IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 3, Ser. I (May. –June. 2022), PP 15-25
www.iosrjournals.org

Machine Learning-Based Sentiment Analysis for Tweets Saudi Tourism: A Review and New Tendency

Sarah M ALrashidi, Fatmh N ALanazi, Hanan A ALbalawi, Ohood M ALbalawi, Awad M Awadelkarim
Faculty of Computers and Information Technology, University of Tabuk
P.O. Box 741, Tabuk 71491, Saudi Arabia

Abstract— *Nowadays, no doubt that social media has an elevated influence on our life, thoughts, and decisions. Consequently, tourists share their feelings, opinions, and experiences about the services provided on their travels through such social networks. So, such social media sites are a huge and very influential source of information that affects all aspects, especially the aspect of tourism in terms of reputation, performance, and improving products and services provided by the concerned authority. Further, Sentiment Analysis (SA) is one of the most important tools that help in understanding and analyzing the polarity of textual data. On the other hand, the concern and development of tourism in Saudi Arabia is the key factor that inspired the recent advancement of the tourism industry, as well as the achievement of Saudi Arabia's Vision of 2030. This makes research in this field a priority and a national value, therefore, this research project contributes to such context. In this regard, this paper presents an exhaustive and state-of-the-art review of sentiment analysis that comprises the related approaches, algorithms, techniques, and applications. The paper also reviews the most substantial and relevant research with more emphasis on tourism sentiment analysis. Finally, it focuses on highlighting the new challenges and methodology of Machine Learning-Based Sentiment Analysis for Twitter Saudi Tourism.*

Keywords— *Sentiment Analysis, Machine Learning, Saudi Tourism, Twitter Saudi Tourism.*

Date of Submission: 06-05-2022 Date of Acceptance: 21-05-2022

I. Introduction

Sentiment analysis (SA) can be defined as a type of natural language processing that is used to detect general feelings about certain issues. Sentiment analysis - also called opinion mining - is performed by creating a system for collecting and analyzing opinions about a particular product. This can be collected from blog posts, comments, reviews, or tweets. Sentiment or opinion analysis is the process of using natural language processing, linguistic computations, and text data analysis to identify positive, negative, or neutral feelings in a given text. Sentiment analysis is used in many areas, such as marketing, customer retention, and others. Polarity rating is a critical step in assessing feelings in texts at the document or sentence level. Polarity assessment is concerned with analyzing whether a document or statement expresses a positive, negative, or neutral viewpoint. Feelings can be categorized in two different ways: the first way the writer chooses words to express feelings and the way the reader interprets the written content. Researchers use aggregates of opinions or large groups of aggregated opinions. Among the obstacles facing the analysis of feelings in the Arabic language are grammar, the use of diacritics, the existence of a group of Arabic dialects, and the presence of various forms of words. Artificial intelligence and machine learning, two important components that have a major role in extracting information that carries the feeling of the writer's within the text and determines the writer's attitude, whether it is positive, negative, or natural.

Saudi Arabia can turn into one of the leading tourist destinations in the future. In addition to the historical and heritage treasures and the natural and cultural diversity, Saudi Arabia has developed increasing support for tourism development. This research works on an Arabic text that will put us in many challenges. Extracting features in the Arabic language may be difficult due to the dialect difference in the Arabic language, as the northern part of Saudi Arabia hardly understands some words from the southern region and the western part. Sentiment Analysis is a solid and great field that focuses on immense text analyzing methods instead of analyzing manually. Dealing with sentiment analysis, a group of limitations will be faced, however, no similar studies can be used as a reference in sentiment analysis for Saudi Arabia tourism. There are various sentiment analysis applications in many areas, including decision-making support, business-related application, and prediction and trend analysis.

- *Approaches*

Research directions have evolved recently to cover many application domains. The sentiment classification is defined on seven important sides namely, subjective rating, sentiment analysis, measurements of the user reviews, lexicon generation, extraction of the emotional words and acceptance of the products, detection of fake messages, and other applications in opinion modeling [1]. On the other hand, sentiment identification is divided into four parts: polarity, expressive text ambiguity, multilingual, and sentiment identification across different domains. The lexicon is a set of words of emotion terms with the value of emotion and the value of strength. Lexicon construction begins with a primitive set of terms known as primaries and the set is expanded with synonyms and antonyms for primitive words.

The sentiment analysis applied methods are divided into three parts. Machine-learning, lexical-based, and mixed approaches to self-rating, sentiment identification, utility measure review, and detection of the spam. Furthermore, modes that are based on ontology and non-ontology-based are decided on lexical construction and aspect extraction. Dealing with the implementation of Sentiment analysis you can use the machine learning models as well as lexical approaches. Machine learning gives the best accuracy while semantic routing gives a good result in the overall work. there are two main parts of Machine learning: supervised learning (classification) and unsupervised learning (clustering). For classification models, we must have training data and testing data. Some of the most used classification algorithms for supervised learning are Decision Tree (DT), SVM, Neural Network (NN), Naive Bayes, and Maximum Entropy (ME). Unexplored angles of the many patterns will affect accuracy. Sentiment analysis across languages is conducted victimization two completely not similar approaches: (a) the lexicography-based approach, during which a private book of the target language is formed by translating an already used lexicon into another language; (b) the Text set-based approach, during which a self-annotated set of the needed language is constructed through projection, and an applied mathematics classifier is trained on the ensuing set. Most Classifications were done using hybrid methods of back-propagation neural network and semantic routing [1]. There are four types of semantic orientation indexes (SO) as input neurons. SO-PMI (AND), SO-PMI (NEAR), semantic correlation, and latent semantic analysis.

Dealing with the semantic orientation of the entire text is assumed to be equal to the total of the individual semantic orientation of words and phrases in a lexicon-based approach [2][3]. On the other hand, and dialing with sentiment analysis there are three levels you can use: document, sentence, and attribute.

