بسم الله الرحمن الرحيم

Sudan University of Science and Technology

Collage of Computer Science and Information Technology

Build Sentiment Analysis Model for the Sudanese Dialect for the Internet Service (Sudani - Case Study)

بناء نموذج لتحليل المشاعر للهجة العامية السودانية عن خدمة الإنترنت (شركة سوداني للاتصالات)

August 2022

A scientific project submitted as one of the requirements for obtaining a Bachelor's degree in Software Engineering and Information Technology

Sudan University of Science and Technology

Collage of Computer Science and Information Technology

Department of Information Technology and

Software Engineering

Build Sentiment Analysis Model for the Sudanese Dialect for the Internet Service (Sudani - Case Study)

بناء نموذج لتحليل المشاعر للهجة العامية السودانية عن خدمة الإنترنت (شركة سوداني للاتصالات)

By:

Athar Abdalgauom (information technology)

Rabab Khalifa (Software Engineering)

Ryan Jafar (Software Engineering)

Under the supervision of:

A. Muhammad Al-Fateh Othman

CHAPTER ONE
INTRODUCTON

INTRODUCTON

1.1Overview

Measuring customer satisfaction and their reactions to the services and products provided has always been an obsession for companies and organizations that provide these services, with the aim of helping them make better decisions in order to improve the quality of their products and services to increase the satisfaction of their customers and gain their trust. Several years ago, this process was difficult and expensive and it followed traditional methods. But with the advancement of Web 0.2 platforms and the proliferation of forums, blogs and social media, the matter has become much easier.

Sentiment analysis is a type of textual analysis also known as mining. It relies on a combination of statistical analysis, natural languages, and machine language to identify and extract information from text files. For example, the feelings of people who write their opinions, ideas, judgments, or evaluation of a particular topic or event, or a company and its activity. This analysis is also known as opinion mining (with focus extraction) or emotional rating. Despite the many and different names, the goal of this analysis is the same, which is to know the opinions of users or the

public on a particular topic by analyzing a large group of texts from multiple sources.

This analysis can be used at multiple levels and these levels depend on the objective of the analysis. For example, you need to average the attitudes and expressions of a group of respondents in order to know what percentage of customers like a particular product. Another example, if you want to know if visitors liked or disliked a set of clothes, and why? Or they are comparing with another product similar to it, you will need to analyze each opinion or phrase focusing on certain aspects and using specific keywords.

Social networking sites constitute one of the most prominent forms of daily life in our contemporary reality, and in it people find complete freedom to express their negative and positive opinions. Therefore, this huge amount of important data must be taken advantage of, collected from the sites, analyzed and used in making decisions.

1.2 Problem Statement and its significance

Sudanese institutions have a lot of data about customers' feedback, comments, suggestions, problems and

The extent of their satisfaction with the services provided by the institution, and because these data are mostly in the colloquial Sudanese dialect, it is not easy to analyze and categorize them, and then extract information and statistics that help in decision-making except by traditional methods. Which are inefficient in terms of time consuming, data unreliability, large workforce, and inaccurate and biased results.

1.3 Research aim and objectives

The aim of the research The research aims to study, design and build a system that analyzes the publications and reviews of customers published in Arabic on social media in order to determine their public opinion, positive/negative/neutral, based on the concepts and techniques of sentiment analysis and opinion mining, which helps the decision maker in institutions (companies). Communication) to take the necessary measures in order to improve the quality of services provided by this institution

1.4 Research scope

1. Building a model for analyzing feelings on the Sudanese dialect into positive, negative and neutral.
2. Use the form on comments from Sudani Telecom's page.
3. Better accuracy than previous models

CHAPTER TWO

THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1 Theoretical Background

2.1.1 Introduction

Sentiment evaluation has emerge as a key science to reap perception from social networks. The area has reached a stage of maturity that paves the way for its exploitation in many special fields such as marketing, health, banking or politics. The modern-day technological advancements, such as deep studying techniques, have solved some of the usual challenges in the vicinity brought about by way of the shortage of lexical resources. In this Special Issue, unique methods that enhance this self-discipline are presented.

In this chapter we discussed Sentiment Analysis and Opinion Mining , social network analysis , algorithms for sentiment analysis and classification algorithms.

2.1.2 SENTIMENT:

Sentiments can be expressed in different ways. It can be expressed with various kinds of emotions, by passing judgments, vision or insight or views of people. An emotion can be expressed as sudden reaction by person consciously or

unconsciously depending on situation. If we study the emotion in text format then it can be seen in two different ways. First it can be studies as impact on writer – the way he chooses word to express certain emotions. And second impact on reader-how he interpret content written depending on his state of mind and his ability to analysis things.[2]

