

Reputation Measurement based on a Hybrid Sentiment Analysis Approach for Saudi Telecom Companies

Bayan Abdullah¹, Nouf Alosaimi², Sultan Almotiri³
College of Computer and Information Systems
Umm Al-Qura Univeristy
Makkah, Saudi Arabia

Abstract—Thousands of active people on social media daily share their thoughts and opinions about different subjects and different issues. Many social media platforms used to express the feeling or opinion and at top of it is Twitter. On Twitter, many opinions are expressed in many fields such as movies, events, products, and services; this data considered a valuable resource for companies and decision-makers to help in making decisions. This study was based on using a hybrid approach to extract the opinions from an Arabic tweet to measuring service providers' reputation. In this study, the Saudi telecom companies used as a case study. This research concentrates on determining peoples' opinions more accurately by utilizing the Retweet and Favorite. The number resulting from positive and negative tweets after applying the polarity equation was used to estimate reputation scores. The result indicated that the STC company represents a high reputation compared to other companies. The proposed approach shows promising results to expand existing knowledge of sentiment analysis in the domain of measure reputation.

Keywords—Reputation; sentiment analysis; Arabic language; social media

I. INTRODUCTION

Data analysis and mining are some of the most important areas studied in computer science [1]. Due to the importance of data in various fields and the amount of data that is generated during the day, many companies have paid great attention to data. The amount of data is increasing day by day due to a large number of social media platforms [2]. A large amount of content on social media is processed periodically. This data can provide information about users' loyalty, complaints, potential customers, or measure the reputation. To deliver several services and better understanding the customers companies rely on different social media platforms such as Facebook and Twitter [3].

Twitter is one of the popular platforms where people express their thoughts and opinions in 140 characters. The number of active users in this micro-blog is more than 328 million users, who vary between presidents, academics, organizations, and bodies [4]. Statistics indicate that the number of Twitter users in Saudi Arabia has reached 41%, which is a large percentage compared to other countries [5]. In addition, the Arabic language is considered one of the rich languages that have broad audiences, so the focus was on it and Saudi society.

With the millions of opinions that are disclosed on social media about some topic, it is out of the question that these opinions will be biased [6]. People tend to refer to people's comments and opinions about products before they buy them, the movies they will watch or the countries they are thinking of traveling to. The opinions expressed by online users are a kind of electronic word of mouth (eWOM) [7]. The opinions are greatly influencing the decisions a customer will make and its behavior [8]. Where these opinions are considered as more trustworthy and credible way because it's independent of marketers' selling efforts [9].

The spread of sentiment on the Internet has created a new field of text analysis. It attracted many studies in many fields such as education [10], stock market [11] and others. User-generated comments are contributing to the establishment of digitalized WOM forms, which in turn contributes to creating an online reputation. Reputation is a factor that influences the desirability of a product or service, so it needs investment and good management [12][13]. Reputation affects companies' capabilities as well as increases their chances of sustainability in the market [14][15].

The objective of this study is to measure reputation based on a hybrid sentiment analysis approach that combines textual and non-textual features. The analysis improved by including the number of retweets and favorite tweets. In the end, the reputation was calculated accurately for the service providers through the Beta reputation equation.

The rest of the paper is structured as: Section 2 investigates the related works of customer feedback analysis in different sectors. Then, Section 3 describes the methodology proposed which added a new step into the SA approach: the polarity equation, and reputation calculation by equation. Section 4 examines the supervised ML algorithm to determine which algorithm was the best and examines the reputation equation with two different experiments. Then describes the results after utilized the proposed methodology. Section 5 concludes the final notes of the work with recommended guidelines for future developments and research.

II. RELATED WORKS

Feedback from customers is important and it may be a complaint, recommendation, evaluation, etc. Different data mining techniques are used on the feedback to provide valuable

information. Analysis of customer feedback has been recently an active area. It covered several areas of the field of airline services, the smartphone industry, hotels, banking services, and many others.

The sources of feedback varied, some researchers resorted to using traditional methods such as the survey. The study of [16] was concerned about the customer satisfaction with the South Korean airline service with the two different customer groups the low-cost carrier and the full-service carrier. To investigate customer satisfaction, they related it to six different factors positive emotion, negative emotion, social words, comparison, risky values, and monetary values. They are analyzing more than 133000 of the customer feedback after they experience the airline service. The positive emotion and the social words were considered the most impactable factor on customer satisfaction while the monetary value has always a negative impact. The limitation of this study was they didn't pay attention to the demographic factor of the customer also they concern only with the English feedback.

