iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** I **Month of publication:** January 2022

DOI: https://doi.org/10.22214/ijraset.2022.39903

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

Opinion Mining Analytics for Spotting Omicron Fear-Stimuli Using REPTree Classifier and Natural Language Processing

Taiwo Olapeju Olaleye[1], Segun Michael Akintunde[2], Chidi Akparanta[3], Tijani Agidibo Avovome[4], Ojugbele Francis Oluyen[45], Ayodele Olubunmi Akparanta[6]
[1, 3, 6]Computer Centre & Services, Federal College of Education, Abeokuta, Nigeria
[2]Computer Science Department, Federal University of Agriculture, Nigeria
[4]Medical Centre, Federal College of Education, Abeokuta, Nigeria
[5]Management Information System, Federal College of Education, Abeokuta, Nigeria

Abstract: Data has indisputably proven overtime to have a better idea and with the surge of big data in the era of coronavirus, research initiatives in the field of data mining continues to leverage computational methodologies. Owing to the dreadful nature of the Omicron-variant, a fight or flight dilemma readily pervades college communities far reaching implications on work ethics of academic front-liners. This study therefore aim to gain insights from academia-sourced data to unravel fear-stimulus in college communities. The predictive analytics is carried out on college-based opinion poll. The Valence Aware Dictionary for Sentiment Reasoning algorithm is deployed for emotion analytics while the Latent Dirichlet Allocation (LDA) and Latent Semantic Indexing (LSI) are employed for topic modelling. The REPTree algorithm models the fear-spotting decision tree using 10-fold cross-validation. Experimental results shows a high performance metrics of 94.68% on Recall and Precision as the hand-washing attribute is returned as the most significant variable with highest information gain. Results of topic modelling likewise returns non-clinical precautionary measures as fear stimulus while the Vader sentiment analysis shows a 22.47%, 25.8%, and 51.73% positive, negative, and neutral polarity scores respectively, indicative of the academic front-liners' pessimism towards effective safety measure compliance with non-clinical regulations.
Keywords: COVID-19, Omicron, Sentiment Analysis, Topic Modelling.

I. INTRODUCTION

Besides the reality of the dreadful Omicron-variant prevalence, fear is another factor that continued to shape behaviour across world communities [1] as the entire universe grapples with the pandemic. The virus was declared as a Public Health Emergency of International Concern (PHEIC) on 30th January, 2020 and, later on 11th March, 2020 was characterized as a pandemic [2]. The COVID-19 pandemic, which is almost convulsing the entire planet, has rendered public health across nations vulnerable and calls for desperate measures to curtail the ugly trend. Proactive measures such as the unprecedented total lockdown of socio-economic activities to curtail the highly infectious and mobile virus was introduced worldwide running into several months. This particular precautionary measure is believed to have contributed immensely towards taming the tide of the ugly trend. However, mixed feelings of apprehension trails reopening of schools by the Nigerian government after the Yuletide break, due to the prevalence of the Omicron-variant. Moreover, the variant has been described as more dangerous than the previous waves. Actually, a more easily contagious Omicron-variant of the virus has raised concern of authorities and citizens alike owing to the recent upsurge of positive cases daily reported in the media.

Fear of community spread of the variant has motivated government across levels to reemphasize the need to observe non-clinical measures including observing physical distance, avoiding gathering of more than 50 people, observance of the use face masks etc. besides the ongoing clamor for vaccination. Whereas, the Nigerian government stipulated certain policy measures as preventive measures to be adhered to its workers including school administrators, the knowledge of the fact that over 65% of the population are upwardly mobile youths of ages below 35 years [3] and some of whom are undergraduates in Tertiary Educational Institutions (TEIs) is indeed a source of concern. Consequently, teaching and non-teaching staff of TEIs who are mostly the vulnerable adults, are at the risk of contracting the virus; hence an expected level of apprehension across Nigeria's universities and college campuses.

