Paper No: SC-19 **Smart Computing**

Detecting human emotions on Facebook comments

E. J. A. P. C. Chathumali*
Faculty of Graduate Studies and Research
Sri Lanka Institute of Information Technology, Sri Lanka
chathujay92@gmail.com

Samantha Thelijjagoda
SLIIT Business School
Sri Lanka Institute of Information Technology, Sri Lanka
samantha.t@sliit.lk

***Abstract*: Human emotion detection plays a vital role in interpersonal relationships. From the early eras, automatic recognition of emotions has been an active research topic. Today, sharing emotions on social media is one of the most popular activities among internet users. However, when it comes to a specific domain like emotion detection in social media, it is still on a research-level. There are less number of applications have been developed to detect emotions online, using online comments and user comments. The aim of this research is to develop a system that identifies human emotions on Facebook comments. Among the different social media platforms, this research specifically focuses on Facebook comments written in the English language to narrow down the problem. The research is based on Semantic analysis, which comes under Natural Language Processing (NLP) and the system development consists of four major steps, including the extraction of Facebook comments via Graph API, preprocessing, classification and emotion detection. To classify the emotions, a classification model was created by using Naïve Bayes Algorithm. When it comes to marketing, emotions are what lead your onlookers to purchase. By using the detected emotions, marketers can promote their campaigns by changing online advertisements dynamically. The results obtained through testing the system show that it is capable of accurately identifying human emotions hidden in Facebook comments with an accuracy level of 80%, making it highly useful for marketing purposes.**

Keywords: *Emotions, Emotion detection, Facebook, Naïve bayes algorithm*

I. INTRODUCTION

Feelings or emotions play a critical role in human life and people express their feelings in everyday communications. There are different types of emotions, having an impact on the lifestyles of people and in their day-to-day interactions. People are mostly ruled by emotions. The choices of people, their actions, and perceptions are all impacted by the emotions they feel at a given time. Relational human correspondence incorporates dialect that is spoken, as well as non-verbal prompts as hand and body motions, the tone of the voice, which are utilized to express feelings and give criticism. Understanding and realizing how to respond to individuals' demeanor incredibly improves associations. Ekman and Freisn having expertise in this area, distinguished six fundamental emotions, including anger, fear, disgust, joy, surprise and sadness which is recognized as a general standard [1].

Today, most of the social media users are accustomed to sharing emotions in their social media platforms. They tend to record their daily life online and frequently share their thoughts, considerations and feelings, and connect with companions through the web. The main reason for this is the ascent of social network websites like Facebook and Twitter that has significantly changed how individuals live. According to the statistics of broadbandsearch.net, in 2018, 2.62 Billion people were using social media platforms [2]. By 2021, this number is expected to rise over 3 Billion. Facebook leads prominent social media platforms with 2.234 Billion active users in 2018 and is the third-highest visited website in the world [2].

According to the statistics of Escale Solutions it has been stated that, for every second on Facebook by January 2019, 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded [3]. With the help of those abundant data, there are great openings for considerate human behavior and more numerous new applications for enhancing individuals' lives. Apart from that, according to brandwatch.com, Facebook marketing statistics illustrates that there are over 60 Million active business pages currently running on Facebook and over 2 Billion people can be reached through these [4]. Therefore, Facebook is a great place for businesses and their marketing campaigns. On the other hand, most of the effective marketing campaigns are driven through human emotions [5]. All the marketing campaigns always focus on touching peoples' hearts and on building an everlasting bond with the customers. For instance, customers always decide on a purchase based on their emotions at that moment [5]. Emotions are the key that leads the onlookers or the customers to purchase. In addition to that, a significant number of the best promoting, marketing campaigns and activities are based on emotions. For example, once a user arrives at Facebook, if the technology can change the marketing advertisements dynamically according to user's emotions, it will be a great advantage and a turning point for businesses [5]. If marketing does not consider emotions of the intended consumers, businesses will be in trouble due to a lack of purchases.