Also, the work of [17] was based on the predictive analysis they investigated the customer intention to return visit the services provided by the airline services. By using the customer feedback comment and satisfaction rating for the previous use of the services they applied a machine learning approach. They used seven classifiers to analyze the sentiment in the comment. Their work was different from the previous work in that they lookup for the actual intention of the customer to return the visit. Also, they analyzed a huge data set what is contain with 309331 customer comment. All the long comment indicates higher accuracy. Their results indicated that customer feedback is very important to determine the actual intention for returned visits and also, it helps to understand the customer behavior over time.

Traditional methods are not recommended due to the sample that may participate in this survey. Also, the survey may be mandatory, which means the respondents may answering only for completion. Therefore, researchers have resorted to some alternative methods, which are based on social media or websites. These methods are more effective as they involve different samples and sectors.

Their work [18] was based on a sentiment analysis which they referred to as opinion mining. They come up with a new idea to analyzing customer feedback from Amazon. So, they designed a framework that received any kind of data from the user and preprocessing it, and then process the data. The result shows the amount of the sales product and the market share as a figure. Also, shows the count of the positive and negative reviews for each feature of the sold phones. Which is give the company a better picture of its product and helps it to succeed in the market.

Their work [19] was focused on the complaints of the customer they analyzed the data set from the Customer Relationship Management system (CRM). Which is a system that deals with the complaints of the customer for the Metropolitan Transportation Authority (MTA). The MTA provides several services that including a subway and buses and a railroad. From the correlation matrix they come up with the agency and subject matter attributes that are positively

correlated and there is no negative correlation. Also, they performed more analysis on details of the complaints to improve the quality of the services. For the classification, they applied different modeling: naive Bayes, KNN, random tree, and ID3. For each model, the input is the complement details that give the agency name as an output. Their work was helpful to determine the factor that led to the customer complaint and also helped the company to improve the quality of the provided services.

The comments that posted in different social media platform could be considered as customer feedback. Mini and Poulouse study [20] to discover the customer intention about the eco-tourism. They took a property in Kerala in India as a case study. They collected the data from Facebook and performed a sentiment analysis on it. They were able to determine the satisfaction and dissatisfaction of the customers which will help improve the quality and improvement process. In their study the focus only on collecting the comment from Facebook that was written in English. Also, the multiple emotion in the collected comments was not easy to extract it.

The study of [21] was based on user-generated content (UGC) to investigate the airport service quality. Their study was in London Heathrow airport services. Sentiment analysis was performing on the collected tweets. They were able to evaluate the quality of the services and determined they good services and the services that need to be improved. UGC is considered a better resource more than the traditional way. The limitation of this study they ignore the demographic characteristics and it's applied to one airport.

UGC on social media plays a significant role in measuring reputation. As the work is done by [22] they measure reputation for four companies Apple, Twitter, Google, and Microsoft by analyzing the tweet. Their proposed approach a distinct work on dealing with the misspelling in the tweet and can apply to all languages also, it does not need a text preprocessing stage. They sentimentally classify each tweet to one of the companies and determine the emotion in this tweet. By relying on using N-gram approaches their results to show better accuracy more than the Bayesian algorithm and a neural network.

The same data set was also analyzed in the study of [23]. Their study was focus on the user reputation by applying some factor to calculate to determine which user have a higher reputation than the others. Their proposed framework was first assigned a score for each tweet the positive tweet was indicated by 1 and the negative was indicated by -1. Then they extracted the ID of the tweet to retrieve the Twitter creator they extracted five factors for each creator: number of followers, number of following, Twitter lists, numbers of retweeted by the followers, and number of tweets posted by that user. Via these five factors, they calculated the reputation of the user. Then they calculate the accumulated weight by multiplying the sentiment score with the user reputation. The tweet from the higher reputation user will have a higher accumulated weight. The proposed method would be helpful to minimize the effort and the size of data by extracting the tweet that only relevant to the higher reputation users.

Online reputation may be useful in determining and measuring profits, according to a study of [24]. They emerged Booking review and Financial Analysis Made Easy (FAME) and used Latent semantic analysis on the textual content. They were able to determine which attributes are associated with financial profitability. From the positive review, they were able to determine that the hotel location and the room quality was indicated the highest impact attribute on the financial profitability. The study was conducted only on the UK hotels and they extract the review from the Booking platform only.

In the tourism sector, there have been several studies and research, such as the work that was done by [25]. They measured Marbella's reputation as a tourist destination. Their study focused on the factors that attract tourists and determine the geographical distribution of the countries from which they came. They analyzed visitor reviews according to sociodemographic characteristics of the city over three years. Their study was focusing only on TripAdvisor reviews.

Also, the study of [26] that took a different sector focused on the political sector. They examine the reputation of the members of the National Assembly (MNAs) of Pakistan on Facebook and Twitter. They were able to determine the credibility of the members and which platforms have a more supportive audience. The comment was chosen based on the systematic random sampling technique. The supportive audience was on Facebook more than Twitter.

