Sukkur IBA Journal of Computing and Mathematical Sciences

Recognized by HEC Pakistan in "Y" Category

E-ISSN: 2522-3003

P-ISSN: 2520-0755

Volume: 4 | No: 1 | Jan - Jun | 2020

Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS) is the bi-annual research journal published by **Sukkur IBA University**, Sukkur Pakistan. **SJCMS** is dedicated to serve as a key resource to provide practical information for the researchers associated with computing and mathematical sciences at global scale.

Disclaimer: The opinions expressed in **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)** are those of the authors and contributors, and do not necessarily reflect those of the journal management, advisory board, the editorial board, Sukkur IBA University press or the organization to which the authors are affiliated. Papers published in **SJCMS** are processed through double blind peer-review by subject specialists and language experts. Neither the **Sukkur IBA University** nor the editors of **SJCMS** can be held responsible for errors or any consequences arising from the use of information contained in this journal, instead errors should be reported directly to corresponding authors of articles.

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Mission Statement

The mission of **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)** is to provide a premier interdisciplinary platform to researchers, scientists and practitioners from the field of computing and mathematical sciences for dissemination of their findings and to contribute in the knowledge domain.

Aims & Objectives

Sukkur IBA Journal of Computing and Mathematical Sciences aims to publish cutting edge research in the field of computing and mathematical sciences.

The objectives of **SJCMS** are:

1. to provide a platform for researchers for dissemination of new knowledge.
2. to connect researchers at global scale.
3. to fill the gap between academician and industrial research community.

Research Themes

The research focused on but not limited to following core thematic areas:

Computing:

- Software Engineering
- Formal Methods
- Human Computer Interaction
- Information Privacy and Security
- Computer Networks
- High Speed Networks
- Data Communication
- Mobile Computing
- Wireless Multimedia Systems
- Social Networks
- Data Science
- Big data Analysis
- Contextual Social Network Analysis and Mining
- Crowdsource Management
- Ubiquitous Computing
- Distributed Computing
- Cloud Computing
- Intelligent devices
- Security, Privacy and Trust in Computing and Communication
- Wearable Computing Technologies
- Soft Computing
- Genetic Algorithms
- Robotics
- Evolutionary Computing
- Machine Learning

Mathematics:

- Applied Mathematical Analysis
- Mathematical Finance
- Applied Algebra
- Stochastic Processes

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Patron's Message

Sukkur IBA University has been imparting education with its core values merit, quality, and excellence since its inception. Sukkur IBA University has achieved numerous milestones in a very short span of time that hardly any other institution has achieved in the history of Pakistan. The university is continuously being ranked as one of the best university in Pakistan by Higher Education Commission (HEC). The distinct service of Sukkur IBA University is to serve the rural areas of Sindh and also underprivileged areas of other provinces of Pakistan. Sukkur IBA University is committed to serve targeted youth of Pakistan who is suffering from poverty and deprived of equal opportunity to seek quality education. Sukkur IBA University is successfully undertaking its mission and objectives that lead Pakistan towards socio-economic prosperity.

In continuation of endeavors to touch new horizons in the field of computing and mathematical sciences, Sukkur IBA University publishes an international referred journal. Sukkur IBA University believes that research is an integral part of modern learnings and development. Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS) is the modest effort to contribute and promote the research environment within the university and Pakistan as a whole. SJCMS is a peer-reviewed and multidisciplinary research journal to publish findings and results of the latest and innovative research in the fields, but not limited to Computing and Mathematical Sciences. Following the tradition of Sukkur IBA University, SJCMS is also aimed at achieving international recognition and high impact research publication in the near future.

Prof. Nisar Ahmed Siddiqui
(Sitara-e-Imtiaz)
Vice Chancellor, Sukkur IBA University
Patron SJCMS

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Editorial

Dear Readers,

It is a pleasure to present to you the seventh issue of (volume 4, issue 1) of Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS). The Higher Education Commission Pakistan recognized SJCMS in **Y-Category**. I congratulate all the stakeholders for this achievement.

During the past few months, lifestyle of the people has been changed due to COVID-19. Advances in science and technology enables the survival in all areas including businesses, medical, education, health, agriculture, communication, transportation, and defense etc. This has been driven by an ever-growing volume of exciting discoveries, largely emanating from research community. In order to highlight the future technology challenges, the SJCMS aims to publish cutting-edge research in the field of computing and mathematical sciences for dissemination to the largest stakeholders. SJCMS has achieved milestones in very short span of time and is indexed in renowned databases such as DOAJ, Google Scholar, DRJI, BASE, ROAD, CrossRef and many others.

This issue contains the double-blind peer-reviewed articles that address the key research problems in the specified domain The SJCMS adopts all standards that are a prerequisite for publishing high-quality research work. The Editorial Board and the Reviewers Board of the Journal is comprised of renowned researchers from technologically advanced countries. The Journal has adopted the Open Access Policy without charging any publication fees that will certainly increase the readership by providing free access to a wider audience.

On behalf of the SJCMS, I welcome the submissions for upcoming issue (Volume-4, Issue-2, July-December 2020) and looking forward to receiving your valuable feedback.

Sincerely,

Ahmad Waqas, PhD
Chief Editor

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Publisher: **Sukkur IBA Journal of Computing and Mathematical Sciences (SJCMS)**
Office of Research Innovation & Commercialization – ORIC
Sukkur IBA University – Airport Road Sukkur-65200, Sindh Pakistan
Tel: (092 71) 5644429 Fax: (092 71) 5804425 Email: sjcms@iba-suk.edu.pk URL: sjcms.iba-suk.edu.pk

Contents

Analysis of Underground Cable Fault Techniques Using MATLAB Simulation
Muhammad Suhail Shaikh, Munsif Ali Jatoi, Muhammad Mohsin Ansari, Zahid Ali Arain, Aleem Ahmed Qader (1-10)

Aspect Based Sentimental Analysis of Hotel Reviews: A Comparative Study
Sindhu Abro, Sarang Shaikh, Rizwan Ali, Sana Fatima, Hafiz Abid Mahmood Malik (11-20)

Multimedia Based e-Learning for Educating Children
Zeeshan Bhatti, Sayed Majid Shah, Muhammad Zahid Tunio, Sadam Hussain Brohi, Anas Iqbal Memon (21-27)

Awareness regarding Coronavirus pandemic among the population of Sindh, Pakistan: A cross-sectional study
Khalida Shaikh, Areeba Memon, Irfan Ahmed Memon, Zulfiqar Ali Laghari, Abdul Moiz memon, *Sarang Shaikh, Sher Muhammad Doudpotta* (28-36)

Average Index Modelling of Campus Safety and Walkability: Case Study of the University of Sindh
Irfan Ahmed Memon ,Saima Kalwar, Noman Sahito, Sabeen Qureshi, Noman Memon (37-44)

Identifying the Machine Learning Techniques for Classification of Target Datasets
Abdul Ahad Abro, Mohammed Abebe Yimer, Zeeshan Bhatti (45-52)

Formatting Guidelines *53*

Vol. 4, No. 1 | January - June 2020

Analysis of Underground Cable Fault Techniques Using MATLAB Simulation

Muhammad Suhail Shaikh[1], Munsif Ali Jatoi[2], Muhammad Mohsin Ansari[3], Zahid Ali Arain[4], Aleem Ahmed Qader[5]

Abstract:

In densely populated areas, underground cables are used for electrical distribution and transmission purposes. Due to this, underground cables have more advantages than the overhead transmission line during abnormal weather conditions such as rain and thunderstorm. However, the calculation and determination of the exact position of fault location occur, have been a significant challenge in the underground transmission and distribution. In this research work, different techniques like sectionalizing, thumper, and time domain reflectometry are analyzed using MATLAB simulation for determining fault location in underground cable systems. Furthermore, the factors which affect the fault location are also discussed. The simulations are carried out for various symmetrical and unsymmetrical configurations for three-phase connections. The current changes with respect to time for various faults are observed. The patterns for current change suggest a nearly sinusoidal behavior with varying amplitude for current when observed for various faults.

Keywords: *Fault Location; transmission line; underground cable; distribution*

1. Introduction

The transportation of electrical energy from production to the consumers is the primary role of the electrical transmission and distribution system. Generally, when any fault is observed in the transmission line, the detection of fault is mandatory to diagnosis it so as to save the power system form further damage. Even though the cable system offers more reliability in the underground as compared to the above headline system, the identification of the fault location is inflexible. The method of fault locating has improved, which increases the demand for reliable service. In recent years, the trend of fault detection and analysis is enhanced with the latest technological application, such as signal processing methods and outcomes in the transient based analysis. It is observed that the transient signals created in the power system are investigated through wavelet transform. Over the years, a lot of research work for fault location and protection in utilizing fault generated traveling-wave techniques [1-2]. In the fault location structure in the traveling-wave current method, the distance is

[1] School of Electrical Engineering ,Yanshan University, Qinhuangdao, China
[2] Department of Biomedical Engineering, Barrett Hodgson University, Karachi, Pakistan
[3] School of Electrical Engineering, Zhejiang University China
[4] Department of Sciences and technologies, Indus University, Karachi, Pakistan
[5] School of Economics and Management ,Yanshan University , Qinhuangdao, China
Corresponding Author: suhail.shaikh@live.com

calculated via the amount of time difference at distribution end amongst an incident wave. The corresponding wave redirected from the fault has been settled for enduring faults in underground low-voltage distribution networks. Though due to the clampdown of the bandwidth of the conventional current transformer (CT) and voltage transformer (VT) which is few GHz and 50 kHz, respectively, the precision delivered by such methods of fault location is far from satisfactory for a power cable [3]. Similarly, for fault location and protection, there have been several actions that are consuming power frequency (low-frequency). Still, in such type of methods which are grounded on power frequency signals, certain valuable data related to high frequencies in the transient state is needed. In [4], artificial intelligence related to wavelet-transform is also utilized in the fault detection on power cable via neural networks. Though, it is much complex along with the increase in the speed of fault detection. In [5-7], different algorithms (for instance, one-end and two-end) algorithms were operating the voltage and current quantities for assessing the location of the fault. An innovative method based on wavelet and artificial neural networks (ANN) towards the un-grounded fault location in a practical 11-kilovolts underground power distribution system has been proposed in [8]. In [9], another off-line procedure is presented to locate the fault in the cable series based on artificial neural networks and signal-processing via wavelet. The four highly dominant types of fault in the distribution system are identified in [10]. The reviews of fault location techniques, according to their classes, are briefly discussed in [11]. Different methods are used for locating the fault in the underground cable depends on the sorts of installations. Under-ground transmission circuits are well-defined paths than the underground distribution circuit, which makes an impact on the methods of fault location application. The information needed to detect the fault location is change with the sort of installations. The literature revealed that many methods are used for fault location.

Faults can be classified into two main categories:

A. *Permanent fault*: In this fault, the electricity is permanently turn off until the fault location is identified [12]
B. *Transient fault*: In this fault, some of the power is lost.

Moreover, faults have various sorts of following types.

A. Open circuit fault.

These types of fault occur due to failure of one or more than the conductor when we can observe this fault using megger when the megger indicates zero, and it means there are no faults. If there is a fault is shows infinite reading during connecting between conductor and earth [13-14].

Short circuit fault

These types of faults occur when two conductors are in contact with each other. We can identify these types of fault using megger. If the value of megger reads zero, it means a fault has occurred. The short circuit fault is subcategorized in the following [13-14].

- **Symmetrical fault**: The symmetrical fault is also called a balanced fault; this fault occurs when all the three-phase are short-circuited simultaneously.
- **Unsymmetrical fault**: In this type of fault, the magnitude of the fault is not equal.
- **Earth fault**: When a conductor of the cable is contact with the earth, it's called earth fault.

In this paper, we have discussed underground cable fault using sectionalizing, thumper, MATLAB, and some other techniques that are also discussed in this paper.

2. Causes of fault

The majority of the faults occur due to moisture entering the insulation of the cable. Thus, atmospheric agents, soil and water, or some other such as mechanically damage the system. Faults increase the amount of voltage and currents, causing damage to the equipment, and the system will be unbalanced due to these faults. So, it is important that when the fault occurs, then isolate the fault section from the system to retain system stability [15].

3. Methods to locate the fault

A. Sectionalizing

This technique involves physically splicing and cutting of the cable to reduce the reliability of the cable and divided cable into the small section. This may enable us to detect the location of the fault, such as if we have 1000-ft cable divided into 500-ft length section, and calculate the reading with the help of high insulation resistance (IR) meter. If the IR shows a very low reading, it indicates that cable is defective. [16]

B. Thumping

In this method, noise is needed for fault detection. When a high voltage is supplied to the faulty cable, the high amount of current arc makes a noise that is enough to hear. Thumping required a high amount of voltage about 25kV to make noise in the underground cable to hear above ground. The high electric current causes heat dissipation, which damages the cable insulation. This damage can be limited by reducing the amount of power supplied [17] as shown in Fig.1, high voltage supply consists of a high voltage DC supply, and a high amount of voltage capacitor. In this experimental setup, voltage is supplied to charge the capacitor into the cable under a faulty condition. The energy stored in the capacitor is discharging into the cable producing the sound and making a vibration. The released energy is defined in Joules is calculated as in Equation. (1).

$$J = \frac{2V * C}{2} \quad (1)$$

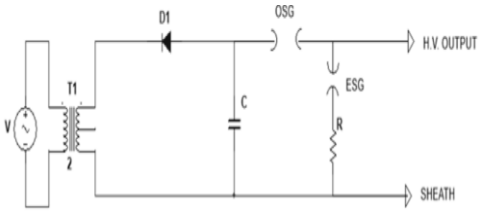

Fig 1. Circuit diagram for thumping

where

- OSG is operational speak Gap
- ESG is earthing particular Gap
- C is Capacitor
- R is Discharge rectifier
- T_1 is a High voltage transformer
- V is the voltage source

C. Time-domain reflectometry (TDR)

Time-domain reflectometry uses a very low amount of signals through the cable to detect and diagnose the location of the fault. When a signal is sent to the cable having no-fault, it returns the signal during the time and through the proper channel. However, on the other hand, a faulty cable does not return the signals. This, we can identify these time and proper channel in TDR.as shown in Fig.2. There is a block diagram of fault detection using Time domain reflectometry (TDR) [17], and the distance can be calculated using Equation. (2).

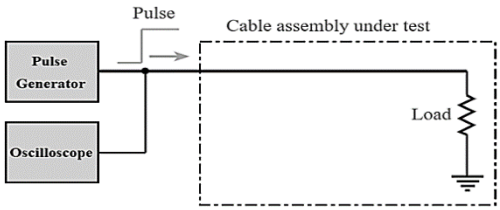

Fig 2. block diagram of TDR

$$D = vp\left(\frac{T}{2}\right) \quad (2)$$

where

- D is the distance in meters
- vp is the velocity of propagation in meter per second

- T is the transit time from the monitoring point in seconds

D. High Voltage Radar Method

High Voltage radar method is categorized as follow:

- *Arc Reflection:* This type of technique requires TDR with a filter and thumper for limiting the surge current and voltage that is sent towards the cable. This causes minimized stress to cable.
- *Surge Pulse Reflection Method:* This technique requires the use of current coupler and storage oscilloscopes.
- *The Voltage Pulse Reflection Method:* This type of voltage technique used to find the fault of more than 25kV voltage.

E. Bridge approaches

Bridge approaches involve a resistive bridge to identify the fault location. In the past years, these bridge techniques are very helpful for the cable fault location, even though the poor precision of common-null gauges (for instance, galvanometer or other means) and which further increases the precision up to 4-5 numbers with the help of digital multi-meters. And other devices, numerous services, are reconsidering the consumption of bridge approaches for locating the fault of cable. There are following two main types of bridges to locating the fault [18].

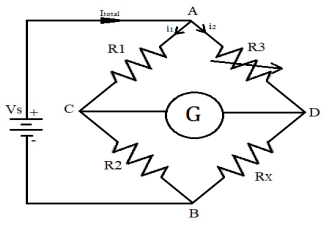

Fig 3. circuit diagram for bridge

- **Murray loop approaches**

Murray loop approach, Varley loop approach, and Pulse-Echo approach are simple basic methods to localize cable fault testing. The Murray loop approach is the fundamental approach used for locating the fault of cables. This technique utilized the elementary components that are easily acquired. These approaches are implemented in either condition, earth fault, as well as the short circuit fault in an underground cable. In these approaches, the acquired results are not affected by the resistance of fault, except the level of fault resistance is raised. The Murray approach and Varley approach are two most commonly used in the cable fault location, which works on the principle of Wheatstone bridge. The location of the fault in the underground cable can be detected via assembling one Wheatstone bridge in it through this approach, and the fault location is easily traced out by contrasting the resistance. However, the proper length of cable should be used in this case. Fig. 2 and Fig.3 revealed the Murray loop approach with proper connection for detecting the location of the fault in both cases, ground fault and short circuit fault [18].

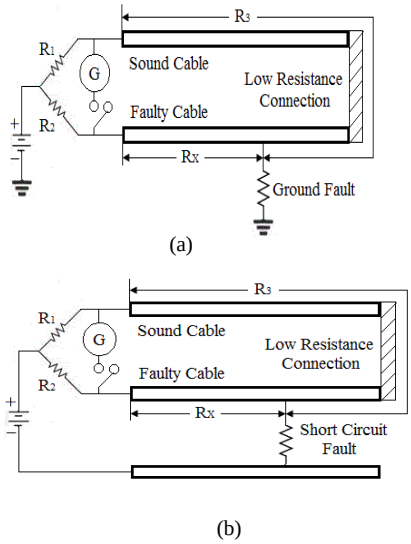

Fig 4. circuit diagram for Murray loop technique (a) ground fault (b) short circuit fault

In this technique, the sound cable is joined with the faulty cable via a low resistance wire, as the total resistance of the cable is not be affected by that resistance, and it should have

the capability to flow the loop current over the bridge circuits without any damage. R_1 and R_2 are the variable resistors creating the ratio arms. The bridge should acquire a balanced state by adequately tuning the variable resistors.

