

Student Feedback Analyzer for E-Learning Platforms

Shashiprabha T.A.S

Department of Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Liyanage I.M

Department of Software Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Lakmal Rupasinghe

Senior Lecturer: Department of Information Systems
Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Chethana Liyanapathirana

Lecturer: Department of Information Systems Engineering
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka

Abstract:- Evaluation of feedback is important in identifying weaknesses and action taking for better systems integration, and in maintaining system efficiency. Most e-learning systems are not configured with a proper user feedback evaluation framework such as e-commerce systems to improve product quality and product reviews by identifying consumer real expectations. In the field of education, most higher learning institutions, universities collect qualitative and quantitative feedback manually and digitally to improve the performance of students learning progress and teaching instructors. But domain-specific e-learning platform feedback evaluations are rare. The sentiment analysis is the most common qualitative feedback analysis. Sentiment analytics systems have become increasingly popular for obtaining data about the perspective learning style of the students. Developing these systems from scratch is a difficult task, so many researches use non-domain-specific, commercial, general-purpose tools. Nevertheless, most existing techniques of sentiment analysis focus only on the abstract level, broadly classifying sentiments into positive, neutral or negative, and lacking the ability to perform fine-grained sentiment analysis. In this study, we propose a supervised ml-based sentiment analysis model that includes five distinct classes (1-5) labeled as strongly negative, negative, neutral, weekly positive and weekly negative in order to provide instructors with rapid and better understanding. Most of the current ml-based Sentiment Analysis models used SVM and Naive Bayes algorithms. We proposed our own ensemble model in this paper by joining three algorithms to boost performance.

And we also included supervised model suggestion mining with the option mining model above mentioned.

To boost the learning material performance of e-learning platforms, this kind of more reliable domain-specific feedback evaluation seems to be very important. In this paper, we propose a web service based on the API to obtain and evaluate feedback through the use of machine learning and AI (Artificial Intelligence) technologies to decide the most suitable way to deliver learning materials (student perspective). In both tasks the program achieves a high accuracy: Extraction of sentiments (84 percent) and analysis

of suggestions (97 percent). Since this method is a micro service we can use this approach with e-learning systems of any kind.

Keywords- Machine Learning, Ensemble, Sentiment, Student Feedback, Accuracy, Suggestion mining, Domain – specific

I. INTRODUCTION

E-learning is a booming sector, and is also rising rapidly. Udemy, Coursera and Lynda are the industry's leading platforms. Statistics show that the Global E-Learning Market is expected to rise at a CAGR (Compound Annual Growth Rate) of about 7.2 per cent over the next decade to reach about \$325 billion by 2025[5]. Considering the protection of time, money, and resources, students now enjoy electrical learning through e-learning platforms. On the other hand, we will see a lot of new e-learning technologies in the years ahead that undoubtedly implies more competition. To accomplish this contest the owner needs determine the students' perception and their expected level of learning materials and course content. Evaluation of student feedback is the best way of determining student perspective. Gathering student feedback for any e-learning platform is a comprehensive task. It enables owners of platforms to listen to their students and understand them, examine their feedback and take appropriate action. Even quantitative feedback is very straightforward and reliable and qualitative feedback can provide a thorough interpretation of the real aspect of the students..

Sentiment analysis - As users we see the rating is a quantitative reflection of qualitative inputs. Anyway, the written text is much richer with detail than a basic rating of stars. The most common method of analyzing qualitative reviews is by analyzing sentiments. Study of sentiment is an analysis of the views of people on a particular object, human, text, etc. This is their thoughts that reflect their feeling about a lone entity whether we like it or not. It is a mechanism by which the views given in a

document are computationally classified and graded to identify whether they are positive, negative or neutral.

Five polarity sentiment analysis - Past researchers have proposed groundbreaking techniques for evaluating the text's polarity. Traditionally they divide the text into three different levels of polarity i.e. Positive, negative and neutral where the positive class includes those documents where the positive language has been used, while the negative class includes those documents in which the user does have some bad experience with a product and the neutral class provides those records which are neither positive nor negative.

Throughout this research paper, we tried to introduce e-learning domain specific sentiment analysis which contains five polarity: Strongly negative, negative, neutral, positive and strongly positive. Binary class labels can be adequate to research large-scale positive / negative opinion patterns in text data such as tweets, product reviews or user feedback, but their limitations do remain. For example: "This python course is excellent. But tutorials are little bit unorganized" when performing a positive review which contains some tiny issue from reviews — a five polarity sentiment analysis can provide more reliable results to an automated system which prioritizes addressing customer complaints than binary classification.

