Sentiment Analysis in Public Health Crisis Applications: A Conceptual Framework

Rafidah Isa[1], Mohammad Fauzan Nordin[2], Roslina Othman[3], Hazwani Mohd Mohadis[4]

[1,2,4] Department of Information Systems, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Gombak, Selangor, Malaysia,

[3] Department of Library & Information Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Gombak, Selangor, Malaysia

rafidah.sict@gmail.com

Abstract— Sentiment analysis, also known as opinion mining, is a Natural Language Processing (NLP) technique. It predicts people's opinions on products or services offered, whether positive, negative, or neutral. As a result, the outcome may assist an organisation in making better decisions. Nevertheless, sentiment analysis in the context of public health crisis apps is still new. A complete framework for the sentiment analysis model in this area is still lacking. In this paper, a comprehensive framework regarding sentiment analysis on user reviews of a public health crisis is proposed. The proposed framework represents a classification model that works as a complete guide for developers, researchers, and firms to create and deploy machine learning models more quickly and efficiently in the field of public health crisis apps. It starts with the data collection of user reviews and is followed by data labelling. The third stage is data pre-processing, where the data will be cleaned to improve the quality of the data and classification effectiveness. The fourth stage is feature extraction, in which the adjectives of the cleaned data will be identified by the TF-IDF technique. The fifth stage is data training and testing. Supervised machine learning classifiers, which are Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbour (KNN) will be used to perform data training and testing. The last stage is performance evaluation and comparison. This stage identifies the best classifier based on the measurement parameters. Based on the experiment conducted, the framework is fit for this study since the experiment produces a good score of the F-measure. The framework will benefit developers, firms, and researchers by assisting them to model their sentiment analysis work.

Keywords— Sentiment Analysis, Public Health Crisis Apps, Supervised Machine Learning, Support Vector Machine, Naïve Bayes, Random Forest, K-Nearest Neighbour.

I. Introduction

Numerous mobile applications and websites allow users to provide their reviews through textual or star ratings on the application or services. The reviews and opinions provided by the users are valuable to a firm as well as end-users. It is one of the user-generated data sources. [1] stated that user-generated content provides the real nature of messages that are available online. It could have a positive or negative impact on the firm. Therefore, understanding the opinions or sentiment polarity of the review is crucially important. The result could assist a firm in understanding people's sentiments. It also helps a firm to make better decisions and improve its performance. Therefore, users' reviews are an ideal data source to understand people's opinions on the products or services offered by a firm.

In the context of public health emergencies, mobile apps play an important role. The mobile phone serves as a medium of communication between authorities and the affected community in handling the crisis. Furthermore, it could provide timely information, surveillance, and rapid crisis response [2]. Apart from that, in the event of a possible health emergency, the usage of mobile apps can simplify the sharing of information amongst health experts [3]. There are various mobile applications developed to help public authorities manage the crisis effectively. For example, Mysejahtera, CovidSafe, BeAware, Immuni and PeduliLindungi are used in Malaysia, Australia, Bahrain, Italy, and Indonesia in managing COVID-19 [4].

Thus, understanding users' sentiment polarity on mobile applications developed is important. The polarity will show users' emotions toward the application being used. It will assist the developers, researchers, and firms to better understand the user's needs and find the gap in the application developed in regards to managing public health crises. Positive polarity will show the strengths of the application, whereby negative polarity may assist the team to further diagnose and improve the application. Hence, to understand the users' sentiments, the users' opinions should be analyzed. It could be done in various ways, and sentiment analysis using the machine learning approach is one of the options. Therefore, a comprehensive sentiment analysis framework specifically for public health applications is needed.

Concerning that, the objective of this study is to propose a framework for sentiment analysis on users' reviews of public health crisis applications. The framework uses supervised machine learning. It comprises six (6) stages, which start with data collection, labelling, data preprocessing, feature extraction, data training and testing, and lastly, performance evaluation. The framework represents a useful study for application developers, researchers, and firms in designing and deploying machine learning models to understand the polarity of people's opinions on the applications developed.

II. RELATED WORKS

Sentiment Analysis has been an active research area since the 2000s. Over the years, various researchers conducted sentiment analysis on healthcare reviews and tweets to understand users' opinions and perceptions. For example, [5] performed sentiment analysis on the Twitter dataset that related to diabetes. Their study utilized the SentiStrength approach to measure the positive and negative sentiment of the dataset. In their study, the sentiment values (positive and negative) of the statement were calculated when the tweet is analyzed using the lexicon sentiment. However, [5] stated that the tool has not yet been validated for use in the health domain.