The galvanometer denoted by points out the balance. The total loop resistance moulded using the sound, and faulty cable both are in the following equations.

$$\frac{R_1}{R_2} = \frac{R_3}{R_x} \tag{3}$$

$$\frac{R_1 + R_2}{R_2} \tag{4}$$

$$\frac{R_3+R_x}{R_x} \tag{5}$$

$$R_x = \frac{R_2}{R_2+R_1}\,(R_3 + R_x) \tag{6}$$

When sound cable and faulty cable are the same in terms of the cross-sectional area, then the conductor's resistance is corresponding to their lengths. Thus, the lengths between test-end and fault-end of the faulty cable are denoted by the L, and if the total range of both cables is L, then L can be expressed as;

$$Lx = \frac{R_2}{R_2 + R_1}\,L \tag{7}$$

The above-discussed approach is valid only where there are the known lengths are used. The fault resistance should not be varied in the Murray loop approach. So, it is a bit hard to fix the bridge to balance. Consequently, the determined position of fault is less accurate. At that point, the level of temperature is raised in the cable because of high current or high voltage current. The balance will fall if the resistance regulates along with the temperature. Therefore, low voltage or current needs to supply to the circuit.

- **Varley Loop**

The test is substantial just when the cable sections are uniform all through the loop. The current flowing across the cable would create a temperature impact. Because of this temperature change, the protection of the link would change. Along these lines, we have to apply relatively less current to this circuit to perform the test. This test is utilized to discover the faulty location in an underground cable by influencing one Wheatstone bridge in it, and by comparing the resistance, we will be able to discover the location that is faulty as opposed to ascertaining it from the known lengths of the cable [18].

MATLAB is a high-level programming language developed by MathWorks, the US which is based on the manipulation of matrices for various sciences as well as business/financial applications. Within MATLAB, an additional package namely Simulink is provided which provides GUI for various functionalities such as control systems, signal processing, etc. In order to study and analyze the cable fault arrangement is used, which is shown in Table-I.

Table 1. Proposed Parameters [19]

S#	Parameters	Value
1	Line Voltage	220 kV
2	Frequency	50 Hz
3	Length of the Cable	200 km
4	Resistance per unit length	0.012 Ω
5	Inductance per unit length	0.9 mH
6	Capacitance per unit length	127 uf
7	Source Resistance	0.9 Ω
8	Source Inductance	0.16 mH
9	Active Power	10 KW
10	Inductive Reactive Power	100 VAR
11	Capacitive Reactive Power	100 R

4. Result and Discussions

In this simulation experiment, we have used a three-phase source power transmission line with three-phase [19], and this transmission line system was simulated to identify the fault on the affected phase [20]. This transmission line system was created in MATLAB 2015a in Fig.5. When symmetrical fault (LLL-LLLG) arises, all three phases are simultaneously short circuits to each other and ground. In this type of fault, the transmission line system remains symmetrical. Fault gives rise to symmetrical current equal to fault current displayed by 120° as shown in Fig.6. respectively, and the transmission model was simulated to identify the faults on the transmission line. It is observed from both waveforms that all the three phases with faults indicated some deviations from the expected waveforms. However, non-symmetrical fault (LG-LL-LLG), the three-phase transmission line unbalance during the fault occurs due to unsymmetrical fault currents (i.e. unequal fault currents in the wires with unequal phase displacement) [21]. Fig.7(a) shows the result of a single line-to-ground (LG) fault. It is observed that phase A is short-circuited and indicated some deviation from the expected waveforms.

Fig.7(b) shows a line-to-line fault (LL) when two lines are short-circuited to each other. It is observed from the simulated waveform that phase C showed some abnormality in the magnitude of the waveform, Fig.7(c) shows the result of double Line-to-ground (LLG) fault when two transmission lines are fall on the ground.

The modeling and simulation of power engineering systems pose serious challenges as it has several complications within various parameters. For ease of implementation, usually, electrical engineering researchers prefer MATLAB for modeling and simulation of electrical networks. The Simulink package available with MATLAB allows dynamic simulation and flexibility such that a controller to be modeled can be varied for various setups as other traditional simulation techniques

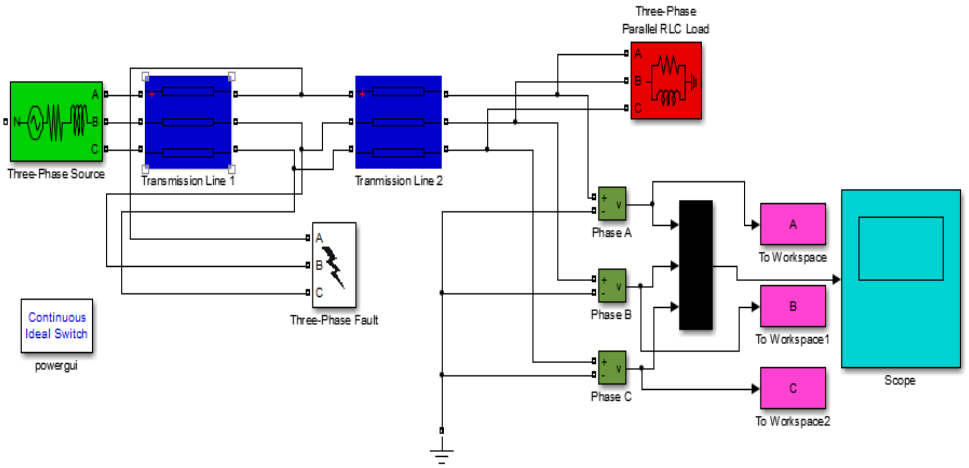

Fig 5. Simulation Modeling [19]

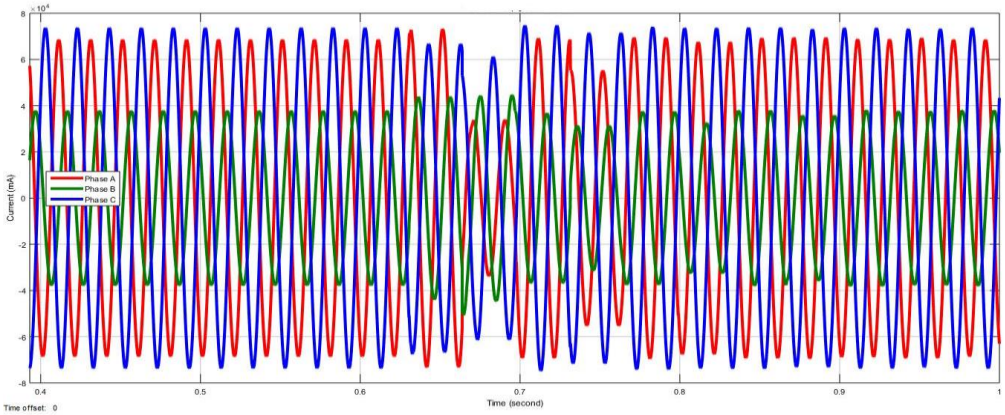

Fig.6(a). LLLG Fault (Symmetrical)

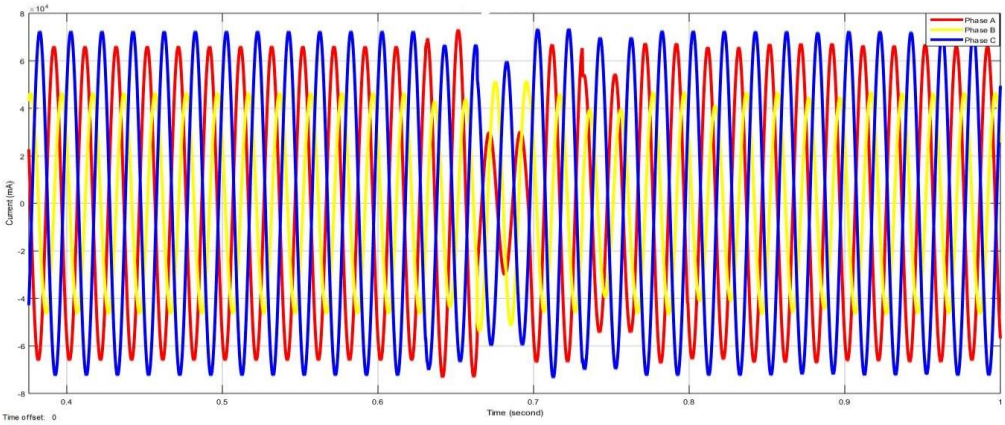

Fig 6(b). LLL Fault (Symmetrical)

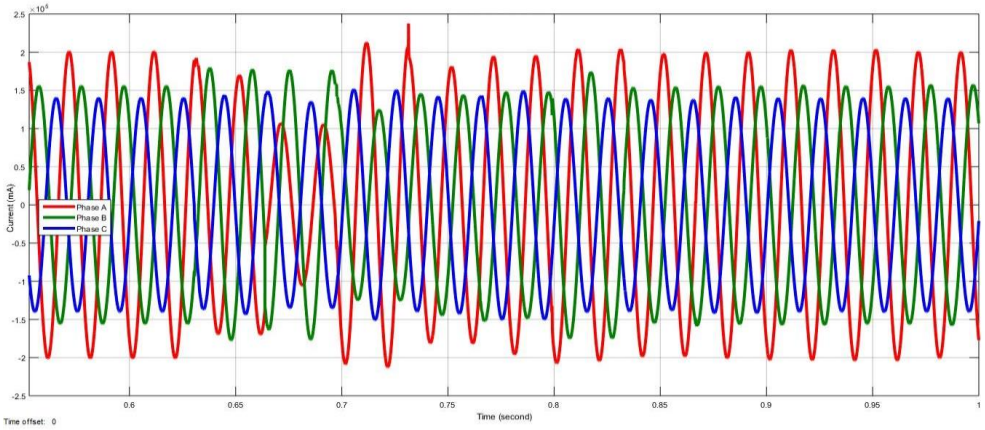

Fig 7(a). LG Fault (Un-Symmetrical)

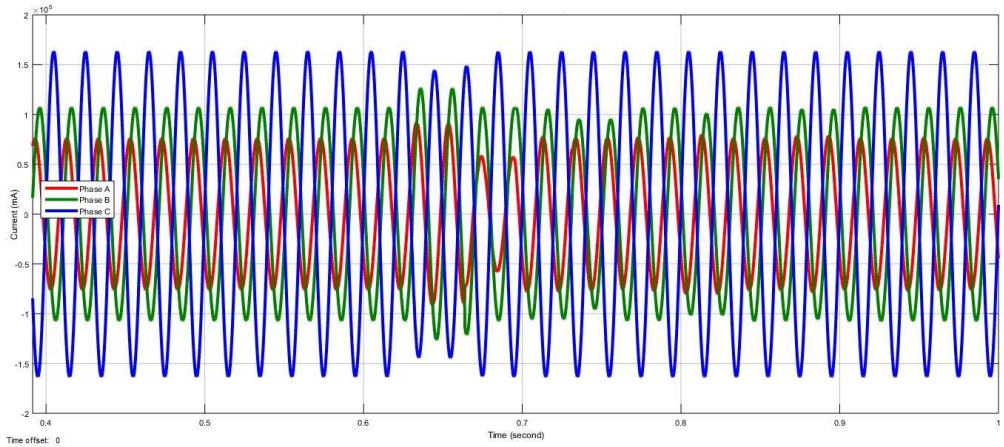

Fig.7(b). LL Fault (Un-Symmetrical)

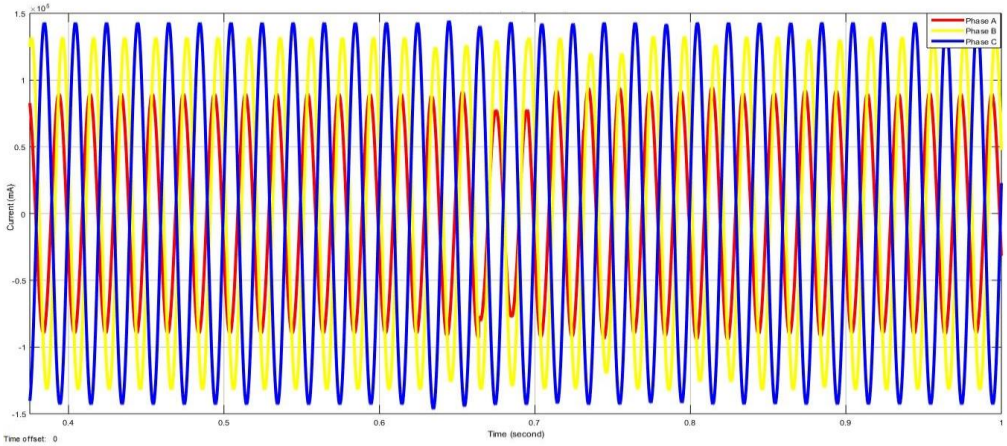

Fig 7(c). LLG Fault (Un-Symmetrical)

Table 2. Comparison of Techniques

S#	Category	Advantages	Disadvantages
1	Sectionalizing	The simplicity of the system	Risk of less reliability of the cable
		Easy maintenance	Labor intensive method
2	Thumping	No splicing of the cable	degradation insulation of the cable
		good result	High voltage
3	TDR	low-energy signals	accuracy only up to 1%
		no insulation degradation	When resistance is more significant than 200 ohms, it stops working
4	High-voltage Radar Method	limits the surge current and voltage	high output surge
		detect difficult faults	more skill compared
5	Varley loop	High sensitivity	R_1 and R_2 are fixed
6	Muray Loop	Very easy	Low sensitivity

5. Conclusion

In this paper, we have discussed and studied different techniques of fault in underground cable and MATLAB simulation of symmetrical and non-symmetrical fault as in Fig.6. It is observed that during the asymmetrical fault, the system will become balanced. The angle between the phase is the same in three phases. This type of fault rarely occurs, while in Unsymmetrical fault, the thee phase will become unbalanced during fault conditions as in Fig. 7. We can remove this fault using symmetrical components like positive, negative, and zero sequences. During this study, we study different types of techniques to find fault in the Underground cable. Moreover, We also performed a simulation on MATLAB. From these waveforms, we can identify the fault in the transmission line.

REFERENCES

[1] Moshtagh, Jamal, and R. K. Aggarwal. "A new approach to ungrounded fault location in a three-phase underground distribution system using combined neural networks & wavelet analysis." 2006 Canadian Conference on Electrical and Computer Engineering. IEEE, 2006.

[2] Moshtagh, Jamal, Parham Jalili, and Hassan Pourvali Souraki. "A New Approach to Series Fault Location in Underground Distribution System Using Combined Neural Networks & Wavelet Analysis."

[3] Qingchao, Zhang, et al. "Transmission line fault location for single-phase-to-earth fault on non-direct-ground neutral system." IEEE transactions on power delivery 13.4 (1998): 1086-1092.

[4] Novosel, Damir, et al. "Unsynchronized two-terminal fault location estimation." IEEE transactions on Power Delivery 11.1 (1996): 130-138.

[5] Girgis, Adly A., David G. Hart, and William L. Peterson. "A new fault location technique for two-and three-terminal lines." IEEE Transactions on Power Delivery 7.1 (1992): 98-107.

[6] Kezunovic, M., and B. Perunicic. "Automated transmission line fault analysis using synchronized sampling at two ends." IEEE Transactions on Power Systems 11.1 (1996): 441-447.

[7] Wang, Chun, et al. "A WAMS/PMU-based fault location technique." Electric Power Systems Research 77.8 (2007): 936-945.

[8] Butler, Karen L., and James A. Momoh. "Detection and classification of line faults on power distribution systems using neural networks." Proceedings of 36th Midwest Symposium on Circuits and Systems. IEEE, 1993.

[9] Gana, N., et al. "A comprehensive review of fault location methods for distribution power system." Indonesian Journal of Electrical Engineering and Computer Science 6.1 (2017): 185-192..

[10] Karan Sanhotra, "Methods of Fault Location in Cable System", Department of Electrical Engineering, Sardar Vallabhbhai National Institute of Technology, Surat 395007, India, International Journal of Electrical and Electronics Research ISSN 2348-6988,Vol. 3, Issue 4, pp: (8-11), Dec-2015.

[11] Kor, Snehal C., and Vivek D. Bavdhane. "Modified Murray Loop Method for Underground Cable Fault Location Detection Using GSM." International Journal of Science and Research 5.6 (2016): 1892-1894.

[12] Darvhankar, G. S., et al. "Study Of 3-ph Underground Cable Fault Locator Using Acoustic Method." Development 2.1 (2015).

[13] Aishwarya *N*. n.gayathri, v.kowsalya, m.kalidas, s.deepika "Automatic underground cable fault detection with sms alert" volume, 05 issue, 02 , feb-2018

[14] www.electronicshub.org/types-of-faults-in-electrical- power systems

[15] www.vishwangelectrospaces.com/blogs/faults-in-underground cables

[16] www.everythingrf.com/community/what-is-a-time- domain reflectometer

[17] Karan Sanhotra" Methods of Fault Location in Cable System" International Journal of Electrical and Electronics Research, Vol. 3, Issue 4, pp: (8-11), Month: October - December 2015

[18] Sahilbhai Vhora, Jaimin Atodaria, Ashutosh Singh, Hardik Kotak, Ankur K. Gheewala" Underground Cable Fault Detection Using Matlab" International Journal of Innovative Research in Science, Engineering and Technology, Vol. 7, Issue 12, December 2018

[19] Hasija, Kunal, et al. "Detection and location of faults in underground cable using Matlab/Simulink/ANN and OrCad." 2014 6th IEEE Power India International Conference (PIICON). IEEE, 2014.

[20] Hasija, Kunal, et al. "Detection and location of faults in underground cable using Matlab/Simulink/ANN and OrCad." 2014 6th IEEE Power India International Conference (PIICON). IEEE, 2014.

[21] Principles of power system V.K Mehta "Revised Edition"

Vol. 4, No. 1 | January - June 2020

Aspect Based Sentimental Analysis of Hotel Reviews: A Comparative Study

Sindhu Abro[1], Sarang Shaikh[1], Rizwan Ali[1], Sana Fatima[1], Hafiz Abid Mahmood Malik[2]

Abstract:

The increasing use of the internet enables users to share their opinion about what they like and dislike regarding products and services. For efficient decision making, there is a need to analyze these reviews. Sentiment analysis or opinion mining is commonly used to detect polarity (positive or negative) of reviews. But it does not show the aspect or orientation of the text. In this study, we have employed state-of-art approaches to perform three tasks on the SemEval dataset. Tasks A and B are related to predicting the aspect of the restaurant's reviews, whereas task C shows their polarity. Additionally, this study aims to compare the performance of two feature engineering techniques and five machine learning algorithms to evaluate their performance on a publicly available dataset named SemEval-2015 Task 12. The experimental results showed that the word2vec features when used with the support vector machine algorithm outperformed by giving 76%, 72%, and 79% off overall accuracies for Task A, Task B, and Task C respectively. Our comparative study holds practical significance and can be used as a baseline study in the domain of aspect-based sentiment analysis.