Many kinds of research have been dedicated to the field of analyzing customer feedback. Regrettably, few studies consider the Arabic language.

The study of [27] was aimed at determining the polarity of the text only. Therefore, it dealt with several topics in politics, sport and education. They collected data from Twitter and Facebook and then applied three classifiers, SVM, Naïve Bayes, and K-NN using Rapid miner. The size of the data set was small, so no promising results were shown.

The study of [28] was also on the same area they concerned about the telecom Saudi company STC, Mobile, and Zain. Their study focused on the assignment of the polarity scores of the tweets by SWN. The determination of polarity in SWN was based on a lexicon as the results showed that most of the tweets tended to be natural. Defining polarity helped them better understand their customers and their needs. They were also able to identify areas of clients who had complained about some problems.

Previous studies have indicated determining the polarity of sentiment in the telecom industry, both of which were based on a lexicon-based approach. All previous studies focused on text analysis only, they did not include the number of retweets and favorites. The following Table I gives an overview of the previous studies, their purpose, source of data, and the approach used in their methodology.

TABLE I. OVERVIEW ON RELATED ARTICLES

Cite	Analysis Objective	Dataset Sources	Approach
[16]	Measuring customer satisfaction	Survey	Structural equation modeling SEM
[18]	Measuring how to succeed in the market	Amazon review	Framework to receive data and sentimentally analyse it
[19]	Understanding the Complement of customer	CRM	Data mining Technique and Machine learning Algorithm for classification
[17]	Determine if the customer will return to visit	Survey	Machine learning for sentiment analysis
[20]	Discover the customer intention about the eco-tourism	Collected tweet	Hybrid-based Approach for sentiment analysis
[21]	Investigate service quality	Collected tweet	Online tools
[22]	Measuring companies reputation	Collected tweet	N-gram
[23]	Measuring user reputation	Collected tweet	Framework to sentimentally analyse tweet and calculate reputation
[24]	Determine the impact of online reputation on hotel profitability	Booking reviews	Latent semantic Analysis
[25]	Measuring the reputation of Marbella city	Trip Advisor reviews	A quantitative analysis
[26]	Measuring leaders reputation	Collected tweet	Quantitative sentiment analysis
[28]	Gain better customer insight	Collected tweet	Sentiment analysis

III. PROPOSED METHODOLOGY

The methodology that followed in this study contains five different phases. Fig. 1 shows an overview of the proposed methodology. Later, in this section, every phase will explain in detail.

The polarity of the collected tweets was determined to positive, negative, or natural. For this study, only positive and negative tweets were included. Most of the sentiment analysis that was performed was focus on polarity only. For more precise a hybrid approach that combines textual and non-textual features was applied. The hybrid approach was based on the polarity equation [29]. The integer number from the polarity equation for both the positive and negative classes was the input for the reputation equation [30].

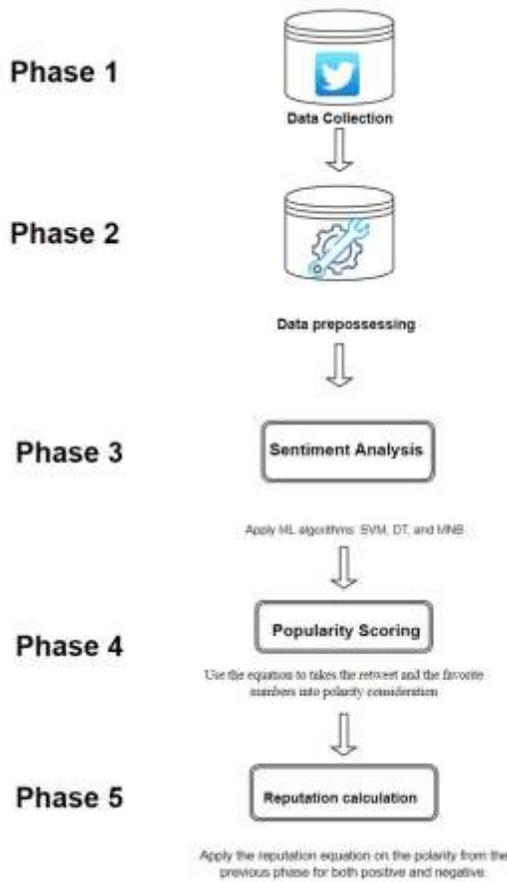


Fig. 1. Overview of the Proposed Methodology with the Five Phases.

A. Phase One: Data Collation

The data for this study was extracted through Twitter API. The analysis focused on the telecom service providers sector in Saudi Arabia. The providers were the data obtained from STC, Mobily, and Zine. The data was collected from the official account of the three providers: @stc_ksa, @Mobily, @ZainKSA. Each of these accounts is active accounts with thousands of tweets exchanged daily.