Furthermore,[6] utilized sentiment analysis on Twitter data to determine how citizens perceive and use various mobile health applications. The data of mHealth apps namely fitness apps, diabetes apps, meditation app, and cancer were mined from Twitter using RStudio. The analysis was carried out to measure the polarity and emotion of the citizen. They used word cloud and bigram for variable frequency identification in their sentiment analysis.

Apart from that [7]conducted sentiment analysis on COVID-19 contact-tracing apps to determine the polarity of end-users sentiment based on the users' reviews in the apps store. AppBot was used to retrieve the reviews. They used word cloud and topic modelling to discover the sentiment polarity from the end-users.

[8] used machine learning to classify health-care-related comments into negative and positive feelings to better understand how patients feel about their care. Weka Data Mining software was used in their study. Whereby, Unigram and bigram approaches were applied for their analysis. Four machine learning algorithms were deployed namely Naïve Bayes Multinomial, Decision Tree, Bagging and Support Vector.

On the other hand, [9] used a machine learning approach to perform sentiment analysis on public opinions about the Human Papillomavirus (HPV) vaccine. They extracted people's opinions on the HPV vaccine on Twitter and analysed them using the SVM model.

Whereby, Weibo data was utilized by [10] to build an SVM classifier to evaluate whether or not a user is in danger of suicide or is experiencing emotional distress. Based on social media data,[11] used a Convolutional Neural Network (CNN) approach to detect several types of medical attitudes that can be derived from users' medical conditions and treatment. Furthermore [12] used sentiment analysis to construct a gender detection technique to better understand how AIDS patients' online chats differ along gender lines. They adopted the deep learning CNN model in their study. On the other hand, [13] classified the sentiment polarity on mental health apps using Stochastic Gradient Descent (SGD) model.

Based on the previous sentiment analysis works on public health-related applications, it can be concluded that the most common data sources were extracted from users' opinions on Twitter, Website and apps store. SVM and CNN were among the supervised machine learning classifier used in their sentiment analysis. Other supervised machine learning classified were not yet utilized in regards to sentiment analysis on health-related apps. Similarly, with the feature extraction techniques, only a few techniques were used. The previous works adopted SentiStrength, Word Cloud and N-gram approach in identifying the feature of the data source. As[14] highlighted that one model does not fit all, so, more classifiers shall have experimented to determine the best classifier with high-performance accuracy. Similarly, with the feature extraction technique, varied techniques shall be explored.

Apart from that, the selected literature also did not explain the detailed steps in conducting their sentiment analysis work except the study done by[9]. A clear framework is necessary to understand the concept and processes of the data under study. Therefore, to fill the gaps, this study will propose a comprehensive framework that explains the detailed process of sentiment analysis work specifically for public health crisis apps. A complete framework may significantly help researchers to design or deploy a sentiment analysis model in their study.

III. CONCEPTUAL FRAMEWORK

A conceptual framework has been developed to detail the implementation process of sentiment analysis on public health crisis apps. Figure 1 consists of six (6) stages; First, using scrapping software, the textual review of the public health crisis application will be scrapped. Secondly, the extracted data will be exported to Microsoft Excel where the data will be labelled as positive, negative, or neutral according to their rating. Thirdly, the data pre-processing will be done using a machine learning tool by performing tokenization, transform cases and filter stop word. Fourthly, feature extraction and selection will be performed to extract the adjective from the data set. Fifthly, data training and testing will be conducted and finally, the classification model's performance will be evaluated and compared to choose the best classifier.