Keywords: *Aspects Based Sentiment Analysis; Sentiment Analysis; Text Classification; Natural Language Processing (NLP); Word2Vec; Machine Learning*

1. Introduction

In recent years, there is a rapid growth of content generated by users on the internet. The web enables users to share their reviews and experiences about services and products. Moreover, it is a growing trend that customers look already available reviews before purchasing any product or service [1]. Therefore, sellers and organizations need to analyze the reviews for effective decision making. The manual process to analyze the reviews is a labor-intensive and time-consuming task. Hence, techniques like sentiment analysis or opinion analysis are commonly used to extract information from reviews. The sentiment analysis, under the domain of natural language processing, used to determine the general opinion (e.g. positive or negative) of the group of individuals regardless of topic or entity (e.g. food, price, location, etc.) [2]. Therefore, it is recommended to use aspect-based sentiment analysis (i.e. ABSA). This concerned with the decomposition of two tasks namely aspect identification and sentiment analysis [3]. In the first task, the aspect of an entity is identified and in the second task, the polarity is estimated for each identified aspect. The sentiment analysis on the aspect level performs an in-depth analysis of reviews [4].

[1]Department of Computer Science, Sukkur IBA University, Pakistan
[2]Department of Computer Science, Arab Open University, Bahrain

For example, when we look at the reviews of the restaurant, ABSA not only returns the overall sentiment of the reviews but also returns for which entity the sentiment is talking about. Such as food, price, location, service, etc. Thus, the results generated from this technique gives a better understanding of what reviewers like and dislike regarding the topic [1]. Moreover, it may help customers to decide on the purchase of the products or using the services. Additionally, ASBA enables manufacturers to improve the quality of their products and services. Therefore, in this study, we have used ABSA to identify the aspects and their polarity of the reviews related to the restaurant.

The proposed solution employed the different feature engineering techniques and ML algorithms to classify restaurant reviews under different entities, attribute, and their polarities. Regardless of this extensive amount of work, it remains difficult to compare the performance of these approaches to classify hotel reviews text. To the best of our knowledge, the existing studies lack the comparative analysis of different feature engineering techniques and ML algorithms regarding the reviews related to restaurants. Therefore, this study contributes to solving this problem by comparing two feature engineering and five ML classifiers on the standard dataset provided by SemEval. This study will serve future researchers in the field of automatic ABSA.

This rest of the paper is organized as Section 2 highlights the related works. Section 3 discusses the methodology. Sections 4, and 5 explain the experimental setup, and results. Finally, Section 6 discusses the conclusion, and future work as well.

2. Related Works

Kiritchenko et al. [5] classified the reviews using the lexicon and linguistic features. Castellucci et al. [6] used a feature based on a bag of words that have been learned from external data. Hu and Liu [7] used an association rule-based system for aspect identification. Additionally, his book [8] highlights the four methods to extract aspects namely, frequent phrases, opinion, and target relations, supervised learning, and topic models. Jakob and Gurevych [9] employed the conditional random fields for aspect term.

Bhattacharyya [11] developed the system which uses dependency parsing rules for opinion extraction. Many researchers used a hybrid approach (i.e. NLP with statistical methods) to improve the performance of the system. In SemEval 2014, Kiritchenko et al. [5] used an entity tagging system named as in-house to extract outside and aspect terms. Toh and Wang [12] used the tagging approach with Wordnet and word clusters. Socher et al. [13] employed grammatical cues with deep learning. Carrascosa [14] study showed that an ensemble learning technique can also be applied in sentiment analysis. In the Aspect Category Polarity Detection task in SemEval 2014, Mohammad et al. [15] achieved the best performance by using different linguistic features, additionally, they also used publically available sentiment lexicon.

Broadly, ABSA methods can be divided into two categories, one that uses domain-independent solutions [16] and second is to use domain-specific knowledge [4] to improve the results. There is a common approach used by researchers that they treat aspect extraction and their polarity classification independently [17], but others also trained one model to solve the two problems [18].

3. Methodology

3.1. Overview

This section represents the overall research methodology that has been followed to perform the ASBA. Fig 1 shows the steps required to train the model. As shown here, our research methodology is composed of six key steps namely data collection, data preprocessing, feature engineering, data selection, classification model construction, and classification model evaluation. The

details of each step are discussed in subsequent sections.

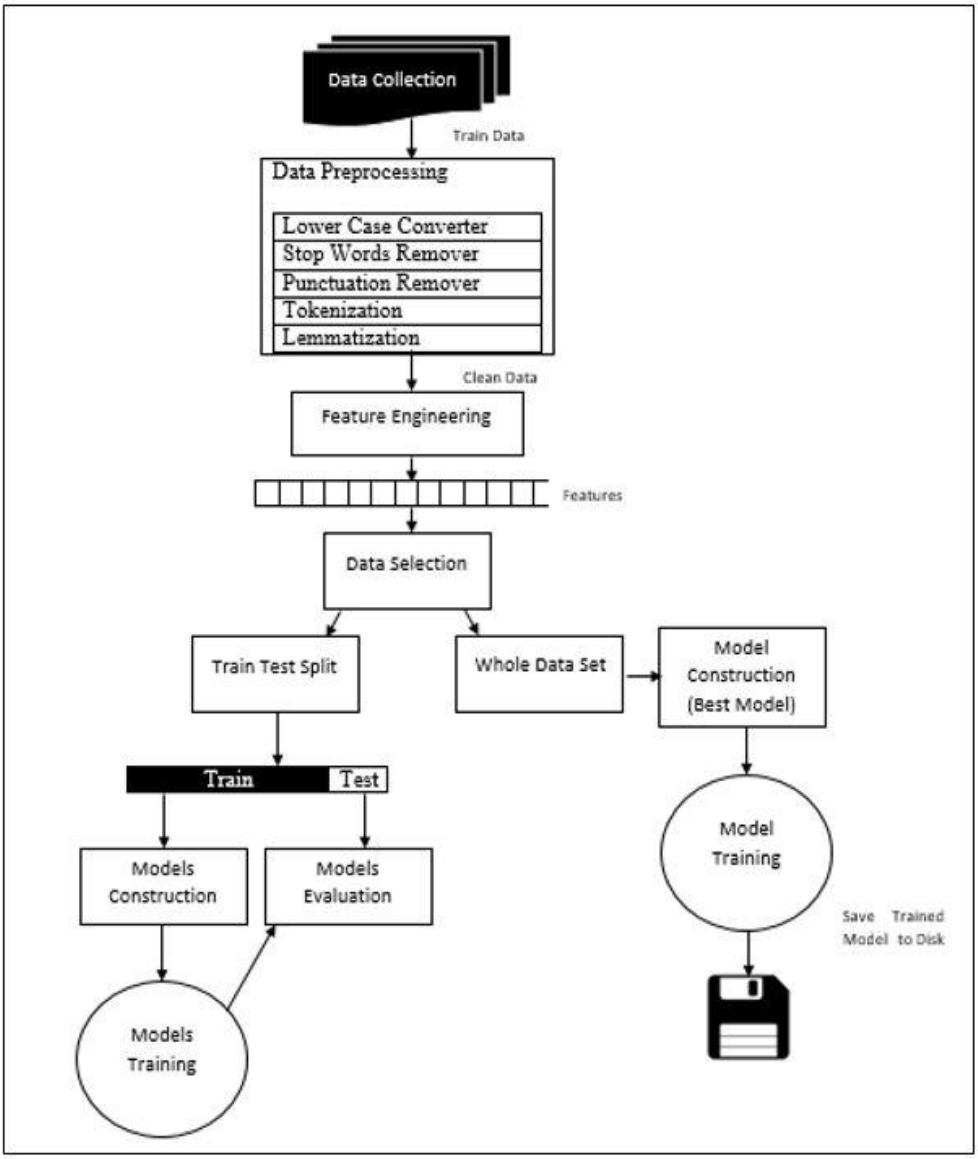

Fig 1. Overall Proposed Methodology

3.2. Data Collection

In this study, we have used publicly available data set from SemEval-2016 - Task 51[3]. This dataset contains reviews for laptops and restaurants. In this study, we will only focus on the reviews related to the restaurant. There is 3658 number of instances for the restaurant; 2799 for training and the remaining 859 for testing. In this dataset, the reviews can be categorized on the basis of aspects (i.e. category, entity, or attribute) and their polarities. By using aspect-based classification, the reviews can be labeled into six distinct classes of entity columns namely, food, restaurant, service, ambiance, drinks, and location. Additionally, the attribute can be labeled as general, quality, prices, style-options, and miscellaneous classes. However, their polarities can be positive, negative, or neutral. The distribution of reviews in training data based on entity, attribute, and polarity is shown in Fig 2, Fig 3 and Fig 4 respectively.

3.3. Text Preprocessing

Several studies show that there is a need to clean data for better classification results [19]. Therefore, we applied several preprocessing techniques to remove features from the data that are not informative. In this step, we have dropped the instances with blank values i.e. 292. Additionally, we have dropped the columns that are not required for text classification i.e. review-id, sentence-id, target, and category. After dropping the empty cells and selecting the required attributes, we converted the text (2507 remaining instance) into a lower case. Using regular expressions and pattern matching techniques, we removed white spaces, punctuation's and stop words. In addition, we have also applied tokenization and lemmatization on the preprocessed text. In tokenization, each sentence is converted into tokens or words, then words are converted to their root forms using WordNet lemmatizer e.g. posts to post

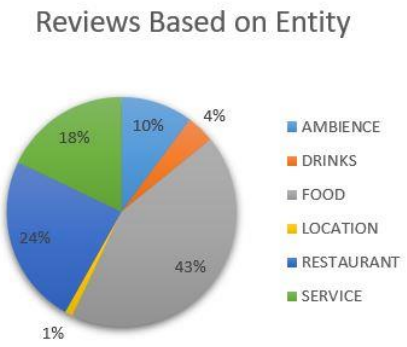

Fig. 2. Entity base distribution

Fig. 3. Attribute Base Distribution

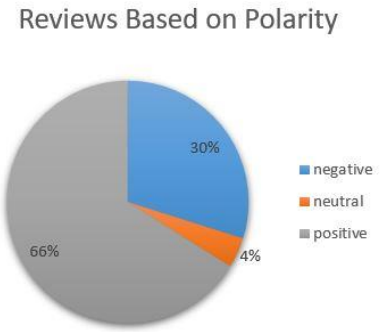

Fig. 4 Review base distribution

[3] The dataset is available at: http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

3.4. Feature Engineering

To learn classification rules, ML algorithms need numerical vectors because they cannot learn from raw data. Therefore, in classification one of the key steps is feature engineering. This step is used to extract the key features from raw text and represents the extracted features in numerical form. In this study, we have performed two types of features engineering techniques namely n-gram [20] with TFIDF [21], and Word2vec [22].

3.5. Data Selection

In this section, we have used two approaches to build the models named as train test split and whole data set. In the first approach, we have used the Pareto Principle. According to this principle, "80% of effects come from 20% of causes" [28]. This principle is also called an 80:20 ratios. In this study, we have split preprocessed data into a previously given ratio i.e. 80% for training and 20% for testing. Table 1, Table 2, and Table 3 show the class-wise distribution on the basis of an entity, attribute, and polarity as well as their train test splitting ratio. The training data is used to train the classification models for learning rules. However, the test data is used to evaluate the trained models.

Table 1: Approach I (Entity)

Class	Label	Total	Train	Test
Ambience	0	255	204	51
Drinks	1	99	79	20
Food	2	1076	861	215
Location	3	28	22	6
Restaurant	4	600	480	120
Service	5	499	359	90
Total		**2507**	**2005**	**502**

Table 2: Approach I (Attribute)

Class	Label	Total	Train	Test
General	0	1154	923	231
Miscellaneous	1	98	78	20
Prices	2	190	152	38
Quality	3	896	717	179
Style_options	4	169	135	34
Total		**2507**	**2005**	**502**

Table 3: Approach I (Polarity)

Class	Label	Total	Train	Test
Negative	0	749	599	150
Neutral	1	101	81	20
Positive	2	1657	1325	332
Total	3	**2507**	**2005**	**502**

In the second approach, we have used the whole data (i.e. 2507 number of instances) to train the model and for evaluation, different test data (i.e. 859 number of instances) were used. Table 4, Table 5 and Table6 show the distribution of data on the basis of entity, polarity, and attribute respectively.

Table 4: Approach II (Entity)

Class	Label	Total	Train	Test
Ambience	0	321	255	66
Drinks	1	137	99	38
Food	2	1467	1076	391
Location	3	41	28	13
Restaurant	4	796	600	196
Service	5	604	449	155

Table 5: Approach II (Attribute)

Class	Label	Total	Train	Test
General	0	1530	1154	376
Miscellaneous	1	131	98	33
Prices	2	238	190	48
Quality	3	1231	896	355
Style_options	4	236	169	67
Total		**3366**	**2507**	**859**

Table 6: Approach II (Polarity)

Class	Label	Total	Train	Test
Negative	0	953	749	204
Neutral	1	145	101	44
Positive	2	2268	1657	611
Total	3	**3366**	**2507**	**859**

3.6. Machine Learning Models

According to "no free lunch theorem" [23], any single classifier cannot outperform better on all types of datasets. Therefore, it is suggested to apply several classifiers on a master numerical vector to see which one achieves better results. Hence, we chose five different classifiers Naïve Bayes (NB) [24], Support Vector Machine (SVM) [25], Random Forest (RF) [26], Logistic Regression (LR) [27], and Ensemble in approach 1. Whereas in approach 2, we have chosen SVM and NB classifiers.

3.7. Classifier Evaluation

In this step, the constructed classifiers were used to predict the class of unlabeled text using test sets. The classifier performance is evaluated by calculating true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These four numbers constitute a confusion matrix as in Fig 5. To assess the performance of the constructed classifiers different performance metrics can be used like precision, recall, F measure, or accuracy. The details of given performance measures are given in [29]. However, in this study, we have used the most commonly used measure i.e. accuracy to evaluate the constructed classifiers. The details of this performance measure are given below.

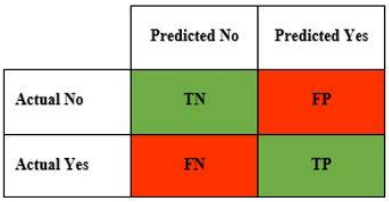

Fig. 5 Confusion Matrix

Accuracy

This evaluation matrix refers to the total number of instances that are correctly classified by the trained model. Refer to (1).

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

4. Experimental Setup

As mentioned in section A, the reviews can be categorized on the basis of aspects and their polarities. In this study, we have performed three tasks. In Task A, we have classified the reviews according to entity type (i.e. food, restaurant, service, ambiance, drinks, and location). In Task B, reviews are categorized according to attributes and labeled as general, quality, prices, style-options, and miscellaneous classes. Whereas in Task C, we have classified reviews according to their polarity like positive, negative, and neutral.

For all these tasks we have used two master feature representations namely n-gram (bigram) with TFIDF [21] and Word2Vec [22]. By using these master feature representations, we have followed two approaches to train the models. In approach 1, we used the train test split to train the five classifiers and evaluated their performance on test data. Whereas in approach 2, we used the

whole dataset to train the models which have outperformed in approach 1 and evaluated their performance by using different test data.

5. Results

This section reports the results of all three tasks. Table 7, Table 8 and Table 9 show the accuracy using approach 1 (i.e. train test split) for Task A, B, and C, respectively. As shown in all three tables, the highest accuracy for Task A (0.71), Task B (0.69), and Task C (0.81) were obtained by SVM with word2vec.

Table 7: Approach I Results (Train-Test Split) - Task A

	Task A	
	Bigram (TFIDF)	**Word2Vec**
LR	0.59	0.70
NB	0.55	0.65
RF	0.58	0.57
SVM	**0.63**	**0.71**
Ensemble	0.61	0.67

Table 8: Approach I Results (Train-Test Split) - Task B

	Task B	
	Bigram (TFIDF)	**Word2Vec**
LR	0.60	0.67
NB	**0.61**	0.58
RF	0.54	0.56
SVM	0.58	**0.69**
Ensemble	0.57	0.66

In text-classification models, the SVM classifier performed exceptionally well among all 5 classifiers. If we evaluate the performance of all classifiers with respect to master feature representation, then we can see in Table 10 and Table 11 that for Task A and Task C the SVM classifiers with both master feature representations outperformed.

Whereas, from Table 8, Task B the NB using bigram with TFIDF (0.61) and SVM with word2vec (0.69) obtained the highest accuracy. Therefore, in approach 2, we have trained 6 models (3 Tasks x 2 master feature representations) on the whole dataset. The detail of all combinations is shown in Table 10.

Table 9: Approach I Results (Train-Test Split) - Task C

	Task C	
	Bigram (TFIDF)	**Word2Vec**
LR	0.73	0.80
NB	0.75	0.74
RF	0.73	0.74
SVM	**0.78**	**0.81**
Ensemble	0.75	0.80

Table 10: Model Selection for Approach II

Task	**Bigram (TFIDF)**	**Word2Vec**
A	NB	SVM
B	SVM	SVM
C	SVM	SVM
Ensemble	0.75	0.80

We have evaluated all these models on test data (i.e. 859). Table 11 shows the results of approach 2. It shows that word2vec obtained the best performance as compared to bigram features with TFIDF.

Table 11: Approach 2 Results for Task A, B & C

Task	**Bigram (TFIDF)**	**Word2Vec**
A	0.70	**0.76**
B	0.67	**0.72**
C	0.78	**0.79**

Furthermore, Fig 6, Fig 7 and Fig 8 show the confusion matrices of best-performing analyses. Fig 6 shows the confusion matrix of the SVM classifier using word2Vec for Task A. As shown here, out of 859 instances, 651 were correctly classified. Of these 651 instances, 47, 11, 341, 140, and 112 were classified as ambiance, drinks, food, restaurant, and service respectively. We can see that all 13 instances of location class were falsely classified.

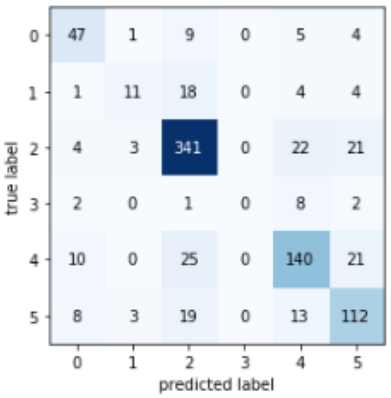

Fig. 6 Task A (Feature: Word2Vec, Classifier: SVM)

However, Fig 7 shows the confusion matrices of the SVM classifier using word2Vec features for Task B. As shown here, 621 instances out of 859 were correctly classified (i.e. General: 336 out of 376, Miscellaneous: 0 out of 33, Prices: 19 out of 48, Quality: 262 out of 335, and Style-options: 4 out of 67).