For each account, 5000 tweets were collected and the total amount of the data was approximately 15,000. Due to time constraints, data were collected within February only. The collected tweet was in the Arabic language mostly, while the tweets that contain some English words or letter were removed later as showed in Section III-B.

For each tweet, the time for its creation, tweet ID, text, the number of their retweet and favorite, user location, screen name, followers, and friends were extracted. For this study, only text, the number of their retweet, and favorite were used. Fig. 2 shows the statistical descriptive for the numerical values in the dataset.

B. Phase Two: Data Preprocessing

To prepared data that has been gathered for the sentiment analysis, it has to be cleaned and preprocessing. Preprocessing data means removing all the noise, unwanted parts, and missing values even the duplicate data. Performing all these procedures leads to get quality data and reliable results. The steps that followed on preprocessing data are:

1) Removed unwanted parts:

- Any English words, letter, or numbers.
- The retweet status, user mentions, URLs.
- Punctuation's, and the Tashkeel.

2) Normalized Arabic words: Arabic letter could be represented in many forms such (ا, إ, آ) all of it refer to (ا) letter. So, the three forms were replaced with (ا). Also, (ة) was replaced with (ه) and (ي) with (ي).

3) Normalizing repeated letters if it was more than two letters.

4) Removed Arabic stop words: Same as any other language the Arabic language contains multilabel stop words. Stop words in Arabic are meaningless same articles (a, an, the) in English. The stop word such (لم عن الذي) removed based on two different corpora.

5) Removed line breaks and extra white spaces.

6) Removed Duplicates tweets.

	count	mean	std	min	25%	50%	75%	max
tweet_id	4656.0	1.360000e+18	0.000000	1.360000e+18	1.360000e+18	1.360000e+18	1.360000e+18	1.360000e+18
rt_count	4656.0	8.488543e+01	597.045859	0.000000e+00	0.000000e+00	0.000000e+00	1.400000e+01	2.822500e+04
fav_count	4656.0	6.432607e+01	22.626225	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.575000e+03
followers	4656.0	2.252855e+04	324997.451510	0.000000e+00	2.800000e+01	1.750000e+02	6.480000e+02	1.438033e+07
friends	4656.0	9.648485e+02	4322.111729	0.000000e+00	1.107900e+02	3.515000e+02	6.890000e+02	1.857690e+05

(a) STC Statistical Descriptive.

	count	mean	std	min	25%	50%	75%	max
tweet_id	4573.0	1.360000e+18	0.000000	1.360000e+18	1.360000e+18	1.360000e+18	1.360000e+18	1.360000e+18
rt_count	4073.0	8.070119e+01	702.810130	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	6.411000e+03
fav_count	4073.0	7.802132e+02	0.843522	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	5.400000e+01
followers	4073.0	1.101250e+03	16066.829386	0.000000e+00	1.700000e+01	6.600000e+01	2.410000e+02	7.823460e+05
friends	4073.0	1.216744e+03	2402.115086	0.000000e+00	1.950000e+02	6.150000e+02	1.653000e+03	1.057230e+05

(b) Mobily Statistical Descriptive.

	count	mean	std	min	25%	50%	75%	max
tweet_id	4947.0	1.362519e+18	4.776207e+10	1.360000e+18	1.360000e+18	1.360000e+18	1.370000e+18	1.370000e+18
rt_count	4947.0	3.259733e+01	1.598934e+02	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+00	2.668000e+03
fav_count	4947.0	1.338549e+01	6.011419e+01	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	3.400000e+01
followers	4947.0	5.917047e+04	2.051923e+05	0.000000e+00	1.800000e+01	1.400000e+02	5.730000e+02	1.545853e+05
friends	4947.0	1.298761e+03	5.695033e+03	0.000000e+00	4.300000e+01	2.090000e+02	6.880000e+02	1.845170e+05

(c) Zain Statistical Descriptive.

Fig. 2. Statistical Descriptive for Numerical Values for the Three Companies.

The collected tweets are considered unlabeled data. To use some of the ML algorithms an online tool was used to label data. The Mazajak tool¹ created by Ibrahim Abu Farha and Dr. Walid Magdy [31] classified the text polarity to Positive, Negative, or Neutral. However, there may be chances where some word polarity is not detected correctly. For example, the appearance of words that contains a prayer to Allah gives a positive labeled result. The tweet (ياربي متي ت عدل ون نكلم؟) labeled as positive, which indicated a negative opinion. Fig. 3 shows the class distribution for the three companies.

After selected Positive and Negative tweets the amount of data became 2,770 for STC, 1,552 for Mobily, and 2,553 for Zain. Therefore, the total amount of the three companies was 6,875.