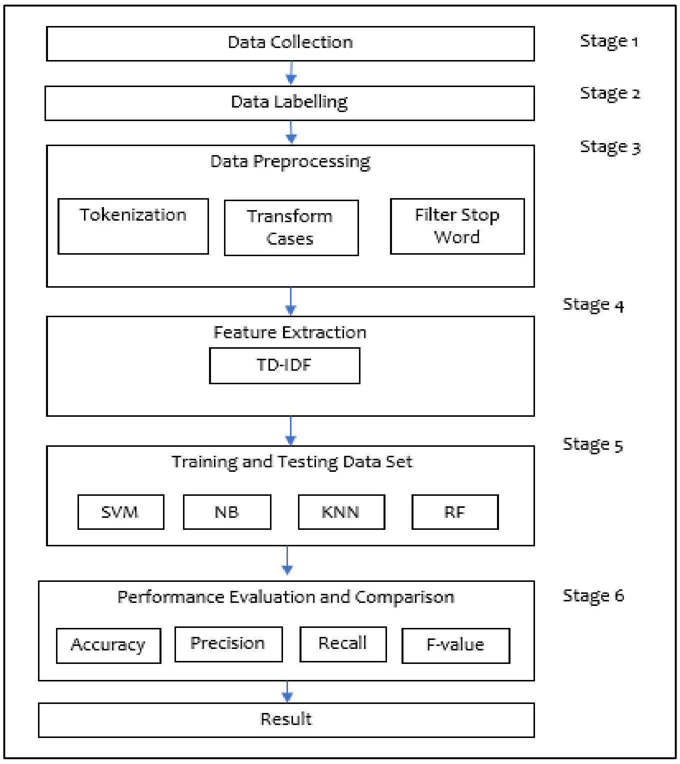

Figure 1: A Conceptual Framework

A. *Data Collection*

This study will focus on a textual review of public health crisis apps. The data will be scraped from the Google Play store using a data extraction tool. The scrapped data will be exported to Microsoft Excel for further steps.

B. *Data Labelling*

The exported data in Microsoft Excel will be labelled as positive, negative, or neutral according to their rating. On Google Play, each user can assign a star rating on a scale of 1 to 5. A scale of "1" represents very dissatisfied and "5" represents very satisfied. We adapted the approach and criteria by[13], [15] and [16]to label the review. Table 1 shows the criteria for data labelling based on users' ratings.

Table 1:Criteria for Data Labelling

Rating	Description	Sentiment Label
1	Very dissatisfaction	Negative
2	Dissatisfaction	Negative
3	Okay	Neutral
4	Satisfied	Positive
5	Very Satisfied	Positive

C. *Data Pre-processing*

Data pre-processing will be done after the data is labelled in Excel. The labelled data will be loaded into a machine learning tool. Data pre-processing will improve the quality of the data and classification effectiveness[19]. This process will clean the data into an understandable format, making it meaningful and informative for machine learning tools. In this study, tokenization, transform cases and filter stop word techniques will be deployed to pre-process the textual data.

Tokenization divides the textual review into individual words known as a token[20]. This technique is important in text mining as it prepares the vocabulary for a machine learning tool to understand the data.

Next, transform cases will be conducted to transform all the characters in the data set to either lower case or upper case. So that the data will be better organized. In this study, the characters will be transformed to lower case.

Afterward, filter stop words will be applied to remove all the stops words for example determiners, conjunction and preposition from the data set. By removing these stop words, machine learning tools could focus on the words that define the meaning of the text and reduce the time it takes to train the classification model as well as improve the classification accuracy. [21]

D. *Feature Extraction*

The fourth stage is feature extraction. It is a task that extracts the adjective from the data set. Later, this adjective will allow the machine learning tool to decide the sentiment polarity as positive, negative or neutral.

In this framework, Term Frequency-Inverse Document Frequency (TF-IDF) will be used to perform feature extraction. TF-IDF approach is the most widely used information retrieval and text mining [22], [23].TF-IDF finds the term (word), quantifies its occurrences in the data set and determines the importance of the word for the document. This technique will produce a numeric word vector which will be used by the classification model to statistically predict the polarity.

E. *Data Training and Testing*

Data training is a process that trains the classifiers to predict the outcome of a study [24]. In this study, the training dataset will train the classifier to predict the sentiment polarity. On the other hand, the testing data set will evaluate the performance of the classifiers based on the performance evaluation metric. At this stage, the pre-processed and vectorized data will be fed into the machine learning classifier as input to perform the process. In this study, the dataset will be divided into a data training and a data testing set with a ratio of 80: 20. The value of the ratio has been adapted from [25].

We will deploy four (4) supervised machine learning classifiers in this framework, i.e Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF) and K-Nearest Neighbour (KNN) to perform data training and testing. The classifiers are widely used in text classification problems [13]

1) Support Vector Machine (SVM):

SVM is a supervised machine learning model that supports both classification and regression analysis. It categorises the labelled data into the classification group.