Moreover, sentences with dual-polarity like "The course was very disgusting ... But the instructor was great there. "Can confuse the binary classifiers of feelings, leading to inaccurate class predictions.

According to our domain (e-learning), instructor and course owner will be able to get very good idea about the stage of quality of the course by using above type of five polarity sentiment analysis. The points above have enough inspiration to tackle this issue!

"The course is very awesome" for example suggests a very powerful positive opinion, while "the course is good" implies a weak positive attitude. This encourages us to conduct research on analyzing university student feedback on five levels than two levels to getting the real aspect of students.

ML classification algorithms—Classification of sentiments is meant to assess the general meaning of a written text and may be a form an admiration or criticism. It can be done by the use of machine learning algorithms such as Naive Bayes and Support Vector Machine. Thus, the problem to be discussed in the project would be:

Most researchers used classification algorithms such as SVM and multinomial naive Bayes ml algorithms to produce more reliable results in machine learning based sentiment analysis models. Here we used our own ensemble model of three ml SVM algorithms, Naive Bayes and logistic regression. What approach of machine learning does best in terms of accuracy on reviews of university student (e-learning)?

II. BACKGROUND

I centered on previous research works related to text analysis, sentiment analysis, and machine learning algorithms in this section.

I centered on works related to text analysis, sentiment analysis, NLP, machine learning algorithms, ensemble models, education domain and data preprocessing techniques relevant for feedback analysis in this section.

Sentiment analysis is an NLP methodology that can derive feelings from a text. This attempts to differentiate the data using controlled or unsupervised approaches into positive negative polarities. Sentiment mining techniques and tools are popular and widely used in consumer-business product review mining. Most researchers frequently find alternative approaches to analyze subjective consumer feedback on e-commerce sites. Salthuri Vanaja has performed some research based on aspect level input study to consider the brand strategies from the consumer perspective [2]. While contrasting two classification algorithms for machine learning she had achieved more precision from the classification of Naive Bayes than the classification of Support Vector Machine and Her method includes Nouns, Pronouns, Verbs, Adjective tags for each word with customer reviews and adjectives derived using classification algorithms [2]. She used ' SentiWordNet ' to generate positive negative, neutral scores for each word [2].

Zarmeen Nasim et al suggested a hybrid model that was trained using TF-IDF and lexicon-based features to evaluate the sentiments conveyed by students in their textual feedback [1].The method presented here is restricted to computing the general sentiment of the student reviews. They contrasted their domain-specific approach to other current APIs at the evaluation stage. Five polarity (fine grained) sentiment analysis for greater accuracy will be their future work [1].

Emitza Guzman et al suggested an ensemble framework for classifying app reviews, and also demonstrated an ensemble model that provides better value than single algorithms for feedback sentiment analysis [3].

Hassan Razal et al proposed a method of sentiment analysis for the scientific document. Various machine learning classifiers, including NB, SVM, DT, LR, KNN and RF, have been used along with various features to process data and improve classification results. Classifier accuracy is measured using different assessment metrics such as F-score and Accuracy score. The findings reveal that SVM is doing higher than other classifiers. Naïve Bayes performs well after SVM. In the case of the macro average, the efficiency of the SVM classifier is better when evaluating the F-score, and the precision is better when measuring the random forest in the case of the micro average. Three polarities (positive, negative and neutral) were used here [4].

P.Sanjay Bhargav et al introduced by Machine Learning Algorithms using Naïve Bayes Algorithm and Opinion mining techniques focused on Natural Language Processing. They used only 2 polarities: positive and negative [6].

Ms Jabeen Sultana et al. proposes a Deep Learning model for the study of sentiment on educational details. In this paper they concentrated on identification of the best model for the accuracy and efficiency of training data collection. The performance of the student model was evaluated through a rank-and - file series: SVM, MLP, Decision Tree, K-Star, Bayes Net, Single Logistics, Multiclass Classification and Random Forest. MLP and SVM have been recognized as outperformance models. Ten-fold-cross-validation (CV) is carried out. Results suggest that, compared to the other classificatory in terms of classification precision, RMSE, sensitivity and expertise, and ROC curve field, both SVM and MLP-deep learning methods have generally granted high performance [7].