C. Phase Three: Sentiment Analysis

This sentiment analysis stage refers to the primary stage. Where the analysis is performing on text-only. The binary sentiment analysis works only with Positive and Negative tweets, Tweets that rated Neutral were excluded from this analysis.

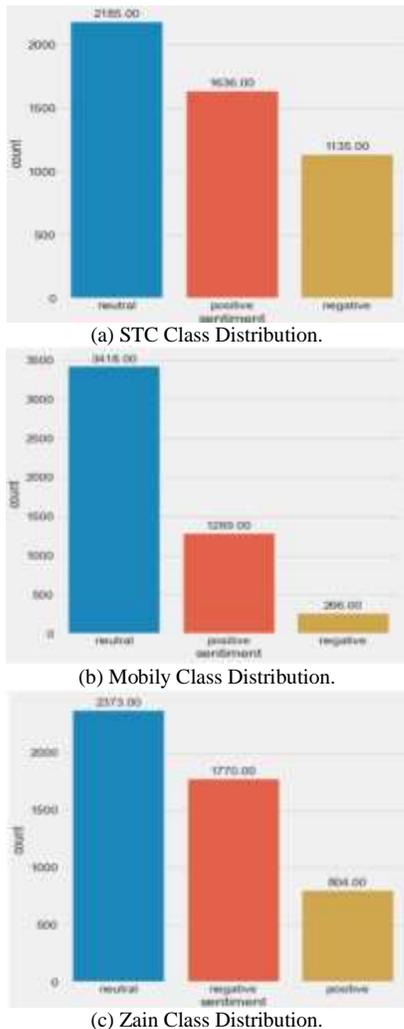


Fig. 3. Class Distribution for the Three Companies.

To prepare the target variable for classification the text must convert into a numerical record. Bag-of-Words representation works fine with a short text, is based on collect unique words in the dataset as features. Therefore, for the tweets in the proposed case Count Vectorizer was used for features extraction.

The dataset was split into training and testing sets. The training part was used to build the classification algorithms model, and the testing part to evaluate the model. Three Machine Learning (ML) algorithms were applied to the dataset Support Vector Machines (SVM), Decision Tree (DT), and Multinomial Naive Bayes (MNB).

D. Phase Four: Popularity Scoring

The popularity scoring equation takes the numbers of retweet and the favorite into polarity consideration. The popularity scoring equation can be considered as an advanced sentiment analysis stage. Since the primary stage dealt with text, the advanced stage deal with a retweet and the favorite.

A retweet means reposting someone’s tweet or message in your timeline [32]. Where the popular tweets are reposted many times. Any person following you was able to see the retweeted tweet. Favorite the tweet means to keep that tweet on your account under the favorite tab. To favorite the tweet only, the user must press the heart-shaped icon. Both retweet and favorite are considered as a sign of agreement and admiration on the content.

The Popularity Scoring Equation that applied adopted from a study of [29]. It consists of two equations dependent on each other. Where the outputs of the first equation are inputs to the second equation.

The first popularity scoring equation 1 takes the sum of the classified tweet Ct (e.g. Positive classified tweet) plus the number of its retweet #RT number plus the number of its favorite #F av.

$$\text{PopularityScoring} = Ct + \#RT + \#Fav \quad (1)$$

The output from equation 1 which is the popularity scoring of the classified tweet (e.g. Positive classified tweet) multiplied by 100 and divided by the total number of the popularity scoring for all the classified tweet 2.

$$\text{DevelopedPopularityscoring} = \frac{\pi 100 * \text{PopularityScoring}}{2 \sum |P \text{opularityScoring}} \quad (2)$$

E. Phase Five: Reputation Calculation

Most reputation measurement systems rely on the opinion generated by customers, where this opinion reflects the general community’s opinion [33]. From various principles of reputation measurement systems, this study is based on Bayesian systems. The Bayesian systems using the binary classes as input [34], and the reputation scores calculated based on statistical updating of beta probability density functions (PDF). To measure the reputation of service providers, the equation presented in a study was adopted [30]. The input in the sentiment analysis system is the positive and negative classes.

¹<http://mazajak.inf.ed.ac.uk:8000/>

In the proposed approach, the output from equation 2 used as input in the reputation equation. After the polarity calculated of both the positive and negative class, with the numbers of retweeting and favorite. The polarity number (e.g. positive class) took as a positive representation in the reputation equation. The equation 3 represent reputation equation:

$$R = \frac{\alpha}{\alpha + \beta} \quad (3)$$

The α represent $\alpha = r + 1$ and $\beta = s + 1$, where r is the output of popularity scoring number for positive classes and s the output of popularity scoring number for negative ones.

