Sentiment Analysis on Students Stress and Depression Due to Online Distance Learning During the COVID-19 Pandemic

Nur Syafirah Izzati bt Samsari
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
2018659762@student.uitm.edu.my

Masurah Mohamad
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch Tapah Campus, Perak, Malaysia
masur480@uitm.edu.my

Ali Selamat
Malaysia Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, Kuala Lumpur 54100
aselamat@utm.my

Article Info

ABSTRACT

Recently, more young people are suffering from mental illness especially stress and depression during the COVID-19 pandemic. The different between how bad the mental health condition depends on the ways each individual handle and overcome it. Besides, factor that leads stress and depression also plays an important roles in order to know the level of the stress and depression. A person's age and background are also among the factors that cause stress and depression. For instance, stress and depression of students mostly happen because of their study especially during COVID-19 pandemic where schools and universities are closed, and online learning need to be implemented. This study analyses on the number of students that were affected from the COVID-19 pandemic during online distance learning. Two keywords "stress and depression" were used to extract the data from the Twitter. Sentiment analysis has been conducted in the analysis process and Naïve Bayes classifier was applied to classify the data into the pre-defined categories. 571 out of 2291 students experienced depression, 872 experienced high levels of stress and 848 were diagnosed with early-stage stress. The obtained results are visualized using polar area diagram to ease the readers' understanding.

Corresponding Author:
Masurah Mohamad
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA, Perak Branch, Tapah Campus
Perak, Malaysia.
email: masur480@uitm.edu.my

1. Introduction

Stress and depression are amongst the mental health problems. Nowadays, in social media especially heard that majority of people were involve in the mental health problems day by day. There are many factors that can lead to mental health problems such as lack of love in life, over stress in a

certain situation, cannot handle problems properly and many more. Basically, this problem depends on the person who faced it on how to handle each of the problems that they are facing through. Stress and depression can be divided in many stages and each stage are different between each other.

One of the factors that lead to student's stress or depression is academic matters. For example, decrease in grades, fear to failure and tough competitions. Students are majority at the young age and sometimes they cannot think wisely when having problems and might harm themselves. Thus, support and guidance from expert, family members, friends and others are really important to help students from doing bad activity due to stress and depression.

Therefore, this project aims to help students by disseminating information that they are not alone in facing stress and depression during the COVID-19 pandemic especially on online distance learning. At the same time, this project helps students to find the best way to overcome the problems by knowing their mental health condition referring to the characteristics of person who were stress and depression that are provided in the system.

2. Literature Review

In modern technology nowadays, students are likely to expose feelings through social media in order to tell people surrounding what they are facing through. The trends also indicate that society is more likely share thought rather than what they think on social media sites [1]. So, from that can know that majority of students are having stress and depression. Stress and depression are feeling that need to avoid by everyone especially students because majority of students are in young age, so they cannot think wisely to make any decision and also to prevent bad things happen at the same time.

Stress makes people become sad, frustrated, nervous or angry. Meanwhile, depression is mood disorders that cause loss of interest and persist of feeling. Depression can lead to physical problems also variety of emotions especially bad emotions. Stress is normal in a part of life as a student. Usually, stress happen when students cannot handle the pressure in life include role as a student and personal matter and ways to overcome it. To know whether students are facing stress and depression are by looking at the habits. Some students prefer to stay alone, talking nonsense such as suicide and want to quit study, late to submit any assignments given and most important is academic performance reduce. Person who are fell burden, saying want to escape, saying they are alone and may talk about desperate and hopeless are people or someone that might have risk of stress and depression [2]. Other than that, number of students who saw the doctors and counselors to get any advice regarding stress and depression also can shows that students is facing stress and depression.