1. Document Level: all of the document is taken and is classified as positive data or negative data.
2. Sentence Level: the whole document is parsed into a sentence and is classified into positive or negative sentences
3. Word or Phrase Level: Product attributes or elements area unit analyzed.

Based on work done in [4] the Authors Extracted the main things that link the words of a document to its categories in their sentiment classification to build the text classifier and model the documents of text into a collection of transactions, any transaction will represent a document of a text and the transaction items is represent the selected terms from the text and the classes of the document. They used data obtained from different positions that commonly need to be pre-processed before full analysis can begin. Some of the most familiar pre-processing steps are filtering, encoding, stop word elimination, tokenization, parts of speech tagging, extraction of the feature, and preparation. Cryptography is used to divide a sentence into meaningful words, phrases, symbols, or other symbols by removing unneeded words.

B. Algorithms and Techniques

Regarding sentiment analysis algorithms and techniques, the machine learning approach includes both supervised and unsupervised learning techniques. Examples of the supervised techniques are Support vector machine, Maximum Entropy, and Naive Bayes. In contrast, the unsupervised techniques comprise Exploit sentiment lexicons, Grammatical analysis, and synthetic patterns.

Naive Bayes and Support Vector Machine (SVM) methods are used to measure the accuracy and precision of the data supervised learning methods. Supervised learning is more reliable in sentiment analysis. In sentiments, analysis texts may be objective facts that differ from personal opinions, and it is essential to distinguish between them. In both supervised and unsupervised techniques, sentiment analysis involves subject-level analysis to identify and differentiate objective and subjective opinions, which is the primary task in sentiment analysis. To classify sentences, whether they are objective or opinion, SVM is often used. In lexicon approaches, the data can be processed linguistically. By extracting features:

1. N-gram Feature: Specifies n number of concatenated words as a feature extraction group, and can take one word at a time (unigram) or two words (bigram).[5]
2. Stemming: It removes prefixes and suffixes and takes the root of speech, which improves indexing and may reduce its accuracy as well.

3. Stop word removal: Remove stop words such as pronouns and prepositions because they do not provide any information.[6]
4. Conjunction handling: Conjugation words like “and” but “or” may change the whole connotation of the sentence so it is important to address them.
5. POS tagging: The "of " is used to specify the part of speech to derive nouns, pronouns, adverbs, etc.

Machine learning approaches and lexicon-based approaches are used to classify polarity in sentiment analysis, the strength of feelings, and the degree of feelings. Fig. 1 illustrates sections of the lexicon-based approach.

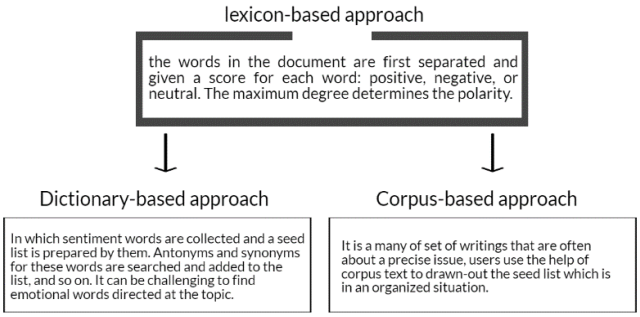

Fig. 1. Sections of lexicon-based approach

Fig. 2. below illustrates sections of the machine-learning approach and algorithms that follow each section.

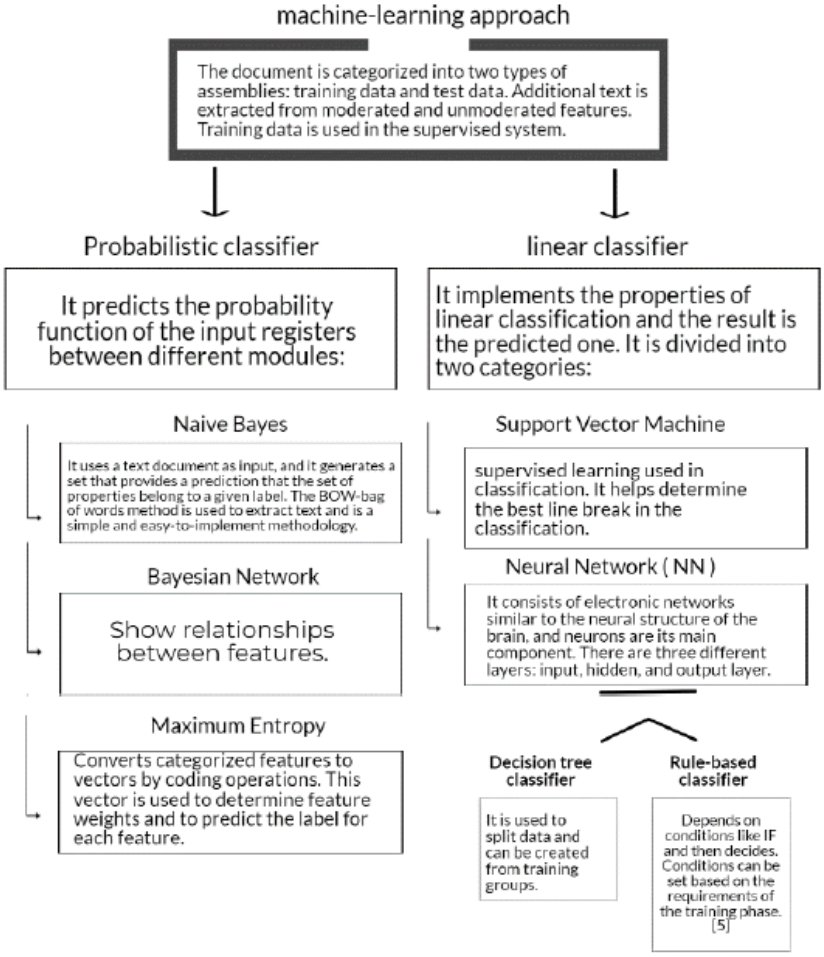

Fig. 2. Sections of Machine-learning Approach and Algorithms.