To clarify the equation let consider this example: if α equals 9 and β equals 3. After applying the equation 3, the reputation score can be calculated as: $R = 9/(9 + 3) = 3/4 = 0.75$.

IV. RESULTS AND EVALUATION

This section indicates the results of the calculated reputation through a proposed methodology. By applying a hybrid sentiment analysis approach on 15,000 collected tweets.

A. Evaluation Metric

The proposed sentiment analysis that was applied was based on two stages: the primary stage and the advanced stage. On the primary stage SVM, DT, and MNB classification algorithms are used to capture the sentiment. For the training model 80% of the data was used and for testing 20% was used. Moreover, to measure the performance of classification four measurements are used in this study, which are the accuracy, precision, recall, and F-score.

Accuracy: is the degree of closeness of the classified outcomes to the true value. Measurement of the accuracy is significantly important because it reflects the percentage to realize the correct pattern and polarity. Where T P indicate True Positive, TN for True Negative, F P for False Positive, and FN for False Negative.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

Precision: The ability of the model to anticipate the positive classes. It's calculated by divided true positives over a total number of true positives and false positives.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall: The values that the model was able to determine correctly. It's calculated by divided true positives over a total number of true positives and false negatives.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F-Score or the F-Measure: conveys the harmonic mean between the precision and the recall. It represents the integration of both precision and recalls into a single score.

$$F - Score = \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

After the primary sentiment analysis stage performed the results of the evaluation model summarize Precision, Recall, F-score, and Accuracy in Tables II, III, and IV.

From the Tables II, III, and IV, the SVM gives a higher accuracy among the three classifiers. The accuracy of the three companies for STC, Mobily, and Zain equal 92%, 95%, and 89% respectively as Fig. 4 showed. DT mostly gives low results compared to the other classifiers because some classes are imbalanced.

TABLE II. THREE CLASSIFIER RESULT FOR STC COMPANY

STC			
Support Vector Machines classifier			
Data label	Precision	Recall	F-Score
Positive	0.95	0.91	0.93
Negative	0.88	0.93	0.90
Accuracy			
0.92			
Decision Tree classifier			
Data label	Precision	Recall	F-Score
Positive	0.86	0.91	0.88
Negative	0.86	0.78	0.82
Accuracy			
0.86			
Naive Bayes classifier			
Data label	Precision	Recall	F-Score
Positive	0.96	0.81	0.88
Negative	0.78	0.96	0.86
Accuracy			
0.87			

TABLE III. THREE CLASSIFIER RESULT FOR MOBILY COMPANY

Mobily			
Support Vector Machines classifier			
Data label	Precision	Recall	F-Score
Positive	0.94	1.00	0.97
Negative	1.00	0.70	0.82
Accuracy			
0.95			
Decision Tree classifier			
Data label	Precision	Recall	F-Score
Positive	0.95	0.98	0.97
Negative	0.91	0.77	0.84
Accuracy			
0.95			
Naive Bayes classifier			
Data label	Precision	Recall	F-Score
Positive	0.96	1.00	0.98
Negative	0.98	0.79	0.88
Accuracy			
0.96			

TABLE IV. THREE CLASSIFIER RESULT FOR ZAIN COMPANY

Zain			
Support Vector Machines classifier			
Data label	Precision	Recall	F-Score
Positive	0.93	0.70	0.79
Negative	0.87	0.97	0.92
Accuracy			
0.89			
Decision Tree classifier			
Data label	Precision	Recall	F-Score
Positive	0.68	0.71	0.70
Negative	0.86	0.85	0.86
Accuracy			
0.80			
Naive Bayes classifier			
Data label	Precision	Recall	F-Score
Positive	0.87	0.65	0.74
Negative	0.85	0.95	0.90
Accuracy			
0.86			



Fig. 4. Accuracy Comparison of SVM for the Three Companies.

B. Evaluation of Reputation

As mentioned before the analysis only included the positive and negative labels. After eliminating the natural labeled, duplicated, and blanked the dataset was reduced to 6,875. Two tests were applied to capture the impact of the Popularity Scoring equation on mustering reputation.

The first test was applied without the Popularity Scoring equation, so the number of retweets and favorites did not include. Only the total number of the classified tweet (e.g. Positive classified tweet) which represents α in the reputation equation. For example, the positive classified tweet for STC is 1636, where the negative is 1134. By applying the reputation equation 3 the reputation scores equal to 0.59.

The second test by applying the Popularity Scoring equation where the number of retweets and favorites counted. The result after applying the reputation equation was higher. The reputation score for STC was 0.97.

However, taking into consideration the number of retweets and favorites for each tweet showed an improvement in

calculating reputation. Table V and Fig. 5, 6 shows the difference in reputation score between the two testing.