The main objective of this research is to design and develop an application that is capable of identifying human emotions of online users in the Facebook social network. To achieve the main objective, the system uses the sub-objectives, including the analysis of existing methodologies of collecting Facebook data, collecting Facebook comments according to the identified methodology, implementing a model that classifies the human emotions and identifying human emotions according to published Facebook comments accordingly.

This system mainly focuses on the comments which have been published in a particular business page. To understand the meaning of the text, it uses the techniques which come under Natural Language Processing (NLP). NLP can be used to process both speech and text. The proposed system uses text processing to capture the meaning of words that have been collected from Facebook comments. In order to identify the real meaning of the text, the system uses Semantic

analysis, which comes under Natural Language Processing (NLP). Semantic analysis describes the method of understanding natural language the way that humans interconnect based on meaning and context [6]. By this foundation, this research utilizes a strategy to categorize emotion based on text instead of the three essential semantic classifications including positive, negative and natural. This has been done with the help of collecting data through automatic semantic analysis. For the proposed research, Facebook social media platform has been used to gather user comments for the particular business page, which was already published by a user. The research specifically focused on Facebook comments written in the English language to narrow down the research problem.

II. BACKGROUND STUDY

Binali, et. al. [7] discuss computational approaches for emotion detection in text. The aim of this project is to present emotion theories that deliver a foundation for emotion models. They present how these models have been utilized by examining computational ways to deal with emotion detection. Apart from that, in order to detect the emotions, authors have proposed a hybrid-based architecture. They have used SVM algorithm for validating the suggested architecture and accomplishes an expectation precision of 96.43% of the web blog information. This hybrid-based architecture is proposed to fulfill two different requirements, including hidden phrase pattern detection and the emotion prediction. In hidden phrase pattern, syntactic and semantic details are used to identify the emotions in the subtle sentences, whereas emotion prediction is used to find the emotion being stated, based on information-rich linguistic resources and trained classifiers.

Yassine et. al. [8] present a framework for examining the relationship between friendship and emotion expression in online social networks. The main objective of this framework is to extract or highlight the emotional content of the text which is already available in online social media. In order to identify the writer's emotions using text, they have performed text mining techniques on comments which are retrieved from social media. The framework is developed using numerous significant steps, including raw data collection, lexicon development, feature generation, data preprocessing, creating data model, text classification and friendship classification. All those features help recognize whether the online post expresses any emotions or feelings. Since the social media comments have informal language, authors have faced difficulties in sentiment analysis. To overcome this issue, they have proposed a framework with special lexicons. Moreover, they have tested this framework with the help of the k-means clustering algorithm and they have got high accuracy outcomes for the model in defining subjectivity of texts and forecasting friendship.

Sewwandi et al., [9] propose a web application to recognize the personality of humans by using linguistic feature analysis. They have used Facebook social media data and information when conducting their research. There are three major dimensions that they have used when developing the application, including ontology-based text classification, questionnaire-based personality detection and linguistic feature vector matrix using Linguistic Inquiry Word Count (LIWC) features such as semantic analysis with supervised machine learning algorithms. According to the authors, the main advantage of this application is its consideration of the personality when they are recruiting or promoting employees. They have conducted a feasibility study, covering the financial, technical and operational areas. From that, they have identified that there are no technical dependencies due to open source technologies. They have used a restful API to extract data from Facebook. As the first dimension of personality detection, the application uses an ontology-based approach. There, ontology is constructed using 'protege' software which has the OWL ontology developing language with the OWL-DL package. Moreover, ontology is designed along with a data set attainable by a recognized Physiologist in Sri Lanka.