For Task C, the confusion matrix is shown in Fig 8. It shows that the SVM classifier with word2Vec features correctly classified 621 out of 859 instances, 124 as negative, and the remaining 557 as positive. As shown here, its performance was lowest in class 1 (i.e. neutral).

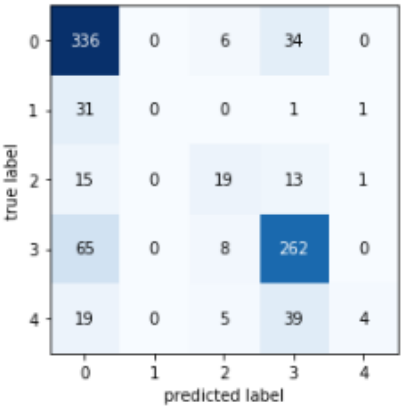

Fig. 7 Task B (Feature: Word2Vec, Classifier: SVM)

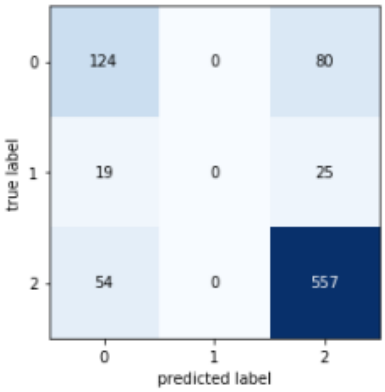

Fig. 8 Task C (Feature: Word2Vec, Classifier: SVM)

6. Conclusion

This study applied automated text classification techniques to classify the restaurant's reviews according to aspect and their polarities. Moreover, this study compared two feature engineering techniques and five ML algorithms to perform three tasks like a) classification of restaurant's reviews according to entity type, b) classification of restaurant's reviews according to their attribute and c) classification of restaurant's reviews according to their polarities. The experimental results showed that the word2vec showed better results for all tasks as

compared to bigram represented through TFIDF feature engineering techniques. Moreover, the SVM algorithm showed better results as compared to NB, LR, RF, and Ensemble for all three tasks. The lowest results were observed in NB, RF, and LR for Task A, Task B, and Task C respectively. The outcomes from our study hold practical significance because these will be used as a baseline to compare future researches within different automatic text classification methods. In the future, the accuracy of the proposed system's classification can be increased by the following two strategies. First, the deep learning-based approaches will be explored and evaluated by comparing it with current state-of-the-art results. Secondly, more instances will be collected and used in the experiments for learning the classification rules efficiently.

REFERENCES

[1] Ekawati, D., and M.L. Khodra. Aspect-based sentiment analysis for Indonesian restaurant reviews. in 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA). 2017. IEEE.

[2] Wang, J., Encyclopedia of Data Warehousing and Mining, (4 Volumes). 2009: iGi Global.

[3] Schouten, K. and F. Frasincar, Survey on aspect-level sentiment analysis. IEEE Transactions on Knowledge and Data Engineering, 2015. 28(3): p. 813-830.

[4] Thet, T.T., J.-C. Na, and C.S. Khoo, Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of information science, 2010. 36(6): p. 823-848.

[5] Kiritchenko, S., et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). 2014.

[6] Castellucci, G., et al. Unitor: Aspect based sentiment analysis with structured learning. in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). 2014.

[7] Hu, M. and B. Liu. Mining opinion features in customer reviews. in AAAI. 2004.

[8] Liu, B., Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 2012. 5(1): p. 1-167.

[9] Jakob, N. and I. Gurevych. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. in Proceedings of the 2010 conference on empirical methods in natural language processing. 2010. Association for Computational Linguistics.

[10] Zhuang, L., F. Jing, and X.-Y. Zhu. Movie review mining and summarization. in Proceedings of the 15th ACM international conference on Information and knowledge management. 2006.

[11] Mukherjee, S. and P. Bhattacharyya. Feature specific sentiment analysis for product reviews. in International Conference on Intelligent Text Processing and Computational Linguistics. 2012. Springer.

[12] Toh, Z. and W. Wang. Dlirec: Aspect term extraction and term polarity classification system. in Association for Computational Linguistics and Dublin City University. 2014. Citeseer.

[13] Socher, R., et al. Recursive deep models for semantic compositionality over a sentiment treebank. in Proceedings of the 2013 conference on empirical methods in natural language processing. 2013.

[14] Carrascosa, R., An entry to kaggle's' sentiment analysis on movie reviews' competition. 2014.

[15] Mohammad, S.M., S. Kiritchenko, and X. Zhu, NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. arXiv preprint arXiv:1308.6242, 2013.

[16] Lin, C. and Y. He. Joint sentiment/topic model for sentiment analysis. in Proceedings of the 18th ACM conference on Information and knowledge management. 2009.

[17] Brody, S. and N. Elhadad. An unsupervised aspect-sentiment model for online reviews. in Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics. 2010. Association for Computational Linguistics.

[18] Jo, Y. and A.H. Oh. Aspect and sentiment unification model for online review analysis. in Proceedings of the fourth ACM international conference on Web search and data mining. 2011.

[19] Shaikh, S. and S.M. Doudpotta, Aspects Based Opinion Mining for Teacher and Course Evaluation. Sukkur IBA Journal of Computing and Mathematical Sciences, 2019. 3(1): p. 34-43.
[20] Cavnar, W.B. and J.M. Trenkle. N-gram-based text categorization. in Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval. 1994. Citeseer.
[21] Ramos, J. Using tf-idf to determine word relevance in document queries. in Proceedings of the first instructional conference on machine learning. 2003. Piscataway, NJ.
[22] Mikolov, T., et al. Distributed representations of words and phrases and their compositionality. in Advances in neural information processing systems. 2013.
[23] Ho, Y.-C. and D.L. Pepyne, Simple explanation of the no-free-lunch theorem and its implications. Journal of optimization theory and applications, 2002. 115(3): p. 549-570.
[24] Lewis, D.D. Naive (Bayes) at forty: The independence assumption in information retrieval. in European conference on machine learning. 1998. Springer.
[25] Joachims, T. Text categorization with support vector machines: Learning with many relevant features. in European conference on machine learning. 1998. Springer.
[26] Xu, B., et al., An Improved Random Forest Classifier for Text Categorization. JCP, 2012. 7(12): p. 2913-2920.
[27] Wenando, F.A., T.B. Adji, and I. Ardiyanto, Text classification to detect student level of understanding in prior knowledge activation process. Advanced Science Letters, 2017. 23(3): p. 2285-2287.
[28] Dunford, R., Su, Q., & Tamang, E. (2014). The pareto principle.
[29] Seliya, N., T.M. Khoshgoftaar, and J. Van Hulse. A study on the relationships of classifier performance metrics. in 2009 21st IEEE international conference on tools with artificial intelligence. 2009. IEEE.

Vol. 4, No. 1 | January - June 2020

Multimedia Based e-Learning for Educating Children

Zeeshan Bhatti[1], Sayed Majid Shah[1], Muhammad Zahid Tunio[2], Sadam Hussain Brohi[1], Anas Iqbal Memon[1]

Abstract:

The education system is rapidly updating with the involvement of technology. The primary education is considering the backbone of the children's study. Normally, the children do not memories their lesson at home and they play games on mobile instead of reading. In this paper, a digital learning approach has introduced based on Sindhi primary education using the mobile application. The children/students can learn easy basics of primary level like; numbers, Sindhi and English alphabets, shapes, and colors along with lessons. A digital approach is consisting of a mobile application that manages the basic learning features with an attractive user-friendly environment together with the Mp3 audio voice and high resolution of images in HD quality. The main features will be appeared on the screen by launching the application in the form of icons as ABC, ابپ, 123, and click them for further process. When a single alphabetical reading task is complete, then it moves into the Arabic form of the Sindhi alphabet with the addition of Zer, Zabar, Pesh. These alphabetical characters connecting to each other for the creation of Sindhi words such as; آهي(Is) {Aahe}, آهن(Are) {Aahin}, هُو (Who) {Hoo}and so on. The next level for children to learning, the basic lessons of Sindhi with audio and the same tasks for basic English for primary education learning. At the initial level of digital learning approach, a Sindhi textbook named "سنڌي ٻاراڻو ڪتاب" (Sindhi Childs Book) {Sindhi Barano Kitab} has been covered through this application.

Keywords: *Multimedia Learning, e-Learning, Cognitive Learning*.

1. Introduction

In the modern century, the advancement of technology devices such as; smartphones, laptops, and computers is a very essential requirement for the educational learning environment. To utilize these devices in educational sectors, schools, colleges, universities, the teachers teach by multimedia with innovative manners, this is also a motivational way for students to capture the information quickly [7]. The innovative learning material design is a key feature of the learning environment [9]. The student's concentration is noticed to be relatively good, during learning using a mobile application [15][10]. Mobile-based learning (MBL) may be the cause of recalling the previous lecturer or discussion and there are chances to reduce the memorization issue for learning [11]. It is observed that students enjoyed learning academia using multimedia devices [8][16].

[1]Institute of information and Communication Technology, University of Sindh, Jamshoro
[2]Department of Computer Systems Engineering, Dawood University of Engineering and Technology, Karachi
Corresponding Author: Zeeshan.bhatti@usindh.edu.pk

However, the development of innovative, informative, logical, and interesting material design based on the mobile application or multimedia-based is one of the challenging tasks together with managing the graded assignments [12]. In the environment of multimedia-based learning, the students develop a critical mind and this factor enables them to become a rapid learner along with a problem-solving approach [13] [17].

2. Related Work

Savannah, a mobile gaming tool was developed for children learning purposes. As the experimental results, there were ten children (5 boys and 5 girls) aged between 11 to 12 years who played gaming and supporting learning using the mobile application. In the ending there were 80% positive findings declared [1]. By the three years of study in Singapore, a Self-Regulated Learning (SRL) model was introduced in the field of educational psychology. The challenges and characteristics were based on ubiquitous learning and mobile learning. There was also a conceptual framework for understanding the learning using mobile from everywhere, anytime [2]. An operational model is discussed with the aim to improve the quality of learning and reducing the time for learning. The introduction of this model in the higher education system, there were 12 universities specific courses had been covered based on multimedia ethics. There were 23 variables used as indicators for capturing the performance and that indicators were categorized into six elements; however, each element shows uniquely representation such as; Confidence, ease of learning, ease of understanding, and so on [3]. Another study, addressed the problem-based learning using multimedia technology. The purpose was that what the teachers have taught to students and what the students have learned from the teachers to analyze that whole scenario and resolve the problems have faced during the teaching [4]. An application of a computer learning based elementary education system was built to monitor the students and the positive results were obtained [5]. Another study was conducted to facilitate the children through a learning app named as Easy Lexia. This was a mobile-based application to face the challenges and solve the learning difficulties. This mobile application consists of some attractive features along with books. The book tasks consisting of storytelling which was developed to follow the research study of dyslexia. The targeted stories according to the age group of 7 to 12 years old age children [6].

3. Methodology

The learning paradigm consists of four basic subjects of primary education i.e. Sindhi, English, Mathematics, and Drawing. The Sindhi subject, covers the single alphabets with the sound of each sign, after that the Arabic script cursive form of Sindhi alphabet will be combined with Zer, Zabar, and Pesh. After finishing with the basic cursive script, the words joining technique will be taught to children. After completing the word learning technique, the Sindhi basic book will start to the next level for the student where they can learn how to use those words inappropriate sentences.

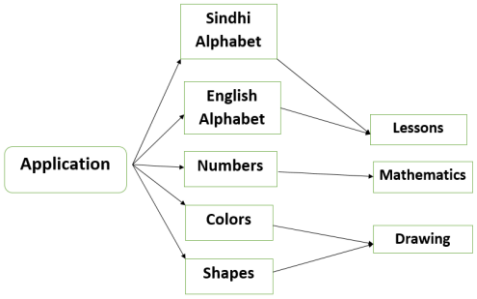

Fig 1: Learning Environment Model Based on Mobile Application

The English learning exercise initially involves basic alphabets then basic spellings

of words with audio voice. After completing the first lesson, the second lesson consists of basic sentences that are commonly used and spoken by children. The Mathematics subject starts with the numbers with a voice. After completing this task, a higher level for students is simple calculations learning which is consists of four operations plus, minus, division, multiplication (+, -, *, /). The fourth subject is Drawing, this starts with identifying colors like red, blue, black, white, pink, etc. Then simple line draw tasks are given, after that joining of that line to each other, then the learning environment draws the shapes like rectangle, oval, arrow, triangle, etc. Then filling the colors to shape.

Algorithm steps:

The algorithm typically involves the following steps:

1. Store the alphabets separately (Sindhi and English) with voice.
2. After completing the alphabet, the 1 digit will be count for learning level with subject and it moves into the table of Arabic form alphabet with Zer, Zabar, Pesh.
3. If the Arabic form table task has been completed, then send the message to the next level for its execution.
4. The complete word is predefined in the table form, so every single joining of alphabetical characters, the joining word matched to the predefined.
5. If the word joining is right or wrong, though the voice message as (good or join again) then the program will be executed.
6. Each word separately stored int the database with voice for lessons.
7. Stord the real numbers in the database with voice.
8. Four functions created for simple calculation (+, -, *, /).
   8.1 take input as the first number,
   8.2 take input for 2nd number,
   8.3 assign the operation task for the calculation by selecting the mathematical symbol.
   8.4 Executes the results.
9. Stored the images with names of its colors.
10. Store shapes with names.
11. If the shape is executed then activate the color mod for fill the color to shapes.

4. Results and Discussion

Each subject clarifies with different images and discussions that are given. Figure 2 shows the main icons of the application. The icon ABC represents for learning the English subject, and the icon اب پ represents the learning for Sindhi subject. However, the 123 icon indicates the learning for the mathematic subject while the green, red and yellow color icon shows the shapes for the representation of drawing subject and the other icon also illustrated as colors which are for the identification of colors.

Fig 2: GUI of Mobile-Based Learning Approach

There is a very attractive background image illustrated together with this and also shows the balloons that children take more interest in learning. The design has been focused to use Multimedia based Learning principles to ensure that the children are able to find the learning interesting and learn as expected [16].

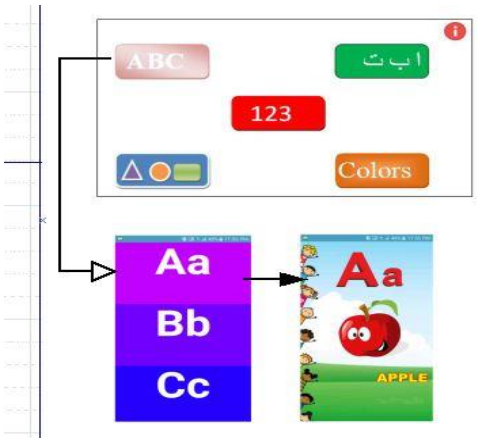

Fig 3: Main activity of learning environment model

Figure 3 represents the main activity of the learning environment where one icon has been selected to start reading and the other icons have been swiped.

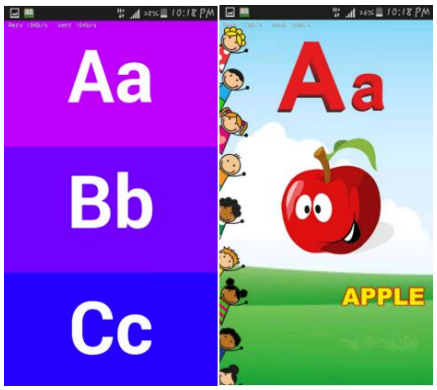

Fig 4: English alphabet with image

Figure 4 shows the English alphabets with dual form capital and small. The selected character indicates the shell of purple color and for the children to understand the image is illustrated with a complete spell as "Apple". However, each character has its own voice as their pronunciation.

Fig 5: English alphabet with image

In Figure 5, there as a sample of another alphabetical character. Where the character "H" is selected with the representation of images for a suitable word as H for Helicopter, while the others are swiped.

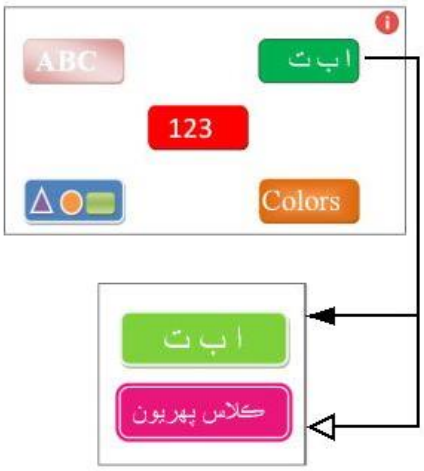

Fig 6: Selection of Sindhi Section

In Figure 6, the Sindhi section for learning has been illustrated. There are five icons mentioned there, but only one icon has been selected for learning and which represents the Sindhi subject. However, others are skipped.

Fig 7: Sindhi section for learning

In Figure 7 the learning environment has moved into Sindhi section which clearly shows the basic characters of Sindhi alphabet اب پ with the title as ڪلاس پهريون (Class One) {Class Pehryon}. There is a very attractive background image also here, which shows the shape of the house with different stylish colors.

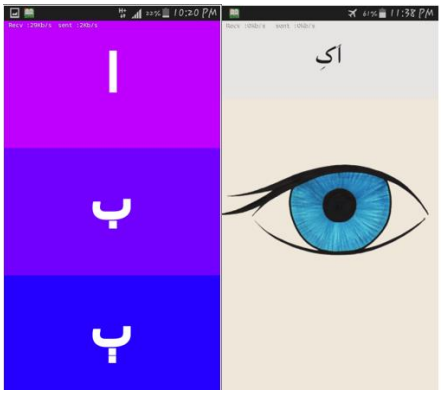

Fig 8: Sindhi alphabet with icons

Figure 8 represents the Sindhi alphabet with appropriate words belonging to that character and also with the shape of that word like اک (Eye){Akh} however, the other characters are not selected for reading, this reading task performed one by one single character.

Fig 9: Sindhi Lesson One

Figure 9 shows the lesson on of a Sindhi textbook named "سنڌي پاراٿو ڪتاب" (Sindhi Childs Book) {Sindhi Barano Kitab}. When the children click the single word, it produces the voice of that word as pronunciation when a user clicks the audio icon the application produces the reading voice of the whole sentence in Sindhi.