2) Naïve Bayes (NB):

NB is a simple probabilistic classifier based on Bayes's Theorem. It calculates the probability of each term corresponding to a label. It stated mathematically as the following equation

P(label|features)= P(label)* P(label|features)/P(features)

3) Random Forest:

Random Forest (RF) is capable of performing both regression and classification tasks. To arrive at a single outcome, it aggregates the output of numerous decision trees.

4) K-Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN) model solves both classification and regression problems as well. It is mostly used to classify a data point based on how its neighbours are classified. KNN algorithm is based on the comparison of data set with k training data sets which are the nearest neighbour of that dataset

F. Performance Evaluation

The measurement parameters that will be used to evaluate the performance of the classifiers are accuracy, precision, recall and F-measure. The equation is given below:

1) Accuracy:

Accuracy is calculated as the number of instances predicted positively divided by the total number of instances. This means accuracy is the percentage of the accurately predicted classes among the total classes.

Accuracy = ((True Positive + True Negative) / Total number of predictions made[20]

2) Precision

Precision(P) is the ratio of the number of total positives (TP) predictions divided by the total number of positive class values predicted. This can be presented as

Precision (P) = TP/(TP+FP)[26]

3) Recall

Recall (R) is the number of positive predictions divided by the number of positive class values in the test data. This can be presented as

Recall (R) = TP/(TP+TN)[27]

4) F-measure

The F-measure is the balance between precision and recall. This can be presented as

F-measure =2*((precision*recall)/(precision+recall)).[28]

The result from the confusion matrix will be used to select the best measurement parameter for determining the performance of the framework. As shown in Figure 2, a confusion matrix is a table that shows how many correct and incorrect predictions a classifier has made.

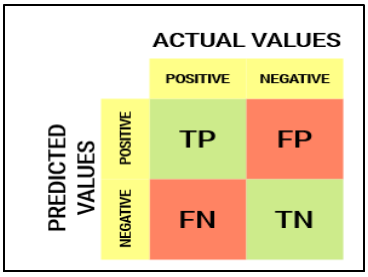

Figure 2: Confusion Matrix

The details of the terms are explained below:

True Positives (TP) is when the actual value is positive and predicted is also positive.

True negatives (TN): when the actual value is negative and prediction is also negative.

False positives (FP): When the actual is negative but the prediction is positive.

False negatives (FN): When the actual is Positive but the prediction is negative.

If the result shows a well-balanced class, then accuracy is a valid choice, but if it is vice versa, then the f-measure is a better choice to evaluate the performance of the framework.

On the other hand, when the classification problem is concerned with a false positive rather a false negative, then precision is suitable to be observed. whereby, recall is to be applied if false negative trumps false positive.

To validate the performance of the classification model, K-fold cross-validation will also be applied in the framework. It will break the dataset into subsets and iterate to train the model. Finally, it gives the average performance of the classifier. In this study, 10 folds of cross-validation will be applied.

IV. EXPERIMENTAL SETUP

A. Dataset

Datasets used to examine the proposed framework are users' reviews of MySejahtera apps. MySejahtera is a mobile app deployed by the Malaysian government to manage the COVID-19 crisis in Malaysia. It works as contact tracing, assisting the users to get treatment once they are infected, as well as a medium of information dissemination between the government and the people.

In this study, the users' reviews labelled with the most relevance from 1st December 2020 to 3rd March 2021 from Google Play were scrapped using Octoparse software. A total of 3070 reviews were analysed using RapidMiner. The study focused on the positive and negative labels for the analysis.

B. *RapidMiner Process Model*

Figure 3 depicts the whole process model to perform the analysis. The model is based on the framework explained in Section III. The upper part shows the processes used to train the model whereas the bottom part tests the performance.

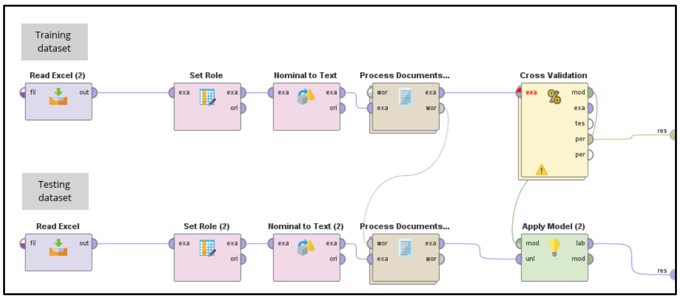

Fig. 3: RapidMiner Process Model

1) *Data preparation and Feature Extraction*

Figure 4 shows the process of data preparation for both training and testing tasks. Read Excel, Set Role, Nominal to Text and Process Document to Data operators were used.