TABLE V. COMPARISON OF THE REPUTATION SCORES OF THE THREE COMPANIES

Without Popularity Scoring	With Popularity Scoring
STC	
0.59	0.97
Mobily	
0.82	0.74
Zain	
0.31	0.19

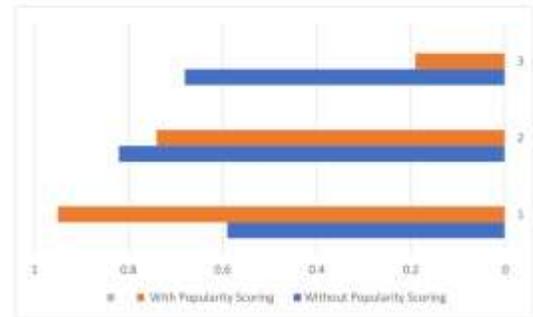


Fig. 5. Comparison of the Reputation Scores of the Three Companies.



(a) STC Reputation Score.



(b) Mobily Reputation Score.



(c) Zain Reputation Score.

Fig. 6. Reputation Scores before and after Applying Popularity Scoring Equation.

From the result, the STC reputation score was higher after applying the Popularity Scoring equation. While Mobily and Zain indicate the lowest reputation score after applying the Popularity Scoring equation. The reputation scores of Mobily and Zain decreased, and this is not due to defective or erroneous results, but after applying a Popularity Scoring equation gave more accurate results that included the re-tweet calculation and favorites.

V. CONCLUSION AND FUTURE WORK

In this study, a new opinion review presented based on a hybrid approach focused on the undertaking of sentence-level sentiment analysis to calculate reputation scores from Arabic tweets. The divergence of opinions between the customers of Telecom service providers in Twitter causes a need for a new approach to determine the reputation score of service providers involving developing a new way to compute the polarity of Arabic sentiment.

A new step was added to the traditional sentiment analysis process to enhance its accuracy. First, the significance value of counting the retweets and favorites numbers in the polarity score explains, which represents a non-verbal opinion. Second, developing a reputation approach based on the polarity score of sentiment. The first step could assist the classifier to understand and make a better accuracy about the sentiment analysis in Arabic text.

Different measures to evaluate the performance and efficacy of the classification were used in this study: the accuracy, precision, recall, and F-score. The result indicates that the SVM is the best-performed classifier, while the lowest-performing classifier is the Decision Tree. Also, the degree of reputation indicates that the STC company represents the highest reputation among its customers compared to other companies.

For future work, the approach will expand to consider multilevel word polarity instead of binary level. Also, there is a need to study the demographic characteristics of customers. In the data label, there may be a possibility where some word polarity is not noticed correctly. Currently, those instances are not handled so, it's better to use a lexicon-based approach besides labeling data by the Maza-jak tool. Moreover, the spam tweets should distinguish and eliminate to deliver more reliable reputation scores.