2.1.3 Sentiment Analysis:

Sentiment analysis or opinion mining is the process to identifying and recognize or categorized the users‘ emotion or opinion for any services like movies, product issues, events or any attribute is positive, negative or neutral. The sources for this analysis is social communication channels i.e Web site which include reviews, forum discussions, blogs, micro-blogs, Twitter etc. This research field is very popular nowadays because of its opinioned data where user can find reviews for any services which are useful for their daily lives. The large amount of opinionated data is stored in digital forms. For particular topic or any opinion the sentiment analysis which relates the mining of the data works and give the output. For sentiment analysis, research works on emotion-based summarization , feeling or mind extraction. Sentiment analysis also known as Opinion mining which uses NLP – Natural Language Processing to following the emotions, feelings of the

public opinion about a particular topic for any product or services. As Sentiment analysis is very famous, it can be also useful in many ways in surveys and advertisement campaign by getting the success rate of any product or services with people‘s opinion or suggestion. .It also give the information about people liking and disliking and company gets much clear idea regarding its product features.[2]

2.1.4 Sentiment Analysis and Opinion Mining

Opinions are central to nearly all human things to do and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the alternatives we make are to a extensive degree, conditioned upon how others see and evaluate the world. For this reason, when we want to make a selection we often are searching for out the opinions of others. This is no longer solely genuine for folks but also authentic for organizations

Opinions and its associated principles such as sentiments, evaluations, attitudes and feelings are the topics of learn about of sentiment evaluation and opinion mining**.(**1**)**

Figure 2.1 below show the processes related to sentiment analysis in social networks.

2.1.5 Sentiment analysis levels

Sentiment analysis has been investigated on several levels: Document Level, Sentence

Level, Phrase Level, and Aspect Level. Sentiment analysis in each level such as document,

sentence and phrase.[1]

2.1.5.1 Document level sentiment analysis

Document-level: Document level sentiment analysis is performed on a whole document,and single polarity is given to the whole document. This type of sentiment analysis is not used a lot. It can be used to classify chapters or pages of a book as positive, negative, or neutral. At this level, both supervised and unsupervised learning approaches can be utilized to classify the document. Cross domain and cross-language sentiment analysis are the two most significant issues in document-level sentiment analysis. Domain-specific sentiment analysis has been shown to achieve remarkable accuracy while staying highly domain-sensitive. In these tasks, the feature vector is a set of words that must be domain-specific and limited.

2.1.5.2 Sentence level sentiment analysis:

Sentence level: In this level of analysis, each sentence is analyzed and finding with a corresponding polarity. This is highly useful when a document has a wide range and mix of sentiments associated with it . This classification level is associated with subjective classification . Each sentence polarity will be determined independently using the same methodologies as the document level

but with greater training data and processing resources. The polarity of each sentence may be aggregated to find the sentiment of the document or used individually. Occasionally, document-level sentiment analysis is insufficient for specific uses . In previous work on sentence-level analysis has been devoted to finding subjective sentences. However, more difficult tasks, such as working with conditional sentences or ambiguous statements . In these circumstances, sentence-level sentiment analysis is critical.

2.1.5.3 Phrase level sentiment analysis:

Phrase level: Sentiment analysis also be performed where opinion words are mined at phrase

level, and classification will be done. Each phrase may contain multiple aspects or single aspects. This may be useful product reviews of multiple lines; here, it is observed that a single aspect is expressed in a phrase It has been a hot topic of researchers in recent times. While document-level analysis concentrated on categorizing the entire document as subjective, either positively or negatively, sentence-level analysis is more beneficial, as a document contains both positive and negative statements. Word is the most basic unit of language; its polarity is intimately related to the subjectivity of the sentence or document in which it

appears. A sentence containing an adjective has a high probability of being a subjective sentence. Additionally, the term chosen for expression represents the demographic characteristics of individuals, such as gender and age, and its desire, social standing, and personality, other psychological and social characteristics . As a result, term serves as the foundation for text sentiment analysis.

2.1.5.4 Aspect level sentiment analysis:

Aspect level: sentiment analysis is performed at the aspect level. Each sentence may contain multiple aspects; therefore, Aspect level sentiment analysis. Primary attention to all the aspects used in the sentence and assigns polarity to all the aspects after which an aggregate sentiment has calculated for the whole sentence .[1]

2.1.6 Challenges of Sentiment Analysis:

sentiment evaluation or classification is viewed a specific case of textual content classification in a herbal language processing. Although the variety of lessons in sentiment evaluation are small, the technique of sentiment classification is extra challenging than the typical theme textual content classification (Liu, 2009). In

subject textual content classification, classification depends on the usage of keywords, however this does now not normally work properly in the case of sentiment evaluation (Turney, 2002).The different difficulties in sentiment evaluation come from the nature of this problem. Sometimes, the poor sentiment would possibly be expressed in a sentence barring the usage of any obvious poor words. Moreover, there is a first-rate line between whether or not a sentence should be labeled goal or subjective. Determining the opinion holder -the one who expresses the sentiment in the text- is one of the most hard duties in sentiment analysis. The sentiment evaluation pretty relies upon on the area of the data. The phrases occasionally have high-quality sentiment in a precise domain, whereas they have any other polarity sentiment in a one of a kind domain. Finally, some different writing patterns such as irony, sarcasm, or negated sentences may want to deliver extra challenges to sentiment evaluation (Liu,2009).**(2)**