Furthermore, with a total number of tested samples put at 3,933,209 and confirmed cases at 249,154, Nigeria is one of the epic center of the virus in Africa accounting for 25,873 active cases, 220,195 discharged cases and 3,086 deaths as at 12th January, 2022 [4]. These figures and aforementioned factors trigger anxieties in the minds of the academia as they re-emphasizes the need for caution and proactive measures. Owing to the dreadful nature of the Omicron-variant, a fight or flight mental condition could pervade TEIs with far reaching implications on work disposition and since the fear of the virus could be more than the virus itself, the pandemic has loaded a huge psychological stress on people around the world, especially medical and educational front liners who have direct contacts with those tested positive due to the nature of their work [5]. Sense of impending danger amongst college staff, feeling of fear, panic and difficulty controlling worry will likewise be counterproductive especially for older staff with pre-existing respiratory conditions [6]. Thus, there is dire need for a study that identifies causes of COVID-19 virus-scare in TEIs front liners not just to ascertain the true state of their mental health with respect to the dreadful Omicron variant, but to establish the links between data points and nosophobia, towards deducing informed conclusions for policy formulation by operators and regulators. This is the thrust of this study because when fear is prolonged or disproportionate, it could become injurious; hence the need to beam attention on the Nigerian workers with daily proximity to human traffic.

Several methodologies have been adopted in literature for related works however, the adoption of machine learning as proposed in this study presents a more accurate predictive analytics of the fear stimuli, owing to its inherent efficient data studying capabilities [7]. Machine learning, through fuzzy logic, likewise helps in decision making processes and use cases including agricultural [8], medical, academics, etc. The results will help to identify most significant fear-stimuli and other germane associated topics in TEI communities would be unraveled through sentiment analysis and topic modelling methodologies. A more in-depth analysis of the academia-sourced dataset, through the instrumentation of machine learning and natural language processing, would give profound insight into opinions expressed by respondents. To the best of our knowledge, this study is the first deploying a three-throng computational approach to identify topical subjects in opinion texts towards unravelling fear-scare in academia-sourced data. Dataset acquired encapsulates feature attributes such as the demography of staff respondents, COVID-19 awareness level, personal health habits information, personal views about the pandemic, their environmental factors etc. The fear-stimuli predictive modelling is carried out on the Java-enabled Waikato Environment for knowledge Acquisition (WEKA, developed at the University of Waikato) and the natural language processing through Orange data mining toolkit (from the University of Ljubljana). Results are evaluated through machine learning performance metrics and discussed. The rest of the paper is structured such that session II discusses existing related works and literature review while session III unveils the methodology used for the proposed model. Session IV discusses the experimental result of the predictive analytics and session V concludes with recommendations.

II. LITERATURE REVIEW AND RELATED WORKS

Fear is commonplace in times of uncertainties like pandemic and it is an adaptive defense mechanism by animals central to survival and which involves several biological processes of grounding as a response to potentially threatening occurrences [9]. Identifying fear-provoking circumstances (stimulus) in work places in a pandemic will help tame anxieties and stress levels in healthy individuals and reduces the symptoms of those with pre-existing health disorders [9]. Similarly, it is opined that an important factor in understanding a population’s response to any threat whatsoever is the fear it elicits [10] since fear is a significant predictor of behavioral changes and health-securing behaviours [11]. Studies on earlier pandemics show that anxiety, or the lack of it, is a significant impulse of behavior [12] as people with slight anxiety about a viral epidemic are less likely to be involved in precautionary hygienic behaviors like periodic hand-washing, seldom observe physical or social distancing stipulations, and are less likely to take vaccinations if available [12]. Conversely, people with excessive anxiety are prone to socially disruptive behaviours including panic buying and frequent visits to hospitals, as minor symptoms are interpreted as signs of serious infections [13].

In [5], authors’ study focuses on assessing the psychological impacts of fear and anxiety amongst health front liners by conducting a single-center, cross-sectional survey through online questionnaires. Elements of fear, worry and melancholy were measured by the numeric rating scale (NRS) on fear, Hamilton Anxiety Scale (HAMA), and Hamilton Depression Scale (HAMD), accordingly. A total of 2299 appropriate contributors were consulted, which include 2042 medical staff and 257 administrative staff. The study observes that fear, worry and melancholy were significantly different between two groups and as compared to the non-clinical staff, front line medical staff with close contact with infected patients exhibit higher scores on fear scale, HAMA and HAMD, and they were 1.4 times more likely to fearful, twice more likely to be worrisome and depressed through melancholy.

Similar to the work of [5], [14] undertook study on the impact of quarantine and physical distancing on mental health of participants as it seeks to find the nexus between the pandemic and the consequent containment measures whose result shows an association between higher levels of depression and anxiety symptoms amongst the surveyed population.

