A SENTIMENT ANALYSIS-BASED SMARTPHONE APPLICATION TO CONTINUOUSLY ASSESS STUDENTS' FEEDBACK AND MONITOR THE QUALITY OF COURSES AND THE LEARNING EXPERIENCE IN EDUCATIONAL INSTITUTIONS

Sarah A. Alkhodair
Assist. Prof. Dr., Information Technology Department, CCIS, King Saud University, Saudi Arabia, salkhudair@ksu.edu.sa

Abstract

The quality of education in a specific educational institution is directly reflected in the outcomes of their system. Higher-quality educational systems continue to deliver better learning experiences to enrolled students and better-developed skills and knowledge. To provide high-quality education, an institution must continually monitor its plans, update its courses' topics and curriculum, and improve teaching facilities and different learning experiences. Students' opinions and feedback regarding different aspects of a course and their personal learning experience, if properly gathered and analyzed, can be strong indicators of the quality of that course and help identify the areas of satisfaction and dissatisfaction with that course. Highlighting the strengths and weaknesses of each course helps faculty members put, execute, and evaluate a course quality improvement plan in the following semester. Such valuable students' feedback and opinions about courses are scattered throughout different social media platforms and managed by different discussion groups, usually students. Thus, gathering honest and freely written comments and opinions in one place is challenging. Furthermore, extracting and analyzing courses' quality and learning experience-related posts is not a trivial task. This study describes the process of designing and developing a smartphone application utilizing Sentiment Analysis techniques to address the problem of gathering, analyzing, and understanding students' feedback and comments regarding different aspects of courses quality provided by an educational institution. The project's primary goal is to benefit from student feedback regarding the institution's courses to continuously assess and monitor the quality of the courses and the students' learning experiences. A sample representative dataset of students' unstructured free-text comments and answers to open-ended questions about five different courses over four consecutive semesters was collected, cleaned, and used to develop and test two sentiment analysis models: Naive Bayes in WEKA and a sentiment lexicon-based model named VADER. To further analyze and assess different aspects of the learning experience and courses along with its overall quality, answers to closed-end questions were also analyzed using the 5-point Likert scale. Preliminary results obtained from evaluating the sentiment analysis models show that the Naïve Bayes model achieved 68.7%, 68.8%, 68.8%, and 68.8%, while the VADER model achieved 72.12%, 72.82%, 72.12%, 71.87%, in terms of accuracy, precision, recall, and F1-score, respectively. Performance testing results of the application show that the maximum usage for the CPU is 44%, for the memory is 119 MB, for sending a request on the network 14.7 KB/s, for receiving a response is 226.5 KB/s, and the maximum energy usage is medium. For stress testing, obtained results show that the application can successfully deal with a maximum of 500 random, fast, and abnormal events. For user acceptance testing, users were surveyed to measure their level of satisfaction with the application using the system usability scale. The results show that 100% of users either agreed or strongly agreed that they would like to use the application and be more engaged in assessing the quality of courses. They also indicated that the application is easy to use, quick, and easy to learn. This paper also highlights various challenges and limitations developers face, along with important recommendations for further improvements and future work directions.

Keywords: Quality of Education, Sentiment Analysis, Course Quality Assessment, Student Feedback.

1 INTRODUCTION

The quality of education in a specific educational institution is directly reflected in the outcomes of their system. Higher-quality educational systems continue to focus on enhancing the quality of education programs they offer and delivering better learning experiences to enrolled students and better-developed skills and knowledge. To provide high-quality education, the institution must encompass a broader multidimensional definition of quality education that involves: learners who are ready to participate and learn; environments that are safe and provide adequate teaching and learning resources and facilities; content that is reflected in relevant courses' topics, curriculum, and materials for the acquisition of basic skills and knowledge; processes where educators use learner-centered teaching and assessment approaches; and outcomes that encompass knowledge, skills, and attitudes (UNICEF, 2000). These dimensions of education quality are interdependent and can impact each other in many ways (UNICEF, 2000). Thus, evaluating and gathering data regarding each of them is crucial to educational institutions focusing on enhancing, maintaining, and providing high-quality education.