C. Applications

There are many applications of sentiment analysis in many fields, including but not limited to [5][6][7]

1. Decision-making support: Creating websites helps in the decision-making process in daily life, such as choosing the next trip, tourist country, the airline, place of residence, etc.
2. Business-related application: By evaluating places, companies can improve the quality of tourist places and achieve tourist satisfaction, thus improving services.
3. Prediction and trend analysis: Users can predict the trends of the trading market through a survey.
4. SPAM detection.

Related work

This section reviews the most substantial and relevant research. The study in [8] reported that much of the current research in sentiment classification targets the large volume of rich online opinion sources such as discussion forums, review sites, social media posts and blogs, and accessible news organizations. In the digital genre due to the large volume of opinions rich in Internet resources. Individuals are expected to create a system that acknowledges and identifies feelings or opinions expressed in textual data. We may extract data from the Internet and predict that online shoppers' preferences are the correct sentiment prediction technique, which can be useful for economic studies or to reinforce studies. So far, this analysis community has focused on two main problems, addition to sentiment classification, feature-based classification, and negation processing. Other research provides a summary of sentiment analysis methodologies and approaches, as well as issues that have arisen in the field. Sentiment ratings are highly dependent on areas or topics. Because various types of features have different distributions, no categorization model consistently beats the other. Different types of variables and rating models are also discovered to be integrated in an efficient way to overcome their drawbacks and make use of each other's benefits, resulting in improved sentiment rating performance. The work hypothesized that human emotions can be categorized into specific primary emotions; As a result, people prefer to group material on halal tourism into the NRC Dictionary's eight emotion categories, and the research found that sentiment analysis is a branch of linguistics that detects the thoughts, sentiments, and moods represented in a text.

The system proposed in [9] showed that the process of categorizing, identifying, and measuring opinions about anything is known as opinion mining (OM) or sentiment analysis (SA). It's a type of Natural Language Processing (NLP) used to track public opinion about a particular law, policy, or marketing campaign, for example. It entails creating a method for collecting and analyzing comments and opinions on legislation, laws, policies, and other topics shared on social media. The process of extracting information is very important because it is a useful method and a difficult task. This means that to extract emotion from an object at the web level, opinion mining techniques must be automated. Machine learning (supervised and unsupervised) and lexical approaches are two well-established methodologies for sentiment analysis. As a result, the researchers' primary goal was to introduce sentiment analysis (SA) survey and opinion mining (OM) methodologies, as well as the many techniques used in this topic. It also examines areas of application and obstacles to sentiment analysis, as well as provides insight into previous studies. The process of extracting information is very important because it is a valuable technique and a difficult task. For this reason, the primary purpose of the paper was to present a study of Belief Analysis (SA) and Methods of Perspective Excavation (OM); Various strategies have been used in this area. It also discusses the areas of application as well as the obstacles to analyzing the presentation with a proper understanding of the scholars' previous work.

As discussed in [10] the field of opinion mining and analysis of the sentiment is fast expanding. On the internet, various e-commerce sites allow people to leave feedback on individual products. These comments are extremely beneficial to both individuals who are interested in purchasing the goods and organizations. We may be able to collect opinions of users from the internet and forecast client preferences if we develop a reliable approach for predicting sentiments. Opinion mining algorithms come in a variety of forms. Pre-processing on feedback is done before using any polarity detecting method. Opinion words and objects on which opinion is created are retrieved from these pre-processed reviews, and any opinion mining approach is used to determine the review polarity. Opinion mining may be broken down into three levels of granularity: level of the document, level of the sentence, and aspect level. Various methods for sentiment analysis are examined in this research, and obstacles and applications in this subject are highlighted. Sentiment analysis has come to be a very popular field of study. Sentiment discovery has a wide array of applications in info systems, consisting of classifying testimonials, summing up reviews, and various other actual-time applications. The thesaurus-based approach takes less handling time than the supervised knowing strategy; however, precision is unqualified from the mark. The monitored learning method supplies much better accuracy. This study can wrap up those supervised techniques that provide much better accuracy than the thesaurus-based method. The approaches that depend on Lexicon consist of wordbook-based and corpus-based. From the different purposes of reading, the recent work on Sentiment classification may be characterized by the technique used, rating level, read of the text, and level

of elaborate text analysis. Machine learning, lexicon-based, applied math primarily based, and rule-based strategies are known for the technical purpose of reading.

Based on [11], The current tourist destination in Indonesia in Yogyakarta has grown significantly due to the technological role, which has a significant impact on the ability to access data. Twitter may also contain information or details about tourist sites that they need. The study was carried out by classifying opinions presented in the form of comments into two groups: positive and negative, with the amount of accuracy being influenced by the training method.

Moreover, the use of Sentiment Analysis is found to be a strong and great field that principally focuses on immense text analyzing methods instead of analyzing manually. The technology of Opinion Mining is widely used for collecting social information and doing the business method productively. The study in [12] applied sentiment analysis to track people's feedback concerning Oman business by using the textual data of social media. For this, they used Twitter textual data for tracking and maintaining tourist opinions concerning this country. also, they suggest an innovative sentiment analysis model supported wisdom knowledge (Domain Specific Ontology). they created a selected Oman business ontology supported by Concept Net. POS tagger sed as a feature selection to select Entities from tweets and entities are compared with ideas using the domain-specific ontology. Also, the sentiment of the selected entities is set by the approach of a merged sentiment lexicon. Lastly, linguistics orientations of domain-specific options are merged regarding the domain. Amongst the surveyed approaches, the Analyzers of the sentiment are dependent on language. There is no existing technique found that it's additional general and appropriate to be language-dependent. withal, the interest in languages apart from English within the field of sentiment analysis or opinion mining is obtaining attention, as there's still a dearth of research and on the market resources in alternative languages.