Method of therapy that guides people through growing one's own motivation for change for positive change is motivational interviewing. Motivational interviewing is one of type of counselling that will help to in mental health problems [3]. Moreover, numbers of people to get any treatment regarding their stress and depression also increase day by day. People with higher income countries were basically to get treatment for their stress problems twice than upper-middle income countries and low-lower middle income. This can be shown by 53.5% for high income countries, 22.8% for low-lower middle-income countries and 28.7% for upper-middle income countries [3]. There are many factors that can affect stress and depression of students such as family problems, learning through online and personalities. These factors are really important in order for students to overcome it as soon as possible and to make sure other factors did not affect them too.

In pandemic era nowadays, stress and depression of students are related to the online distance learning. Online distance learning is a learning that go through online. Online learning can be held using many platforms such as Google Meet, Webex Meetings, Google Classroom, WhatsApp group and Zoom Meetings. All of these platforms required internet in order to make it run smoothly. Online learning required more effort from students as to make sure they can attend class as planned without any problems with internet collection, hardware or things that will be used to go online learning such as laptop and telephone. In other words, students need to focus on other situations rather than online distance learning only.

Basically, online distance learning will be held in order to make sure syllabus of this learning can be follow in any situations and students can still learning even though online. As example, during pandemic now, most universities and schools had their learning through online. Ministry of Education

in Republic of Indonesia had change to study at home from schools in order to prevent the virus from spreading and this situation also changes from face-to-face learning to online learning [4]. Based on study that carried out by Support et al. [4], it shows that online learning started during the pandemic session. Approach that teachers or lecturers and student used during online distance learning are from classroom to zoom, from personal to virtual and from seminars to webinars. It shown that this approach transforms the traditional method to the modern method of teaching and learning [5]. Figure 1 shows the example of journal and statistics of students' stress and depression that happened during the COVID-19 pandemic Mental health of students become worse as pandemic occurs and there are also a few negative effects raised from this situation.
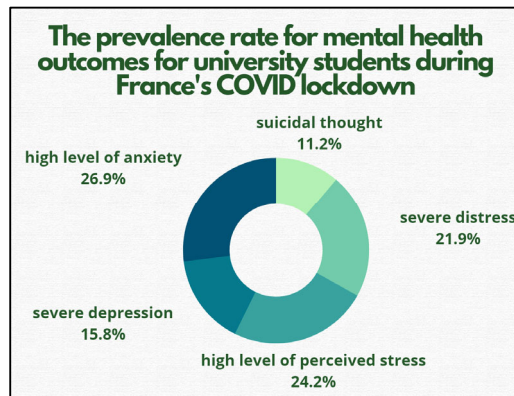

Figure 1: Example of statistics students' stress and depression

Sentiment analysis is one of text data analysis. Sentiment analysis is analysis that have been made in order to know opinions, feeling or emotion, decisions, view of people through online based on word or text and also that was a private statement of individual. Some words are insignificant in the sentiment analysis while predicting the sentiment as positive and negative [6]. It is a very powerful application of natural language processing (NLP) and finds usage in a large number of industries. To obtain valuable insights of people's opinion about products or services is primary aim of sentiment analysis [1]. Hence, categorizing people's opinion into positive, negative and neutral can be made.

Naïve Bayes are basically based on Bayes theorem [7]. It is one of the most used classifiers in sentiment analysis process. When the size of training set is less, Naïve Bayes were used. By using Bayes' theorem, every pair of features being classified in Naïve Bayes are independent of each other. Using the concept of conditional probability, Bayes' theorem describes the outcome probabilities of related events [8]. Bayes' theorem is basically can find probability of events while the other events are occurred. Furthermore, Naïve Bayes are fast and easy to implement but the requirement of predictors to be independent. In the other words, the predictors are dependent, and this situation hinders the performance of the classifier. Besides that, Naïve Bayes are commonly used algorithm. So, it is become easier to people learn for the first time and look for solutions for any problems regarding this because for sure there are a lot of tutorials on the internet.