Nowadays, students' feedback and opinions have become a vital component of the quality assurance and control process of developing high-quality educational courses and programs and "the provision of quality feedback is widely perceived as a key benchmark of effective teaching" (Beaumont, O'Doherty, & Shannon, 2011). Qualitative feedback in the form of freely-written textual data is the students' way to express their opinions and sentiments toward different aspects of their personal experience of the learning process, course content, facilities, and events (Gottipati, Shankararaman, & Lin, 2018). Sentiment refers to the polarity of a text, such as a word, sentence, or paragraph. This polarity of text can be either positive, negative, or neutral (Taboada, 2016) and students' feedback often falls within the positive or negative categories (Gottipati, Shankararaman, & Lin, 2018). This qualitative data, if properly gathered and analyzed, can become rich sources of valuable information regarding different aspects of course's content, education process, facilities, environment, outcomes, and students' learning experiences; therefore, it is considered a strong indicator of the quality of that course. Analyzing such data helps identify the areas of satisfaction and dissatisfaction, and highlights each course's strengths and weaknesses. Moreover, it helps faculty members and educators understand the students' learning styles and the difficulties they face, design and execute courses' quality improvement plans, continuously monitor and evaluate the quality of the courses, and update the plan accordingly. Furthermore, educational institution administration can actively monitor courses' quality over time and uses this information and students' feedback to enhance the quality of the programs, the outcomes, and the facilities and environment it offers. More importantly, it helps educators and institutions become more responsive to student's needs and adapt to them when it counts. Consequently, students in such institutions feel supported, more engaged, and partially responsible for shaping their learning environment and experience to their liking and preferences.

Gathering and analyzing students' feedback and sentiments to gain powerful insights into the quality of a course, its evolution over time, and its strengths and weaknesses is not easy for many reasons. First, valuable students' feedback and sentiments about courses are scattered throughout many social media applications and websites, and managed by different discussion groups that are usually run by students. Thus, gathering honest and freely written comments and opinions about courses in one place is challenging. Also, gathering such data from these social media applications and websites without compromising the students' privacy is difficult. Furthermore, faculty members face the challenge of accessing such valuable data that can significantly help them get honest feedback and comments from their students. Moreover, extracting and analyzing courses' quality and learning experience-related posts from such data is not a trivial task because students tend to mix discussion, questions, tips, jokes, and unrelated topics within the same discussion group. Manually cleaning and dealing with such noisy unstructured textual data is a complicated and lengthy process (Taboada, 2016). Also, the feedback-gathering surveys and techniques used by institutions to collect quantitative data regarding a predefined set of course learning outcomes (Abedin, Taib, & Jamil, 2014) limit the types of data and analysis to be performed and, consequently, the amount and value of information and insights that can be extracted.

This study primary goals are: to design and develop a smartphone application utilizing sentiment analysis techniques, to evaluate the sentiment analysis models in terms of accuracy, precision, recall, and F1-score, to evaluate application performance in terms of maximum usage for the CPU, memory, energy, and sending or receiving a response over the network, to evaluate the application performance under stress and abnormal situations, and to assess users' satisfactions with the developed application. This study outcome would help to address the problem of gathering, analyzing, and understanding students' feedback and comments regarding different aspects of courses provided by an educational institution. The significant contributions of the application are: gathering student's opinions and feedback regarding different aspects of the courses content, teaching process, learning experience, environment, and outcomes at the end of each semester; providing the faculty members and the institution's management with a tool that can help assess the quality of the courses, highlighting the aspects of strengths and weaknesses of each course, and monitoring how the overall course's quality is changing over time. The proposed application gathers students' feedback and comments in one platform accessed by all: students, faculty members, and management of the institution without compromising the students' privacy. Data and Sentiment analysis techniques will be utilized by the application to tackle the problem of extracting the sentiments and opinions of students from unstructured freely-written comments and textual data. Furthermore, the application addresses the limitation of the type and extent of feedback gathered through the course learning objectives surveys by collecting feedback in two forms: quantitative ratings for a predefined set of questions and qualitative freely written comments and feedback related to vital aspects related to the education process, content, environment, and outcomes.

2 BACKGROUND KNOWLEDGE

2.1 Quality Education

2.1.1 Quality of Learning Environment

Positive learning outcomes, pursued by educational institutions, generally occur in quality learning environments. A physical learning environment refers to the actual place where the formal process of education occurs (UNICEF, 2000). The quality of the physical learning environment provided by an institution and its facilities indirectly affects the quality of learning, and it is hard to measure using empirical evidence (Fuller, et al., 1999). Nevertheless, several studies have reported that the quality of the physical learning environment and the institution facilities are strongly correlated to learners' academic achievements (Carron & Chau, 1996) (Willms, 2002) (Pennycuick, 1993). This includes the availability of adequately well-equipped buildings and laboratories, libraries, classroom maintenance, space, and services, which can have a critical impact on other quality dimensions of education, such as the education process and content. For example, it can affect the ability of educators to adopt some instructional techniques and approaches and the availability of sufficient learning materials and adequate working conditions for learners and educators (UNICEF, 2000). High-quality learning environments in educational institutions set the stage for learning to happen. This learning begins with high-quality content.