It is observed that the pandemic and quarantine have a disadvantageous impact on mental health as an increase of psychiatric symptoms and of mental health problems in the general population is expected as a fallout of the measures since most health professionals in isolation units are seldom trained nor supported for their mental health care in Italy. Mental health services worldwide are not prepared to manage the short- and long-term consequences of pandemic.

With a Coronavirus anxiety scale, [15] came up with a brief mental health screener to measure COVID-19 related anxieties premised on the basis that mental health issues of people impacted by the pandemic have not been adequately addressed by relevant authorities. The 5-item scale screener, based on 775 adults with anxiety over the virus demonstrated compact dependability and validity as elevated CAS scores were found to be connected with COVID-19 diagnosis, impairment, alcohol/drug coping, negative religious coping, extreme hopelessness, suicidal ideation etc. by discriminating well between persons with and without dysfunctional anxiety using an optimized cut score of ≥ 9 (90% sensitivity and 85% specificity).

The work of [10] seeks to discover the implications of fear emotion on students' and teachers' technology adoption during COVID-19 pandemic. The study adopts Google Meet as an educational social platform in private higher education institutes and data from the study were analyzed with partial least squares structural equation modeling (PLS-SEM) with machine learning supervised algorithm. The J48 classifier performed better in predicting the dependent variable and the introduction of the fear of COVID-19 was an improvement on existing literature which similarly seeks to understudy adoption of technologies amidst the pandemics. A cross-national longitudinal study to predict fear and perceived health towards COVID-19 was conducted by [10] using factors such as worrying about shortages in supplies, perceived vulnerability to disease (PVD) and sex etc. Result shows case counts does not elicit adaptive responses to environmental threats while [16] investigated the pervasiveness of nosophobia, and readiness of people to seek medical attention amidst the COVID-19 epidemic in Calabar of Cross River State in Nigeria. The study shows that nosophobia is associated with age and healthcare seeking attitude while gender and education seldom play significant role in its pathology. It also discovered that fear varied with respect to the type of diseases hence the need for sensitization of the public. In [17], the author identified exposure to avalanche of news coverage about diseases and its risks with recurrent exposure to people with severe illnesses as factors that trigger nosophobia amongst citizens of India and in [12], authors conceptualized and designed a novel COVID-19 stress scale to efficiently determine the stress severity of the virus by measuring fear of getting infected, fear of contact with contaminated articles or planes, disease-related xenophobic anxiety, socio-economic anxiety of the outbreak (e.g., loss of job), compulsive examination and reassurance-seeking regarding possible pandemic-related threats, and traumatic stress symptoms about the pandemic like hallucination and disturbing thought; hence the design of the Stress Scales to quantity the aforementioned features. The work of [18] investigated the predictors of COVID-19 fear and observed that incremental fear is related to risks around loved ones and anxieties related to their health while regular social media use is another fear threshold with respect to the pandemic.

In [19], a topic modelling approach and sentiment analysis was implemented on information flow on twitter during the coronavirus outbreak using the LDA topic modelling preprocessor identifying the most relevant and accurate subjects related to the virus. The model cofirms the prevalence of fear in negative sentiments and the positive sentiments depicts trust in government establsihments to tame the trend. A noun only approach for topic modelling was the thrust of [20] by comparing three topic models trained on news datasets generated from news corpus, lemmatised scope of the news item, and noun only corpus. the experiemntal result shows that excluding all other words except now improved topic semantic occurrence.

III. RESEARCH METHODOLOGY

The research methodology framework implemented in this study is as presented in this section comprising of REPTree supervised machine learning and opinion mining through sentiment analysis and topic modelling. The data analytics is executed on the content of data captured from the opinions gathered from teaching and non-teaching staff of tertiary educational institutions. The study adopted LSI and LDA methods for the topic modelling and sentiment analysis encapsulated in a 6-phase approach as presented in Fig. 1 which includes data acquisition, data preprocessing, machine learning, Topic modelling, sentiment analysis, and result analysis phases.

A. Data Capturing

A total of 2376 data features acquired through Google form survey tool, constituting 10232 text corpus, is captured for this study, representing opinions of teaching and non-teaching staff of twenty-one (21) TEIs in Nigeria. A total of 15-no feature attributes (survey responses) is acquired per respondent while the last question (*SN 15 on* Table 1) is an open-ended question to capture textual phrases for the topic modelling and sentiment analysis phases described in sections 0 and 0 respectively.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

The remaining 14-no attributes is prepared in an Attribute-Relation File Format (*arff)* for the REPTree machine learning phase described in section C. The description of the data attributes is as presented in Table 1 indicating the data type attributes.