Mashal et. al. [10] propose an approach to identify emotion, intensity using social media data. It consists of four major steps, including training data set, preprocessing, feature vector construction and emotion intensity prediction. As the first step they have trained the data set which covered four emotion categories including fear, anger, joy and sadness. Training data set included tweets and it has used real-valued score in the range of 0 to 1. That range is defined by the authors to represent a degree of emotion. After that they have done the preprocessing. According to the authors, preprocessing is very important because tweets are one hundred and forty words long in length and they state that there can be elements that mislead the classification process. They have done the preprocessing with the help of major steps, including conversion of words to lowercase, replacement of user mentioning with @user, replacement of letters and punctuations that are repeated and removing hashtags. As the third step they have constructed a feature vector. There, they have used WordNet files for each and every emotion type along with the physically interpreted words and a list of opposing words. To invert the meaning of the sentence negating words are used in the feature vector creation. Apart from those, they have trained the prediction model in the preprocessing. As the last step they have predicted the emotion by using the feature vector with the help of intensity which is given to the prediction algorithm as an input.

Rabeya, et. al. [11] have conducted a research to detect emotions in Bengali text at the sentence level. The research has used a backtracking approach which is based on the lexicon. In order to conduct the research, the authors have considered only two basic emotions; happiness and sadness. Apart from that, the research uses the most commonly used labeled databases; ISEAR and SemEval. The implementation of the approach is carried out in a few steps including data gathering and extraction, preprocessing, analyzing sentiment and emotion classification. To extract data for processing, authors have used sources such as journals, news, lines, textbooks, blogs, Facebook posts etc. Once the collected data has been extracted preprocessing is done for the tokenization of sentences. Tokenization process has been done for each and every input sentence generating the expression which can be used to calculate the sentiment as the next step of the methodology. To find out the emotions in the Bengali text, sentiment of each sentence has been identified by using the backtracking technique. There, the authors have used the expression from the back to find the sentiment. For instance, researches have considered that people have expressed their emotions at the end of the sentence and not in the position's

length -1 to -3. After that, the particular emotion is identified based on the sentiment associated with them. According to the authors, the proposed method resulted in an accuracy of 77.16%.

In a research conducted by Muljono, et. al. [12] identifies an approach to recognize the emotions using Indonesian text and they have evaluated the performance of four techniques including J48, K-Nearest Neighbor (KNN), Support Vector Machine-Sequential Minimal Optimization (SVM–SMO) and Naïve Bayes (NB). Authors have found that most Indonesians communicate with each other by using their own local languages, including Javanese, Sundanese and Balinese, but in formal situations Indonesian is used as the language. When conducting the research, they have collected data from different websites such as www.dongengceritarakyat.com, www.dongeng.org, and www.pendongeng.com. Their datasets consisted of one to three sentences. WordNet Affect List has been used to gain the emotional tag in the sentences. Also, they have manually labeled sentences which are not included in the WordNet Affect List. To increase the accuracy of the research they have collected 1000 sentences and stored in text format file. As the next part of the research they have done the preprocessing. They have used several steps, including tokenization, case normalization, stop word removal, and stemming to achieve the preprocessing. There, Lucene library has been included to develop the code from scratch. After that they have done the term weighting by using the TFIDF (Term Frequency Inverse Document Frequency) which is a popular term weighting scheme. As the next step classification has been done by using the two techniques; machine learning and supervised learning, and unsupervised learning. This research compares four classification algorithms, including Naïve Bayes (NB), J48, K-Nearest Neighbor (KNN) and Support Vector Machine-Sequential Minimal Optimization (SVM–SMO) to identify the best classification algorithm for textual emotion recognition. The experiment of these classification techniques is carried out using WEKA (Waikato Environment for Knowledge Analysis). Finally, they have done the evaluation with the help of 10-fold and split validation. Moreover, data have been divided in to testing and training by cross validation. There, the authors have defined that 90% of the total data have been used for training and 10% have been used for testing. Based on the experiments, they have stated that SVM-SMO classifier gives the best performance when compared to other classification algorithms.