Data visualization is the graphical representation of information and data. Usually, data visualization is represented in graphs, charts and maps and these provide an accessible way to understands pattern, outliners and trends of something. Data visualization makes it easy to gain insight into and analyse a lot of data [9]. One of the advantages of data visualization is able to grab people interest and keep people's eyes on the message. When people see a chart, people quickly see trend and internalize it. Usually if people see large spreadsheet of data, people cannot see the trend. In data visualization there are many tools such as tableau, d3.js and Power BI.

3. Research Method

A well-organized methodology plays significant roles to complete the project. A waterfall model well known as System Development Life Cycle (SDLC) used to understand the users' need and to ensure that the project is on the right track [10]. This project starts with a planning stage then followed by information gathering stage, the requirement about web-based analysis system,

sentiment analysis and data visualization. The third stage is data collection and analysis phase. After that proceed with designing phase. Then the development stage is done by embedding the program using coding as for data visualization. Finally, the last stage is testing phase which test the functionality of the project and features of the project.

The system architecture is a conceptual model that describes the structure, behavior and overall views of the system [9]. Figure 2 illustrates the architecture of the proposed work. There are four phases in developing the web-based analysis system which are data extraction, data cleaning, sentiment analysis and data visualization.
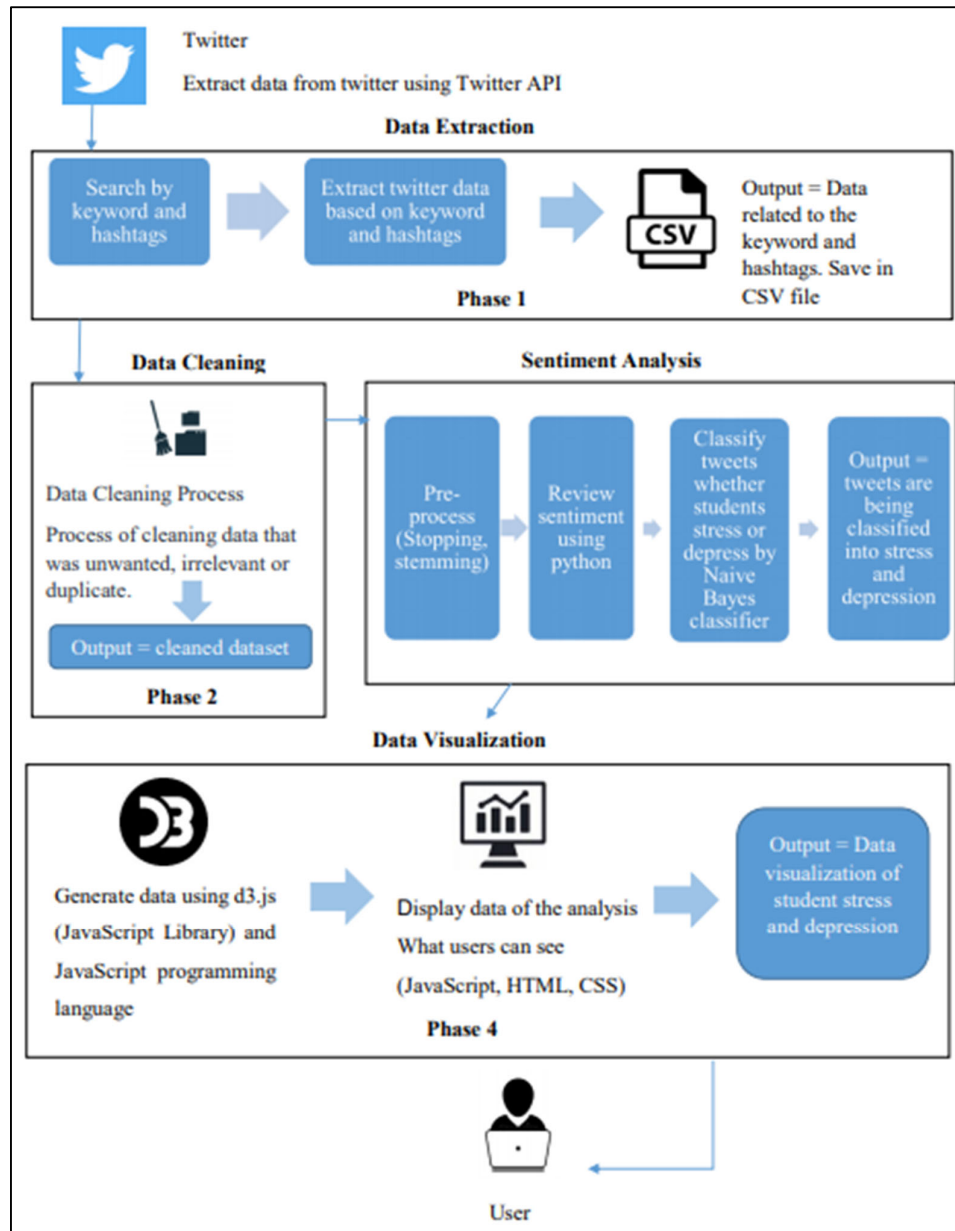