B. Data Preprocessing

This study applies preprocessing tasks to remove irrelevant contents, as proposed in [21] through the following steps:

1) *Transformations:* Including conversion of SN 15 opinion to lower case
2) *Noise Removal:* Elimination of punctuations, white spaces etc.
3) *Tokenization:* Includes tokenization of texts with *Regexp* approach. A uni-gram approach of word-tokenization is implemented on the opinions.
4) *Filtering:* Exclusion of stop words including articles, conjunctions, and prepositions that do not carry enough discriminative content needed for the opinion-mining task.

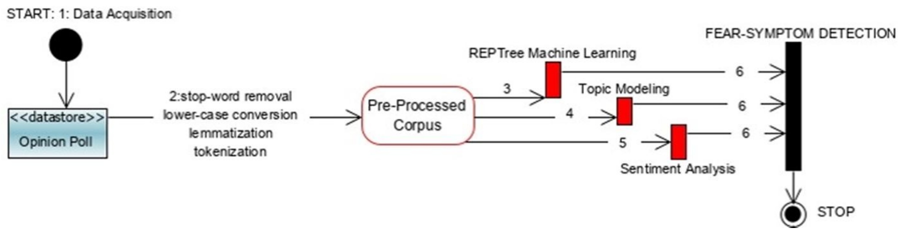


Fig. 1. Proposed study framework

Table 1. Nature of dataset attribute

S/N	Questions/Attribute	Attribute_id	Response options	*Arff* response_id
1.	Staff Category	s_cadre	Teaching; Hostel porter; Non-Teaching; Medical Staff	TS;HP;NT;MS
2.	How often do you use your nose/face mask?	mask_wearing	Always-Often-Sometimes-Rarely-Never	M-A;M-O;M-S;M-R;M-N
3.	What is your salary level?	staff_cat	Below #120,000; #120,000 and above	SS; JS
4.	Have you had COVID-19 test and or vaccination before?	covid_test	Yes; No	Yes; No
5.	How well is COVID-19 precautions being handled in your College?	college_handling	Not at all satisfied; slightly satisfied; moderately satisfied; very satisfied; completely satisfied	CH-1;CH-2;CH-3;CH-4;CH-5
6.	How well are your students complying with COVID-19 safety measures?	std_compl	Not at all compliant; slightly compliant; moderately compliant; very compliant; extremely	CC-1;CC-2;CC-3;CC-4;CC-5

			compliant	
7.	Rate your fear level of contracting the Omicron-variant at your college	fear_rate	1-2-3-4-5	F-1;F-2;F-3;F-4;F-5
8.	Do you know of anyone who tested positive for COVID-19?	casePositive	Yes; No; Prefer not to say	+Yes; -No
9.	Which of the following is correct?	awareness	There is a drug to treat COVID-19; there is a vaccine for COVID-19;there is both a drug for treatment and a vaccine for COVID-19;I am unsure which is correct	Awareness+; Awareness-
10.	How would you rate your Omicron-variant worry-level when resuming to work?	worry_level	1; 2; 3;4; 5	W-1; W-2; W-3; W-4; W-5
11.	In which zone of the country is your college of education located	Zone	North-East; North-West; North-Central; South-West; South-South; South-East	NE; NW; NC; SW; SS; SE
12.	Which of these best describes the community where your college is located?	college_loc	City; town; village/rural area	City; town; village
13.	How often do you wash your hands daily?	hand_wash	Always-often-sometimes-rarely-never	W-A;W-O;W-S;W-R;W-N
14.	Are you working under the fear of contracting COVID-19 from office?	GROUND TRUTH	Yes; No	F-Yes; F-No
15.	What do you think about the reopening of colleges for academic activities	Opinion	Open ended	Not applicable