2.1.2 Quality of Content

In education, content refers to courses' and programs' intended and the taught curriculum. The starting point for developing and implementing a course curriculum is translating the program and course learning outcomes and goals into measurable objectives (UNICEF, 2000). These outcomes should be clearly defined, grade-level appropriate, and properly sequenced (UNICEF, 2000). A high-quality curriculum should be student-centered, and have standards-based structures. It should also be adaptive to current and future societal needs and responsive to emerging needs, fields, and issues (UNICEF, 2000). Furthermore, a high-quality curriculum emphasizes deeply on critical areas of knowledge to be covered and the contextualized concepts and problems of study and stresses life-skills development, such as problem-solving, knowledge acquisition, assertion and refusal skills, goal setting, decision-making, conflict resolution, self-awareness, and cooperation and communication (Walczowski, Dimond, & Waller, 2000) (UNICEF, 2000). However, high-quality content must be situated in the context of high-quality education processes to be most effective.

2.1.3 Quality Processes

Educational processes refer to how institutions' educators and administrators utilize the educational system inputs to achieve meaningful learning experiences for learners. The limited perspective of teaching as a rigid, teacher-centered, static presentation of knowledge is no longer valid (Carron & Chau, 1996) and education processes should ensure ongoing support for student-centred learning. Educators are the most significant pillar of any education process. The availability high-quality educators who are most qualified to aid their students in learning and have a profound mastery of their subject matter (Darling-Hammond, 1998) significantly impacts learning quality. Students' achievements depend highly on the educators' expertise (Mullens, Willett, & Murnane, 1996), their ability to utilize that knowledge and expertise to assist students in learning (UNICEF, 2000), and mentor them by providing constant support and feedback. Furthermore, the use of traditional and non-traditional methods of instruction by educators and the efficient use and management of class time significantly impact student learning (Fuller, et al., 1999). Educators instructional methods and skills must be adaptive to the evolving learning styles and needs of students, help them build on prior knowledge, expand their knowledge base, and develop solid cognitive skills. Educators must also investigate different feedback gathering, and assessment practices and techniques to better understand the challenges, difficulties, and needs of their students. A high-quality process must also provide students with periodic and challenging projects and assignments, monitor and evaluate their performance on a regular basis, and present them with opportunities to engage, participate and take responsibility for various activities (Simon, 2013).

2.1.4 Quality Outcomes

The learning environment, content, and processes lead to a mix of intended and non-intended results for students. Intended results are the quality outcomes expected and evaluated by the educational institution. Assessment of students' academic achievement outcomes has constantly been achieved via testing information. These outcomes are easily measurable using standardized tests. Thus, students' academic achievements are often used as indicators of the quality of education provided by the institution. This is not the case with other outcomes that may be more challenging to measure and less tangible such as the development of skills. These outcomes can be indirectly assessed by measuring the degree to which the students master these skills and are capable of successfully applying them in real-life situations to solve different problems; their ability to demonstrate curiosity and autonomy; and the degree to which students show responsibility and commitment to each other and the community (UNICEF, 2000).

2.2 Sentiment Analysis

Sentiment Analysis is the task of classifying text polarity regarding a specific entity, aspect, or event as positive or negative by employing supervised or unsupervised machine learning methods and heuristic techniques (Taboada, 2016). Sentiment Analysis techniques have been applied effectively in many application domains such as brand and social media monitoring. Two well-known approaches of Sentiment Analysis of textual data are: the lexical approach and the machine learning approach.

2.2.1 The Lexical Approach – VADER (Valence Aware Dictionary for sEntiment Reasoning)

The lexical approach builds a lexicon or a dictionary of sentiments and uses it to map words to sentiments. This approach assesses the sentiment of a text by summing up the sentiment scores of each dictionary-listed word in that text. The major advantage of this approach is that it does not need labeled data to build and train the model. It depends on the sentiment dictionary, which can be customized to include domain-specific terminologies. VADER (Hutto & Gilbert, 2014) is based on the lexical approach. It is a sentiment analysis model introduced in 2014 for classifying and detecting textual data polarity and the intensity of this polarity. VADER calculates the sentiment score of a text as follows. For each word in the text, if it exists in the VADER-dictionary, the model will map this word to the corresponding sentiment intensity using the dictionary. The total sentiment score of text is calculated by summing up the sentiment scores of each word in the text that is listed in the VADER-dictionary (Hutto & Gilbert, 2014).