Figure 2.1.2 below show the challenges in sentiment analysis

2.1.7 Application of Sentiment Analysis:

People share knowledge, experiences and ideas with the world by way of the usage of Social Media like blogs, forums, wikis, review sites, social networks, tweets and so on. This has modified the manner in which human beings speak and impact social, political and monetary conduct of different humans in the Web Indeed the Web approves all and sundry having a voice ,promising to enhance human collaboration skills on a worldwide scale, enabling men and women to share opinions by means of read-write Web and user's generated contents. According to an opinion "is certainly a wonderful or negative sentiment, view, attitude, emotion, or appraisal

about an entity or an issue of the entity" from an opinion holder at a specific time . The entity can be a product/service, event, person, organization, or subject matter consisting of aspects (features/attributes) that represents each elements and attributes of the entity. With the explosion of person generated opinions there is the need by way of companies, politicians, carrier providers, social psychologists, researchers and different actors to analyze them in order to put into effect higher selection choices. The literature on sentiment evaluation targeted on different domains, from administration sciences to pc science, social sciences and commercial enterprise due to its significance to society as complete and special duties such as: subjective expressions sentiments of words, subjective sentences**()**.

2.1.8 Social Networks Analysis

Social media is media used for social interaction. They are enabled with the aid of conversation technologies such as the net and smartphones and they flip conversation into an interactive dialogue

(Wikipedia, 2011). Interactions on social media being fairly distributed, decentralized and occurringin actual time, they supply the indispensable breadth and immediacy of data required in instances of emergencies (Palen & Vieweg, 2008). Since social media provide a uniquely fast and effective way to disseminate information, correct and inaccurate, top and terrible unfold equally alike as incorrect

information can unfold like wild fire. However, there is indication that social networks have a tendency to favor valid statistics over rum ours (Castillo, Mendoza, & Poblete, 2011).Twitter and Facebook are accurate examples of social media beneficial in disaster conditions because they provide vital records as they are happening. Twitter is a micro-blogging service, a structure of lightweight chat permitting customers to put up and change brief 140-character-long messages regarded as tweets. Although most tweets are dialog and chatter, they are additionally used to share applicable information and file information (Castillo, et al., 2011). Twitter is turning into a precious device in catastrophe and emergency situations as there is growing proof that it is no longer simply a social network, it is additionally a information service(Yates & Paquette, 2011).

Social Network Analysis (SNA) is a sociological method for analyzing patterns of relationships and interactions between social actors in order to find out underlying social shape such as: central nodes that act as hubs, leaders or gatekeepers; tremendously related groups; and patterns of interactions between businesses (Wasserman & Faust, 1994). SNA has been used to find out about social interplay in a wide range of domains. Examples include: collaboration networks (Newman, 2001), administrators of companies (Davis & Greve, 1997; Davis, Yoo, & Baker, 2003), organizational conduct (Borgatti & Foster, 2003), inter-organizational family

members (Stuart, 1998), computer-mediated communications (Garton, Haythornthwaite, & Wellman, 1999), and many others.(**4**).