Fig 10: Sample of Sindhi Lesson

Figure 10 describes the sample lesson of the Sindhi section which already discussed above. There are three lines as sentences and twelve words are also there mentioned for the reading.

Fig 11: Numbers representation

Figure 11 describes the mathematics portion. There are three numbers in sequence that are illustrated and the number one 1 selected for reading and it also shows in the figure as shape and written the same as 1 and however the voice of every single number by clicked the number.

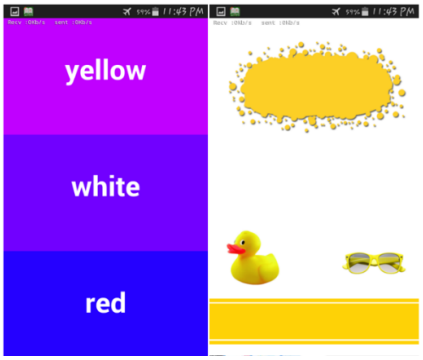

Fig 12: colors identification

Figure 12 shows the color identification for children with voice activities. There are three colors mentioned yellow, white, red. However, the yellow color has been selected and its shape also displayed for the children's color identification learning purpose.

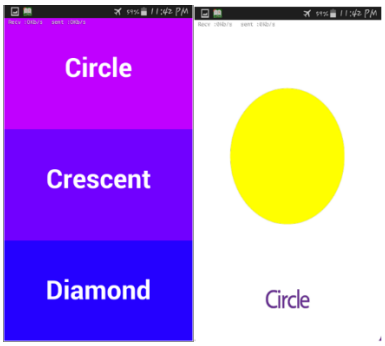

Fig 13: Shape understanding

Figure 13 shows the, shapes understanding for children. There are three shapes displayed as Circle, Crescent, Diamond. While only the circle has been selected and its shape also illustrated together with voice for the understating of shapes.

5. Conclusion

The aim of this work is to develop a child-friendly application with multimedia principles to ensure that children learn while

they are using mobile and reaming interested in using the application. For the digital learning paradigm, four main exercises were developed to facilitate multimedia-based learning for children. Each exercise developed contain vibrenet colors with legible text and voice narrations to ensure that children find the application interesting and keep learning from it.

REFERENCES

[1] Facer, K., Joiner, R., Stanton, D., Reid, J., Hull, R., & Kirk, D. (2004). Savannah: mobile gaming and learning?. Journal of Computer assisted learning, 20(6), 399-409.

[2] Sha, L., Looi, C. K., Chen, W., & Zhang, B. H. (2012). Understanding mobile learning from the perspective of self-regulated learning. Journal of Computer Assisted Learning, 28(4), 366-378.

[3] Nazir, M. J., Rizvi, A. H., & Pujeri, R. V. (2012). Skill development in multimedia based learning environment in higher education: An operational model. International Journal of Information and Communication Technology Research, 2(11).

[4] Neo, M., & Neo, K. T. (2001). Innovative teaching: Using multimedia in a problem-based learning environment. Journal of Educational Technology & Society, 4(4), 19-31.

[5] Rachmadtullah, R., Ms, Z., & Sumantri, M. S. (2018). Development of computer-based interactive multimedia: study on learning in elementary education. Int. J. Eng. Technol, 7(4), 2035-2038.

[6] Skiada, R., Soroniati, E., Gardeli, A., & Zissis, D. (2014). EasyLexia: A mobile application for children with learning difficulties. Procedia Computer Science, 27, 218-228.

[7] Aldalalah, M., Fong, S. F., & Ababneh, W. Z. (2010). Effects of multimedia-based instructional designs for Arabic language learning among pupils of different achievement levels. International Journal of Human and Social Sciences, 5(5), 311-317.

[8] Faryadi, Q. (2006). Bye, Bye Verbal-Only Method of Learning: Welcome Interactive Multimedia. Online Submission.

[9] Singh, V. K. (2003, June). Does Multimedia really improve learning effectiveness. In Asia Pacific Conference on Education, Re-envisioning Education: Innovation and Diversity, Singapore.

[10] Teoh, B. S., & Neo, T. K. (2006, July). Innovative teaching: Using multimedia to engage students in interactive learning in higher education. In 2006 7th International Conference on Information Technology Based Higher Education and Training (pp. 329-337). IEEE.

[11] Eysink, T. H., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches.

[12] Leacock, T. L., & Nesbit, J. C. (2007). A framework for evaluating the quality of multimedia learning resources. Journal of Educational Technology & Society, 10(2), 44-59.

[13] Neo, M., & Neo, T. K. (2001). Innovative Teaching: Integrating Multimedia into The Classroom In A Problem-Based Learning (Pbl) Environment. Multimedia University, Malaysia.

[14] Bhatti, Z., Abro, A., Gillal, A. R., & Karbasi, M. (2018). Be-Educated: Multimedia Learning through 3D Animation. arXiv preprint arXiv:1802.06852.

[15] Bhatti, Z., Mahesar, A. W., Bhutto, G. A., & Chandio, F. H. (2017). Enhancing Cognitive Theory of Multimedia Leaning through 3D Animation. Sukkur IBA Journal of Computing and Mathematical Sciences, 1(2), 25-30.

[16] Bhutto, G. A., Bhatti, Z., Rehman, S. U., & Joyo, S. (2018). Multimedia based learning paradigm for School going children using 3D Animation. University of Sindh Journal of Information and Communication Technology, 2(4), 202-207.

[17] Hassan, Z., Bhatti, Z., & Dahri, K. (2019). A Conceptual Framework Development of the Social Media Learning for Undergraduate Students of University of Sindh. University of Sindh Journal of Information and Communication Technology, 3(4), 178-184.

Vol. 4, No. 1 | January - June 2020

Awareness regarding Coronavirus pandemic among the population of Sindh, Pakistan: A cross-sectional study

Khalida Shaikh[1], Areeba Memon[2], Irfan Ahmed Memon[3], Zulfiqar Ali Laghari[4] Abdul Moiz memon[5]

Abstract:

Methods: This is a cross-sectional descriptive study and was carried out in Sindh, Pakistan. 588 people from different cities of Sindh were asked to fill a questionnaire that tested their attitudes, knowledge, and practices related to the SARS-CoV-2 pandemic. Results: Television and other social media platforms have led to increased dissemination of knowledge among the population. Their increased understanding has led them to be more compliant with adopting precautionary measures, as frequent handwashing was seen in 89.8% and social distancing in 87.2%. The results also showed a greater prevalence of myths among the population. Conclusion: Increased access to correct knowledge can help to dissipate misconceptions and help spread accurate knowledge about the role that the public should play in reducing disease transmission.

Keywords: *SARS-CoV-2; Pandemic; COVID-19; Awareness; Severe acute respiratory syndrome coronavirus-2*.

1. Introduction

The appearance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) triggered a severe global pandemic and is a major public health issue [1]. It came in realization in December 2019 in Wuhan, China, when a group of patients was verified to have novel coronavirus infection. Politically and geographically Pakistan and China are closely connected. Several Chinese people are working on multiple developmental projects in Pakistan, Pakistan took strong measures such as closing its border with China, scrutinizing flights from China and arranging for its workforce and students in affected cities to remain there so that potential spread could be avoided. But the situation got out of control when the virus spread in Iran, many Pakistanis are involved in Iran for work and regularly visit the country for religious purposes. For many reasons, a controlled exit from Iran was not possible also Iranians asked other countries to call back

[1]Department of Physiology, Liaquat University of Medical & Health Sciences, Pakistan.
[2] Aga Khan University Hospital, Karachi, Pakistan.
[3] Department of City and Regional Planning, Mehran University of Engineering & Technology, Jamshoro, Pakistan.
[4] Department of Physiology Sindh University, Jamshoro, Pakistan
[5]Mechanical Engineering Department, Mehran University of Engineering & Technology, Jamshoro, Pakistan
Corresponding Author: khalidashaikh1@gmail.com

their nationals. All these factors led to an evident spread of this outbreak into Pakistan affecting many people. According to COVID 19 visualizer, the number of total cases reported globally as of 10th July 2020 was 12,434,830 with 558,416 deaths out of which 243,599 were reported in Pakistan with 5,058 deceased and 149,092 recovered. The number of cases keeps increasing daily causing a huge problem.

According to current evidence, the COVID-19 virus is primarily transmitted between people through respiratory droplets and contact routes [2]. The mean COVID-19 incubation time was a little different in recent publications. Wang et al, reported 5 days, 7 days, and 8 days for median durations from first signs to dyspnea, hospital admission, and acute respiratory distress syndrome (ARDS) respectively in 138 cases [3]. The mean incubation period of this infection was estimated to be 4.6 days and 95% of disease onset happened within ten days [4]. The symptoms occurring in the disease are non-specific and range from being asymptomatic to the development of severe pneumonia and even death with fever, cough, myalgia, and fatigue being the most common ones [5-7].

As of now, there's no approved antiviral treatment or vaccine present for this disease.[8] Thus, proper management of such patients and strict practice of precautions by the general public is critical in saving the lives of people affected and stopping the transmission of the disease to those unaffected by it respectively. The government along with global public health bodies is trying hard to increase the awareness regarding the disease, it's the transmission and precautionary methods amongst the general public to stop the transmission of this communicable disease [9].

Public health and infection prevention initiatives are desperately required to reduce the damage associated with COVID-19 and minimize the global dissemination of the virus.[10] But, Due to limited data being available on how to increase public awareness and what methods are effective there is a lack of evidence to devise appropriate strategies to increase the public's awareness. Due to inappropriate usage of technology, false information is being circulated around and as a result, people with limited knowledge are adapting practices that might prove dangerous. This research aims at providing an idea of how aware the people in Pakistan about safe precautionary practices are and help the policymakers in making an effective plan towards spreading the right knowledge regarding this global pandemic.

2. Materials and Methods

This is a cross-sectional, descriptive study carried out in Sindh, Pakistan. The study lasted three months (i.e.) January to April 2020. 588 participants were included in the study. All the chosen subjects had a secondary or higher level of education and were aged ≥ 20 years.

For data collection, a detailed Questionnaire was prepared, and the pilot tested amongst 10 participants to determine the clarity and acceptability of the questions. These responses were not included in the final study. The final questionnaire was distributed amongst the selected group of volunteers. The volunteers were each assigned a city and they approached every participant individually to ensure standardized filling of the questionnaire. Written consent was also obtained from all participants.

The questionnaire was designed in English and included some well-known myths that were taken from WHO's site and restructured to see if people respond with true or false for each statement. Further, it inquired about precautionary measures being taken during this pandemic and their city of residence. Each participant's satisfaction level regarding the steps being taken by the government to handle the current health crisis and their stress level related to the lockdown being practiced was also included in the questionnaire. The

questions from the Perceived Stress Scale (PSS) were asked to determine whether the participant was stressed or not.

All the responses collected were then divided into three categories based on the city of residence and its population. If the city of residence had a population of ≥ 0.7 million it was marked as category 1(large cities). If the city population was <0.7 and ≥ 0.3 million, it was marked as belonging to category 2(medium cities). And a population of <0.3 million belonged to category 3(smaller towns).

The data collected, was then transferred to an Excel sheet and further coded and analyzed using IBM SPSS version 23.0. Chi-Square test was used to compare categorical groups and check whether a difference exists between categories. The data was presented using Frequencies and Proportions via Tables and Bar graphs.

3. Results

The results showed that people in all the cities got informed about the latest knowledge on COVID via multiple sources. It shows that

nose, and mouth (49.2%) along with disinfecting shared surfaces such as doorknobs and table-tops, etc. (47.2%). (Figure 2).

Figure 3 shows whether the participants practicing social distancing find it to be stressful or not. Most people in medium

television and social media remain the major source of information overall. In larger cities, 390 people out of 474 (82.27%) had social media as one of their sources of information thus catering the majority with television being the second major source (71.5%). In medium cities, 30 out of 42 people (71.4%) had access to the information being shown on television and 27 out of 30 (64.2%) people got their information via social media. In smaller towns, the responses were quite variable with television being the major source of information (62.5%). Government/NGO volunteers contributed the least in all the cities. Moreover, in smaller cities contribution via websites/blogs were also quite low. *(Figure 1)*

The most frequently observed precautionary measure amongst the study population was regular handwashing (89.8%). Moreover, many participants (82.7%) practiced social distancing. In addition to that, most participants reported avoiding shaking hands (78.7%) and taking precautionary measures in public (69%). Participants also avoided touching, eyes,

(64.3%) and smaller cities (58.3%) didn't find social distancing to be stressful. Although, the number of those finding social distancing to be stressful is significant enough and cannot be ignored. In larger cities, however, 243 out of 474 people (51.2%) found social distancing to be stressful.

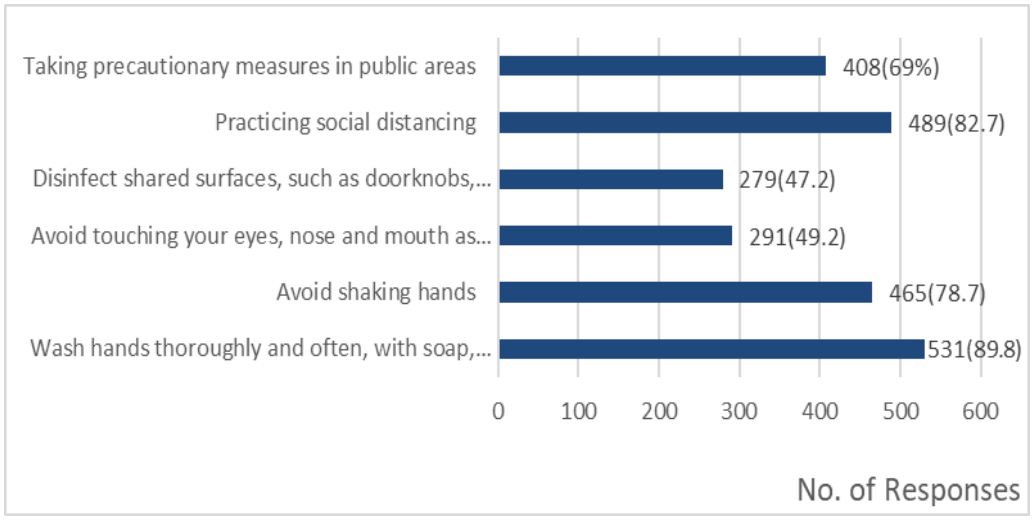

Fig 1**.** source of information in small, medium and large cities

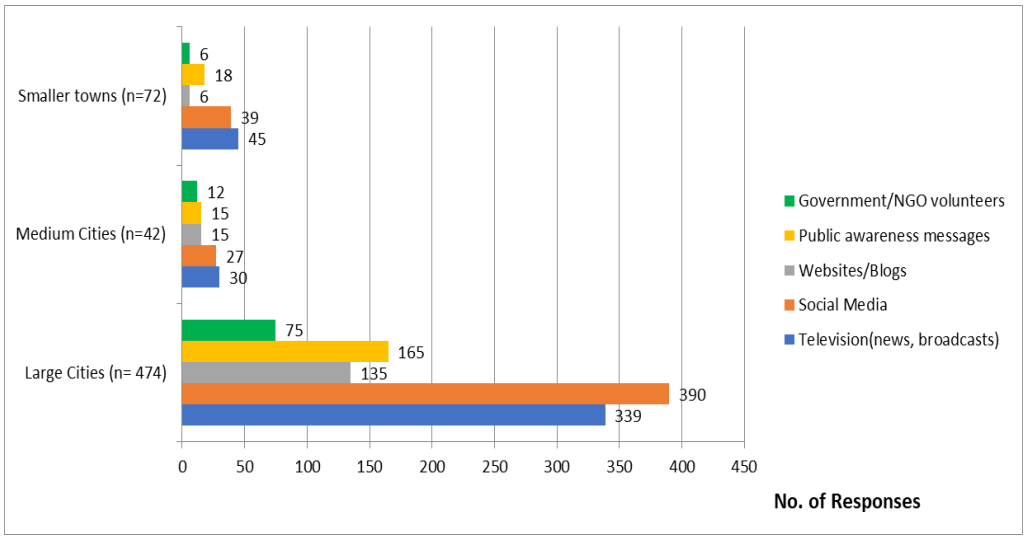

Fig 2. Percentage of (%) of precautionary measures being taken against SARS-COV-2

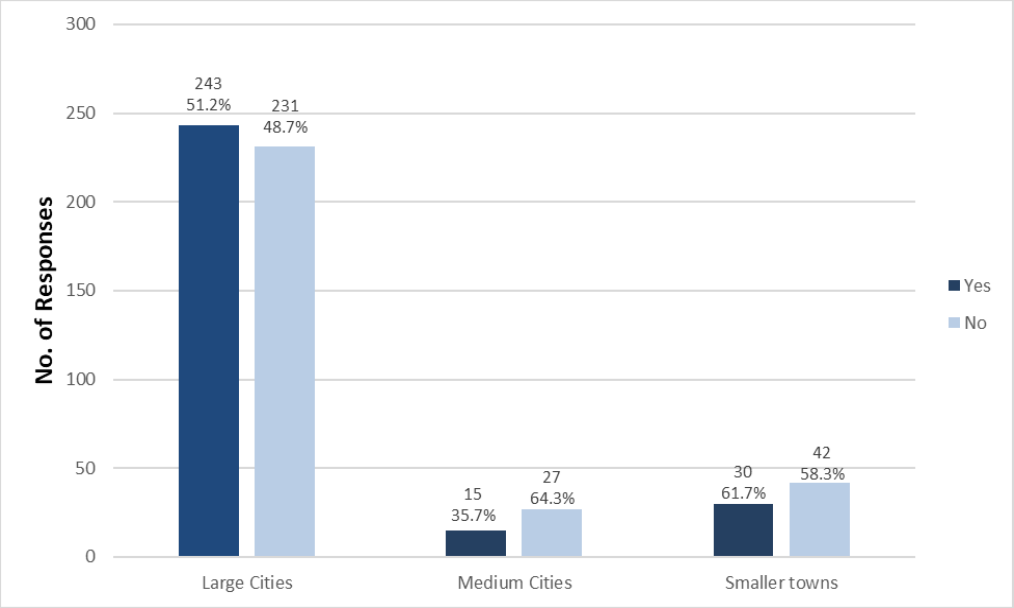

Fig 3**.** Bar-Graph showing whether people find social distancing stressful (Yes) or not (No)

Table 1 shows many people believe in myths related to COVID 19, which can result in wrong precautions and prevention procedures. Spraying alcohol or chlorine all over your body will not kill viruses was the most common myth among the study population as the largest number of people responded to this question wrongly, followed by thermal scanners that can detect infected people.