2.2.2 The Machine Learning Approach – Naïve Bayes

The Machine learning approach uses labeled data to learn the task of classifying a word or a group of words into one sentiment class: positive or negative. The advantage of machine learning models is that their performance usually improves when trained on more volumes of data. Naive Bayes is a simple and powerful probabilistic supervised machine learning algorithm used for classification tasks. The core of Naive Bayes classifier is based on the Bayes theorem with the assumption of independence between the features. A Naive Bayes classifier uses the Bayes theorem to predict the membership probability of a text for each output class. The class with the highest probability is then assigned as the class of that text (Webb, Keogh, & Miikkulainen, 2010).

3 METHODOLOGY

This section briefly describes the methodology followed in this study. The design and development process consisted of five phases: (1) requirement elicitation, (2) deciding on the scope and features of the application, (3) application design, (4) application development, and (5) application testing.

3.1 Requirement Elicitation

The system's requirements elicitation process consisted of three stages: understanding the current methods for course quality assessment, interviewing intended application users, faculty members, and students, and conducting a benchmarking analysis. The first stage of this process involved understanding the current course quality assessment methods in the educational institution under study and gaining strong and sufficient background knowledge of the domain. This stage helps define the application's primary purpose, objectives, and initial scope, including the specific strategies for course quality assessment and the potential data to be collected. In the second stage of the requirement elicitation process, interviews with the system's stakeholders and potential users were conducted. Interviews are one of the most popular requirements elicitation techniques used to verify facts, clarify processes, engage end users and stakeholders by soliciting their opinions and ideas, and identify the system requirements and features. During the interviews, a list of predefined questions regarding the goal, objectives, and features of the proposed application, and steps and difficulties of existing course's quality assurance process were used in addition to open-ended questions. This process helps understand the current process and the needs and expectations of each user. The third stage of the requirement elicitation process aimed to understand the competitors and the targeted market. In this stage, a benchmarking analysis was conducted on powerful applications and websites similar to the proposed application in terms of provided features, targeted problem, or expected users. Table 1 shows the benchmarking analysis results focusing on the main features proposed for the application.

Table 1. Benchmarking analysis results

Features \ Application	piHAPPINESS (2018)	ziplet (2015)	vivi (2015)	FeedbackPanda *(2019)*	Zonka (2014)
Register, log in/out, edit profile, reset password	✓	✓	✓	✓	✓
List of all courses	X	X	X	✓	X
Search for a course	X	X	X	✓	X
Post comments and feedback	✓	✓	✓	✓	✓
Edit own comments and feedback	X	X	X	X	X
Delete own comments and feedback	X	✓	✓	X	X
Display anonymous comments and feedback to students	X	X	✓	✓	X
Sentiment Analysis techniques	✓	X	✓	X	X
Generate course quality report dynamically	X	X	X	X	X
Graphical representation of Sentiment Analysis results	✓	X	X	X	X

The results show that some essential features were provided to users by all competitors, such as registration, editing personal profiles, posting comments and providing feedback regarding courses, and logging in and out of the system. Other important features were only provided by some competitors, such as listing all courses, searching for a specific course, allowing a user to delete his/her comments/feedback on a course, applying sentiment analysis techniques, and displaying the course feedback to students.

3.2 Deciding on the Scope and Features of Application

The main goal of this study is to design and develop a smartphone application utilizing Sentiment Analysis techniques to analyze and understand students' feedback and comments regarding different aspects of courses provided by an educational institution. The requirement elicitation process resulted in the decision to design the application to gather students' opinions and feedback at the end of each semester in two forms: quantitative ratings for questions and qualitative comments related to quality education dimensions. Table 2 shows a list of essential aspects of the different quality education dimensions considered in this study.