Figure 2. System architecture of the proposed work.

3.1 Data Extraction

The data will be extracted by Twitter Application Programing Interface (API). The Twitter API framework gathers Twitter streaming tweets that also store tweet scores along with their timestamp [10]. Only public of tweets can be accessed through Twitter API. Plus, while creating the request, POST method will be used and GET method will be used to retrieve results. Search keywords and

hashtags is the first step before extract data from twitter. Keywords such as *ODL is really stressful, depress of ODL and stress and depress ODL and hashtags such as #stressODL, #depressODL and #ODLsucks* will be included. Data related to the keyword and hashtags will be the output for this phase. The extracted data then will be saved in comma-separated values (CSV) file.

3.2 Data Cleaning

Process of cleaning datasets are including removing the username or @ sign, Unicode characters, emoji, links and hashtags using Jupyter Notebook.

3.3 Sentiment Analysis

Basically, sentiment analysis has five processes which are data collection, text preparation, sentiment detection, sentiment classification and presentation of output [11]. Each process has owned roles to make sure process of sentiment analysis can be done in the correct way and efficiently. Firstly, data collection is the process to collect data from various Twitter resources. Secondly, text preparation process will remove the unrelated data such as urls or links, numbers and stop words which has no analytical significance. Thirdly, sentiment detection process will be executed. This process will extract only opinions, reviews, and feedback meanwhile information about facts or knowledge will be removed. Finally, Naïve Bayes classifier will be used to classify the data according to the specified categories either positive, neutral, or negative. The polarity will define whether the tweets in stress or depress category by using getTextPolarity function. If the score equals to 0 and 1, the tweets are in stress category that divide into higher level of stress and early stage of stress. Meanwhile if score equals to -1, the tweets are categorized as depress. Figure 3 illustrates the workflow of the sentiment analysis process explained above.
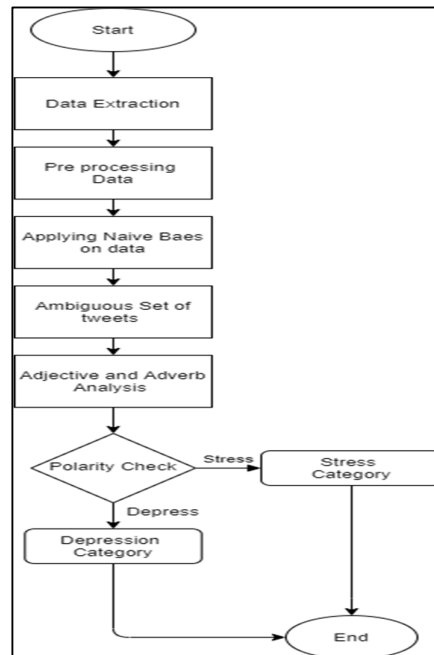

Figure 3. Sentiment analysis workflow.