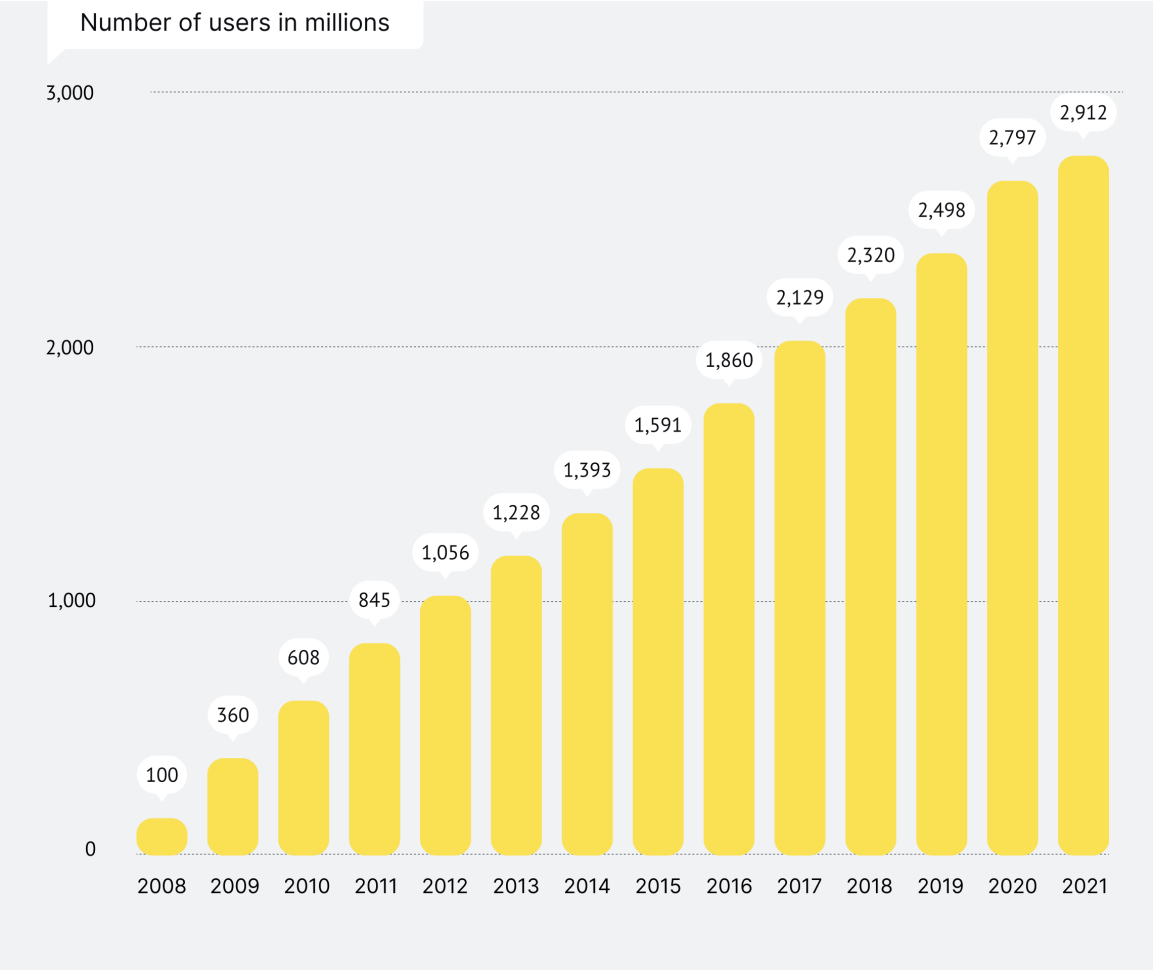


Figure below 2.1.3 Number of Monthly Active Facebook Users 2008-2021

2.1.9 Arabic Language

The focal point is in the direction of sentiments expressed in the Arabic language due to the developing populace of Internet customers that use the Arabic language; it is estimated about 5% of global customers. Also, in the ultimate few years, it is regarded one of the most rising languages on the web. The morphological complexity and the lexical ambiguity of the Arabic language can pose a project when working with it. Another trouble is that the Arabic language has three one-of-a-kind varieties: classical Arabic (CA), present day trendy Arabic (MSA) and dialectical Arabic (DA). Throughout this article The identification of fantastic and terrible opinions is no longer an effortless task; it requires data retrieval, linguistic knowledge, herbal language processing and a profound comprehension of the textual context .

A majority of the SSA research have targeted solely on the English or European languages [5]. Some research analyzed the morphologically-rich languages like Urdu, Arabic, and Turkish [6]. The dialects in the Arabic language are very challenging. This trouble is extra extraordinary if people, who submit texts on social networking sites, use an casual fashion of writing, or if the bi/multi-lingual folks use a combination of extraordinary languages. For instance: فالنتاين – مرهم .(lol (laugh out loud راديو – كريم للبشرة- لول Any tweet can show the variability and complexity of the

sentiments. Though the twitter messages are short, they can existing a large array of statistics in a compressed format[7]. Tweets deliver sarcasm, or an unclear/mixed polarity as follows: يعم والشر يخص الخير. high-context cultures characterize a challenge in tweets, now not solely due to sarcasm however additionally due to totally unique meanings of one phrase e.g. the phrase “Eshta”, "**;**إشطة "**l** for Egyptian dialect and the phrase "Zabit" " زابط " in

Sudanese dialect. (8).

2.1.10 Sentiment Analysis Methodologies

Sentiment evaluation performed a superb position in the vicinity of researches carried out with the aid of many, there are many techniques to carryout sentiment analysis. Still many researches are going on to discover out higher picks due to its significance in this scenario. Some of the strategies are mentioned in this research.