Table 2. A list of essential aspects of quality education dimensions considered in this study

Education Quality Dimension	Data to be collected
Quality Learning Environments	The quality and availability of adequately well-equipped buildings and laboratories, libraries, classroom maintenance, space, and services
Quality Content	Curriculum and material with a deep emphasise on critical areas of knowledge, concepts, and skills such as problem-solving, knowledge acquisition, assertion and refusal skills, goal setting, decision-making, conflict resolution, self-awareness, cooperation and communication.
Quality Processes	Faculty member's mastery of subject and delivery style, mentoring of learners and ability to motivate them, continuous feedback on learning, effective use of traditional and non-traditional instructional methods, posting of periodic and challenging projects and assignments, presenting learners with opportunities to engage, participate and take responsibility for various activities
Quality Outcomes	Tests and formative assessment of knowledge and assessment of acquired skills and the ability to apply them

Furthermore, the application will provide the courses' coordinators and the department with a tool that can help assess the quality of the taught courses, highlight the aspects of strengths and weaknesses of each course, and monitor how the overall course's quality is changing over time. Table 3 shows a list of the primary features considered to develop the application.

Table 3. A list of primary features considered to develop the application

Feature
Registration
Logging in/out of the system
Updating personal information in profile
Password Change
Reset a Forgotten Password
View feedback and opinions of students (anonymously) regarding the courses.
Search
View list of all courses
View the course page and the overall quality report
Provide course feedback by answering specific questions related to the quality of the course.
Post a comment on a course

Delete his/her own comments
Edit his/her own comments
Add a course
Edit a course details
Delete a course
Delete a comment
Block a user
Dynamic analysis of the overall quality of the course using sentiment analysis
Automatic generation of the course quality report
Display the course quality progress graphically

Furthermore, essential non-functional requirements related to the application's reliability, performance, availability, security, usability, privacy, and ethical issues were considered throughout the design and development process of the application.

3.3 Application Design

In the application design phase, Object-Oriented Analysis and Design (OOAD) approach was adopted to increase usability and productivity and simplify the system integration process. Intended users of the application can be categorized into faculty members, students, and administrators. All users must have basic technical skills and familiarity with smartphone applications. The application was designed for mobile phones running the Android operating system (OS). Well-known UX guidelines were followed in designing the application to ensure delivering an appealing product to the end users. Figure 1 shows sample screenshots of the application.

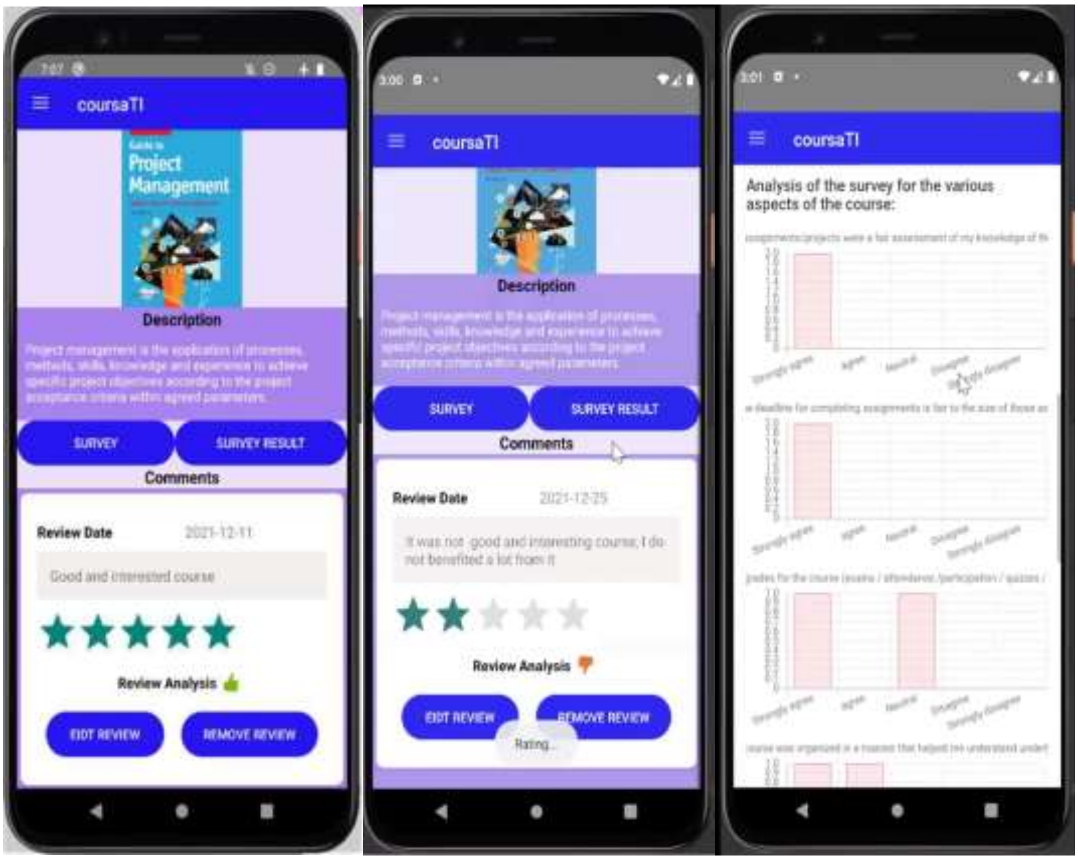

Fig. 1. Sample screenshots of the application prototypes.