3.4 Data Visualization

After the classification process is completed, the process of counting the number of students based each category is conducted for results presentation. The number of people who were facing stress or depression will be counted and display in the polar area diagram at the main page of the system. This polar area diagram was created by using D3.js. D3.js is a JavaScript library for data-driven document manipulation. Using Hypertext Markup Language (HTML), Scalable Vector Graphics (SVG), and Cascading Style Sheets (CSS), D3 allows data to bring to life. Plus, D3.js is a JavaScript framework that allows web browsers to create dynamic, interactive data visualization [12]. Usually, D3.js were used to visualize data in order to help for better understanding [13].

4. Results and Discussion

The main point of this section is to analyze the result that corresponds to the implementation of the proposed work in terms of functionality and interface design. Data that were used to conduct this project were extracted from Twitter based on hashtags and keywords. Twitter API that needs twitter developer account is needed. It has one attribute which named as "Tweets" that contains 2291 rows collected from February until April 2021 that need to be processed. Out of 2291 students, Naïve Bayes has classified 872 (38.1%) students in higher level of stress, 848 (37%) in early stage of stress and 571 (24.9%) students are suffered from depression problem.

4.1 Results Analysis

The functionality testing is conducted in order to know whether the project was successfully executed and achieved the requirements or not [14],[15]. Table 1 shows the result of functional testing using black box testing approach of the web-based analysis system of students' stress and depression due to online distance learning during the COVID-19 pandemic.

Table 1. Functional testing of the system

No.	Module	Expected Function	Test Results
1.	Main Page & Visualization Page	Users can click on the navigation bar button and can see the result of the analysis on polar area diagram.	Yes
2.	Characteristics Page Interface	Users can see the clear picture and details the characteristics of people who were in stress and depression listed along with the button to go home function well.	Yes
3.	Ways to overcome Interface	Users can see the ways to overcome stress and depression clearly and button to go home can function well.	Yes

4.2 Interface Design

This web-based analysis system has two different user type i) students, parents or workers and ii) administrator. Each user has a unique role. Figure 3 shows the main page of the proposed system. Several information that consists of scrapping month that shows the period of the data were extracted, navigation bar to go to another page and the total number of students for each category either stress or depress.

Users can navigate to the visualization page to view more information on stress and depress statistic via polar area diagram as shown in Figure 5. In addition, users also able to visit the characteristics page that explain the characteristics of person who were stress and depression based on certain symptoms as depicted in Figure 6. Besides, users could also find out the ways to overcome the stress or depression in a correct way as presented in Figure 7.
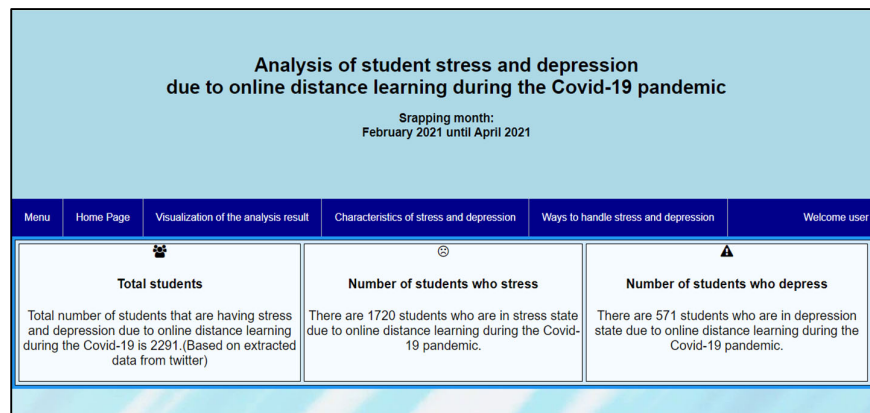