The study based on [13] states that when dealing with emotion analysis, there will be many difficulties. Research work has indicated that the development of emotion classification algorithms or opinion exploration remains an open analysis space. In addition, Naive Bayes and Support Vector Machines (SVM) are common methods used in supervised learning curricula to mine opinions or analyze emotions. Most approaches used in the classification process are Machine Learning Approaches and Lexicon-based Approaches. The Machine Learning models belong to supervised learning and classification of text especially. So, it's known as—Supervised Machine Learning. It is composed of many methodologies like Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighborhood, Maximum Entropy, and Neural Networks.

Several lexicons are on the market to produce sentiment analysis, as well as the final work on [14], the Sent WordNet, the LIWC dictionary, the lexicon of sound judgment Clues, the Q-WordNet, and the Sentiment-based Lexicon. In this work, they choose to utilize the NRC lexicon since it has been proven to work in comparable studies. Moreover, the NRC lexicon, which was constructed manually through crowdsourcing, is capable of classifying emotions into eight primary categories: trust, anticipation, surprise, anger, joy, fear, sorrow, and disgust, among others. The method of recognizing and choosing "subjective data from enormous amounts of textual data by merging data processing techniques, machine learning, linguistic communication process, data retrieval, and data management" is known as sentiment analysis. Tourists' large information and data mining technology are established at an unmatched rate. In the significantly fierce tourist market, it will obtain more valuable details as well as even more market opportunities with modern tourism data technology. Tourist huge data makes the allocation of tourist resources a lot more practical, finds possible customers and provides better-individualized solutions to the existing clients. With the growth of huge tourist information, much more information analysis innovations will be made use of in the tourist market. To use the huge data innovation in the tourist sector, the authors presented the principle of large information and also the development demand of tourism big information, summarizes the standard modern technology of data mining and also the mining innovation of huge tourist information, and also lastly gives the application instructions of data mining in tourist.

Based on studies in [15] found that sentiment analysis helps to select "subjective data from massive volumes of textual data by merging machine learning, data processing techniques, linguistic communication process, collecting of the data and data management." Sentiment analysis defines texts as positive, negative, or neutral. it's also delineated by the range between 1 (clearly positive) and -1 (clearly negative).

The study based on [16] aimed to explore the sentiments of tourists about the travel resort of Cancun that a Mexican city through Twitter. But instead of evaluating sentiments on a one-dimensional scale as either positive or negative feelings, they analyzed the feelings on a larger area so that they are multidimensional and visualized them based on the passage of time through a type of artificial neural network: self-organizing maps (SOM) through which they discovered peaks and valleys in the visualization. Kohnen's map specifically. Used a neural network to model the sentiment that expresses travel, using a previous methodology developed by Minqing Hu. Used Twitter API to collect 70,570,800 tweets with an area of 20,042 GB, over some time, from Oct. 30, 2009, to May 21, 2010, the data carries a lot of information, however, only used the text comment and the date. Then they filtered it so that tweets containing the word "Cancun" only. Used a hybrid of a keyword-

based algorithm to measure sentiment, which in turn used a binary choice algorithm that was designed according to a main standard Naive Bayes algorithm and tested through that main standard measurement, and they Reach a correlation coefficient of R = 0.63, and they were able to visualize feelings during 200 days. In the preprocessing stage, the emotional words were collected and subjective words that express feelings of travel from the web and prepared a list of 87 travel words that use these words as a dictionary through the vector space model. After the pre-processing stage, using a software package called Viscovery SOMine for visual ensemble analysis, statistical data extraction, segmentation, classification, etc. Based on SOM. The maximum and minimum relative opinions were extracted during the specified periods and visualized through multi-dimensional self-organizing maps to provide a greater understanding of the data for companies and travel destination researchers as well.

Based on the study in [17], which aims to introduce data mining and the most important stages that this technique passes through. The process of data mining passes through three main stages: The first stage is pre-modeling, in this stage, the work difficulty is classified and then the work difficulty is transferred through data mining applications, then the resulting data is evaluated and data mining is prepared by providing the necessary data and dealing with its problems and repairing them. Then is the modeling stage, in which data mining tools and techniques are used in data analysis. Evaluate models and determine the final model. Finally, post modeling. Concerning the decisions that will be taken after the finish of the data analysis. One of the most important data mining tools is visualization, which helps in summarizing the data to understand it and present the results. Then there is predictive modeling in which a neural network is used to understand the complex relationships of data. And the use of decision trees to divide the data into subgroups that are easy to interpret and use. Through data mining, visitors can be categorized and their preferences through recognition trees, which is the best option for understanding these preferences.

The study [18] aims to propose an application that can analyze the sentiments of tourists from ratings and selection of destinations in Indonesia. The data was collected by TripAdvisor from 10 reviews of travel destinations in Indonesia, which were in Indonesian. The text was categorized into positive emotions and negative emotions by manual tagging to be used as training data in SVM. In the preprocessing stage, the raw text data was prepared and converted into a basic text form and the noise was removed. Features were also extracted such as derivation, encoding, and removal of stop words. Also, (TF-IDF): Inverse Document Frequency term weighting and n-gram, unigram were used to extract the features. SVM algorithm analyzed the data by classifying it into two parts (positive and negative) and giving positive emotions symbol "1" and negative "-1". The software used uses PHP and the Libsvm library, which supports SVM functions such as training and compilation. The data entered for the supporting vector machine is in the form of a matrix. The prediction of negative and positive emotions was achieved with an accuracy of 85%. The positive emotion recall value is 80% while the negative is 100%. They need to improve training data to improve accuracy. It is also preferable if they categorize the reviews into several categories according to the needs of the travel such as hospitality, weather, etc.