REFERENCES

- [1] G. Dubey, A. Rana, and N. K. Shukla, "User reviews data analysis using opinion mining on web," in 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 603–612, IEEE, 2015.
- [2] H. Akkineni, P. Lakshmi, and B. Babu, "Online crowds opinion mining it to analyze current trend: A review," International Journal of Electrical and Computer Engineering, vol. 5, no. 5, 2015.
- [3] W. He, S. Zha, and L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry," International journal of information management, vol. 33, no. 3, pp. 464–472, 2013.
- [4] K. Arun, A. Srinagesh, and M. Ramesh, "Twitter sentiment analysis on demonetization tweets in india using r language," International Journal of Computer Engineering and Information Technology, vol. 9, no. 6, p. 119, 2017.
- [5] T. ministry of communication and information system, "Saudi arabia is the most twitter-crazy country in the world: Business insider."
- [6] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," Expert Systems with Applications, vol. 40, no. 10, pp. 4241–4251, 2013.
- [7] E. Jeong and S. S. Jang, "Restaurant experiences triggering positive electronic word-of-mouth (ewom) motivations," International Journal of Hospitality Management, vol. 30, no. 2, pp. 356–366, 2011.
- [8] P. Phillips, S. Barnes, K. Zigan, and R. Schegg, "Understanding the impact of online reviews on hotel performance: an empirical analysis," Journal of Travel Research, vol. 56, no. 2, pp. 235–249, 2017.
- [9] J. Nieto, R. M. Hernández-Maestro, and P. A. Muñoz-Gallego, "Marketing decisions, customer reviews, and business performance: The use of the top rural website by spanish rural lodging establishments," Tourism Management, vol. 45, pp. 115–123, 2014.
- [10] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," IEEE Access, vol. 5, pp. 2870–2879, 2017.
- [11] G. A. A. J. Alkubaisi et al., "The role of ensemble learning in stock market classification model accuracy enhancement based on naive bayes classifiers," International Journal of Statistics and Probability, vol. 9, no. 1, pp. 1–36, 2020.
- [12] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, "Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews," Journal of Hospitality Marketing & Management, vol. 25, no. 1, pp. 1–24, 2016.
- [13] G. E. Bolton, E. Katok, and A. Ockenfels, "How effective are electronic reputation mechanisms? an experimental investigation," Management science, vol. 50, no. 11, pp. 1587–1602, 2004.
- [14] Z. Xiang, U. Gretzel, and D. R. Fesenmaier, "Semantic representation of tourism on the internet," Journal of Travel Research, vol. 47, no. 4, pp. 440–453, 2009.
- [15] G. Viglia, R. Minazzi, and D. Buhalis, "The influence of e-word of-mouth on hotel occupancy rate," International Journal of Contemporary Hospitality Management, 2016.
- [16] E. Park, Y. Jang, J. Kim, N. J. Jeong, K. Bae, and A. P. Del Pobal, "Determinants of customer satisfaction with airline services: An analysis of customer feedback big data," Journal of Retailing and Consumer Services, vol. 51, pp. 186–190, 2019.
- [17] S. Hwang, J. Kim, E. Park, and S. J. Kwon, "Who will be your next customer: A machine learning approach to customer return visits in airline services," Journal of Business Research, vol. 121, pp. 121–126, 2020.
- [18] M. Gupta and S. Sebastian, "Framework to analyze customer's feedback in smartphone industry using opinion mining," International Journal of Electrical & Computer Engineering (2088-8708), vol. 8, no. 5, 2018.
- [19] A. Ghazzawi and B. Alharbi, "Analysis of customer complaints data using data mining techniques," Procedia Computer Science, vol. 163, pp. 62–69, 2019.
- [20] U. Mini et al., "Opinion mining for monitoring social media communications for brand promotion," International Journal of Advanced Research in Computer Science, vol. 8, no. 3, 2017.
- [21] L. Martin-Domingo, J. C. Martín, and G. Mandsberg, "Social media as a resource for sentiment analysis of airport service quality (asq)," Journal of Air Transport Management, vol. 78, pp. 106–115, 2019.
- [22] H. S. Manaman, S. Jamali, and A. AleAhmad, "Online reputation measurement of companies based on user-generated content in online social networks," Computers in Human Behavior, vol. 54, pp. 94–100, 2016.
- [23] A. Bukhari, U. Qamar, and U. Ghazia, "Urwf: user reputation based weightage framework for twitter micropost classification," Information Systems and e-Business Management, vol. 15, no. 3, pp. 623–659, 2017.
- [24] S. C. Anagnostopoulou, D. Buhalis, I. L. Kountouri, E. G. Manousakis, and A. E. Tsekrekos, "The impact of online reputation on hotel profitability," International Journal of Contemporary Hospitality Management, 2020.

- [25] M. J. Carrasco-Santos, A. M. Ciruela-Lorenzo, J. G. Méndez Pavón, and C. Cristófol Rodríguez, "An online reputation analysis of the tourism industry in marbella: A preliminary study on open innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 2, p. 111, 2021.
- [26] S. Shami, A. Ashfaq, and S. N. Khan, "Reputation management on social media: Analysis of audience feedback on posts and tweets of pakistani mnas.," *Journal of Political Studies*, vol. 26, no. 2, 2019.
- [27] R. M. Duwairi and I. Qarqaz, "Arabic sentiment analysis using supervised classification," in *2014 International Conference on Future Internet of Things and Cloud*, pp. 579–583, IEEE, 2014.
- [28] S. Khatoon, "Real-time twitter data analysis of saudi telecom companies for enhanced customer relationship management," *International Journal of Computer Science and Network Security*, vol. 17, no. 2, pp. 141–147, 2017.
- [29] A. S. Almalki, *A New Opinion Review Methodology for Arabic Twitter Sentiment Analysis*. The Catholic University of America, 2020.
- [30] H. Al-Hussaini and H. Al-Dossari, "A lexicon-based approach to build service provider reputation from arabic tweets in twitter," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 4, pp. 445–454, 2017.
- [31] I. A. Farha and W. Magdy, "Mazajak: An online arabic sentiment analyser," in *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pp. 192–198, 2019.
- [32] G. F. Khan, *Seven layers of social media analytics: Mining business insights from social media text, actions, networks, hyperlinks, apps, search engines, and location data*. Gohar Feroz Khan, 2015.
- [33] A. Jøsang and S. Pope, "Semantic constraints for trust transitivity," in *Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling-Volume 43*, pp. 59–68, 2005.
- [34] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision support systems*, vol. 43, no. 2, pp. 618–644, 2007.