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:

$$R = \frac{\alpha}{\alpha+\beta} \tag{3}$$

The α represent α = r + 1 and β = 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.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

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.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

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

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

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.

$$F-Score = \frac{Precision*Recall}{Precision+Recall} \tag{7}$$

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

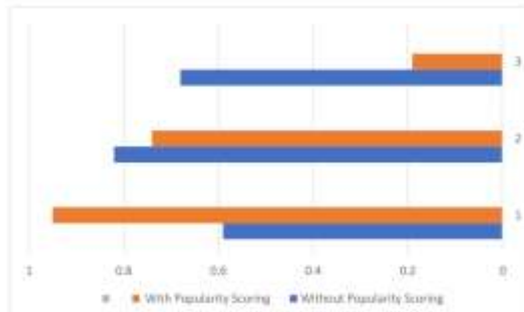

Fig. 5. Comparison of the Reputation Scores of the Three Companies.

(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 opinionmining 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 toprural 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 Pobil, "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.