III. METHODOLOGY

The proposed approach (review analyzing) is outlined in this section. The diagram below (figure 1) illustrates the proposed API demonstrating purposeful student input research methodology to high-level diagram, and describing each phase of the high-level diagram.

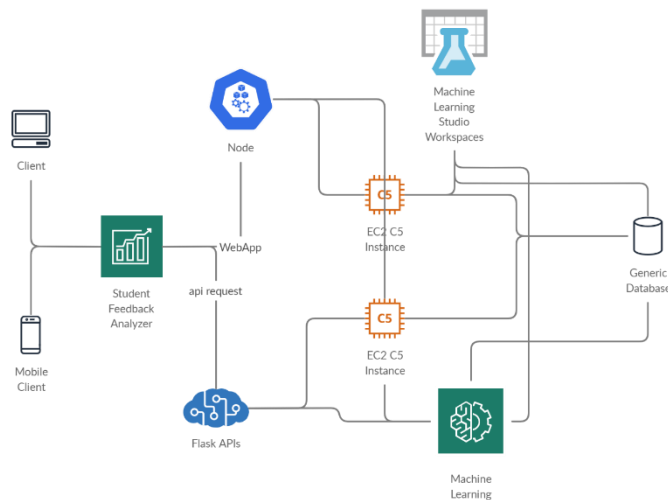


Figure 1- High Level Diagram of Proposed API

There are two EC2 instances .These two instances are specially stored in servers in which both backend and frontend will run on these servers .And flask will be the back-end API server .Node will be the front end server to front-end.

When student feedback analyzer request for a API of sentiments back end flask will response with the help of the machine learning model. Machine learning model is periodically improved through machine learning studios jupyter.Gragual changes are stored in the generic database with versioning.

When the web application is access the mobile client for a sentiment flask will response to the client feedback analyzer with

the latest machine learning model. Different sectors data set are checked.

The proposed approach (review analyzing) is outlined in this section. Firstly, the student review dataset creation is prepared, and is then used for the study of sentiments.

Methodology on ml model building is shown in below figure 4 .The suggested technique consists basically of three steps i.e. Data preprocessing and normalization, feature engineering and classification that are discussed below:

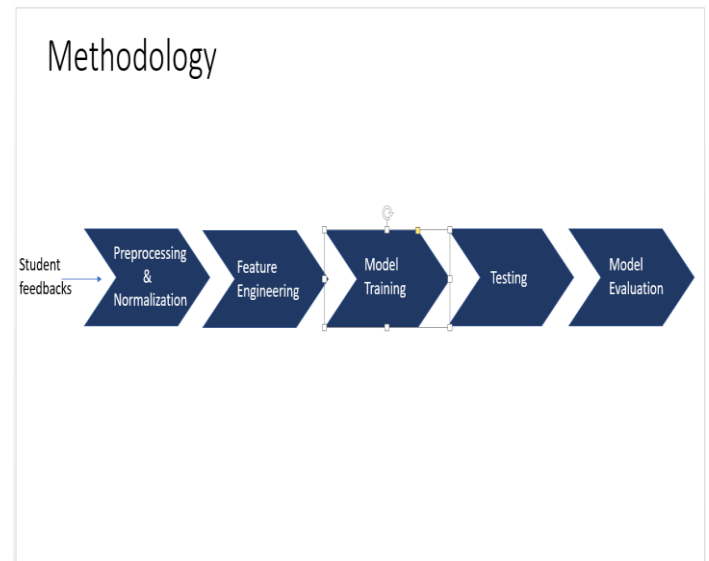


Figure 2- ML Model Pipeline

PHASE 1: Data Pre-Processing and Normalization

Data Set - The subject of our dataset is feedback from IT university students who have taken Coursera courses that are allocated to IT undergraduates. Initially we evaluated specifically the data collection that we had used. We had chosen the dataset from the Coursera e-learning platform. I obtained this dataset from Keggale machine learning repository. The whole collection of data had about 20000 sample numbers. Sample has the following fields: course id, scores, review (by a course user), and sentiments (1-5 score) suggestions (whether or not a suggestion is offered by review).It has been divided to two samples as test and train (75%-25%).

The raw data must be pre-processed to increase the consistency and efficiency of the classification process. The role of pre-processing deals with the process of preparation which eliminates repetitive words, non-English characters and punctuations. It improves data skills and appropriateness. It involves eliminating non-English letters, eliminating stop words, removing accented characters, expanding contractions, removing repeated characters, removing URLs, removing hashtags, handling negations, and handling emoticons.

