 4, p. 319.340, 1999.

[32] C. Vapnik and C. VladimiCSuppon.vectorr. networks," Machine Learning, vol. 20, no. 3, p. 273-297,1995.

[33] S. Reschke, "Naive Bayes and Text Classification I - Introduction and Theory; arXiv:1410.5329v4 [es.LG], p. 20, 14 February 2017.

[34] Gurney, K. "An introduction to neural networks, UCL Press Limited 11 New Fetter Lane London EC4P 4EE, SBN 0-203-45151-1 Master e-book ISBN." (1997).

[35] Quinlan, J. Ross. "Induction of decision trees." Machine learning 1, no. 1 (1986): 81-106.

[36] Turney, Peter D. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews." arXiv preprint cs/0212032. 2002 Dec 11.

[37] U. N. G. o. E. O. G. Names, "UNGEGN list of country names," 2 May 2011. [Online]. Available:https://unstats.un.org/unsd/geoinfo/UNGEGN/docs/26th-gegn-docs/WP/WP54_UNGEGN%20WG%20Country%20Names%20Document%202011.pdf

[38] A. B. Al-Saleh and M. E. B. Menai, "Automatic Arabic text summarization: a survey," Artificial Intelligence Review, vol. 45, no. 2, p. 203–234, 2016

[39] Badaro, Gilbert, Ramy Baly, Hazem Hajj, Wassim El-Hajj, Khaled Bashir Shaban, Nizar Habash, Ahmad Al-Sallab, and Ali Hamdi. "A survey of opinion mining in Arabic: A comprehensive system perspective covering challenges and advances in tools, resources, models, applications, and visualizations." ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) 18, no. 3 (2019): 1-52.

[40] S. Al-Azani and E.-S. M. El-Alfy, “Combining emojis with Arabic textual features for sentiment classification,” in Proceedings of the 2018 9th International Conference on Information and Communication Systems (ICICS), pp. 139– 144, IEEE, Irbid, Jordan, April 2018.

[41] A. Hamdi, K. Shaban, and A. Zainal, “A class-specific sentiment analysis framework,” ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP),vol. 17, no. 4, p. 32, 2018.

[42] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, “Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels’ reviews,” Journal of Computational Science, vol. 27, pp. 386–393, 2018

[43] s. Khoja, "Research Interests," Pacific University | 2043 College Way • Forest Grove, Oregon 97116, [Online].Available: http://zeus.cs.pacificu.edu/shereen/research.htm. [Accessed 11 July 2017].

[44] M. N. Al-Kabi, I. M. Alsmadi, R. T. Khasawneh, and H. A. Wahsheh, “Evaluating social context in Arabic opinion mining,” @e International Arab Journal of Information Technology, vol. 15, no. 6, pp. 974–982, 201

[45] Altyeb Altaher “Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting” International Journal of Advanced and Applied Sciences, Volume 4, Issue 8 (August 2017), Pages: 43-49

[46] A. Elouardighi, M. Maghfour, H. Hammia, and F.-z. Aazi, “A machine learning approach for sentiment analysis in the standard or dialectal Arabic facebook comments,” in Proceedings of the 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), pp. 1–8, IEEE, Rabat, Morocco, October 2017.

[47] A. Aliane, H. Aliane, M. Ziane, and N. Bensaou, “A genetic algorithm feature selection based approach for Arabic sentiment classification,” in Proceedings of the ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp. 1–6, IEEE, Agadir, Morocco, November 2016.

[48] H. K. Aldayel and A. M. Azmi, "Arabic tweets sentiment analysis – a hybrid schema," Journal of Information Science, vol. 42, no. 6, p. 782–797, 2016.

[49] Semantria, L. L. C. "Semantria Out-of-the-Box Reliability". (2015). https://semantria.com/case-studies

[50] Oscar Araque, Ignacio Corcuera, Constantino Roman, Carlos Iglesias, and J. Fernando Sanchez-Rada. “Aspect based Sentiment Analysis of Spanish Tweets.” TASS 2015, September 2015, pp 29-34

[51] Rehab M. Duwairi, Islam Qarqaz , “Arabic Sentiment Analysis using Supervised Classification”The 1st International Workshop on Social Networks Analysis, Management and Security (SNAMS - 2014), August 2014, Barcelona,spain.

[52] Soliman, Taysir Hassan, M. A. Elmasry, A. Hedar, and M. M. Doss. "Sentiment analysis of Arabic slang comments on facebook." International Journal of Computers & Technology 12, no. 5 (2014): 3470-3478.

[53] Arabic Sentiment Analysis. (n.d.). Retrieved September 20, 2021, from https://saifmohammad.com/WebPages/ArabicSA.html

[54] RapidMiner | Best Data Science & Machine Learning Platform. (n.d.). Retrieved September 20, 2021, from https://rapidminer.com/

[55] Nguyen, Giang Hoang, Abdesselam Bouzerdoum, and Son Lam Phung. "Learning pattern classification tasks with imbalanced data sets." Pattern recognition (2009): 193-208.

[56] Akosa, J. (2017, April). Predictive accuracy: a misleading performance measure for highly imbalanced data. In Proceedings of the SAS Global Forum (pp. 2-5).

الملخص

ينتج الانتشار السريع لوسائل التواصل الاجتماعي كمية هائلة من البيانات كل يوم.ويعد فهم هذه البيانات والتنقيب عنها لتحديد موقف المستخدمين تجاه المنتجات والخدمات والأحداث والموضوعات الأخرى مفيدًا جدًا للأفراد وكذلك أصحاب المصلحة. اللغة العربية هي إحدى لغات مشاركة المحتوى الشائعة عبر الشبكات الاجتماعية ، ويواجه تحليل المحتوى المكتوب باستخدام اللغة العربية العديد من التحديات ، لا سيما في حالة اللغة العامية التي تستخدم على نطاق واسع في وسائل التواصل الاجتماعي. اقترحت هذه الرسالة نهجًا هجينًا لتعزيز تحليل المشاعر لمحتوى وسائل التواصل الاجتماعي المكتوب باللغة العربية باستخدام العامية الفلسطينية.

يعتمد النهج المقترح على استخدام التعلم الآلي والمفردات في تحليل المشاعر للمحتوى العربي.

يتكون من مرحلتين ، يتم التعامل مع المرحلة الأولى بطريقة التعلم الآلي باستخدام المصنفات الثلاثة SVM و KNN و Naïve Bayes ، بينما يتم التعامل مع المرحلة الثانية بطريقة تعتمد على المعجم باستخدام المصنفات المختلفة SVM و Naive Bayes.

يتم تنفيذ المرحلة الثانية من النهج لتعزيز نتيجة المرحلة الأولى. تم استخدام مخرجات المرحلة الأولى لتدريب المصنفين SVM و NB. بينما تم استخدام المعجم لتحديث أوزان SVM الناتجة عن المرحلة الأولى.

يتم اختبار النهج المقترح باستخدام مجموعة من البيانات Dataset المخصصة مستخرجة من صفحات Facebook لبعض مزودي الخدمات العامة في فلسطين ، وتتألف مجموعة البيانات من آلاف التعليقات والمنشورات حول مواضيع مختلفة.

أدى النهج القائم على المعجم إلى تحسين دقة الكشف عن قطبية التعليق حيث زادت نتائج الدقة من 90.57٪ إلى 90.68٪ باستخدام النهج المقترح ، بينما زادت نتائج قياس F من 94.50٪ إلى 94.58٪ ، أشار أيضًا إلى أن خوارزمية SVM كانت أفضل خوارزمية ML مقارنة بخوارزمية NB و KNN لمشكلة البحث هذه .