The Read Excel operator loads the labelled data from Excel format into RapidMiner whereby the Set Role operator is used to set the attribute label. Nominal to Text operator functions to convert nominal data to text format. Process Document to Data operator is a nested operator. The sub operators are tokenize, filter stop word (English) and transform case. The sub operators are shown in Figure 5. These sub-operators carry out pre-processing tasks, as described in section III(C).The outer operator will perform feature extraction using the TF-IDF technique.

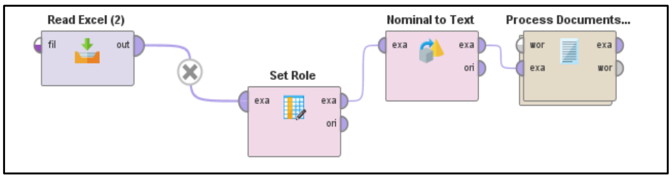

Fig. 4:Data Preparation Operators

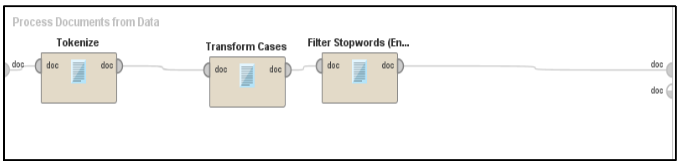

Fig. 5:Data Pre-processing Operators

2) *Training and Testing*

The selected classifier and Apply Model operators are responsible to perform the training and testing tasks. Figure 6 shows the process model for the task. The task is a sub-process of the cross- validation operator.

Simultaneously, the cross-validation operator partitions the dataset set and iterates the training and testing task. It runs 10 folds to validate the performance of the classifiers.

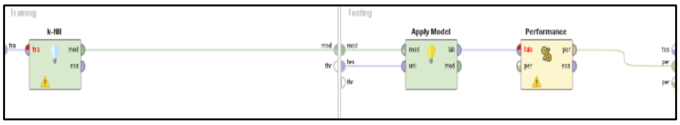

Fig. 6:Testing and Training Operators

3) *Performance Evaluation*

The performance (Binomial Classification) operator in Figure 6 is responsible for evaluating the performance of the model employed. The parameters used to measure the performance of the model are based on the performance metric discussed in section III (F).

Steps 2 and 3 were repeated to experiment with different classifiers used in this study by changing the classifiers for testing and training.

V. RESULT

Figure 7 depicts the training dataset loaded into RapidMiner. It demonstrates an imbalance of positive and negative reviews. Therefore, an f-measure has been observed to measure the performance of the model.

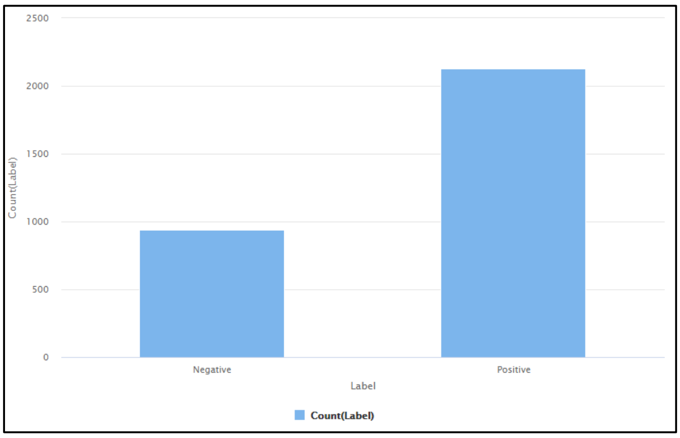

Fig. 7 Dataset Loaded into RapidMiner:

Figures 8 to 11 are the confusion matrix of the SVM, NB, RF and K-NN respectively. The meaning of the confusion matrix is explained in *Section III (F) (4)*. For example, in the SVM confusion matrix below, 417 reviews are labelled as positive and predicted as positive. Similarly, 2020 reviews are labelled as negative are predicted to be negative.