3.4 Application Development

3.4.1 Development Environment and Tools

The application was developed using Android Studio (2013) version 2020.3.1 and NetBeans (1996) version 12.5. Other software and tools used for the development of the application include MySQL (1995), phpMyAdmin (1998), and Weka (2005). Incremental development and integration of all components and features in table 3 was adopted throughout the application development phase.

3.4.2 Building the Sentiment Analysis Models

The major steps in building the sentiment analysis models are illustrated in the Figure 2.

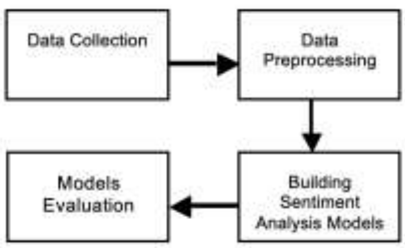

Fig. 2. Major steps in building the sentiment analysis models.

Data Collection

A comprehensive survey covering the most critical aspects of the quality of courses and the associated learning experience was designed and used to collect a representative sample dataset of students' unstructured free-text comments and feedback about five courses over four consecutive semesters. Participants were senior students at the Department of Information Technology who successfully passed the five courses in the last two years. For each course, participants were asked to provide feedback using a five-point-scale rating and express their opinions freely concerning essential aspects of the course's teaching process, content, and learning experience in addition to the difficulties, challenges, usefulness, skills learned, and suggestions.

Data Pre-processing

The responses of 111 participants were then cleaned and preprocessed by applying several techniques. Lowercasing, where all words in the dataset were modified to contain lowercase characters only. Noise Removal, which removes all punctuations, numbers, emojis, and other unnecessary characters from the text. Stop words removal, tokenization, and stemming techniques were also applied to further preprocess the unstructured textual data and prepare it to train the sentiment analysis models. The collected dataset was then manually annotated by assigning a label of positive or negative class to each comment.

Building the models

The preprocessed dataset was used to develop two sentiment analysis models: Naive Bayes sentiment classifier developed using Weka, and VADER, a sentiment lexicon-based classification model.

3.4.3 Quantitative Data Analysis

To further analyze and assess the significant aspects of the quality of courses and the learning experience, answers to closed-end questions were analyzed using the 5-point Likert scale. The results were used to identify areas of dissatisfaction and satisfaction with each course. The analysis results were also used to dynamically generate and update the course's overall quality report, visualize its quality evolution over time, and highlight the strengths and weaknesses as pointed out by students.

3.5 Application Testing

During the development process of the application, several rounds of unit testing and system integration testing were conducted to ensure the proper development and integration of the different components of the application. The sentiment analysis models were tested using a held-out test set, and the classification performance of each model was reported in terms of accuracy, precision, recall, and F1 score. Furthermore, performance testing, stress testing, and user acceptance testing were performed to assess the application's overall performance.

4 RESULTS

4.1 Sentiment Analysis Classifier Evaluation

The preliminary results of evaluating the sentiment analysis models on the collected dataset is shown in table 4. The obtained results show that the classification performance of VADER was better than the Naïve Bayes model. This is expected due the size of the dataset used to train the Naïve Bayes model. However, it is expected that the Naïve Bayes model will outperform VADER when more data is collected by the application, labeled, and used to train the model.

Table 4. The preliminary results of evaluating the sentiment analysis models

Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	68.7%	68.8%	68.8%	68.8%
VADER	72.12%	72.82%	72.12%	71.87%

4.2 Performance Testing

Performance testing is mainly used to assess the application's performance under normal circumstances by measuring the application's speed and consumption of the device's resources such as memory, processing power, and energy. The application was installed and tested on a Pixel 5 API 29 device. To test the application's performance on the device, Android Profiler, by Android Studio (2013), was employed, and the results of the profiler for the device's CPU, memory, network usage, and energy were reported in Figure 3. The figure shows that the maximum usage for the CPU is 44%, for the memory is 119 MB, for sending a request on the network 14.7 KB/s, for receiving a response is 226.5 KB/s, and the maximum power usage is medium.