Sarakit, et. al. [13] has conducted a research to classify the emotions in Thai YouTube comments. The main purpose of this research is to add more value to YouTube. They have used comments written in Thai language for a particular video to identify the emotions. Authors have considered the six basic emotions, including anger, fear, happiness, disgust, sadness and surprise. Their research consists of several steps such as collecting YouTube comments, pre-processing, feature selection, term weight and classification. As the first step, the authors have collected data from the YouTube. To do that, they have considered 50 Thai video clips. Among them, 20 video clips are music videos and 30 are video clips from commercial advertisements. In order to collect the data, they have used the YouTube 2.0 API. According to the authors, they have gathered a total of 2771 comments from Music videos and 3077 comments from advertisements. After that they have performed the pre-processing. Pre-processing includes the procedures segmenting comments into a set of word tokens, removing Thai and English stop words, Removing HTML tags and stemming words. Lexitron dictionary from NECTEC and LexTo: Thai Lexeme Tokenizer have been used for segmentation. As the third step, feature selection has been done by assigning a frequently threshold value for selected keywords. If the word has appeared more than the minimum threshold value, it is considered a keyword. If it has not, it is removed from the feature list. After that, they have assigned term frequency and term frequency inverse document frequency. Finally, emotion classification has been done by the standard classification methods including SVM, multinomial naive Bayes and decision tree namely J48. According to the authors, the highest accuracy is returned from the support vector machine (SVM). It generates an accuracy of 76.14% for commercial advertisements and 84.48% of music videos.

III. Methodology

In order to classify human emotions, the developed system includes a few major steps. As the first step, the system collects Facebook comments on a particular business page, which is already published. The next step is to preprocess the collected comments according to the common pre-processing procedures. As the third step, the system processes a critical step which is called classification. To do that, a classification model is created using the machine learning algorithm; Naïve Bayes. Apart from that, the model is trained using an existing data set. Finally, by comparing the keywords, the system classifies human emotions. The classification process is done for every hour and generates the highest occurring emotion as the output. Likewise, this process is continued for each and every hour to classify the emotions.

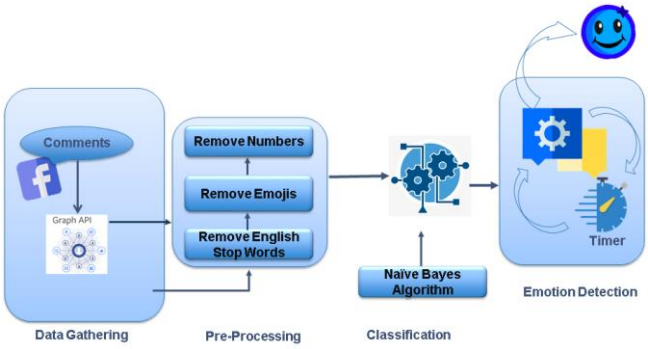

Fig. 1. High-level diagram

A. Data collection

The very first step of the research was to collect the data or comments of users from Facebook. To do that, Facebook Graph API has been used. Graph API, is an inbuilt API provided by Facebook for their developers. It is a primary way of getting data in and out from the Facebook platform. Moreover, Graph API [14] is based on HTTP and it can used for the applications to programmatically query data, manage advertisements, upload photos etc. According to Facebook, they have allowed developers to extract the comments from the API only if he/she is an admin of a particular page or their

own comments on their own posts. Since Facebook wants to protect user data, they have followed this mechanism. To extract any information through the Graph API, developers need to log in to "Facebook for developers" website and get their access token.

B. Pre-processing

Once the data have been collected through Graph API, next step was processing them according to the common pre-processing procedures. Those procedures include, eliminating English stop words, removing numbers, removing emoji's, removing double quotations, slashes, underscores etc.

C. Classification

As the third step, the system has performed the critical process of classification. To do that, a classification model was created by using the machine learning algorithm Naïve Bayes. Naïve Bayes belongs to the probabilistic algorithm family and it uses Bayes's Theorem to predict the text category. Naïve Bayes classifiers are extremely scalable and require a number of parameters in a learning problem. The model was trained using the existing data set. When creating the model, tags or the labels should be created as the first step. The research has used the emotion tags, including Happiness, Sadness, Anger and Surprise.