The goal [19] is to gather data on the feelings of international tourists in Bangkok to develop and promote tourism in the city. Rapid Minder version 7.4 was used to collect information from Twitter and retrieved 10,000 random English language tweets about the sentiments of foreign tourists in Bangkok. The purposes of tourism were classified into five categories: travel, business, visiting relatives, education, and health. Places inside Bangkok are also categorized into nature, country culture, temples, historical sites, local cuisine, nightlife activities, and shopping. In the pre-processing stage, the text was analyzed as hidden shapes, patterns, and trends in the native text based on statistics and mathematics. Natural language processing (NLP) was used to reduce unnecessary text while retaining the important points of the original document. Then classify the document and categories to simplify the additional process in terms of management and sorting. Documents with similar contents are grouped for quick extraction and retrieval of information and filtering. They classified the emotions into positive and negative and were represented by mathematical models giving positive emotions number “1” and negative feelings number “0”. Four machine learning techniques were focused on: ANN, which achieved the highest accuracy of 80.32 %, then SVM, with an accuracy of 80.11% Then the decision tree with an accuracy of 83.79%, and finally Naive Bayes with an accuracy of 55.66%.

Based on the study in [20]. which aims to propose a system that reduces the time spent by readers of travel blogs and user reviews by summarizing and visualizing the information intended by the authors. So that information can be obtained easily and in less time. Tourist reviews were collected from Ctrip and TripAdvisor, while tourist blogs were collected from Ctrip, MaFengow, and Tuna. In the preprocessing stage, a special morpheme analyzer for Chinese known as THULAC: THU Lexical Analyzer for Chinese, with accuracy reaches 92.9% with the parts of speech of the Chinese language. The expanded property dictionary was created by collecting words that correspond to each property manually and creating a property dictionary. Then collect synonyms from the Yaudao dictionary and related words from the Baida dictionary. By morpheme analyzer, only names are extracted from categories and properties from other categories such as “price” because it does not contain many nouns but adjectives to describe price in language Chinese and then compare these names to the

dictionary and if they exist, they are classified as corresponding property. After classifying the categories and classifying the words by categories, sensitivity was analyzed using the Random Forest algorithm which contains many trees for decision making. Where the decision is taken according to the majority, if the decision tree shows more positive results than negative, the result is considered positive. The Random Sensitivity Analyzer was evaluated with a performance score of 0.82. Using the time-series graph, the number of revisions was visualized and the data were represented in them by time-series, Radial, and Network graphs. The system achieved less time understanding blog content and reviews with an average of 91.6% through a survey with 10 users.

The study based in [21] aims to exploit machine learning techniques to benefit from them in data analysis to make various changes in the field of industry in the future. ML in tourism has three phases: Before the travel: the next destination can be predicted. It uses algorithms such as neuro-fuzzy and discrete hidden Markov NN - based econometric model, neural network, modeling incorporating neural networks to gray - Markov models and NN enhanced hidden Markov model. During the trip: The most important tools for this stage are the recommendation system using the mobile phone recommendation system. Many machine learning algorithms are used to make a more accurate model. Hence a more accurate recommendation. After the trip: It is possible to analyze the feelings and opinions of tourists towards the places of visit using machine learning methods as well. At this stage, data is collected through internet networks and social networking sites such as Twitter and then analyzed through written content.

In Big Information Era, Statistical Tourism Observatory requires to be revised. Authors in [22] presented a conceptual version of the Digital Tourism System (DTS) where different types of conventional and non-standard data can be refined by stars and also spectators in the tourist market. Especially, big information can be beneficial as well as the number of Data researchers within the tourism market becomes popular. DTS enabled to stress four knowledge areas of interest for various functions, particularly location management, research study and innovation, market analysis, and labor market to enhance tourism management and research. Trick actions of the understanding discovery pyramid were manipulated to offer an included value in decision-making based on analytical understanding techniques. Two examples were revealed, mining online textual as well as photo information respectively.

Point Of View Mining (OM) or Sentiment Evaluation (SA) can be specified as the task of identifying, drawing out, and classifying viewpoints on something. It is a type of natural language processing (NLP) to track the public mood to a certain regulation, plan, advertising, etc. The advancement of social networks has contributed exceptionally to these tasks, thus offering us a transparent platform to share views throughout the globe. These Electronic Word of Mouth (EWOM) statements shared online are much more common in the business and solution market to make it possible for the customer to share their point of view. Hereof, the authors in [1] provide a rigorous survey on belief evaluation, which depicts ideas presented by over one hundred write-ups released in the last years relating to needed tasks, approaches, as well as applications of sentiment evaluation.

The research explored halal tourism tweets from social media and using sentiment analysis, the authors found that there were more additional positive feelings than negative feelings among the tweets collected. The results showed that halal tourism can be a global market and not only for Muslim countries. Hence, business players should seize the opportunity to use social media to their advantage to market their halal business venture packages as it is an effective methodology for communication during this decade. Based on studies in [2], opinion mining and sentiment analysis have a wide range of applications. In addition, there are many challenges to research with a focus on this area. Therefore, this has been an active research space in recent years. It is not possible to consider one model a higher rating than the other. No model is systematically superior to the other. Different classification strategies will be used in combination with each other to overcome the disadvantages of each. Efficiently combining different algorithms improves sentiment rating performance. Based on the described studies the dictionary-based approach takes less time to process than the supervised ML approach but the accuracy is not the best. The use of a supervised ML approach provided the best accuracy. From this survey, it can be concluded that supervised techniques provide higher accuracy compared to primarily dictionary-based approaches.