The high Chi-square values for two of the myths i.e. "The new coronavirus CANNOT be transmitted through mosquito bites" (X^2=12.688, P<0.01) and "Thermal scanners CAN detect people who are infected" (X^2=6.672, P<0.05) suggest that there is a significant association between them and division of cities that are statistically significant. For other myths, the situation is more serious as these exist indifferent to any category and are existent in a larger number of people.

4. Discussion

This study revealed that the ongoing pandemic has increased the stress levels amongst the population. There were several studies conducted in the past, assessing the awareness, knowledge, and practices about such infectious disease outbreaks that took place in the past [11]. But the literature search has not found any such study regarding this recent outbreak of SARS-CoV-2. Therefore, the study can provide data to the public health policymakers necessary to devise effective prevention strategies. This research showed that the majority of people are aware of the coronavirus situation.

The participants got their basic necessary information via television and social media, as they remain the top two sources for every category. This was expected as in a developing country like Pakistan, where television and smartphones are a cheap source of entertainment and thus a common source of

acquiring knowledge and information. The pandemic situation and subsequent increase in the number of deaths have resulted in a limited number of people volunteering in government/NGO awareness campaigns and thus their contribution remains the lowest amongst each category.

Table 1. Knowledge about SARS-COV-2 amongst participants of small, medium and large cities

Myths	*Category 1* Large Cities n(%)	*Category 2* Medium Cities n(%)	*Category 3* Smaller Towns n (%)	X^2	P-value
COVID-19 virus can ONLY be transmitted in areas with cold and dry climates. (True) (False)	108 (22.8) 366 (77.2)	9 (21.4) 33(78.6)	21(29.2) 51(70.8)	0.507	0.776
The new coronavirus CANNOT be transmitted through mosquito bites. (False) (True)	63 (13.3) 411 (86.7)	21 (50) 21 (50)	12 (16.7) 60 (83.3)	12.688	0.002
Hand dryers ARE effective in killing the new coronavirus. (True) (False)	99 (20.9) 375 (79.1)	9 (21.4) 33(78.6)	15 (20.8) 57 (79.2)	0.002	0.999
Ultraviolet disinfection lamp CAN kill the new coronavirus. (True) (False)	147 (31) 327 (69)	12 (28.6) 30 (71.4)	21 (29.2) 51 (70.8)	0.063	0.969
Thermal scanners CAN detect infected people. (True) (False)	232 (48.94) 242 (50.6)	21 (50) 21 (50)	42 (58.3) 30 (41.7)	6.672	0.045
Taking a hot bath does NOT prevent the new coronavirus disease. (False) (True)	189 (39.9) 285 (60.1)	18 (42.9) 24 (57.1)	33 (45.8) 39 (54.2)	1.756	0.416
Spraying alcohol or chlorine all over your body will NOT kill viruses. (False) (True)	243 (51.3) 231 (48.7)	27 (64.3) 15 (35.7)	36(50) 36 (50)	0.919	0.632
Vaccines against pneumonia, such as pneumococcal vaccine and Haemophilus influenza type B (Hib) vaccine, do NOT protect against the new coronavirus. (False) (True)	84 (17.7) 390 (82.3)	9 (21.4) 33(78.6)	6 (8.3) 66 (91.7)	1.539	0.463
Eating garlic CAN help prevent infection with the new coronavirus. (True) (False)	108 (22.8) 366 (77.2)	6 (14.3) 36 (85.7)	27 (37.5) 45 (62.7)	3.252	0.197
New coronavirus infects ONLY older people or people with preexisting medical conditions. (True) (False)	57 (12) 417 (88)	9 (21.4) 33(78.6)	12(16.7) 60 (83.3)	1.263	0.532
Antibiotics CAN be used as a means of prevention or treatment. (True) (False)	111 (23.4) 363 (76.6)	15 (35.7) 27 (64.3)	24 (33.3) 48 (66.7)	1.904	0.386

Social distancing and frequent handwashing are the most emphasized behavioral interventions by health professionals to reduce the risk of transmission. [12] The results of the population seem consistent with this as handwashing (89.8%) and social distancing (82.7%) are practiced the most. A significant number of people practicing social distancing find it stressful. The proportion of those in larger cities was greater than those in medium and smaller cities. Mostly, daily wage workers work in major cities and are most affected by lockdown, which has the primary goal of social distancing, hence for them social distancing is stressful and a difficult pill to swallow [13, 14].

Many people find it to have negative effects on their health. While technology has helped people communicate with their friends and families, for the majority, the main part of their everyday experience is nonverbal contact, and being unable to do so triggers stress in many including those residing in medium and smaller cities. Most women and children in smaller cities are usually home-bound and that may have contributed to lower figures than those in larger cities.

An important finding in this study is the high prevalence of myths amongst participants. We tested the knowledge of participants regarding the most commonly encountered misconceptions that are being circulated on different media platforms. These myths are addressed by WHO and all the information is available on their website. Despite that, at least 20% of participants in each category believe that this virus can only be transmitted in cold and dry climate although there is no indication that COVID-19 cases will decline as the weather gets warmer[15] or humid [16]. In middle cities, people were equally split as to whether mosquitoes act as a vector for this viral disease or not. Studies in the past have shown that it spreads from human to human transmission and that mosquitoes have no role in its transmission [5, 17]. Moreover, ultraviolet disinfection lamps and hand dryers do not play any part in killing this new coronavirus (SARS-CoV-2) and the ultraviolet light can irritate the skin and is harmful [18]. A vast number of people believe that thermal scanners are used to identify people infected with the new coronavirus. The role of thermal scanners, however, is restricted to the diagnosis of the people with fever, which is one of the symptoms of this disease [19]. However other people can have the disease and yet go undetected as it takes 2 to 10 days before individuals get ill and develop a fever [20]. Alcohol and chlorine can be used to disinfect surfaces but they have no role in killing the viruses that have already entered the body and these chemicals can prove harmful to the body [21]. Supportive treatment is currently the only treatment option available for patients affected with SARS-CoV-2 virus and although several attempts to develop a vaccine have been made, the effectiveness of pneumococcal vaccines in protecting people against this disease has not yet been proven in any research [22, 23]. In addition to that, the study showed that many people believe antibiotics to be a cure for this disease. Antibiotics are to be used against bacterial infections and they have no role in preventing viral infections. Antibiotics are to be used only when appropriate and not as a treatment of viral infections as this can exacerbate the antibiotic resistance issue [24]. The prevalence of myths in this society shows that people also get the wrong information via some of these platforms and that can be risky because people continue to follow such beliefs and some of them can be detrimental and may cause new issues in these challenging times.

Therefore, this study demonstrates that awareness programs are actually effective in spreading knowledge, but attention must be paid to the attitudes of the local population and their perceptions should also be taken into account when planning such awareness campaigns.

5. Conclusion

The knowledge regarding this pandemic isn't adequate as the misconceptions continue to exist among a significant percentage of the population. The data gathered in this study could be used to track public opinion and attitudes to develop new health policies and awareness-raising strategies. Increased access to correct knowledge can help to dissipate misconceptions and help spread accurate knowledge about the role that the public should play in reducing disease transmission

REFERENCES

[1] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. J. I. j. o. a. a. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges," p. 105924, 2020.

[2] J. Liu *et al.*, "Community Transmission of Severe Acute Respiratory Syndrome Coronavirus 2, Shenzhen, China, 2020," (in eng), *Emerg Infect Dis*, vol. 26, no. 6, Jun 17, 2020.

[3] D. Wang *et al.*, "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," *JAMA*, vol. 323, no. 11, pp. 1061-1069, 2020.

[4] C. A. Donnelly *et al.*, "Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong," vol. 361, no. 9371, pp. 1761-1766, 2003.

[5] J. F.-W. Chan *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," vol. 395, no. 10223, pp. 514-523, 2020.

[6] N. Chen *et al.*, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study," vol. 395, no. 10223, pp. 507-513, 2020.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," vol. 395, no. 10223, pp. 497-506, 2020.

[8] L. I. D. J. T. L. I. d. The "Challenges of coronavirus disease 2019," vol. 20, no. 3, p. 261, 2020.

[9] A. Zumla, D. S. Hui, E. I. Azhar, Z. A. Memish, and M. J. T. L. Maeurer, "Reducing mortality from 2019-nCoV: host-directed therapies should be an option," vol. 395, no. 10224, pp. e35-e36, 2020.

[10] F. Song *et al.*, "Emerging 2019 novel coronavirus (2019-nCoV) pneumonia," p. 200274, 2020.

[11] J. Shi *et al.*, "Knowledge, attitudes, and practices of nonpharmaceutical interventions following school dismissals during the 2009 influenza A H1N1 pandemic in Michigan, United States," vol. 9, no. 4, 2014.

[12] T. Wise, T. D. Zbozinek, G. Michelini, and C. C. Hagan, "Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States," 2020.

[13] U. W. Chohan, "Forecasting the Economic Impact of Coronavirus on Developing Countries: Case of Pakistan," 2020.

[14] B. Krishnakumar, S. J. J. o. M. Rana, Immunology, and Infection, "COVID 19 in INDIA: Strategies to combat from combination threat of life and livelihood," 2020.

[15] Y. Zhu and J. J. S. o. T. T. E. Xie, "Association between ambient temperature and COVID-19 infection in 122 cities from China," p. 138201, 2020.

[16] P. Shi *et al.*, "The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak-evidence from China," 2020.

[17] Q. Li *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," 2020.

[18] W. H. Organization. *Coronavirus disease (COVID-19) advice for the public: Myth busters*. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

[19] G. Winter, "COVID-19 and emergency planning," vol. 25, no. 4, pp. 184-186, 2020.

[20] S. A. Lauer *et al.*, "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application," 2020.

[21] A. Shankar *et al.*, "Cancer Care Delivery Challenges Amidst Coronavirus Disease–19 (COVID-19) Outbreak: Specific Precautions for Cancer Patients and Cancer Care Providers to Prevent Spread," vol. 21, no. 3, pp. 569-573, 2020.

[22] R. Ralph *et al.*, "2019-nCoV (Wuhan virus), a novel Coronavirus: human-to-human transmission, travel-related cases, and vaccine readiness," vol. 14, no. 01, pp. 3-17, 2020.

[23] F. Yu, L. Du, D. M. Ojcius, C. Pan, S. J. M. Jiang, and infection, "Measures for diagnosing and treating infections by a novel coronavirus responsible for a pneumonia outbreak originating in Wuhan, China," 2020.

[24] W. Dodds, "Disease Now and Potential Future Pandemics," in *The World's Worst Problems*: Springer, 2019, pp. 31-44.

Vol. 4, No. 1 | January - June 2020

Average Index Modelling of Campus Safety and Walkability: Case Study of the University of Sindh

Irfan Ahmed Memon[1] , Saima Kalwar[1], Noman Sahito[1], Sabeen Qureshi[1], Noman Memon[1]

Abstract:

A pedestrian-friendly area encourages the inhabitants to experience the place enjoyably on foot. A pedestrian's sense of comfort is defined by the environmental qualities that encourage walking. The city's environmental qualities also make them better for walking, encouraging both physical and social activities. This Study focuses on safety and walkability in the University of Sindh. According to a survey done by researchers at the University of Sindh, the important issue at the university is lacking pedestrian facilities. Therefore, this study covers major facets of walkability. It is envisioned for policymakers and planners who want to improve the walking spaces for their communities with the best current information on pedestrian facilities. This research identified basic details about the specific walking plans and project creation. It includes trail preparation, roads, street lighting, street repairs, path and track repair, public protection, personal security, etc. This research explores the improvement of university walkways to accommodate pedestrians through the Average Index Model. Besides, the study presents recommendations on the issue of pedestrian safety. Also, discusses how the University of Sindh's road network aspects and addresses how campus roadway system could include sustainable transportation choices for inhabitants.

Keywords: *Pedestrian safety; pedestrian walkability; average index model*

1. Introduction

Universities around the world are committed to providing their students and staff with conducive living and learning environments, so campus manipulators' mobility is a problem that is faced by many major universities and needs to tackle as part of their environmental efforts on campuses. Campus walkability is a significant accessibility element of on-campus mobility. Travelers need access to a network of routes that are connected directly and easy to follow, connecting the hostel, open fields, faculties, public transit stops, and other amenities. Thus, it enhances their campus experience, focused on health, safety, accessibility, enjoyment, and learning [1, 2].

Walking is also correlated with other benefits, varying from air quality control, road congestion, and health problems such as reducing obesity. It is also encouraging neighborly interactions and a more comfortable and healthy way to live in an urban environment. There are enough justifications on the association between the built environment and walking [3, 4]. The key aim when measuring walkability on a university campus is to promote safe living, reduce congestion, and sustainable alternative

[1]Mehran University of Engineering & Technology, Jamshoro, Sindh, Pakistan
Corresponding Author: irfan.ahmed@faculty.muet.edu.pk

of mode choice. Walkability is the main component of sustainable alternative mode choice and offers societal advantages and contributions to public wellbeing, environmental prosperity, and preserving the climate. A pedestrian-friendly environment can decrease vehicle use and encourage walking [1].

The University of Sindh, the oldest and the largest university in Pakistan. The total number of students is almost 35000 in 2020. The total number of teachers is 800 and 65 Departments. There is insufficient space for walking due to that accidents are common inside the campus. Erstwhile, Sindh University student died after being hit by campus transport [5, 6]. Meanwhile, there is unattractive roads, unsafe, and uncomfortable environment. On the other hand, people use to walk on the roads rather than pedestrians, it is because of a lack of awareness. Therefore, there is a need for the pedestrian in the University of Sindh.

Therefore, this study identified the existing condition of the campus for pedestrian facilities and to evaluating a walkability aspect i.e. its characteristics and potential.

2. Literature Review

2.1. Walkability

Walkability applies to the urban area that is comfortable with a variety of activities [7]. Walkability is the degree of accessibility and protection for pedestrians such as; lack of casual observations, gaps between pedestrians and automobiles as well as high-quality linked footpaths [8].

The walkability is the degree to which walking is readily accessible as a secure, linked, affordable and fine form of transportation. Walking can be a first-class mode of travel if the surroundings provide the quality of walking conditions, safety, remedy, and convenience [9].

2.2. Connectivity and Accessibility

Connectivity and accessibility are important courtesy to pedestrians [10]. Connectivity is about the easy movement from origins to destinations and it is the key element affecting the route assignment of pedestrians [11].

To encourage human beings to wander across the area, road networks need to be better linked with appropriate sidewalks to a specific holiday location. A well-planned road network has streets that are well connected to other mode choices. The streets also offer a variety of facilities and spaces, a correct fine of paths, sufficient walking distance, clear signs. The drawing of the road provides artistic pastime for the urban community. Both of these considerations are important for ease of walking [12].

Accessibility is an essential component of walkability. A community is nice because everyone can access public spaces. A regular citizen can be impaired if connectivity is not accessible. Simultaneously, either a user with a wheelchair, a blind person, or an elderly is not affected in an on-site situation. Street with a frontage of parking makes pedestrians difficult to get inside in any building or structures it forces pedestrians to get involved in the parking lot [3, 4].

2.3 Safety of Walking

Safety is one of the main factors for walkability. Safe pedestrian environments enable pedestrians to walk safely and reduces incident or crime concern experiences. In 2005, Southworth believed that the exceptional atmosphere of taking walks could affect pedestrian intensity [12]. Pedestrian complexity is needed to improve protection as the places expand to be extra translucent. The thing regarding pedestrian security and safety is correlated with vehicle activity and crossing visibility and protection. Good street network virtually defines better walkability for the pedestrian to create more comfortable and safer circumstance.

3. Methodology

This research investigated key issues through various approaches and tools. The survey participants posed closed-ended questions concerning the following topics:

• walkability indicators (Sidewalk width, Sidewalk Maintenance, Streetscape, Shading Devices, etc.)

• Walking experience from their origin to the faculties based on a scale as the need for the indicators 5 (need) to 1 (no need) indifferent aspects: safety, security, comfort, convenience, and interestingness.

Conclusively, two major data collection methods were used to obtain knowledge from the identified field of research. These are discussed as follows:

3.1. Primary Source

In this regard, the researcher selected the Sindh University Jamshoro main campus as a case study. From the primary source, the data was collected through various techniques i.e. quantitative method detailed survey, questionnaire, personnel observation. Moreover, the quantitative method was used to get the detailed information regarding pedestrian problems faced by people of the study area.

3.1.1. Questionnaire

The requirement for analysis work has generated a need for an effective method of sample sizing to reflect a defined demographic condition [13]. A new approach called "small sample technique" published a formula for exploring a sample size by the research division of the National Education Association [13]. By this approach, it shows the connection between sample size and total population. It should be recalled that as the population grows the sample size will decline and stay constant.

A sample size of 384 was chosen in this survey out of the 33523 population. The questionnaire was completely close-ended and filled from respondents in the study area, i. e Sindh University Jamshoro. The survey questionnaire consisted of two separate sections. The first section focused on personal information from the respondents. The second part was about research data which is to be collected from the respondents about pedestrian facilities and their interest in the improvement of the pedestrian in the university. A self-administered questionnaire was filled from students from different departments of the university This primary data set helped the researcher to achieve the objectives of the study.

3.1.2. Traffic and Pedestrian Count Survey

Traffic count is a count of vehicular or pedestrian traffic. It is conducted along a particular path or intersection. In this survey, the researcher surveyed pedestrians and vehicular traffic from the three different locations of the University of Sindh Jamshoro. As there were different main points for traffic count that were selected such as; Main gate, Central library (Zero-point), Hostel road. The number of vehicles was collected on manual sheets from 7:30 am to 8:30 am and peak hours from 1:30 pm to 2:30 pm. Traffic, as well as pedestrian count, were done from all 3 points.

4. Average Index Model

The study applied the Average Index model developed by Abd Majid and McCaffer which provides means to verify the validity of a criterion by the respondents [14, 15]. State the expectations of the students and the powerful physical-environmental factors. Data analysis was carried out through content analysis to assess the students' efficacy of cycling and walking physical-environmental variables and the person analysis on encouraging students to do walking and cycling.

Program processing tools were used for data analysis and IBM SPSS was supported as help applications for data interpretation and performance categorization. Levels of the significance of physical-environmental variables for motivating students to bike and

walk were tabled dependent on the number of responses for every physical-environmental attribute.