Pre-processing is an intermediate step in the classification of text and emotion. A vast range of methods are used to increase the

efficiency of classification. This makes it easier to standardize across a corpus of documents that allows to create meaningful features and decreases dimensionality and bruising created by several factors such as irrelevant symbols, special characters, XML and HTML tags etc.

- **Stripping html tags**

Our text often includes needless elements such as Html tags that don't bring much meaning to sentiment analysis. Therefore, before extracting features, we should be confident we delete them. This is an excellent job for the BeautifulSoup library to provide the necessary functions.

- **Removing accented characters**

We deal with English feedback in our dataset, so that characters in some other format, and in particular accented characters, are translated and standardized into ASCII characters.

- **Contractions expanding**

Contractions of words or syllables, in English, are simply shortened. In standard text, contractions pose a challenge since we have to work with special characters such as the apostrophe and often have to transform each contraction into its initial and extended form.

- **Special Characters Removal**

Simple regex function used for achieved this.

- **Text to lemmatize**

Word stems are also the basic shape of possible words and can be formed by adding prefixes and suffixes to the stem in order for new words to be produced. It's called inflection. The reverse process is known as stemming when acquiring the base form of a phrase. A number of rods like Porter Stemmer and Lancaster Stemmer are included in the nltk kit. The lemmatization is quite like stemming, where we eliminate word affixes in order to obtain the basic form of a word. In this case, though, the base form is known as the root, but not as the root. The difference is that the word root is still a word that is lexicographically right, but the root stem cannot be accurate. We used lemmatization only for lexicographically accurate terms in our normalization process.

- **Stopwords Elimination**

Words with little or no significance in particular when creating meaningful text characteristics are often known as stopwords or words. There are typically terms with the highest frequency in a text corpus whether you use a simple term or word frequency. Words such as a, an, the ... are known as stopwords. There is no common stopword list, but a standardized list of stopwords is provided by nltk in English.

PHASE 2: Feature Engineering

We had used some pre-training formulas, like Term Frequency Inverse Document Frequency and Count Vectorizer, to convert the analysis texts into numeric data. Since we decided to encode the value of the existence of such words, the explanation for using

these algorithms is that these algorithms encode the value and existence of words in different forms. We used term frequency-Inverse Document Frequency embedding to determine the numeric frequency matrix for each word t in each review text. When, the term frequency $tf(t, d)$ measures the number of times the term t initially appeared in document d . Document d builds the vocabulary $V(d) = \{t | n(t, d) > 0\}$. So if a word w doesn't exist in a document d , the term frequency $tf(t, d)$ will be zero in this case. The definition of the term frequency is in principle the same as CountVectorizer.

$$tf(t, d) = \frac{n(t, d)}{V(d)}$$

$n(t, d)$ = occurrence of the word t in the document d

The inverse frequency of the subject granted a document set D $idf(t, D)$ is the log of document number N divided by $df(t, D)$, the quantity of documents $d \in D$ containing the term t . Popular terms in D will also have a low-term frequency score, while rare words will have a high-term frequency.

$$idf(t, D) = \log \left(\frac{N}{df(t, D)} \right)$$

To summarise, TF-IDF score $w(t; d)$ increases with its count for a word, but would be counteracted if the word exists in too many documents.

$$w(t, d) = tf(t, d) \times idf(t, D)$$

Equally, count vectorizer is given by a value matrix with each value reflecting the word's count frequency within that document (review). This matrix reflects the one hot encoded description of the numerous words found in the corpus. Entry a_{ij} = the cumulative number of times the word j th appears in the document i th.

We did this individually with each of the various circumstances because we wanted to validate how words would play a major role in the study of sentiments under different conditions.

PHASE 1: Model Training

We had planned to attend to test these following algorithms accurate analysis of sentiments using the above numeric representation techniques. We used such algorithms like svm and logistic regression with voting method for create our ensemble model.

- **SUPPORT VECTOR MACHINE**

Support vector machines are probably one of the most well-known and spoken about algorithms for machine learning. It remains the standard at the time it was developed in the 1990s and is now the best methodology for a high-performance, small-tuned algorithm. It is a discriminating classifier, since the algorithm produces the optimal hyper plane that categorizes new instances, given labeled training data (supervised learning). The algorithm is able to forecast uncertain data on the basis of this training.