D. Emotion Detection

Finally, preprocessed data has been sent to the model via API and by comparing the key words, the model the classified human emotions as the output.

IV. Experimental Results

This chapter illustrates the results of the above steps. As the first step, the system has collected data via Facebook Graph API and comments have been returned as a Json object. After gathering the data, Json object should be deserialized in order to covert the Json string to a particular custom object. After that, the collected data has been assigned to a list which includes comments put up during the last hour. As the next step, preprocessing has been performed. The resulted values have been sent to the model as a byte array. Then, the model has classified the comments one by one and resulted in the output. Finally, among the resulted values, the system has returned the maximum emotion with an image, and all the results, including sadness, happiness, anger and surprise are displayed in a pie chart. The process is being repeated until it classifies the emotions in the next hour. Similarly, this has been continued until the user exits from the system.

The following figure shows the most recent hourly data gathered via Facebook Graph API. It has returned the unprocessed comments and stored them inside a Json object. After that, the Json object has been deserialized and converted to a particular custom format.

Once the data have been collected a list has been created by assigning all the comments which are published in the last hour. The comments have been preprocessed before sending to the model. Then, resulted values have been sent to the emotion classification model in order to classify the output. Once the model has returned the output, the system calculates the number of emotions expressed during the last hour.

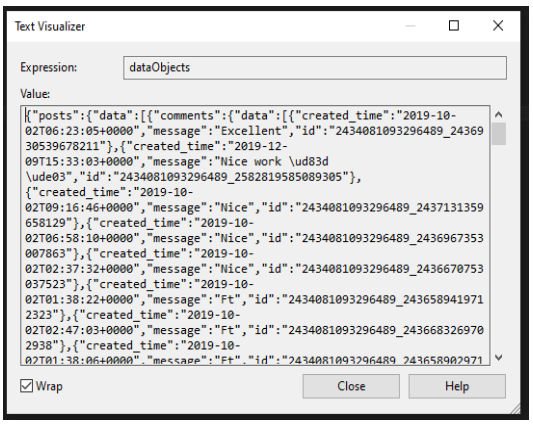

Fig. 2. Collected Facebook comments

Finally, among the values the system has returned, the emotion having the maximum count with an image and all the results, including sadness, happiness, anger and surprise have been displayed in a pie chart.

Fig. 3. Graphical representation of detecting emotion during the last hour

Once all the outputs have been generated, the process repeats the same steps and continues emotion detection until the user exits from the system.

According to the final outcome of the research, an accuracy level of 80% is achieved. To find the accuracy of the results, confusion matrix has been used. It has given the accuracy by comparing the actual and predicted classes. Figure 4 given below shows the classes of the confusion matrix.

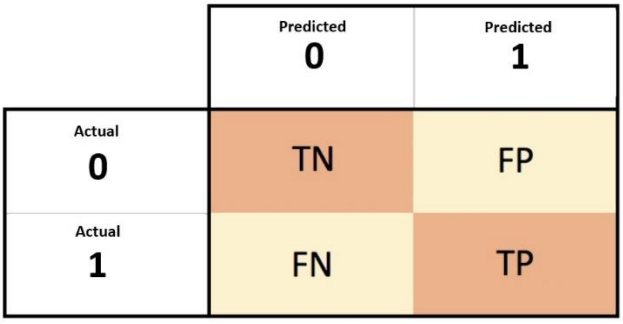

Fig. 4. Graphical representation of confusion matrix

TP (True Positive) which describes the predicted values are correctly predicted as actual positive. FP (False Positive) defines the predicted values are incorrectly predicted an actual positive. For an example, negative values predicted as positive. FN (False Negative) defines that the Positive values are predicted as negative.TN (True Negative) describes the predicted values are correctly predicted as an actual negative.