C. REPTree Predictive Modelling

Machine learning is a branch of artificial intelligence that trains algorithms with data, in either supervised or unsupervised learning approach, and thereby gives capabilities to learn from data without being explicitly programmed and in turn make informed decisions after the learning process in what is referred to as testing the knowledge gain [22]. REPTree is reputable as a quick decision tree learner-algorithm which constructs a decision or regression tree with information gain as the splitting methodology, and prunes it with reduced error pruning method as it results in a more accurate classification tree; size of training and testing notwithstanding. In this work, a 10-fold cross-validation approach as described in Fig. 3 is adopted on the WEKA machine learning software for both the REPTree training and the testing dataset in *60:40* ratio. In the 10-fold cross-validation, the 14:2376 (attribute: instance) sample is randomly apportioned into 10 equal size sub-samples out of which one subsample is reserved as the validation set for testing the model, and the remaining 9 sub-samples are used for training the REPTree algorithm in section Fig. 2. The cross-validation process is iterated in 10 clocks (the folds), with each of the 10 sub-samples used precisely once as the testing data while the 10 results from the folds is then averaged to generate a distinct estimation. This approach ensures all attributes: instances are deployed for both training and validation while each subset is used for validation once each.

D. Topic Modelling of Preprocessed Data

Topic modelling abstract subjects in data corpus based on word clusters alongside their corresponding frequency contained in each text document [23]. It is often deployed for natural language processing (NLP) to uncover topical subjects and thereby extract semantic meaning from unordered text documents in applications such as opinion and text mining and in information retrieval systems. This study aims to deploy topic modelling as part of mechanisms to facilitate in-depth understanding of emotions expressed in the opinions of respondents concerning COVID-19 in an academic environment. The topic modeling widget of the orange data mining tool including the Latent Dirichlet Allocation (LDA), Latent Semantic Indexing (LSI), and Hierarchical Dirichlet Processing (HDP) algorithms are implemented on the attribute 15 text-responses earlier presented on Table 1. LDA is a three-level ranked Bayesian system such that each item of a document is modeled as a fixed mixture over a primary set of topics. LSI however yields topics with negative and positive keywords with negative and positive weights on the topic. The positive weights represent words highly representative of the topic and which influences its occurrence while for negative weights, the topic seldom occur if they appear less in it. Multidimensional scaling (MDS) is used to visualize the modelling, as a low-dimensional projection of the topics represented as points by fitting distances between the points. The proposed topic modelling and sentiment analysis framework is presented in Fig. 4.

E. VADER-based Lexicon Sentiment Analysis

The sentiment analysis uses polarity to classify the opinion expressed by respondents into three categories of *Positive, Neutral & Negative* while its VADER-approach scores each word-token, as contained in each text-opinion. The approach computes the sentiment scores assuming sentiment is related to the presence of certain known-words or phrase *(for bi-gram or more)* in an opinion structurally represented. Therefore, opinions are assigned certain sentiment value referred to as lexicon. In line with the work of [24], the occurrence frequency of each word in the dictionary, as used by the respondent, determines the computations of its positive, negative or its neutral state. This study employ the polarity score calculator of the VADER model. The model yields the fourth attribute, the compound score, for each opinions expressed by correspondents in answering poll question 15. The compound score represents an accumulated description of the first three negativity, positivity, neutrality scores, and is computed for each opinion expressed as:

$$x = \frac{x}{\sqrt{x^2 + \alpha}} \quad (1)$$

where x is the sum of valence scores of component words, and α is default to 15 as the normalization constant. The compound score is then regularized between -1 and +1 which represents the positivity and negativity of each opinion. The VADER-based sentiment analysis framework of this study is as described in Fig. 4.