- **NAIVE BAYES CLASSIFIER**

Naive Bayes is a classification algorithm based on probability commonly used by the scientific community. Naive Bayes classification completely relies on the hypothesis that the presence of a specific attribute is unassociated with the presence of any other feature in a class. Model of Naive Bayes useful for large collections of data. Naive Bayes is considered to be much more advanced methods of classifying with ease. Google is currently using it for spam or not to classify an email. Some new organizations often use this tool to identify news into various categories, such as technology, politics, entertainment and sports etc.

- **RANDOM FOREST CLASSIFIER**

Random Forest is a flexible method of machine learning which enables regression and classification tasks. It also uses dimensional reduction techniques, tackles missed values, outliers and other main data discovery measures and performs a very successful work. It's a kind of ensemble learning strategy, which mixes a set of weak models with a strong one. A tree provides the classification of a new entity based on attributes and we say "votes" to this class for a tree. The forest selects a class with the most votes (over all forests) and it classifies takes contributions from multiple trees on average.

- **LOGISTIC REGRESSION CLASSIFIER**

Logistic regression is an algorithm for classification which assigns observations to a collection of discrete groups. E-mail Spam, Web Deception or not Deception, Tumor Malignant or Benign are some of the examples of labeling concerns. Logistic regression transforms its output by means of the logistical sigmoid function, which returns a probability value. Logistic regression is a machine learning algorithm used to address classification problems based on the concept of probability.

- ❖ **ENSEMBLE MODELS**

Ensemble learning refers to the methods used to prepare and integrate the results of many apprentices, considering them as a decision maker team. The idea is that the decisions of the committee should be better, on average, than any particular committee member with individual forecasts properly combined. Many theoretical and observational experiments have found that ensemble models are more reliable than single models most often. The ensemble members might forecast true quantities, class names, posterior probabilities, classifications, clusters, or some other quantity. Their judgments can also be mixed using a variety of different approaches, including averaging, voting and

probabilistic processes. The bulk of ensemble learning approaches are common and can be utilized across wide-ranging frameworks and learning activities.

Ensemble methods are strategies that create and then merge several models to achieve better performance. Methods together provide typically more precise solutions than does a single model.

- **VOTING CLASSIFICATION**

The voting classification is a machine learning model focused on the highest likelihood of a chosen class as the result, which trains in a category of several models and forecast outcomes (class). It essentially aggregates the outcomes of each voting classifier and predicts the output class based on the largest vote majority. The concept is to build a common model that trains these models and forecasts output by the cumulative majority of voting for each output class instead of making different individual models and testing correctly for each.

PHASE 5: Model Evaluation

The evaluation metrics provide a deeper view into a multi classifier's output features. **Accuracy** (A_i) is also used as a categorization metric. However, consistency attributes are far less reluctant than specific and reminder can differ in the number of correct decisions:

$$A_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$$

Precision (ai) is calculated as a criterion for a random text (d) being listed as a sub-category or as a right one. It reflects the capacity of the classificatory to identify a document as correct and incorrect in contrast to other documents in the category:

$$\pi_i = \frac{TP_i}{TP_i + FP_i}$$

The precision of a classifier is calculated by accuracy. Precision means less false positive, thus poorer accuracy means more false positive. This is also in conflict with recall, decrease recall is a convenient way to improve precision.

Recall (pi) is defined as the probability that this decision will be made if a random document dx is listed under category (ci)

Recall tests the classifier's completeness or awareness. Higher recall means fewer false negatives, and less recall means more false negatives. Improved recall will also minimize precision as it grows more and more more difficult to be accurate with expanded sample size.

F-measure Metric: precision and recall can be merged in order to create the single, weighted mean of precision and recall, metric called the F-measure. The findings will be analyzed in the final step of this work to identify the challenges, changes and how the study can be expanded. There will also be a description of the progress reached and the scope for the future. In accordance with the previous articles a comparative description of the work actually under way and the work planned is produced.

IV. RESULTS & DISCUSSION

There are some ruled based sentiment analyzing tools like Vadar, TextBolb that can be customized for five state sentiment analysis. In here we can use their compound value for the analysis. But thing is they are not more accurate and domain specific. In the below table, I compare our application with above mentioned tools.

Approach	Accuracy
Textbolb	37.30%
Vadar	57.62%
Proposed Approach	80.84%

TABLE I

SUMMARY OF ACCURACY SCORES OBTAINED FOR DIFFERENT CURRENT TOOL AND OUR APPROACH FOR STUDENT FEEDBACK DATASET

In most of previous sentiment analysis ml based applications, researchers have been used naïve bayes, random forest and SVM algorithms for most accurate results. In here we used our own ensemble model using SVM, naïve bayes and logistic regression together. In below table 2, we compare our ensemble model with other existing models.