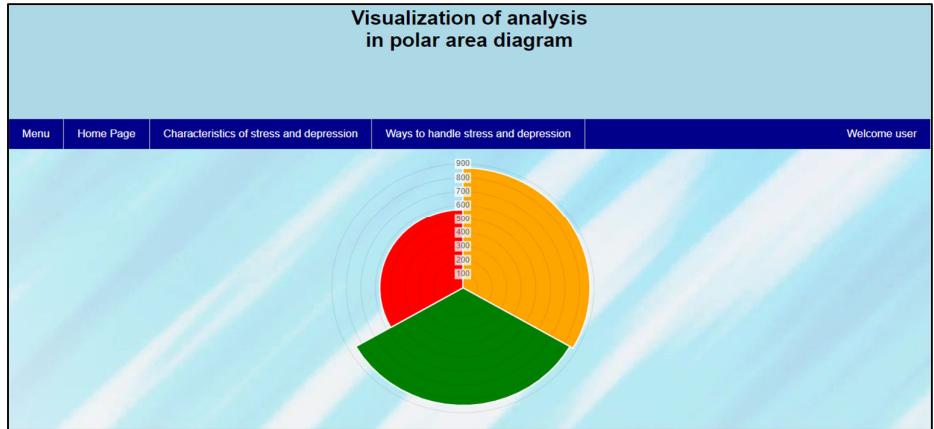Figure 4. Interfaces for Main Page of Web-Based Analysis System

Figure 5. Interface for the Visualization page of the web-based analysis system
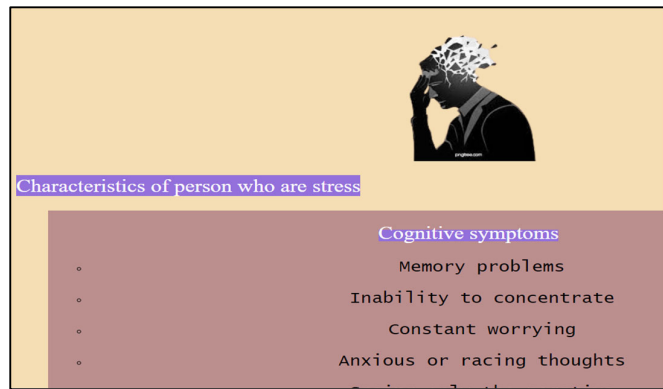

Figure 6. Interface for the characteristics page of the web-based analysis system

Figure 7. Interface for the ways to overcome page of the web-based analysis system

5. Conclusion

Nowadays, pandemic situation that had happened all over the world had been affected many people including students, parents and workers. From this situation, most of the people were suffered from mental health illness as many activities and daily routine are totally change. Thus, this project aims to analyze on the students' mental health so that early prevention could be taken and to spread the awareness to the authorities or anyone regarding to this issue. However, there are a few items can be enhanced to increase the functionality of this project such as employing real time data so that, reader can always update with the latest statistics and information. Several variables or attributes such as students background also can be added during the data extraction phase to provide more meaningful information to the readers.

Acknowledgements

Conflict of Interest
The authors declare no conflict of interest in the subject matter or materials discussed in this manuscript.

References

[1]   S. Shayaa *et al.*, "Sentiment analysis of big data: Methods, applications, and open challenges," *IEEE Access*, vol. 6, pp. 37807–37827, 2018.

[2]   A. Beautrais, "Stress and suicide in medical students and physicians," *New Zeal. Med. Student J.*, no. 30, pp. 11–14, 2020.

[3]   T. McFadden, M. S. Fortier, and E. Guerin, "Investigating the effects of physical activity counselling on depressive symptoms and physical activity in female undergraduate students with depression: A multiple baseline single-subject design," *Ment. Health Phys. Act.*, vol. 12, pp. 25–36, 2017.

[4]   D. Pajarianto and others, "Study from home in the middle of the COVID-19 pandemic: analysis of religiosity, teacher, and parents support against academic stress," 2020.