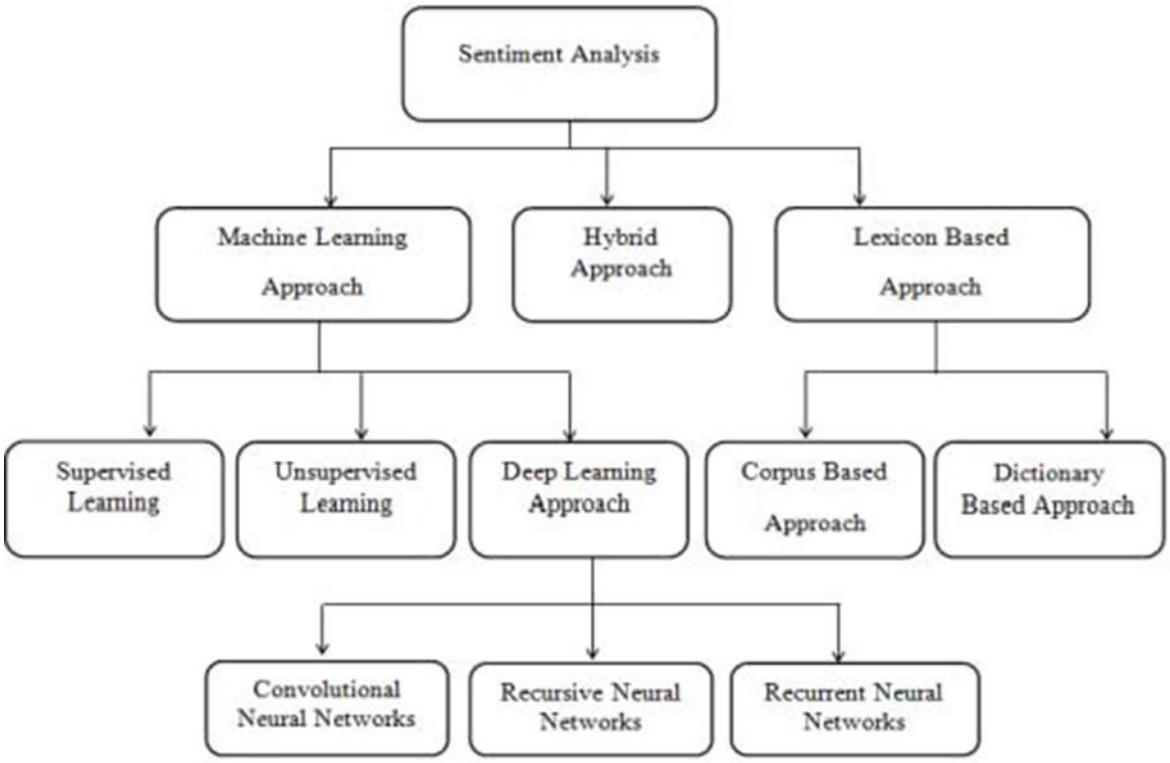


Figure below 2.1.3 Sentiment Analysis methodologies

1. Data collections:

Data can be collected from the internet via web scraping, social media, news channels,E-commerce websites, Forums, Weblog, some other websites shown in Fig. 2. Data Collection is the first stage in the Sentiment Analysis. Depending on task sentiment analysis of findings, text data can be combined with other types of data like video, audio, location, etc.[1]

1.1. Social media:

Social data refers to information gathered via social media networks. It demonstrates how consumers interact with the product by accessing, posting, and exchanging over. Academic study on individual, group and behavior uses social media as a dynamic data source. It refers to Internet apps that are web or mobile-based that enable users to create, access, and trade user generated content.[1]

2. Feature selection:

It is important to remember that developing a classification model requires first identifying relevant features in dataset.Thus, a review can be decoded into words during model training and appended to the feature vector. For single word is considered, the technique is called a “Uni-gram”; when two words are considered, the technique is called a “Bi-gram”; and if three words are considered, the technique is referred to as a “Tri-gram.” combination of unigram and bigram helpful for analysis; the context feature which helpful for getting results most accurate.[1]

3. Feature extraction:

Feature extraction is a key task in sentiment classification as it involves the extraction of