It is located that belief classifiers are severely dependent on domain names or topics. it is evident that neither classification model continually exceeds the other; various functions have distinctive distributions. It is also discovered that different kinds of functions and category algorithms are integrated into a reliable method to overcome their private disadvantages, benefit from each other's values, and finally improve the sentiment category efficiency [4].

II. Problem Motivation and New Tendency

Saudi Arabia can turn into one of the leading tourist destinations in the future, in recent years has witnessed unprecedented development in tourism in several regions. Which is one of the emerging sectors and represents one of the axes of Vision 2030. The land of Saudi Arabia is the cradle of the Islamic religion, with the presence of the "ALHARAM ALMAKKY" and "ALHARAM ALNABWY," which is the destination for

millions of Muslims annually, so Saudi Arabia will be the first Arab Islamic destination preferred by Muslim tourists. In addition to the historical and heritage treasures and the natural and cultural diversity, as a result of which Saudi Arabia has developed increasing support for tourism development. Therefore, it was important to measure the extent of tourists' satisfaction, feelings, and reactions toward these recent developments. With the advancement of information technology and mobile computing, many innovative concepts are enforced to facilitate improved expertise within the tourist's journey. This analysis proposes a unique plan in providing details regarding travel spots in a summarized manner to tourists by following the sentiment of Saudi Arabia's and non-Saudi tourists from their textual data on social media. There is a lack of research on the position of Saudi Arabia's tourism sentiment analysis. Hence, we propose a model that can track and improve this area. Travelers use social media platforms not only to find places that meet their desires and expectations but also to share their experiences and reviews about how they planned their trip, where they purchased their tickets, the places they visited, the buildings they stayed in, and a variety of other information on these websites and blogs. Through these apps, anyone may offer their real-time thoughts on a topic, service, or location. Among these applications, Twitter is the most used text-based social media application. Proceeding from the saying that the first method of solving the problem is to identify it first. We had to use the text data of tourists from their accounts on Twitter, which gives us a lot of information about the percentage of tourists' satisfaction with the services provided by Saudi Arabia. Data management in a rational touristy necessitates analyzing the content of social media applications. One of the most effective ways to provide feedback on a service or product is to use sentiment analysis. The material in sentiment analysis is usually separated into positive and negative categories.

In this research, Twitter is used as a data source for textual data. Tweeter data helps to grasp the tourist's behavior patterns in Saudi Arabia to transform their methods and assist travelers in meeting their needs. The proposed model targets the textual data of the Saudi Arabia tourist to identify and track the tourist sentiment, improve tourist satisfaction by eliminating any problems faced by them, solving the issues of the hospitality and tourism sector affected by social media, and measuring the fulfillment of the tourists towards visited places and the proposed services in Saudi Arabia. In addition to supporting and developing new, unprecedented ways in Saudi Arabia to support tourism that depend on sentiment analysis.

A. Problem Motivation

As stated by the above-mentioned intensive critical literature review and according to the best of our knowledge, there is no single similar work in the sentiment analysis of Saudi Arabia tourism Twitter data up to date. On the other hand, there are several limitations when dealing with sentiment analysis: The primary one is an opinion word that is thought to be positive in one scenario and could be thought of as negative in another scenario. A second challenge is that people do not invariably have categorical opinions during the same approach. People make contradictory remarks. Most reviews contain both good and negative remarks, which can be managed by examining phrases one at a time. However, some may struggle to grasp what someone thought supported a little amount of text due to a lack of context. Conditional sentences are in Sentiment mining is also making identical problems like interrogative sentences. Spam sentiments are those sentiments that are denoted by the opposite or rival organization for increasing their product worth or their organization worth among the users. It is a difficult task to figure out which objects are relevant and features on which the opinion is made. A lot of times the same feature of a place is addressed by using different words. The same sentence can be considered positive in one situation and negative in the other. It's even harder to build a machine system which understands can differentiate the sarcastic comments from the non-sarcastic ones.

B. Limitations of Arabic sentiment analysis on Saudi Arabia tourism

Sentiment analysis is increasingly being performed in the literature, specifically as the production of social information increases. However, the majority of research study assignments reported in the disquieting literature of presentation analysis relate to information generated in English more than in any other language due to the lack of similar works in this situation. There are no similar studies that can be used as a reference in sentiment analysis for Saudi Arabia's tourism. This research will work on an Arabic text that will put us in many challenges. Extracting features in the Arabic language may be difficult in some cases, due to the dialect difference in the Arabic language, as the northern part of Saudi Arabia hardly understands some words from the southern part, as well as the western part. Likewise, all parts of Saudi Arabia have words that cannot be understood in other regions. Additionally, it will have a big impact, not just relying on classical Arabic analysis, and there will be more data and different opinions from across Saudi Arabia. All of the above puts us in a challenge and reinforces the importance of the work presented in this research. This introduces a requirement to give additional attention to Arabic sentiment analysis research. Previous analysis within this field is restricted and mostly focused on Standard Arabic.

C. Goals and Expected Contributions

- Contribute to the development of the Kingdom's 2030 vision by the development of tourism which can contribute to the progress and prosperity of countries.
- Works on one of the foremost standard microblogs known as Twitter could be a good way to figure out how individuals feel concerning the popular travel websites using Twitter information.
- Measure the satisfaction of the tourists towards visited places and the proposed services in Saudi Arabia.
- Supporting and developing new, unprecedented ways in Saudi Arabia to support tourism that depend on sentiment analysis.