TABLE 2

SUMMARY OF ACCURACY SCORES OBTAINED FOR DIFFERENT MODEL FOR REGISTRY STUDENT FEEDBACK DATASET

	CV				TF-IDF			
	Accuracy %	Precision %	Recall %	F-Score %	Accuracy %	Precision %	Recall %	F-Score %
SVM	78.35	74.33	78.35	75.36	80.59	74.49	80.59	80.59
Logistic Regression	79.97	75.64	79.97	77.07	79.02	70.25	79.02	69.97
Naïve Bayes (Multinomial)	80.14	75.30	80.14	76.42	78.93	62.29	78.93	69.63
Random Forest	79.37	72.48	79.37	72.12	78.57	73.80	79.57	72.20
Proposed Ensemble Model	79.30	72.25	79.30	72.05	80.84	75.08	80.84	76.66

V. CONCLUSION & FUTURE WORK

We observed our ensemble model gives best performance than other single machine learning algorithms. And also purposed ensemble model gives best accuracy for tfidf features than CV features. The CV primarily reflects the number of words present in the feedback, while the significance of the words inside the review is defined by TFIDF. In here we achieved emoticons detection and negation handling challenges of sentiment analysis. Future work would be sarcasm detection and spam review detection. The paper outlined an ensemble method for sentimental analysis on reviews from students. The methodology

proposed involved machine learning approaches, along with 5 subjects of polarity emotions. It analyzed other APIs for sentiment analytics and contrasted the findings with the collection of five label to the study of polarity sentiment. It was found that TF-IDF and domain-specific e-learning student feedback analysis ensemble model was used to achieve the best results.

ACKNOWLEDGEMENT

We hereby acknowledge that the research work submitted to the Sri Lankan Institute of Information Technology under the direction of our Supervisor, Dr. Prabath Lakmal Rupasinghe and our Co-Supervisor Miss Chethana Liyanapathirana, titled "Student Feedback Analyzing Assistant Using Micro Services Architecture" is really a documentation of an original creation performed by me. This research study is applied in conditional fulfillment of the Bachelor of Science Special (Honors) Degree in Information Technology curriculum. The findings contained in this article have not been applied for grant of any degree or qualification to any other university or organization. Knowledge extracted from others ' published or unfinished research was mentioned in the article, and a list of references is given.

REFERENCES

- [1] Z. Nasim, Q. Rajput and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, Langkawi, 2017, pp. 1-6, doi: 10.1109/ICRIIS.2017.8002475.
- [2] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data", *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2018. Available: 10.1109/icirca.2018.8597286 [Accessed 19 February 2020].
- [3] E. Guzman, M. El-Haliby and B. Bruegge, "Ensemble Methods for App Review Classification: An Approach for Software Evolution (N)," *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, Lincoln, NE, 2015, pp. 771-776, doi: 10.1109/ASE.2015.88.
- [4] Raza, H., Faizan, M., Hamza, A., Mushtaq, A. and Akhtar, N., 2019. Scientific Text Sentiment Analysis using Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications*, 10(12).
- [5] R. Markets, "Global E-Learning Market to Reach \$325 billion by 2025 - Rapid Growth in Online Content & Digitization / Innovations in Wearable Technologies are Flourishing the E-learning Industry", *GlobeNewswire News Room*, 2020. [Online]. Available: https://www.globenewswire.com/news-release/2017/02/06/914187/0/en/Global-E-Learning-Market-to-Reach-325-billion-by-2025-Rapid-Growth-in-Online-Content-Digitization-Innovations-in-Wearable-Technologies-are-Flourishing-the-E-learning-Industry.html?utm_source=emojics&utm_campaign=How%20to%20Collect%20User%20Feedback%20for%20Your%20E-Learning%20Platform. [Accessed: 21- Feb- 2020].
- [6] N. Altrabshah, M. Cocea and S. Fallahkhair, "Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback", *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, 2014. Available: 10.1109/ictai.2014.70 [Accessed 21 February 2020].
- [7] J. Sultana, N. Sultana, K. Yadav, and F. Alfayez, "Prediction of Sentiment Analysis on Educational Data based on Deep Learning Approach," *2018 21st Saudi Computer Society National Computer Conference (NCC)*, 2018.