Contrary, 109 reviews are False Positive (FP) in which the reviews are labelled positive but are predicted negative and 524 reviews are False Negative(FN). In this case, the reviews are labelled negative but are predicted to be positive.

f_measure: 86.46% +/- 1.32% (micro average: 86.45%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	417	109	79.28%
pred. Positive	524	2020	79.40%
class recall	44.31%	94.88%	

Fig. 7 Confusion Matrix of SVM

f_measure: 73.01% +/- 2.45% (micro average: 73.06%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	719	776	48.09%
pred. Positive	222	1353	85.90%
class recall	76.41%	63.55%	

Fig. 8 Confusion Matrix of NB

f_measure: 81.90% +/- 0.07% (micro average: 81.90%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	0	0	0.00%
pred. Positive	941	2129	69.35%
class recall	0.00%	100.00%	

Fig. 9:Confusion Matrix of RF

f_measure: 86.41% +/- 1.61% (micro average: 86.41%) (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	562	221	71.78%
pred. Positive	379	1908	83.43%
class recall	59.72%	89.62%	

Fig. 11 Confusion Matrix of KNN

Table 2 shows the overall performance of each of the classifiers in terms of the F-measure. The results obtained from the analysis showed a promising result. The F-measure of the classifiers ranges from 86.46% to 73.01%. SVM had the highest overall f-measure score of 86.46%, followed by KNN, RF, and NB, who had 86.41%,81.90%, and 73.01%, respectively. Thus, the model can predict the sentiment polarity of the reviews with a very low error rate.

TABLE 2:
PERFORMANCE OF CLASSIFIERS

Classifier	F-Measure (%)
SVM	**86.46**
NB	73.01
RF	81.90
KNN	86.41

VI. CONCLUSIONS

The proposed framework deploys supervised machine learning techniques. The framework comprises six (6) stages. It starts with the data collection of user reviews from Google Play. The second is data labelling. Data labelling is important in supervised machine learning since good labelling will result in more accurate predictions. The third stage is data pre-processing, this process will produce a better-quality dataset for further processes. Fourthly, the frequency and importance of the words in the dataset will be performed by using the TF-IDF technique. Later, the pre-processed and vectorized data will be trained and tested using the supervised machine learning classifiers. SVM, NB, RF and KNN classifiers will be used in this framework. Lastly, the performance of the classifiers will be evaluated and compared to determine the best classifier. The F-measure scores has been observed since the result showed an imbalance in training dataset. The SVM classifier performs well compared to the other classifiers, with an F-measure score of 86.46%.

The analysis result obtained from the proposed framework indicates that the model is fit for this study as the f-measure shows a good score. Thus, the framework enables developers, researchers and firms to create and deploy machine learning models more quickly and efficiently. It will help them choose the best classifier to understand the polarity of people's reviews of the application developed.

Certainly, in the future, it will be interesting to look into different feature extraction techniques in the framework. So that the model performance of different techniques could be compared and evaluated. Hence it will improve the performance of the proposed framework.

REFERENCES

[1] E. Fersini, *Sentiment Analysis in Social Networks*. 2017.

[2] World Health Organization, *Communicating Risk in Public Health Emergencies*. 2017. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/259807/9789241550208-eng.pdf?sequence=2

[3] D. S. and L. Buis, "Mobile Health Apps for Medical Emergencies: Systematic Review," *Journal of Medical Internet Research*, 2020.

[4] P. H. O. archive pageTate R.-M. archive pageBobbie Johnson, "A flood of coronavirus apps are tracking us. Now it's time to keep track of them.No Title," *MIT Technology Review*, 2020.

[5] E. Gabarron, E. Dorronzoro, O. Rivera-Romero, and R. Wynn, "Diabetes on Twitter: A Sentiment Analysis," *Journal of Diabetes Science and Technology*, vol. 13, no. 3, pp. 439–444, May 2019, doi: 10.1177/1932296818811679.

[6] R. R. Pai and S. Alathur, "Assessing mobile health applications with twitter analytics," *International Journal of Medical Informatics*, vol. 113, pp. 72–84, May 2018, doi: 10.1016/j.ijmedinf.2018.02.016.

[7] K. Rekanar *et al.*, "Sentiment Analysis of User Feedback on the HSE Contact Tracing App," 2020, doi: 10.21203/rs.3.rs-96174/v1.

[8] F. Greaves, D. Ramirez-Cano, C. Millett, A. Darzi, and L. Donaldson, "Use of sentiment analysis for capturing patient experience from free-text comments posted online," *Journal of Medical Internet Research*, vol. 15, no. 11, 2013, doi: 10.2196/jmir.2721.