The proposed research works on one of the foremost standard microblogs known as Twitter which could be a good way to figure out how individuals feel concerning the popular travel websites using Twitter information. To do so, we use the Twitter API to stream tweets, filter relevant tweets using a search query, do sentiment analysis on Twitter textual data from Saudi Arabia tourism to determine how people view each travel site and save the tweets for further study. This study continues to be within the initial stages in terms of touristy business in Saudi Arabia.

Hospitality and also the touristy sector are greatly stricken by social media and have generated tons of interest among researchers. Social media has utterly evolved during an approach that which folks analyze travel destinations on travel websites and read many reviews before making selections. They talk over with immense quantity of information obtainable through the type of reviews, images, videos, small blogs, and honest travel experiences shared by fellow travelers. This word of mouth has become the foremost necessary type of selling as a result of folks trusting these opinions and reviewing quite the complete promotional material. The value of user-generated content outweighs the value of comprehensive data. this will facilitate in making ''a virtual online community of individuals having the same language and same interest wherever they'll share their complete experiences and supply valuable data. And this fashion, folks can feel connected though they need not see one another or meet. Hence the proposed system targets the textual data of the Saudi Arabia tourist to:

- Identify and track the tourist sentiment.
- Improving tourist satisfaction by eliminating any problem faced by them.
- Solving the problems of the hospitality and tourism sector affected by social media.
- Measure the satisfaction of the tourists towards visited places and the proposed services in Saudi Arabia.
- Supporting and developing new, unprecedented ways in Saudi Arabia to support tourism that depend on sentiment analysis.

Based on all of the above, we can contribute to the development of the Kingdom's 2030 vision, as the preservation and development of tourism greatly contribute to the progress and prosperity of countries.

D. The Gathered Data

The practice of accessing the internet and searching for information about travel destinations, hotels, places to visit, and restaurants, as well as making reservations, has become extremely prevalent in the travel and tourism industry. People prefer to share their experiences, reviews, and images on social media, which plays a significant part not only in the planning of the trip but also during the journey. Following their vacation, individuals frequently submit evaluations on well-known websites and social media platforms. Travel companies should respond to these evaluations and seek to retain current consumers methodically. If unfavorable reviews and comments exist, businesses should address them by contacting dissatisfied consumers to protect their brand's reputation. The proposed research will use Twitter as a data source for the textual data. Tweeter data helps to grasp the behavior pattern of the tourist serving Saudi Arabia to transform their methods and assist the travelers to meet their needs. A study is finished on the Twitter platform and folks who travel a lot or are additionally concerned with touristy services have a high frequency of posting. This data has additional credibility such as any famous person has its popularity. Twitter user selections are stricken by supply information, frequency, and dependability. Saudi Arabia can use a skilled person who will often post positive reviews, and suggestions and respond to visitors' queries and feedback. this can facilitate in observance of complete names on social media and negative postings are handled to extend tourism credibility.

Social media platforms enable nations to immediately communicate with and interact with thousands of tourists. Through these social media channels, travelers may communicate and share their experiences, views, and encounters with a variety of services and locations. These social media sites provide comprehensive information about tourists, including their habits, views of their interests, likes, and dislikes. The proposed system collects the textual data of Saudi Arabia tourists from their accounts on Twitter. The collected data contain the plain text of the Tweets and the replies on it. All the collected tweets need to be cleaned first in the preprocessing stage of our work. the proposed system used many filter procedures to scrap the more related data

to the proposed system. One of the filtering procedures is to search in a specific tag and select the tweets that are related to Saudi Arabia tourists.

E. Data Collection Mechanism

This research uses a Twitter API to collect the textual data from Twitter using the Tweepy library. API stands for Application Programming Interface, which is an interface that allows anybody to access a server's core functionality and simplifies the interaction of computer programs with online services. many net services offer APIs to assist users to move with their services and access knowledge in a programmatic approach. users build an invitation for a particular form of knowledge —filtered by keyword, usernames, locations, named places, etc. and because the tweets keep matching the keywords, they're pushed on to the user. to start the method, they produce an app to start this method, an app is created to access the Twitter API so we can scan and write Tweets from the app that exploit keys and tokens. a python library that hooks up with the Twitter API is used to access the Streaming API to alter the interaction between Python and Twitter's API. one of all these python libraries is Tweepy which helps python to move with Twitter and use its API will be collected the tweets from Twitter API, will be mine these tweets by querying the Search API, therefore, determining necessary data regarding users' opinions and emotions from unstructured text. Then Will be performed sentiments analysis on collected tweets from Twitter employing a Python library for process matter knowledge that provides a straightforward API to hold out simple language process (NLP). However, fetching tweets with Tweepy has some constraints based upon account tier and a limitation of the number of tweets. Twint is an open-source Python collection that is developed to accumulate tweets. It is a free device that can be established conveniently with no demand for a Twitter account or the setting up of a Twitter API and it can collect almost all readily available tweets and allow to scrape a user's followers. following, and more while evading most API limitations. As is the case with Tweepy, Twint can gather data based on search phrases, hashtags, dates, as well as location. this study benefits from Twint and Tweepy tools together.

F. Methodology Overview

This section exhibits an overview of the proposed methodology; thus, Fig. 3 below displays the phases of such a proposed approach.