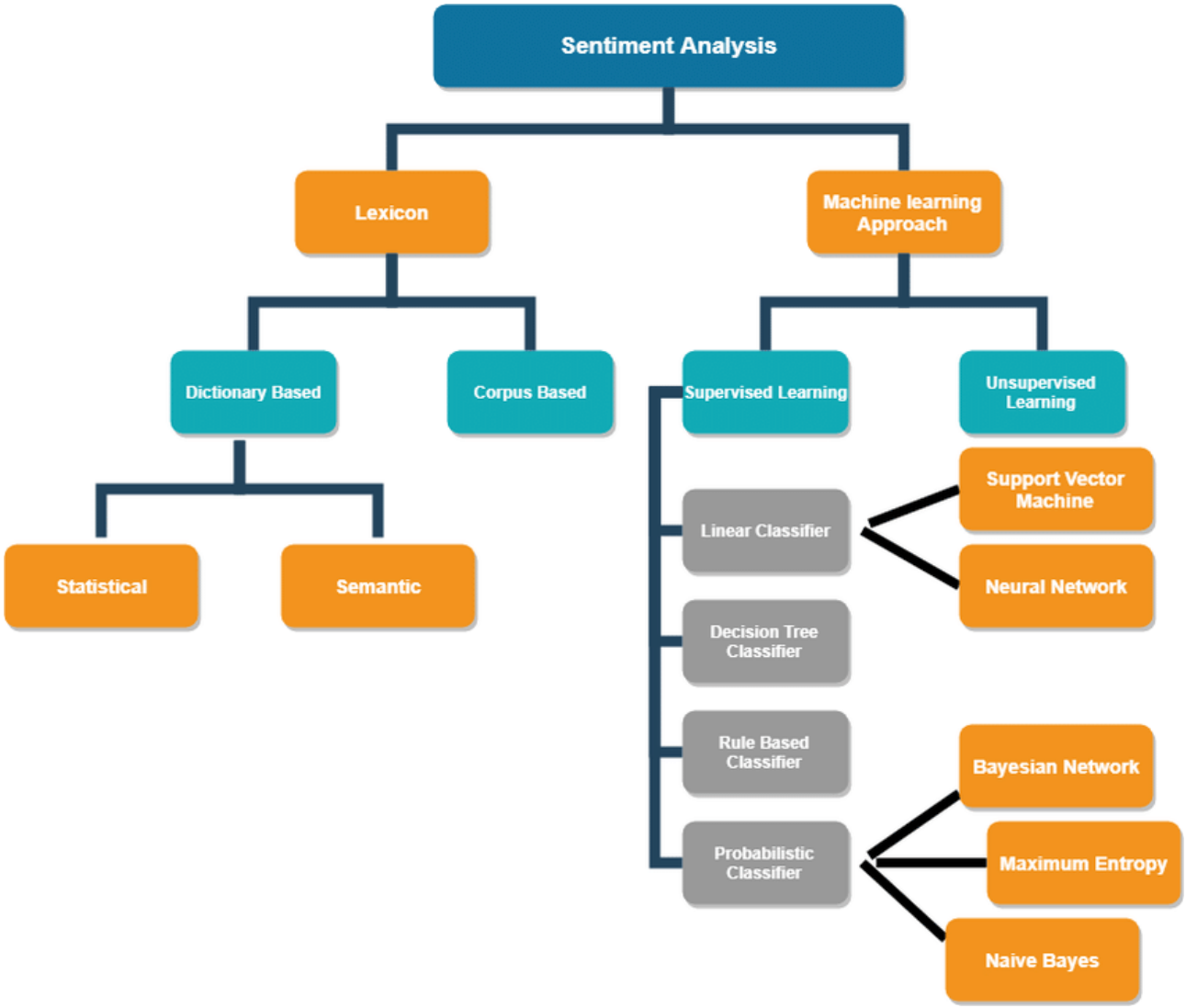
valuable information from the text data, and it will directly impact the performance of the

model. The approach tries to extract valuable information that encapsulates the text's most

essential features;commonly used techniques for feature extractions are Terms frequency,Parts of Speech tagging,Negations and Bag of Words.[1]

4. MATERIAL AND METHOD AND APPROACHES:

Numerous methodologies are available for opinion mining, but two main groups are used. The problems of Sentiment Analysis will be solved by the first group using by implementing the machine learning approach. The second group uses lexicon-based method which is a linguistically-inclined method.[2]



4.1 Machine learning Approach:

Machine learning strategies work by training an algorithm with a training data set before applying it to the actual data set. Machine learning techniques first trains the algorithm with some particular inputs with known outputs so that later

it can work with new unknown data . Some of the most renowned works based on machine learning areas follows[3]:

4.1.1 Support Vector Machine

It is a non-probabilistic classifier in which a large amount of training set is required. It is done by classifying points using a (d-1)-dimensional hyper plane. SVM finds a hyper plane with largest possible margin . Support Vector Machines make use of the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class membership[3].

4.1.2 N-gram Sentiment Analysis:

In the fields of linguistics and probability, an n-gram is a contiguous sequence of n items from a

given sequence of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The n-grams typically are collected from a text or speech corpus. When the items are words, n-grams may also be called shingles[3].

4.1.3 Naïve Bayes Method:

NB technique is utilized for both categorization and training. NB is a Bayesian classifi-

cation approach based on the theorem of bayes. NB is a probabilistic classifier that uses

Bayes theorem to predict the probability of a given set of features as part of any particular

label. The conditional probability that event A occurs given the individual probabilities of

A and B and conditional probability of occurrence of event B[1].

4.1.4 Logistic regression (LR):

A machine learning technique known as logistic regression works by multiplying an input value by a weight value. It is a classifier that learns which input

properties are most helpful in identifying positive and negative classes. Logistic regression is a probabilistic regression analysis used for classification tasks. For binary classification applications, logistic regression is commonly deployed. When there are multiple explanatory variables, logistic regression calculates the ratio of odds. Logistic regression uses Maximum-likelihood to calculate best parameters[1].

4.1.5 Decision tree (DT)

DT Classifier is a supervised learning technique where a tree is built using the training example to classify the polarity of the text. DT uses a condition to divide data into parts recursively. RF are used frequently than DT which combines multiple DT to avoid over fitting and improve accuracy[1].

4.1.6 Maximum Entropy Classifier:

A Maximum Entropy (ME) classifier, or conditional exponential classifier, is parameterized by a set of weights that are used to combine the joint-features that are generated from a set of features by an encoding. The encoding maps each pair of feature set and label to a vector. ME classifiers belong to the set of classifiers known as the exponential or log-linear classifiers, because they work by

extracting some set of features from the input, combining hem linearly and then using this sum as exponent. If this method is done in an unsupervised manner, then Point wise Mutual Information (PMI) is made use in order to find the co-occurrence of a word with positive and negative

words. The ME Classifier is one of the models which do not assume the independent features[3].