[5]   L. Mishra, T. Gupta, and A. Shree, "Online teaching-learning in higher education during lockdown period of COVID-19 pandemic," *Int. J. Educ. Res. Open*, vol. 1, p. 100012, 2020.

[6]   T. Bi, P. Liang, A. Tang, and C. Yang, "A systematic mapping study on text analysis techniques in software architecture," *J. Syst. Softw.*, vol. 144, pp. 533–558, 2018.

[7]   M. D. Devika and A. Sunitha Cªand Ganesh, "Sentiment analysis: a comparative study on different approaches," *Procedia Comput. Sci.*, vol. 87, pp. 44–49, 2016.

[8]   J. L. Puga, M. Krzywinski, and N. Altman, "Points of significance: Bayes' theorem," *Nat. Methods*, vol. 12, no. 4, pp. 277–278, 2015.

[9]   L. Loos, K. Verbeeck, and L. De Laet, "Data visualisation as a tool for informed structural design," *Comput. Des.*, vol. 115, pp. 267–276, 2019.

[10]   R. Scroggins, "SDLC and development methodologies," *Glob. J. Comput. Sci. Technol.*, 2014.

[11]   M. Trupthi, S. Pabboju, and G. Narasimha, "Sentiment analysis on twitter using streaming API," in *2017 IEEE 7th International Advance Computing Conference (IACC)*, 2017, pp. 915–919.

[12]   C. K.-S. Leung, C. L. Carmichael, P. Johnstone, R. R. Xing, and D. S. H.-C. Yuen, "Interactive visual analytics of big data," in *Ontologies and Big Data Considerations for Effective Intelligence*, IGI Global, 2017, pp. 1–26.

[13]   T. M. Pearce, M. N. Nikiforova, and S. Roy, "Interactive browser-based genomics data visualization tools for translational and clinical laboratory applications," *J. Mol. Diagnostics*, vol. 21, no. 6, pp. 985–993, 2019.

[14]   T. F. Efendi, "ANALYSIS OF THE IMPLEMENTATION OF THE SIMPLE SALARY SIM APPLICATION IN GROGOL DISTRICT, SUKOHARJO DISTRICT," *Int. J. Econ. Bus. Account. Res.*, vol. 4, no. 4, 2020.

[15]   I. Rana, P. Goswami, and H. Maheshwari, "A review of tools and techniques used in software testing," *Int. J. Emerg. Technol. Innov. Res*, vol. 6, no. 4, pp. 262–266, 2019.

Biography of all authors

Picture	Biography	Authorship contribution
	Nur Syafirah Izzati bt Samsari is currently an alumnus of Universiti Teknologi MARA Perak Branch, Tapah Campus (UiTM Perak). Studied in Bachelor Degree of Computer Science. Research interests include data visualization and sentiment analysis. She can be connected at 2018659762@student.uitm.edu.my.	Design the research work, data collection, data analysis and interpretation, drafting article.
	Dr. Masurah Mohamad 🆔 🔠 SC 🅿 is currently a Senior Lecturer at Universiti Teknologi MARA Perak Branch, Tapah Campus (UiTM Perak) which is an education institute established by the Ministry of Higher Education Malaysia for 15 years in the Department of Computer Science. Her research interests include data sciences and analytics, data mining and information retrievals, Artificial Intelligence and machine learning, recommender systems, soft-computing and data visualization. She can be contacted at masur480@uitm.edu.my.	Data analysis and results interpretation and article validation.
	Prof. Ali Selamat is currently a Dean of Malaysia Japan International Institute of Technology (MJIIT) which is an education institute that is established by Ministry of Higher Education Malaysia with the aim of enhancing Japanese Oriented Engineering Education in Malaysia and Asian with the support from Government of Japan through Japanese International Cooperation agency (JICA) and Universiti Teknologi Malaysia together with 29 Japanese University Consortium (JUC).	Results analysis and article validation.