[9] J. Du, J. Xu, H. Song, X. Liu, and C. Tao, "Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets," *Journal of Biomedical Semantics*, vol. 8, no. 1, Mar. 2017, doi: 10.1186/s13326-017-0120-6.

[10] Q. Cheng, T. M. Li, C. L. Kwok, T. Zhu, and P. S. Yip, "Assessing suicide risk and emotional distress in Chinese social media: A text mining and machine learning study," *Journal of Medical Internet Research*, vol. 19, no. 7, 2017, doi: 10.2196/jmir.7276.

[11] S. Yadav, A. Ekbal, S. Saha, and P. Bhattacharyya, "Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System." [Online]. Available: http://sideeffects.embl.de/

[12] S. Park and J. Woo, "Gender classification using sentiment analysis and deep learning in a health web forum," *Applied Sciences (Switzerland)*, vol. 9, no. 6, 2019, doi: 10.3390/app9061249.

[13] O. Oyebode, F. Alqahtani, and R. Orji, "Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews," *IEEE Access*, vol. 8, pp. 111141–111158, 2020, doi: 10.1109/ACCESS.2020.3002176.

[14] T. Puyalnithi, M. V. V, and A. Singh, "Comparison of Performance of Various Data Classification Algorithms with Ensemble Methods Using RAPIDMINER," vol. 6, no. 5, pp. 1–6, 2016.

[15] W. R. Hersh *et al.*, "Recommendations for the use of operational electronic health record data in comparative effectiveness research," *EGEMS*, vol. 1, no. 1, 2013.

[16] S. McIlroy, N. Ali, H. Khalid, and A. E. Hassan, "Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews," *Empirical Software Engineering*, vol. 21, no. 3, pp. 1067–1106, 2016, doi: 10.1007/s10664-015-9375-7.

[17] R. M. Duwairi, M. Alfaqeh, M. Wardat, and A. Alrabadi, "Sentiment analysis for Arabizi text," *2016 7th International Conference on Information and Communication Systems, ICICS 2016*, no. April, pp. 127–132, 2016, doi: 10.1109/IACS.2016.7476098.

[18] P. Pang and L. Lee, "Sentimental education: extract subjective sentences from reviews," *ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, p. 8, 2004, [Online]. Available: file://localhost/Volumes/alexvanvenrooij 1/Elektronische Artikelen/Papers/Unknown/2004-8.pdf

[19] Ryan Miller, "Data Preprocessing: what is it and why is important," Dec. 13, 2019.

[20] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment analysis and classification based on textual reviews," *2013 International Conference on Information Communication and Embedded Systems, ICICES 2013*, pp. 271–276, 2013, doi: 10.1109/ICICES.2013.6508366.

[21] S. Singh, A. Pareek, and A. Sharma, "Twitter Sentiment Analysis using Rapid Miner Tool," *International Journal of Computer Applications*, vol. 177, no. 16, pp. 44–50, 2019, doi: 10.5120/ijca2019919604.

[22] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," in *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, Nov. 2016, pp. 61–66. doi: 10.1109/ICEEOT.2016.7754750.

[23] A. Madasu and S. E, "A Study of Feature Extraction techniques for Sentiment Analysis," Jun. 2019, [Online]. Available: http://arxiv.org/abs/1906.01573

[24] Sajid Lhessani, "What is the difference between training and test dataset?," Sep. 28, 2019.

[25] E. Zuo, H. Zhao, B. Chen, and Q. Chen, "Context-Specific Heterogeneous Graph Convolutional Network for Implicit Sentiment Analysis," *IEEE Access*, vol. 8, pp. 37967–37975, 2020, doi: 10.1109/ACCESS.2020.2975244.

[26] M. A. Ghaderi, N. Yazdani, B. Moshiri, and M. Tayefeh Mahmoudi, "A new approach for text feature selection based on OWA operator," *2010 5th International Symposium on Telecommunications, IST 2010*, pp. 579–583, 2010, doi: 10.1109/ISTEL.2010.5734091.

[27] G. N. A. S. Whitelaw C, "Using appraisal groups for sentiment analysis," 2005, pp. 625–631.

[28] X. Yiming Yang, "A re-examination of text categorization methods," 1999, pp. 42–49.