The average index was calculated based on the frequency analyses to determine the ranking of each factor to be considered. The average index modeling is computed to determine the following:

$$\text{Average Index} = (\sum aixi)/(\sum xi) \quad (1)$$

Where, a = constant, weighing factor for i, (i = 1, 2, 3.....n), Xi = frequency of respondent. The classifications of the rating scales are as follows to assess the degree of significance of the constructability concepts considered in this analysis:

Extremely effective	$3.50 \leq I \leq 4.00$ *or* $87.5 \leq I \leq 100$
Very effective	$2.50 \leq I \leq 3.50$ *or* $62.5 \leq I < 87.5$
Moderately effective	$1.50 \leq I \leq 2.50$ or $37.5 \leq I < 62.5$
Ineffective	$0.50 \leq I \leq 1.50$ or $12.5 \leq I < 37.5$
Extremely ineffective	$0.00 \leq I \leq 0.50$ or $0.00 \leq I < 12.5$

5. Results and Discussion

As researchers have analyzed the data by using appropriate techniques according to the nature of the study to get the most accurate and reliable results. So, researchers have used different techniques including traffic count surveys, for measuring the volume of current traffic as well as pedestrians. Descriptive analysis, for analyzing the data in SPSS frequency/percentage and creating graphs in excel while the average index model was also used as for analyzing the needs of different elements in terms of safety and walkability in the University of Sindh.

The traffic count survey is done in the different timing in a day for taking the average of the passing vehicles and pedestrians. In this survey researchers have selected three different points in the university of Sindh from these points vehicles and pedestrians were counted in peak hours from 7:30 am to 8:30 am and 1:30 pm to 2:30 pm for 3 days. Table 1 describes the traffic count survey on three said road locations at the university. The total number of vehicles from the main gate to different points in university from 7:30 AM to 8:30 AM per hour is 1852.

Table 1. Peak hour traffic count from7:30 AM to 8:30 AM

Date	**Points**	**Time AM**	**No. Vehicles**
3/9/2019	CL (Zero Point)	7:30 to 8: 30	504
4/9/2019	Main Gate	7:30 to 8: 30	1121
5/9/2019	Hostel Road	7:30 to 8: 30	227
Total			1852

Table 2 shows the traffic count survey on the above-highlighted points of the campus. The total number of vehicles from the main gate to different points in university from 1:30 PM to 2: 30 PM per hour is 1212.

Table 2. Peak hour traffic count from 1:30 PM to 2:30 PM

Date	**Points**	**Time PM**	**No. Vehicles**
3/9/2019	CL (Zero Point)	1:30 to 2:30	302
4/9/2019	Main Gate	1:30 to 2:30	720
5/9/2019	Hostel Road	1:30 to 2:30	190
Total			1212

Table 3 express the pedestrian count survey at a different point in the University of Sindh Jamshoro from 7:30 AM to 8:30 AM, so the number of pedestrians at the main gate University of Sindh was larger than the other selected points.

Table 3. Peak hour pedestrian count from 7:30 AM to 8:30 AM

Date	**Points**	**Time AM**	**No. Vehicles**
18/9/2019	CL (Zero Point)	7:30 to 8: 30	523
19/9/2019	Main Gate	7:30 to 8: 30	776
20/9/2019	Hostel Road	7:30 to 8: 30	223
Total			1522

Table 4 represents the pedestrian count survey at a different point in the campus at afternoon peak hours from 1:30 PM to 2:30 PM, so the number of pedestrians at the central library the University of Sindh was greater than the other selected points.

Table 4. Peak hour traffic count from 1:30 PM to 2:30 PM

Date	**Points**	**Time PM**	**No. Vehicles**
18/9/2019	CL (Zero Point)	1:30 to 2:30	720
19/9/2019	Main Gate	1:30 to 2:30	320
20/9/2019	Hostel Road	1:30 to 2:30	89
Total			1111

The researcher has conducted a descriptive analysis of 384 responses by percentage and frequency, the result shows in the data tables and different graphs for understanding the overall scenario of responses for the improvement of pedestrians in the University of Sindh Jamshoro. The number of respondents was 384, so n=384.

The average duration of walking per day, as 50.5% people walk one hour per day, 42.7% walk 2-3 hours per day,22% people walk 4-5 hours per day, and 4% people walk 6 or more than 6 hours per day. The result shows that most people walk one hour per day as shown in Figure 1.

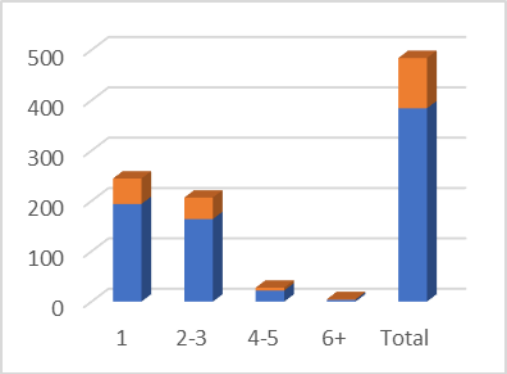

Fig. 1. Duration of walking per day

According to survey it was identified that different sort of barriers was faced by students while walking, according to 14.6 % students too far distance is a barrier in walking, 40.9% roadside condition is the main barrier,15.4% students harsh weather condition is the main barrier in walking and according to 29.2%, unavailability of the footpath is the barrier. Most of the students consider roadside walk conditions as the main barrier in walking as discussed in Figure 2.

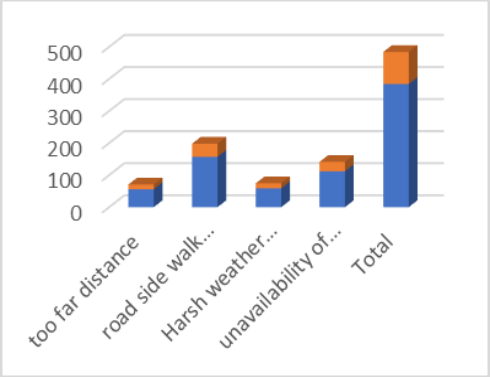

Fig. 2. Duration of walking per day

Likert scale 2.50 average total points mean there is satisfaction condition. If the value increases more than 2.50 it means, there is a need for that factor /there is a problem. If the

A.T.P (Average Total points) decrease less than 2.50 it means, there is no need for that factor/there is no problem.

Simultaneously, Average Point Index analysis in Table 5 shows the overall average satisfaction of university student about the factors which are mentioned below.

Furthermore, the findings of Average Index Analysis concluded that the most needed elements which scored as, Personal safety (fencing between roads and footpath) 4.09, Shade 4.25, Garbage cans/ Recycling 4.11and remaining elements. According to the survey, the majority walk around an hour per day. From the Analysis Students always walk with their friends on the pedestrian.

Table 5. Walkability elements average index modeling through Likert scale

Likert scale	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**TP/N**	**ATP**
Points	5	4	3	2	1		
Picnic tables/Benches	122	150	54	30	28	1460//384	3.80
Personal safety (fencing between roads and footpath)	132	133	41	47	31	1573/384	4.09
Shade	114	165	51	31	23	1633/384	4.25
Garbage cans/ Recycling	130	130	63	33	29	1582/384	4.11
Ramps and curbs	122	160	67	20	15	1506/384	3.92
Sign and road markers	119	135	52	43	35	1412/384	3.67
Gardens/flower beds	110	98	76	60	40	1330/384	3.46
Green space/ Nature / Scenery	125	111	55	58	35	1385/384	3.60
Access to washroom	136	132	46	43	27	1459/384	3.799
Playgrounds	90	74	131	40	49	1268/384	3.30
Side walks	145	115	70	33	21	1482/384	3.85

The survey result shows that roadside walking condition is the main barrier for cyclists and for pedestrian, absence of sidewalks in many areas of the University of Sindh forces the pedestrian to walk on the road which results in a pedestrian crash. Consequently, the direct observation existing sidewalks are uneven, unpaved and some are covered with tall grasses due to poor maintenance. Like the survey, the result shows that people depend on university buses as compare to pedestrians and people don't feel safe at peak hours. From the analysis, it has been evaluated that most of the residents live inside the campus i.e hostels. It was also found that major factor for road accident is poor enforcement by law and over speeding.

According to the survey sufficient number of signboards, speed breakers, road marking are not available, this item of aesthetic on

roadways which help to direct the flow and direction of traffic need to be obeyed by road users, they could be in the form of solid white lines used to indicate parking, pedestrian, bicycle lanes and other features.

A survey result shows that pedestrian safety is the responsibility of university administration, so the university should take measures to enhance the walkways and footpaths, so people prefer walking instead of using an automobile.

6. Conclusion

The research here indicates the pedestrian's safety and walkability in the University of Sindh Jamshoro. The purpose of the research is to full fill the needs of the pedestrians and proposing the pedestrian in the university of Sindh Jamshoro. The data were collected from all students and staff of the university. Different techniques were used to analyze the data including the average index model, in the result that shows, there is a need for walkways, shades, ramps and curbs, garbage cans. A traffic count survey was done for knowing the volume of traffic as well as pedestrians at different points. Most of the pedestrians were facing problems while walking as they had to use roads for walking due to lack of walkways beside the roads on the campus. Descriptive analysis was also done, as a result, pedestrians were facing barriers while walking including, maintenance of walkways, unavailability of footpaths, too far distance, lack of shades on walkways, etc.

Policy and investments provide momentum to renovate public universities, encourage pedestrianization and allow people to enjoy better mobility and quality of life, so there must be valid and authentic policies for the improvement of the pedestrian at the University of Sindh Jamshoro.

The Administration of the University of Sindh Jamshoro must be involved in the development/improvement of the pedestrian under planning rules and pedestrian standards. Pedestrians' walkability and safety can be achieved by developing pedestrian services including sidewalks, shades, landscaping, Signboards, road marking, Ramps and curbs, Garbage cans, etc.

Traffic Rules must be imposed on vehicles that are violating the laws. Sindh university must possess multi/short, and long walkways between different departments for the encouragement of people's interest towards pedestrians.

Develop crosswalk marking in the university for providing easiness for pedestrians to cross the roads inside the university. Decrease crosswalk distance. In the case of pedestrian safety, there must be the installation of pedestrian lights and fencing on footpaths.

REFERENCES

[1] L. K. Keat, N. M. Yaacob, and N. R. Hashim, "Campus walkability in Malaysian public universities: A case study of Universiti Malaya," *Planning Malaysia*, vol. 14, no. 5, 2016.

[2] I. A. Memon, "Factors influencing travel behavior and mode choice among Universiti Teknologi Malaysia employees," Universiti Teknologi Malaysia, Faculty of Built Environment, 2010.

[3] S. Handy, L. Weston, and P. L. Mokhtarian, "Driving by choice or necessity?," *Transportation Research part A: Policy and Practice*, vol. 39, no. 2–3, pp. 183-203, 2005, DOI: http://dx.doi.org/10.1016/j.tra.2004.09.002.

[4] S. Handy, X. Cao, and P. L. Mokhtarian, "Self-selection in the relationship between the built environment and walking: Empirical evidence from Northern California," *Journal of the American Planning Association*, vol. 72, no. 1, pp. 55-74, 2006.

[5] D. News. "Accredited universities, institutions' list issued." Dawn News. (accessed 11th July 2020).

[6] Z. Ali. "Student dies after being hit by Sindh University bus on campus." The Express Tribune. https://tribune.com.pk/story/1328909/student-dies-hit-sindh-university-bus-campus (accessed 15th July 2020).

[7] H. Sung and S. Lee, "Residential built environment and walking activity: Empirical

evidence of Jane Jacobs' urban vitality," *Transportation Research Part D: Transport and Environment,* vol. 41, pp. 318-329, 2015.

[8] S. Shamsuddin, N. R. A. Hassan, and S. F. I. Bilyamin, "Walkable environment in increasing the liveability of a city," *Procedia-Social and Behavioral Sciences,* vol. 50, pp. 167-178, 2012.

[9] T. Litman, "Online TDM Encyclopedia," in *Online TDM Encyclopedia*, ed. Victoria, BC, Canada: Victoria Transport Policy Institute, 2014.

[10] J. Rotmeyer, "Can elevated pedestrian walkways be sustainable," *e SustainableCityIV: UrbanRegenerationandSustainability,* pp. 293-302, 2006.

[11] S. T. D. E. M. Kumar, "Review Paper On Study of Pedestrian Flow on Grade Separated Crossing," 2014.

[12] M. Southworth, "Designing the walkable city," *Journal of urban planning and development,* vol. 131, no. 4, pp. 246-257, 2005.

[13] V. R. Krejcie and W. D. Morgan, "Determining sample size for research activities," *Education and Psychological Measurement,* vol. 30, pp. 607-610, 1970.

[14] M. Z. Abd Majid, Y. Bigah, A. Keyvanfar, A. Shafaghat, J. Mirza, and H. Kamyab, "Green Highway Development Features to Control Stormwater Runoff Pollution," *Journal of Environmental Treatment Techniques,* vol. 4, no. 4, pp. 173-175, 2015.

[15] M. A. Majid and R. McCaffer, "Assessment of work performance of maintenance contractors in Saudi Arabia. Discussion," *Journal of Management in Engineering,* vol. 13, no. 5, 1997.

Vol. 4, No. 1 | January - June 2020

Identifying the Machine Learning Techniques for Classification of Target Datasets

Abdul Ahad Abro[1], Mohammed Abebe Yimer[2], Zeeshan Bhatti[3]

Abstract:

Given the dynamic and convoluted nature of numerous datasets, the necessity of enhancing performance outcomes and handling multiple datasets has become more challenging. To handle these issues effectively and improve the quality of multiple approaches, the capabilities of various Machine Learning techniques such as K-Nearest Neighbor (KNN), Logistic Regression (LR), Naive Bayes(NB) and Support Vector Machine (SVM) have been utilized in this study. In this paper, the binary classification method using five different datasets, and many predictor variables have been utilized. Moreover, this research has mainly focused on determining the classification of data into the subsets that share the standard designs. In this regard, many approaches had been studied extensively and used to achieve better yields from the existing literature; however, they were inadequate to provide efficient outcomes. By applying four Supervised ML classification algorithms along with the UCI Datasets of ML Repository, the robustness of the method is progressed. The proposed mechanism is assessed by adopting five performance criteria concerning the accuracy, AUC (Area Under Curve), precision, recall, and F-measure values. The current study experimental results revealed that there is a significant improvement in the confusion matrix rate compared with a similar study and this method can also be used for machine learning problems such as binary classification.

Keywords: *Machine Learning, Data Mining, K-Nearest Neighbor, Logistic Regression, Naive Bayes, Support Vector Machine*

1. Introduction

The combination of classifiers is now an active research area in the ML and Pattern Recognition [1][2][3]. Many theoretical and empirical studies have been published which show the advantages of the combination paradigm over the individual classifier models [4][5]. A significant number of researches have been conducted to design multiple classifier systems based on the same classifier models trained on different data or feature subsets. ML has been widely used in a variety of industries, such as Remote Sensing, Image Classification, and Pattern Recognition.

ML can learn and improve automatically from experience, without explicit programming. It is the primary aim to automate learning without human intervention. ML algorithms use statistics to find patterns in massive amounts of data [6]. Whereas the algorithms which are used in this research are briefly described below:

[1]Department of Computer Engineering, Ege University, Turkey.
[2]Department of Computer Engineering, Dokuz Elyul University, Turkey.
[3]Department of Information Technology, University of Sindh, Pakistan
Corresponding Author: zeeshan.bhatti@usindh.edu.pk

Firstly, KNN: a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. It is used for statistical estimation and pattern recognition[7][8].

Secondly, LR: a standard statistical approach that is ideal for performing regression analysis where the dependent variable is binary. It is used to describe the data and to explain the relationship between one dependent binary variable with one or more independent nominal, ordinal, interval, or ratio-level variables [9].

Thirdly, the NB classifier: combines the Bayes paradigm with the decision rules like the hypothesis, which provides satisfactory results. It applies Bayes theorem, with the naive assumption of conditional independence between each pair of features given the value of the class variable. In [10], proposed the NB learning framework for large-scale computational efficiency and multi-domain platform classification.

Fourthly, SVM: is a paradigm that uses classification algorithms for two-group problems. It is accuracy and predictive performance on the survival of traumatic brain injuries performed significantly better than LR [11].

On the other hand, this paper has structured with several sections. In section 2, previous related work is described briefly. The methodology adopted for performing different experiments is explained in Section 3. Section 4, provides experimental work, datasets detail, evaluation of experiments is performed to obtain different results. Lastly, certain conclusions are drawn based on the outcomes and future work is suggested in Section 5.

2. Related Work

Classifications based on KNN, LR, NB, and SVM has recently witnessed a surge of research efforts. In this paper, we have used the classification of supervised learning. Moreover, in the literature, classification algorithms could be affected significantly or negatively by some features [1][2]. The goal of classification is to accurately forecast the target class for each case in the data. Whereas in the model build training process, a classification algorithm co-ordinates between the values of the predictors and the values of the target. Different classification algorithms execute different procedures for discovery associations. These associations are model, which can function to a different dataset in which the class is unidentified [12] [13] [14].

In [15], KNN is the slowest classification technique because the classification time is directly proportional to the number of data. When the data size is more prominent, more extensive distance calculation should be performed to make it extremely slow. Moreover, it uses the number of nearest neighbors "k" as one of the parameters in classifying an object, and the value of k influences the classifier performance [16].

In [17], Cubic SVM, Quadratic SVM, and Linear SVM have better performances in predicting the outcome of traumatic brain injury as compared to LR.

In [18], NB is the most popular data mining algorithms. Empirical results indicate that the selective NB demonstrates superior classification performance while retaining the simplicity and flexibility at the same time.

SVM is a useful method for solving classification and regression problems. In [19], the SVM approach can substantially improve prediction accuracy and would help to mitigate the adverse impact on urban expansion.

3. Methodology

This section presents an overview of the proposed method, which describes the pre-processing stage of data and classification algorithms.

3.1. Overview of the Proposed System

An overview of the proposed system is given in Fig. 1. This system consists of numerous phases: datasets, base learners, and comparative analysis of the results. Besides, the generalization performance of the system, 10-fold cross-validation is used for all classifier learners and datasets.