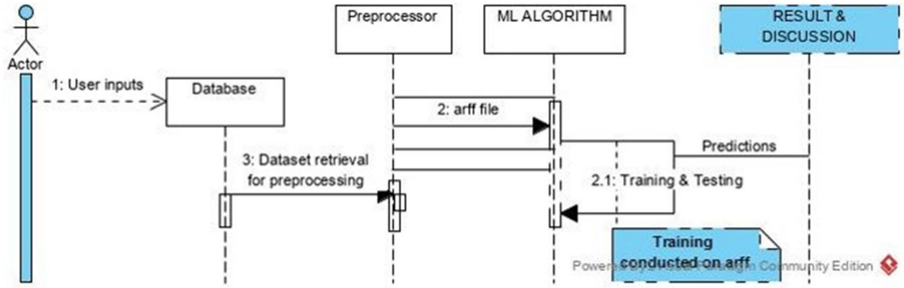


Fig. 2. Description of stages for the REPTree fear-stimuli detection model

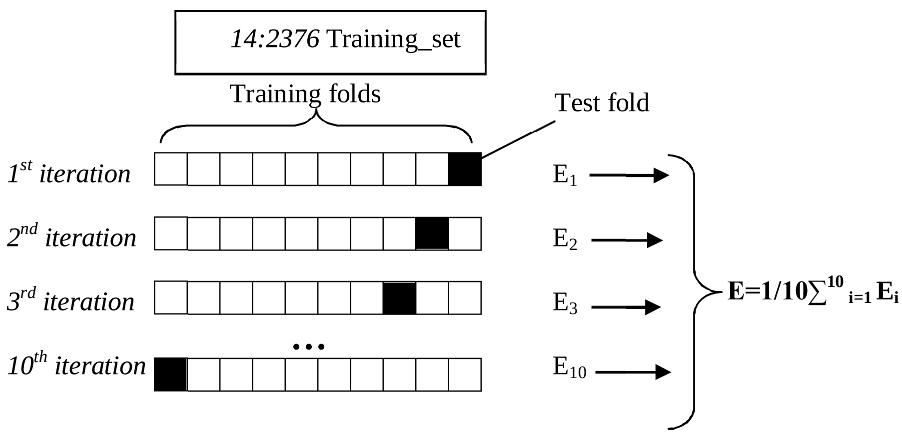


Fig. 3. Concept of machine learning 10-fold cross-validation adopted in the study

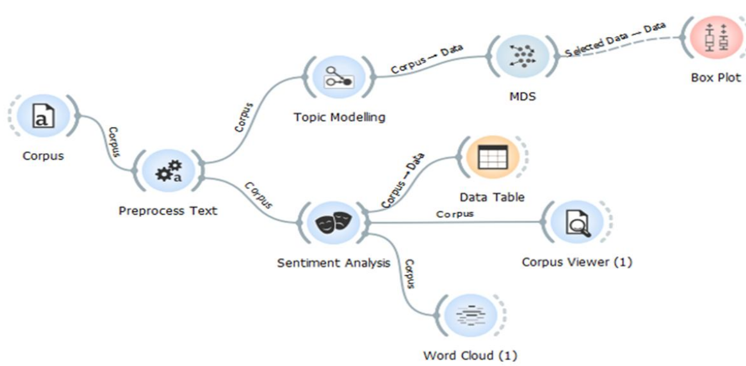


Fig. 4. Framework of the topic modelling and sentiment analysis

IV. RESULT & DISCUSSION

Results obtained are discussed in this section. The machine learning phase of the study, using REPTree learning algorithm, is implemented on the preprocessed data to produce the proposed fear-stimuli detection system in the WEKA data mining software on a 1.7 GHz Intel Core i3 CPU with 4 GB RAM. The resulting decision tree model, predicting the Omicron fear-stimuli from the 14-no data attribute, is presented in Fig. 5 returning attribute_id '*hand_wash*' as the major cause of fear in the academic community. The result shows a tree of size 18 built in 0.02 seconds with the 13th attribute at the root node, haven possessed the highest information gain out of the entire 14-no attribute data table. The predictive accuracy of the experimental model is with a *94.68%* accuracy level hence the encouraging *Precision* and *Recall* weighted average figures. While the Receiver Operator Characteristic (ROC) weighted average shows how the number of accurately classified fearful/fearless instances varies with the number of incorrectly classified instances for the binary problem, the precision and recall weighted averages underscores the efficient performance of the model.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue I Jan 2022- Available at www.ijraset.com

The true positive rates of the class labelling, noted as Recall with 0.947 and same value for instances truly classified as positives, representing the Precision, further validates the reliability of decisions made by this proposed model. Experimental results from the natural language processing of opinions expressed in attribute 15 with the question, '*What do you think about the reopening of TEIs for academic activities*', returns results across the different stages of the framework. The tokenization of the text-corpus computed in a uni-gram approach, claculates the weights of each word contained in the corpus. A document frequency of 0.00-1.00 was set for the scheme and its visualization is shown in Fig. 6 through a word cloud. The size of each uni-gram shows their frequency in the corpus, returning words like hand, wash, washing, important, students, hardly, etc. as those with profound emphasis in the corpus. Result from the LDA and LSI topic modelling, with 9-topic configuration, is presented in Fig. 7. As could be observed, LSI and LDA contains consistent words in topics 4 and 3 respectively with tokens namely: government, college, hardly, wash, hand, water, and nose, which distinctly models topical issues in the minds of respondents. The sentiment analysis part of the framework returns the positive, negative, neutral and compound polarities for each text-respnse to the same attribute 15, indicating emotions expressed therein. In the result, only 25.8% of the respondents expressed opinions clustered as negative sentiments, indicating the state of mind of the majority as either positive or neutral while expressing their opinion about the Omicron-variant and school reopening. Opinions clustered as negative sentiments has compound scores <= -0.05 while compound score >= 0.05 are clustered as positive sentiments with neutrality sentiment captioned between compound score > -0.05 and < 0.05. Fig. 8 shows the outcome of the sentiment analysis of the most negative opinions expressed by correspondents.