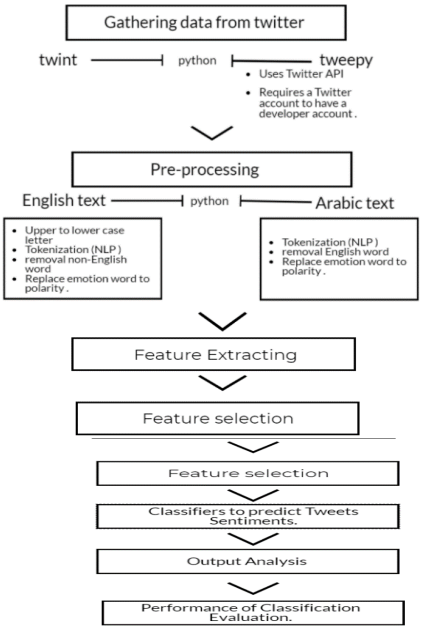

Fig. 3. The Proposed Methodology

The proposed approach consisted of seven phases, the first stage is gathering data, Gathering the data according to the proposed mechanism in section E. Data Collection Mechanism. In the second stage The Tweets dataset will be preprocessed for English text and Arabic text to remove symbols, Stemming and Lemmatization, and so on. In the third stage, features will be extracted to choose useful words from the tweets. In the fourth stage, features will be selected to choose the appropriate feature to identify the characteristics and increase the accuracy of classification. In the fifth stage, a variety of classifiers may be used to predict the emotion of a tweet, which might be (positive or negative). Then in the sixth stage, the output will be analyzed. and in the final stage, some of the common techniques will be used to represent works to evaluate the automated classification method as described in Fig. 3.

III. Conclusion

Sentiment Analysis is a solid and great field that focuses on immense text analyzing methods instead of analyzing manually. In this regard, this paper presented an exhaustive and state-of-the-art review on sentiment analysis that comprises the related approaches, algorithms, techniques, and applications. The paper also reviewed the most substantial and relevant research with more emphasis on tourism sentiment analysis. Finally, it focused on highlighting the new challenges and methodology of Machine Learning-Based Sentiment Analysis for Twitter Saudi Tourism. Conclusively, this research study can contribute to the Kingdom's 2030 vision attainment and development, as the preservation and development of tourism significantly contribute to the progress and prosperity of countries.

References

[1]. Kumar Ravi, Vadlamani Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowledge-Based Systems, Volume 89, 2015.
[2]. Raisa Varghese and M Jayasree, "A Survey On Sentiment Analysis and Opinion Mining", International Journal of Research in Engineering
[3]. Hemamalini, Dr.S LITERATURE REVIEW ON SENTIMENT ANALYSIS, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 04, APRIL 2020.
[4]. Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. International Journal, 2(6), 282-292.
[5]. S. Pandya and P. Mehta, "A Review On Sentiment Analysis Methodologies, Practices And Applications", 2020.
[6]. "LITERATURE REVIEW ON SENTIMENT ANALYSIS", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 9, no. 04, 2020.
[7]. S. Mukherjee and P. Bhattacharyya, "Sentiment Analysis: A Literature Survey", 2013.
[8]. Hajmohammadi, M., Ibrahim, R. and Ali Othman, Z., 2012. Opinion Mining and Sentiment Analysis: A Survey. INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY, 2(3), pp.171-178.
[9]. Saad, S. and Saberi, B., 2017. Sentiment Analysis or Opinion Mining: A Review. International Journal on Advanced Science, Engineering and Information Technology, 7(5), p.1660.
[10]. M., V., Vala, J. and Balani, P., 2016. A Survey on Sentiment Analysis Algorithms for Opinion Mining. International Journal of Computer Applications, 133(9), pp.7-11.
[11]. Hermanto, D., Ziaurrahman, M., Bianto, M. and Setyanto, A., 2018. Twitter Social Media Sentiment Analysis in Tourist Destinations Using Algorithms Naive Bayes Classifier. Journal of Physics: Conference Series, 1140, p.012037.
[12]. V. Ramanathan and T. Meyyappan, "Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism," 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), 2019, pp. 1-5, DOI: 10.1109/ICBDSC.2019.8645596.
[13]. Adl, A., & Elfergany, A. K. (2020). Tracking How a Change in a Telecom Service Affects Its Customers Using Sentiment Analysis and Personality Insight. International Journal of Service Science, Management, Engineering, and Technology (IJSSMET), 11(3), 33-46. http://doi.org/10.4018/IJSSMET.2020070103.
[14]. Wenjie Xiao and Changguo Xiang, "Overview of Tourism Data Mining in Big Data Environment, " 7th International Conference on Education, Management, Computer and Medicine (EMCM 2016), vol. 59, 2017.
[15]. Feizollah, A., Mostafa, M.M., Sulaiman, A. et al. Exploring halal tourism tweets on social media. J Big Data 8, 72 (2021). https://doi.org/10.1186/s40537-021-00463-5
[16]. H. Dinh, "Naive Bayes and Unsupervised Artificial Neural Nets for Caneun Tourism Social Media Data Analysis", 2010.
[17]. M. Karathiya "DATA MINING FOR TRAVELS AND TOURISM", 2012.
[18]. I. Pertiwi Windasari and D. Eridani "Sentiment Analysis on Travel Destination in Indonesia", 2017.
[19]. T. Kuhamanee, N. Talmongkol, K. Chaisuriyakul, W. San-Um, N. Pongpisuttinun, and S. pongyupinpanich, "Sentiment Analysis of Foreign Tourists to Bangkok using Data Mining through Online Social Network", 2017.
[20]. Y. Hyeon Gu, S. Joon Yoo, Z. Jiang, Y.Jin Lee, Z. Piao, Helin Yin, and Seogbong "Sentiment Analysis and Visualization of Chinese Tourism Blogs and Reviews", 2018.
[21]. Y. A. Al-mulla "Machine Learning in Tourism",2020.
[22]. Carmela Iorio, Giuseppe Pandolfo, Antonio D'Ambrosio and Roberta Siciliano, "Mining big data in tourism," Quality & Quantity, vol. 54, no. 5-6, pp. 1655-1669, 2019.

Sarah M ALrashidi, et. al. "Machine Learning-Based Sentiment Analysis for Tweets Saudi Tourism: A Review and New Tendency." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 24(3), 2022, pp. 15-25.