Fig. 3. Performance test results show the application's maximum usage of the device's CPU, memory, network, and energy.

4.3 Stress Testing

Stress testing aims to assess the application's performance under abnormal circumstances. Exerciser Monkey (2013) was used to test the application. It is a program that performs stress testing on applications by initiating numerous random events at high speed. The goal is to estimate the maximum number of fast, random, and abnormal events the application can handle at any given time before it crashes. Figure 4 shows the obtained result of the stress testing indicating that the application can successfully deal with a maximum of 500 abnormal events at any given time.

```
Terminal: Local  +
    // Sending event #400
:Sending Touch (ACTION_DOWN): 0:(717.0,1874.0)
:Sending Touch (ACTION_UP): 0:(694.5706,1812.0166)
:Sending Trackball (ACTION_MOVE): 0:(2.0,-5.0)
:Sending Trackball (ACTION_MOVE): 0:(1.0,0.0)
:Sending Touch (ACTION_DOWN): 0:(490.0,128.0)
:Sending Touch (ACTION_UP): 0:(474.88968,100.62714)
:Sending Trackball (ACTION_MOVE): 0:(-1.0,-5.0)
:Sending Touch (ACTION_DOWN): 0:(612.0,89.0)
:Sending Touch (ACTION_UP): 0:(613.04553,72.57654)
:Sending Trackball (ACTION_MOVE): 0:(4.0,-3.0)
:Sending Trackball (ACTION_MOVE): 0:(1.0,-4.0)
:Sending Touch (ACTION_DOWN): 0:(816.0,2144.0)
:Sending Touch (ACTION_UP): 0:(807.56683,2148.0)
:Sending Touch (ACTION_DOWN): 0:(991.0,1017.0)
:Sending Touch (ACTION_UP): 0:(1005.73535,1012.829)
:Sending Trackball (ACTION_MOVE): 0:(4.0,-3.0)
Events injected: 500
:Sending rotation degree=0, persist=false
:Dropped: keys=0 pointers=2 trackballs=0 flips=0 rotations=0
## Network stats: elapsed time=2685ms (0ms mobile, 0ms wifi, 2685ms not connected)
// Monkey finished
```

Fig. 4. Stress test results show that the application can successfully handle 500 abnormal events at a time.

4.4 User Acceptance Testing

For user acceptance testing, users were asked to try the application and fill out a questionnaire to measure their level of satisfaction with the application. The satisfaction level was measured using the System Usability Scale (SUS). The results showed that 100% of users either agreed or strongly agreed that they would like to use the application and be more engaged in assessing the quality of courses, that the application is easy to use, and that it is quick and easy to learn. While 83.3% either agreed or strongly agreed that the design is friendly and the navigation through the application is easy.

5 LIMITATIONS AND FUTURE WORK

5.1 Limitations

The main limitation of the application development was the small size of the dataset used to build the sentiment analysis models. This is due to several difficulties faced during the data collection and cleaning phases. First, no social media platform could be used to collect freely written comments and feedback without compromising the privacy of the students and violating other security and legal policies and regulations. Also, students were hesitant to participate and express their opinions freely regarding the courses. These difficulties yielded a small dataset that was further reduced after the cleaning and preprocessing steps on the unstructured short comments. This hinders the Naïve Bayes model from learning the hidden patterns effectively and achieving high classification performance. However, when put into use, the application can solve these issues by providing a single platform where students can freely and anonymously express their opinions on different aspects of the education quality which can be then used to increase the size of the initial dataset.

5.2 Future work

The application can be further improved to include advanced features. For example, adding a recommender system to provide course enhancement suggestions to faculty members based on other courses with similar weaknesses and strengths. Adding a planning features to automatically generate a course quality improvement plan. Allowing faculty member to design his/her own feedback questions and surveys. This will be an additional survey customized to a specific course and instructor. Schedule a feedback survey to be available for students in a specific date during a semester which will help evaluate the intermediate outcomes of the course and implement required adjustment, if any. Automatically send Notifications by email to course owners and administrators for whenever a course's overall quality score dropped.