For example, if the user has commented “Excellent job, poor work, amazing talent, prices are high and great work”, the system returned, the output emotions as happy, sad, surprise, sad and happy respectively. Therefore, those values have belonged to the TP class. Furthermore, there were some failed outputs as well. For example, if the user has published comments ‘Not happy, not good’, the system generates the output emotion as ‘happy’ instead of the actual result ‘sad.’ Therefore, this kind of outputs belong to the FP. The following equation (1) has been used to calculate the accuracy level.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

V. Conclusion

Feelings or emotions play a critical role in human life and people express their feelings in everyday communications. Almost all the people in the world share their emotions on social media. The purpose of the research is to identify emotions of online users by using Facebook comments. Selecting Facebook from numerous social media platforms, the research specifically focuses on Facebook comments written in the English language in order to narrow down the problem. When it comes to marketing, emotions are what lead the onlookers to purchase. Therefore, identification of user emotions can be helpful for businesses in various scenarios. For example, when a user arrives on Facebook and if the technology can change the marketing advertisements dynamically according to user’s emotions, it will be a turning point for a business. Therefore, the main objective of the developed system was to identify emotions of the online users by extracting their comments. The system uses the techniques in semantic analysis and the major steps include data collection from Facebook, preprocessing, classification and emotion detection. To identify the emotions, a classification model has been created by using the Naïve Bayes algorithm. Finally, the output is graphically represented by using a pie chart.

REFERENCES

[1] A. Varghese, J. P. Cherian and D. J. Kizhakkethottam, "Overview on Emotion Recognition System," in *International Conference on Soft-Computing and Network Security*, India, 2015.

[2] "broadbandsearch," [Online]. Available: https://www.broadbandsearch.net/blog/internet-statistics. [Accessed: 14-01-2019].

[3] B. Bot, "Escale Solutions," Top 20 Facebook Statistics in January 2019 - Escale Solutions, [Online]. Available: https://www.escalesolutions.com/blog/top-20-facebook-statistics-in-january-2019/. [Accessed: 28 -01-2019].

[4] "Brandwatch," [Online]. Available: https://www.brandwatch.com/blog/facebook-statistics/. [Accessed: 28-11-2019].

[5] "The Ultimate Guide to Emotional Marketing," hubspot, [Online]. Available: https://blog.hubspot.com/marketing/emotion-marketing. [Accessed: 11-01-2019].

[6] "Expertsystem.com," Natural Language Process semantic analysis: definition, [Online]. Available: https://www.expertsystem.com/natural-language-process-semantic-analysis-definition/. [Accessed: 02-02-2019].

[7] H. Binali, C. Wu and V. Potdar, "Computational Approaches for Emotion Detection," in *4th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST)*, Dubai, United Arab Emirates,2010.

[8] M. Yassine and H. Hajj, "A framework for emotion mining from text in online social network," in *IEEE International Conference Data Mining Workshops*, USA, 2010, p. 1136–1142.

[9] D. Sewwandi, K. Perera, S. Sandaruwan, O. Lakchani, A. Nugaliyadde and S. Thelijjagoda, "Linguistic Feature Based Personality Recognition Using Social Media Data," in *6th National Conference on Technology and Management (NCTM)*, Malabe, 2017.

[10] S. X. Mashal and K. Asnani, "Emotion Intensity Detection for Social Media Data," in *IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC)*, Tamil Nadu, 2017.

[11] T. Rabeya, S. Ferdous, H. S. Ali and N. R. Chakraborty, "A Survey on Emotion Detection A lexicon based backtracking approach for detecting emotion from Bengali text," in *20th International Conference of Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2017

[12] Muljono, N. A. S, Winarsih and C. Supriyanto, "Evaluation of Classification Methods for Indonesian Text Emotion Detection," in *International Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia, 2016.

[13] P. Sarakit, T. Theeramunkong, C. Haruechaiyasak and M. Okumura, "Classifying Emotion in Thai YouTube Comments," in *6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES*), Hua-Hin, Thailand, 2015.

[14] "Facebook for developers," [Online]. Available: https://developers.facebook.com/docs/graph-api/overview/. [Accessed: 25-05-2019].