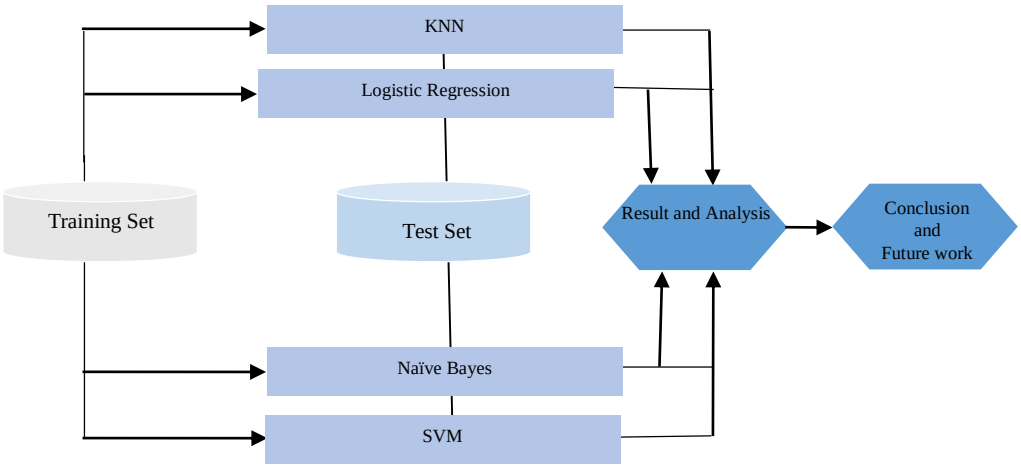

Fig.1. The framework of the method.

3.2. Data Preprocessing

In this phase, the ranges of the values of the data from different ML datasets may be high. In such a scenario, certain features can significantly or negatively affect algorithms for classification accuracy. Therefore, the data values are normalized to [0,1] range using min-max normalization technique [20].

3.3. Classification of Algorithms

In this study, four base learners, including KNN, logistic regression, NB, and SVM, are employed.

There are numerous phases of methods related to the datasets and classifiers focused on ML. In this work, four ML classifiers, along with several datasets, are experienced for binary classification.

LR classifier relies on feature extraction. Typically, it delivers more authentic results than KNN, NB, and SVM. The primary aim of this analysis is to establish the classification accuracy and performance evaluation of multiple datasets.

The KNN classifier does not have a specialized training phase and uses all the data for training during classification and it does not assume anything about the underlying data [15].

LR classifier is another method borrowed by ML from the field of statistics. It is a statistical model and used when the dependent variable is categorical.

NB is a probabilistic ML model. It requires linear parameters in the number of functions of the variables and highly scalable [18].

SVM is an ML algorithm that can be used for classification problems as well as for regression. It is segregated in two classes and co-ordinates the individual observation.

4. Experimental Design

In these subsections, we describe and present the experimental process, evaluation measures, and experimental results.

4.1. Experimental Process

In the experimental process, five datasets have been used from the UCI ML Repository [21]. All experiments are performed on a total of 4 ML classifiers by using WEKA (Waikato Environment for Knowledge Analysis) ML toolkit and JAVA programming language

[22]. On the one hand, we have utilized default parameter values for all the classifiers in WEKA.

On the other hand, we have carried out 10-fold cross-validation to all datasets to yield reliable results. The 10-fold cross-validation is imposed on the original dataset randomly partitioned into 10 equally sized sets, one of which is used as test validation, while the remaining sets are used for training operations. The process is repeated 10 times and calculates the averages of the results.

Dataset characteristics are evaluated concerning the attributes and the number of instances. These datasets are typically used to solve ML related issues. There are various numerical attribute descriptions illustrated in Table 1. The number of instances, attributes, and classes for each dataset are presented in Table 1. The datasets are selected from the UCI ML Repository according to their distinct parameters. It is determined by investigating the appropriate data or datasets which are being utilized for binary classification problems.

Table 1. Characteristics of the five Datasets Used in This Study

Datasets	Instances	Attributes	Classes
Annealing	898	39	6
Breast Cancer	286	10	2
Hepatitis	155	20	2
Vertebral	240	7	2
Yeast	1484	9	10

In this work, four different ML approaches have been carried out along with the five datasets, which are considered suitable for the classification. However, the performance metrics are calculated according to the binary classification problems based on the confusion matrix.

4.2. Assessment of Measures

This section describes the five performance evaluation measures of the proposed method, consisting of accuracy, AUC, precision, recall and F-measure.

Accuracy reflects how close an agreed number is to a measurement. It is specified further in Eq.1.

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

In equation 1, TN, FN, FP and TP show the number of True Negatives, False Negatives, False Positives and True Positives.

AUC represents the area under the ROC Curve. AUC calculates the whole two-dimensional area beneath the whole ROC curve from (0,0) to (1,1).

Precision is a positive analytical value [1][23]. Precision defines how reliable measurements are, although they are farther from the accepted value.

The equation of precision is shown in Eq.2.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The recall is the hit rate [1][23]. The recall is the reverse of precision; it calculates false negatives against true positives. The equation is illustrated in Eq. 3.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F-measure can be defined as the weighted average [1][24] of precision and recall. This rating considers both false positives and false negatives. The equation is illustrated in Eq. 4.

$$F - measure = \frac{2}{1/precision + 1/recall} \quad (4)$$

These criteria are adjusted proportionally in the data by the reference class prevalence in the weighting operation.

4.3. Experimental Results

Tables 2-6 for all datasets present accuracy, AUC, precision, recall, and F-measurement weighted values with ML algorithms. In Table 2-6, high Acc, AUC, Precision, Recall, and F-measure are shown in Bold, while the greyed shows insufficient results.

To sum up, Tables 2-6, has been designed in terms of different specifications according to

the multiple datasets relating to the numerous approaches of ML. In Table 2, LR has better outcomes, which provides 99.1091% Acc when comparing with others.

Probably, in Table 3, KNN indicates 72.3776% Acc adequate consequences. Similarly, in Table 4, the NB presents 84.5161% Acc effective results. Whereas, in Table 5, the SVM illustrates the 92.9167% Acc productive outcomes. However, in the end, LR shows a 58.6253% Acc result in Table 4.

The annealing, hepatitis, and vertebral datasets have significant outputs concerning the accuracy, AUC, precision, recall, and F-measure parameters in Table 2, 4, and 5; however, breast cancer has somehow satisfactory output in Table 3 and yeast shows lower outcomes in Table 6.

Furthermore, it is analyzed that LR for annealing dataset in Table 2, provides a more accurate outcome. Likewise, KNN in breast cancer dataset concerning Table 3, indicates adequate consequences and in Table 4, NB presents effective results in the Hepatitis dataset. In addition, in Table 5, Vertebral dataset SVM provides positive findings. Finally, LR indicates the progressive result in Table 6, yeast dataset.

Table 2: Weighted Values for Annealing Dataset

Annealing					
Methods	**Acc (%)**	**AUC**	**Precision**	**Recall**	**F-Measure**
KNN	99.1090	0.985	**0.991**	**0.991**	**0.991**
LR	**99.1091**	**0.992**	0.991	0.991	0.991
NB	86.3029	0.957	0.933	0.863	0.882
SVM	39.3096	0.646	0.703	0.393	0.433

Table 3: Weighted Values for Breast Cancer Dataset

Breast Cancer					
Methods	**Acc (%)**	**AUC**	**Precision**	**Recall**	**F-Measure**
KNN	**72.3776**	0.628	0.699	**0.724**	0.697
LR	68.8811	0.646	0.668	0.689	0.675
NB	71.6783	**0.701**	**0.704**	0.717	**0.708**
SVM	66.0839	0.596	0.662	0.661	0.661

Table 4: Weighted Values for Hepatitis Dataset

Hepatitis					
Methods	**Acc (%)**	**AUC**	**Precision**	**Recall**	**F-Measure**
KNN	80.6452	0.653	0.794	0.806	0.799
LR	82.5806	0.802	0.814	0.826	0.818
NB	**84.5161**	**0.860**	**0.853**	**0.845**	**0.848**
SVM	79.3548	0.731	0.814	0.794	0.802

Table 5: Weighted Values for Vertebral Dataset

Vertebral					
Methods	**Acc (%)**	**AUC**	**Precision**	**Recall**	**F-Measure**
KNN	85.4167	0.660	0.852	0.854	0.853
LR	92.5	**0.930**	0.919	0.925	0.920
NB	77.9167	0.854	0.886	0.779	0.812
SVM	**92.9167**	0.788	**0.924**	**0.929**	**0.925**

Table 6: Weighted Values for Yeast Dataset

Yeast					
Methods	**Acc (%)**	**AUC**	**Precision**	**Recall**	**F-Measure**
KNN	52.2911	0.685	0.524	0.523	0.522
LR	**58.6253**	**0.825**	**0.585**	**0.586**	0.577
NB	57.6146	0.816	0.585	0.576	0.566
SVM	58.2884	0.785	0.489	0.583	**0.602**

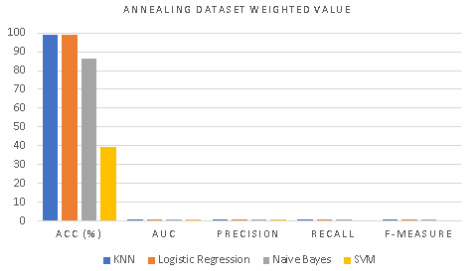

Fig.2. The chart is showing the effects of the Annealing dataset

In Fig. 2-6, indicates the enhanced classification and performance evaluation based on the datasets provided in the following mentioned charts. The LR, Annealing dataset has higher accuracy followed by KNN, NB, and SVM, in Fig. 2. Moreover, in Fig. 3, KNN, Breast Cancer

dataset, provides better outcomes after LR, NB, and SVM. Likewise, in Fig. 4, NB efficiency, Hepatitis dataset, yields efficient outputs as compared to LR, KNN, and SVM sequentially. Whereas, SVM, vertebral dataset, has higher accuracy in contrast to LR, KNN, and NB in Fig.5. Lastly, in Fig. 6, the LR, Yeast dataset, has outperformed than SVM, NB, and KNN.

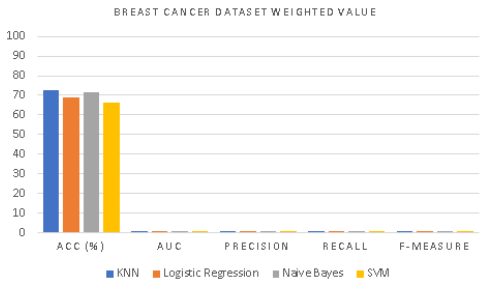

Fig.3. The chart is showing the effects of the Breast Cancer dataset.

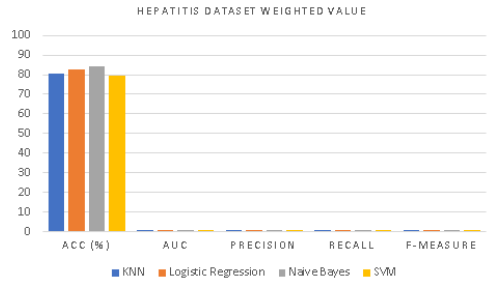

Fig.4. The chart is showing the effects of the Hepatitis dataset.

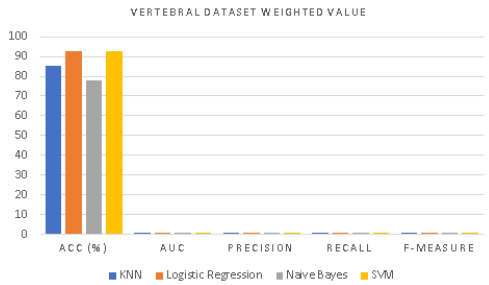

Fig.5. The chart is showing the effects of the Vertebral dataset.

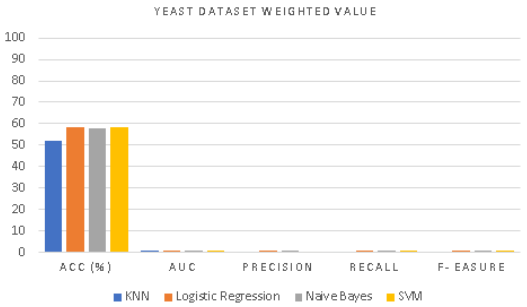

Fig.6. The chart is showing the effects of the Yeast dataset.

5. Conclusions And Future Work

Based on the experimental and numerical results, the main findings of this research work can be summarized as follows:

In this paper, we have examined the implementation of four ML algorithms which are named as k-nearest neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machine (SVM) to classify multiple datasets. The efficiency of algorithms is further demonstrated in terms of precision, recall/sensitivity, accuracy, and F - score. Whereas many ML algorithms are unable to provide satisfactory results as they are dependent on the datasets. The sensitivity of the same algorithm can be severely affected by analyzed varying sizes of training and test sets.

Generally, LR has more successive consequences than KNN; whereas, in most datasets, the NB delivers more effective outputs than SVM. There is no winner outright in terms of the performance outcomes; it depends on the characteristics of the datasets, the simulation, and the circumstances.

In the future, we plan to reform our study of classification models by introducing the Intelligent ML algorithms, which are more useful to an extensive collection of real-life datasets.

REFERENCES

[1] A. A. ABRO, E. TAŞCI, and A. UGUR, "A Stacking-based Ensemble Learning Method for Outlier Detection," Balk. J. Electr. Comput. Eng., vol. 8, no. 2, pp. 181–185, 2020.

[2] A. A. Abro, M. Alci, and F. Hassan, "Theoretical Approach of Predictive Analytics on Big Data with Scope of Machine Learning."

[3] A. A. Aburomman, M. Bin, and I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," vol. 38, pp. 360–372, 2016.

[4] S. S. Tabrizi and N. Cavus, "A hybrid KNN-SVM model for Iranian license plate recognition," Procedia - Procedia Comput. Sci., vol. 102, no. August, pp. 588–594, 2016.

[5] E. Ayyad, "ScienceDirect Procedia Procedia Computer Procedia Procedia Computer Science Computer Computer Science Science Science Performance evaluation of intrusion detection based on machine Performance evaluation of using intrusion detection on machine learning Ap," Procedia Comput. Sci., vol. 127, pp. 1–6, 2018.

[6] H. Mohammad and D. Science, "Performance Evaluation of Machine Learning Algorithms in Ecological Dataset OF MACHINE LEARNING ALGORITHMS," no. March, 2019.

[7] J. Xiao, "SVM and KNN ensemble learning for traffic incident detection," Phys. A Stat. Mech. its Appl., vol. 517, pp. 29–35, 2019.

[8] R. Kalyan, D. Mishra, and A. Kumar, "A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices," Appl. Soft Comput. J., vol. 35, pp. 670–680, 2015.

[9] S. S. Panesar, R. N. D. Souza, F. Yeh, and J. C. Fernandez-miranda, "Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database," World Neurosurg. X, vol. 2, p. 100012, 2019.

[10] F. Xu, Z. Pan, and R. Xia, "E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework," Inf. Process. Manag., no. February, p. 102221, 2020.

[11] J. Feng, Y. Wang, J. Peng, M. Sun, J. Zeng, and H. Jiang, "Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries," J. Crit. Care, vol. 54, pp. 110–116, 2019.

[12] C. Technology and J. Vol, "Design And Develop Network Storage Virtualization By Using GNS3," vol. 13, pp. 20–24.

[13] S. Sabahi and M. Mellat, "International Journal of Production Economics The impact of entrepreneurship orientation on project performance : A machine learning approach," Int. J. Prod. Econ., no. April 2019, p. 107621, 2020.

[14] I. Kayijuka, "computational efficiency of singular and oscillatory integrals with algebraic singularities," no. November, 2018.

[15] A. Ashari, "Performance Comparison between Naïve Bayes , Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," vol. 4, no. 11, pp. 33–39, 2013.

[16] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of Roman-Urdu opinions using Naı ̈ ve Bayesian, Decision Tree and KNN classification techniques," J. King Saud Univ. - Comput. Inf. Sci., vol. 28, no. 3, pp. 330–344, 2016.

[17] J. zhou Feng, Y. Wang, J. Peng, M. wei Sun, J. Zeng, and H. Jiang, "Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries," J. Crit. Care, vol. 54, pp. 110–116, 2019.

[18] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," Knowledge-Based Syst., vol. 192, p. 105361, 2020.

[19] F. Karimi, S. Sultana, A. S. Babakan, and S. Suthaharan, "Computers , Environment and Urban Systems An enhanced support vector machine model for urban expansion prediction," Comput. Environ. Urban Syst., vol. 75, no. January, pp. 61–75, 2019.

[20] T. Classification and B. K. Singh, "Investigations on Impact of Feature Normalization Techniques on Investigations on Impact of Feature Normalization Techniques on Classifier ' s Performance in Breast Tumor Classification," no. April 2015, pp. 10–15, 2017.

[21] UCI Machine Learning Repository, 2018, https://archive.ics.uci.edu/ml/index.php.

[22] T. A. Engel, A. S. Charão, M. Kirsch-Pinheiro and L. A. Steffenel, "Performance improvement of data mining in weka through

GPU acceleration," Procedia Comput. Sci., vol. 32, pp. 93–100, 2014..

[23] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, 2006.

[24] L. A. Bull, K. Worden, R. Fuentes, G. Manson, E. J. Cross, and N. Dervilis, "Outlier ensembles: A robust method for damage detection and unsupervised feature extraction from high-dimensional data," J. Sound Vib., vol. 453, pp. 126–150, 2019.

0.7” Margin Top

Paper Formatting Guidelines

Title: 14pts Bold

Author Name: 11pts Bold

Affiliation: 11pts Italic

Abstract: Single Paragraph (Min: 150 – Max: 250 words)

Paper Length: Formatted as guided (Min: 4,000 – Max: 8,000 words)

Font: Times New Roman 10pts

Font Color: Black

Line Spacing: 1.0 (single line space throughout the paper)

Indentation: Justify

References & In-text Citations: Follow the IEEE & Use EndNote X7 for In-text citations and Bibliography.

Headings: 12pts Bold

1. **12pts Bold**
   1.1. **11pts Bold**
      1.1.1. ***11pts Bold Italic***

Margin Left **0.8”**

Margin Right **0.5”**

6.5”x10”

Page Size: 6.5” (Width) x 10” (Height)

Submission: Formatted paper as guided can be submitted through our online submission system at http://sjcms.iba-suk.edu.pk

1.9” Margin Bottom

Sukkur IBA Journal of Computing and Mathematical Sciences

SUKKUR IBA UNIVERSITY
Merit - Quality - Excellence

SUKKUR IBA UNIVERSITY
Airport Road, Sukkur -65200
Sindh, Pakistan
Tel: +92-71-5644233
Fax: +9271-5804419
Email: sjcms@iba-suk.edu.pk
URL: sjcms.iba-suk.edu.pk

HEC Recognized Journal