6 CONCLUSIONS

Students' opinions and feedback regarding different aspects of a course and their personal learning experience, if adequately gathered and analyzed, can be strong indicators of the quality of that course. This study described designing and developing a smartphone application utilizing Sentiment Analysis methods to benefit from student feedback regarding the institution's courses and continuously assess and monitor the courses' quality. Sentiment analysis models' performance was measured and reported in terms of accuracy, precision, recall, and F1-score. Performance testing, stress testing, and user acceptance testing were also performed to assess the application's overall performance. Preliminary results obtained from evaluating the sentiment analysis models showed that the Naïve Bayes model achieved higher indicators than the VADER model in terms of accuracy, precision, recall, and F1-score, respectively. Performance testing and stress testing results show that the application has very good performance and can successfully deal with a maximum of 500 random, fast, and abnormal events. 100% of users either agreed or strongly agreed that they would like to use the application and be more engaged in assessing the quality of courses. They also indicated that the application is easy to use, quick, and easy to learn.

REFERENCE LIST

Abedin, N. F. Z., Taib, J. M., & Jamil, H. M. T. (2014). Comparative study on course evaluation process: Students' and lecturers' perceptions. *Procedia- Social and Behavioral Sciences, 123*, 380-388. doi:10.1016/j.sbspro.2014.01.1436

Android Profiler - Android Developers [Computer Software]. (2013). Retrieved from https://developer.android.com/studio/profile/android-profiler

Android Studio & App Tools - Android Developers [Computer Software]. (2013). Retrieved from https://developer.android.com/studio

Apache NetBeans [Computer Software]. (1996). Retrieved from https://netbeans.apache.org/

Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education (Dorchester-on-Thames), 36*(6), 671-687. doi:10.1080/03075071003731135

Carron, G., & Chau, T. N. (1996). *The quality of primary schools in different development contexts*. Paris: Unesco.

Darling-Hammond, L. (1998). *Doing what matters most*. New York: National Commission on Teaching & America's Future.

UNICEF. (2000). D*efining quality in education.*

FeedbackPanda: Student Feedback Templates for Online Teachers [Computer Software]. (2019). Retrieved from https://www.feedbackpanda.com/

Frank, E., Hall, M. A., Holmes, G., Kirkby, R., Pfahringer, B. & Witten, I. H. (2005). Weka: A machine learning workbench for data mining.. In O. Maimon & L. Rokach (eds.), *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers* (pp. 1305--1314). Springer.

Fuller, B., Dellagnelo, L., et al. (1999). How to raise children's early literacy? *Comparative Education Review, 43*(1), 1-35. doi:10.1086/447543

Gottipati, S., Shankararaman, V., & Lin, J. R. (2018). Text analytics approach to extract course improvement suggestions from students' feedback. *Research and Practice in Technology Enhanced Learning, 13*(1), 6-19. doi:10.1186/s41039-018-0073-0

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the ... International AAAI Conference on Weblogs and Social Media, 8*(1), 216-225. doi:10.1609/icwsm.v8i1.14550

John E. Mullens, John B. Willett, & Richard J. Murnane. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from belize. *Comparative Education Review, 40*(2), 139-157. doi:10.1086/447369

Mwaura, P. A. M., & Marfo, K. (2011). Bridging culture, research, and practice in early childhood development: The madrasa resource centers in east africa. *Child Development Perspectives, 5*(2), 134-139. doi:10.1111/j.1750-8606.2011.00168.x

MySQL [Computer Software]. (1995). Retrieved from https://www.mysql.com/

Pennycuick, D. (1993). *School effectiveness in developing countries - A summary of the research evidence*

PhpMyAdmin [Computer Software]. (1998). Retrieved from https://www.phpmyadmin.net/

PiHappiness [Computer Software]. (2018). Retrieved from https://www.pihappiness.com/

Simon, E. (2013). *Teacher development in higher education* (1. publ. ed.). New York [u.a.]: Routledge.

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics, 2*(1), 325-347. doi:10.1146/annurev-linguistics-011415-040518

UI/Application Exerciser Monkey - Android Developers [Computer Software]. (2013). Retrieved from https://developer.android.com/studio/test/other-testing-tools/monkey

Vivi [Computer Software]. (2015) Retrieved from https://www.vivi.io/empower-your-teachers-to-deliver-personalized-learning/

Walczowski, L. T., Dimond, K. R., & Waller, W. A. J. (2000). A digital engineering curriculum for the new millennium. *International Journal of Electrical Engineering & Education, 37*(1), 108-117. doi:10.7227/IJEEE.37.1.10

Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve bayes. *Encyclopedia of Machine Learning, 15*, 713-714.

Willms, J. (2002). *Standards of care: Investments to improve children's educational outcomes in latin America*

Ziplet | The Easy Way To Check In With Your Students [Computer Software]. (2015). Retrieved from https://ziplet.com/

Zonka Feedback [Computer Software]. (2014). Retrieved from https://www.zonkafeedback.com/