4.1.7 K-nearest neighbours (KNN)

KNN algorithm is not extensively used in sentiment analysis but has shown to produce good results when trained carefully. It operates on the fact that the classification of a test sample will be similar to nearby neighbours. The K value may be selected on any hyper parameter tuning algorithms like Grid search or Randomized search cross validation. The polarity may be hard voted based on K nearest neighbors values, or soft addition may be done to find overall polarity[1].

4.1.8 Semi-supervised learning:

In this case, where the training dataset contains both labelled and unlabelled data, semi supervised learning appears to be a viable option . It is motivated that while gathering unlabelled data is relatively easy in many real-world applications, such as collecting articles from various blogs, labelling is expensive or labour-

consuming because labelling the training dataset is typically performed by humans[1].

4.2 Lexical Based Approach:

Lexicon Based techniques work on an assumption that the collective polarity of a sentence or documents is the sum of polarities of the individual phrases or words. This method is based on emotional research for sentiment analysis dictionaries for each domains. Next, each domain dictionary was replenished with appraisal words of appropriate training collection that have the highest weight, calculated by the method of RF (Relevance Frequency) . The word-modifier (increases or decreases) the weight of the following appraisal word by a certain percentage. Word-negation shifts the weight of the following appraisal word by a certain offset: for positive words to decrease, for negative to increase. The procedure of the text sentiment classification was carried out as follows. First weights of all training texts the classified text is calculated. All the texts are placed into a one dimensional emotional space. The proportion of deletions was determined by the cross-validation method. Then the average weights of training

texts for each sentiment class were found. The classified text was referred to the class which was located closer in the one-dimensional emotional space[1].

2.2 Literature Review

2 2.1 *Introduction*

In the past few years, social media platforms have become the most popular for individuals to share their views and experiences towards different services and products. Therefore, it attracts many researchers to use it as a reference for opinion extraction and sentiment analysis. Most of the previous research studies in this field use machine learning-based and lexical-based approaches to classify the emotional states of English-language tweets. Limited research work has been done to determine opinion trends for tweets in other languages such as Arabic. In addition, machine learning and deep learning approaches have recently achieved remarkable results compared to traditional methods of analyzing a huge amount of data as is the case with social networking data. In the department, we seek to shed light on these researches and studies.

2.2.2 Literature Review

1. In 2016, Afrah Mutawakil and Malak Adel conducted a research for obtaining a bachelor's degree based on the data available from social networking sites (Twitter) The aim of their study is to treat Sudanese tweets and categorize them into (political,

religious, sports, entertainment, other) and Find the most frequent words in Sudanese tweets for the most frequently discussed topics. means where it achieved the best measurement accuracy of 68.6% (for religious opinions 36.41%, political opinions 27.20%, entertainment opinions 13.03%, sports opinions 11.51%, other 11.85%).

2. In 2021, Islam Seif and Al- Samani built a model for analyzing feelings on the Sudanese dialect based on data collected from the social networking sites Facebook. The aim of their study is to evaluate how the processing steps and their library, such as legitimization, affect the model results.

They relied in testing the proposed model on two different classifiers, namely SVM and NAIVE BAYIS NB to classify comments based on their polarity, whether they are positive, negative and neutral. The work was evaluated by four different measures ( ) The results showed that the SVM classifier excelled with the two libraries of legitimization, as it achieved the best measurement accuracy of 68.6%, while it achieved NB

63.1%.

3. In 2017, a research was presented by the Taliban King Abdullah and Mirsad Hadzkadic building a model for analyzing sentiments in Arabic tweets f in order to discover people's feelings and opinions based on data collected from social networking sites (Facebook, Twitter). About 95%.

4. In the year 2020, a research was presented by students at the University of Hassan I from the Moroccan authorities and the University of Moulay Slimane in Beni Mellal to analyze Moroccan feelings based on the means of communication (Twitter) through experiments that were conducted on four methods of classification, as well as the classification of polarity, positive and negative, through algorithms. Different machine learning results showed a similarity in accuracy provided by both classification algorithms, and the highest scores were achieved using the NAIVE BAYES model.

5. In 2019, Rehab Hicham , Joudi Mohieldin and Zitouni Abdelhafid published a paper for the purpose of analyzing opinions on newspaper comments in Algeria with the aim of developing an approach to classify the comments into positive and . negative

NAIVE BAYIS THE best results were obtained if the classifier was used.

6. In 2013, the student Izzat Muhammad Al-Smadi , College of Information Technology , International Islamic University of Science and Education presented a research to analyze sentiments in Arabic and English , based on available data from social networking sites (Twitter , Facebook , YouTube) in order to extract

feelings and determine their polarity (positive-negative-neutral - spam) were based . on testing the proposed model on the SVM classifier

7. A review in 2019 Prof. _ Moez Bin Haj Hamida and Prof. _ Omaima Oueslati at the University of Tunis Al-Manar presented a review of Arabic sentiment research with the aim of detailing problems related to feelings and studying the challenges of the Arabic language scale.They applied supervised machine learning algorithms to the user's social

8. In 2019, a PhD research was presented by student Hoda Jamal at Sudan University of Science and Technology entitled Sentiment Analysis of Arabic tweets written in Sudanese colloquial with the aim of classifying emotions and comparing current machine learning techniques applied to sentiment analysis for data collected fromThe social network Twitter has adopted machine learning algorithms for model testing The results of the first experiment showed The support algorithm that achieved the best accuracy, retrieval and measurement accuracy of the test is 95.1, %76.5

respectively. While the naive BES algorithm achieved the best accuracy of ,%84.4 As the results showed .%85.1

The second experiment showed that the support algorithm that achieved the best accuracy and measurement accuracy of the test, which is equal to 75.2% and 83.9%,

respectively. While the naive BES algorithm achieved the best accuracy, which is It has also been achieved The best return by decision tree algorithm is .%88.41 In addition, the ratios were calculated .%99.9

Percentage of positive and negative opinions about Sudanese government services. It was found that 9.4% represented positive opinions towards Government services, while 90.6% represent negative opinions

the study	***Objectives***	***data set***	***Tools and tec hniques***	***The result***
PhD research entitled Application of sentiment analysis to comments in the Sudanese dialect (1)	***Evaluate how processing steps and libraries*****affect legitimization** ***On the general Sudanese text***	***User comments from Sudani Telecommute Facebook page***	**Machine learning**	***Use with libraries*** **legitimization** ***Improves accuracy of sentiment categorization***
Graduation research entitled Analyzing Opinions from Sudanese Comments (2)	***Classification of Sudanese comments into religious, entertainment, political, sports and other***	***Sudanese tweets on Twitter***	**K means & Hierarchical R language .**	***The percentage of frequency of vocabulary in each classification is given***
Research from the University of North Carolina	***Reviewing Efforts to Build Sentiment Analysis***	***Applications and systems created to***	***The techniques used in each of the systems***	***There is no single system until the date of the study that can deal***

entitled Sentiment Analysis in Arabic Tweets : The Challenges of Language Anatomy (3)	***Systems for Arabic***	***analyze Arabic Twitter data***	***under study are described***	***with all Arabic dialects with high accuracy***
Research titled Twitter data set Moroccan feelings (4)	***Measuring the Twitter data set through experiments on four classification methods***	***Tweets Twitter Moroccan***	***PYTHON SKLEARN***	***accuracy provided by each of the classification algorithms is similar, and the highest score was achieved using a model***
Research published in the Journal of King Saud University (5)	***Developing an approach to classify comments in the Algerian Arab online press into positive and negative***	***Comments in the Algerian electronic press***	***Machine learning***	***NB classifier , the best results were .obtained***
A research entitled Sentiment Analysis in Arabic and English (6)	***Extracting feelings and determining their poles positive,) negative, (neutral***	***Twitter, Facebook, YouTube reddit***	***Machine learning***	***Give the percentages for experimental accuracy to determine the polarity of each suspension***
A review of feelings	***Identify important gaps in reading and***	***Twitter , Yahoo, News Sites, Face book***	***Machine learning***	***Experiments revealed that per formance is hi***

research in Arabic (7)	***writing and suggest future directions***			***ghly dependent on the quality of emotional resources***
PhD research entitled* Sentiment Analysis of Arabic Tweets Written by Sudanese Dialect *(8)	***classifying emotions and comparing current machine learning techniques applied to sentiment analysis for data collected from The social network***	***Twitter data***	***Machine learning***	

References:

1. A survey on sentiment analysis methods, applications,

   and challenges

2. A Review On Sentiment Analysis Methodologies,

   Practices And Applications

3. Sentiment Analysis:A Comparative Study On Different Approaches

4. ng Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012..

5. Sentiment Analysis of Arabic Tweets Written by Sudanese Dialect , Huda Jamal Abdelhameed Musa September 2019

6. 9 July 2011 Social Media Data Mining: A Social Network Analysis Of Tweets

   During The 2010-2011 Australian Floods France Cheong ,Christopher.

7. Abo, M.E.M., et al. Sentiment analysis algorithms: evaluation performance of the Arabic and English language.
8. in 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). 2018. IEEE.

9. Rosenthal, S., N. Farra, and P. Nakov. SemEval-2017 task 4: Sentiment analysis in Twitter. in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017.

10. Kiritchenko, S., X. Zhu, and S.M. Mohammad, Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 2014. 50: p. 723-762.

11. Subjectivity and Sentiment Analysis of Sudanese Dialect, Arabic April 201 , https://www.researchgate.net/publication/332963263_SSA-SDA_Subjectivity_and_Sentiment_Analysis_of_Sudanese_Dialect_Arabic