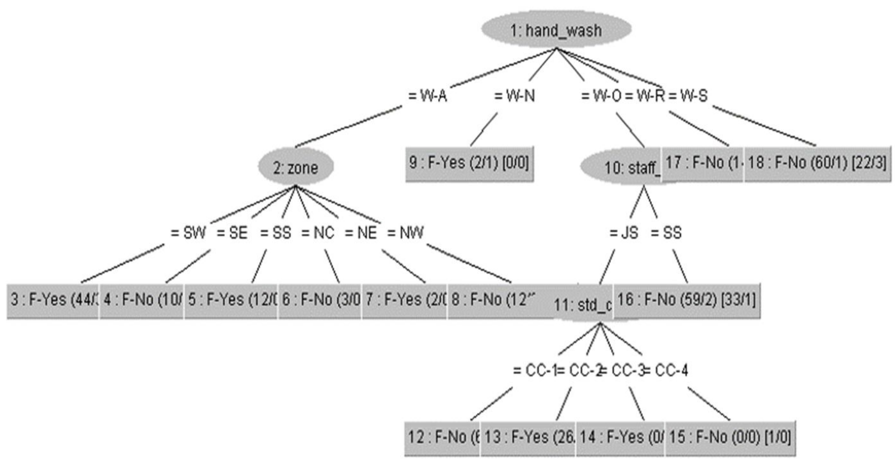


Fig. 5. Prediction result of the REPTree fear-stimuli detection model

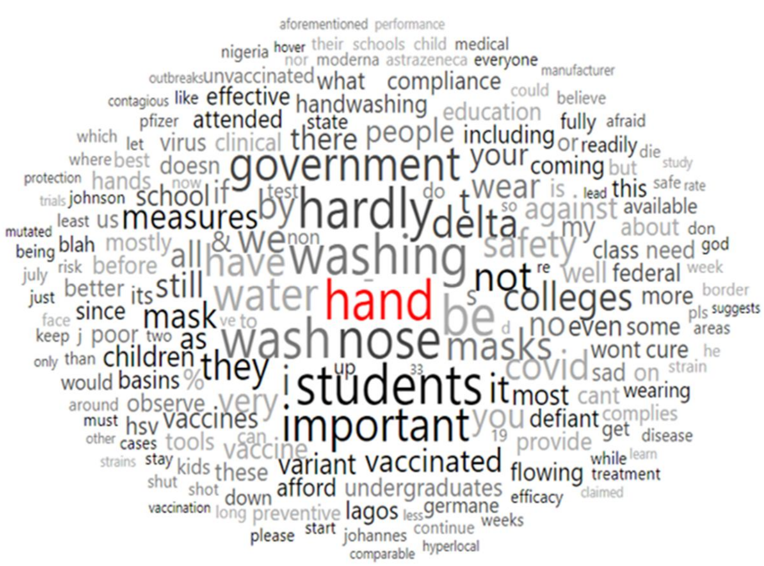


Fig. 6. Word cloud of text corpus

Topic	Topic keywords
1	hand_NN washing_VBG important_JJ student_NNS mask_NNS hardly_RB nose_JJ water_NN nose_NN wash_NN
2	hardly_RB, important_JJ, mask_NNS, nose_JJ, student_NNS, washing_VBG, college_NNS, government_NN, water_NN, nose_NN
3	wash_NN, government_NN, mask_NNS, nose_JJ, student_NNS, water_NN, washing_VBG, hand_NN, safety_NN, college_NNS
4	government_NN, college_NNS, wash_NN, hardly_RB, mask_NNS, nose_JJ, water_NN, student_NNS, important_JJ, wash_VBP
5	student_NNS, safety_NN, water_NN, mask_NNS, wash_VBP, nose_JJ, college_NNS, measure_NNS, washing_VBG, delta_NN
6	water_NN, college_NNS, safety_NN, hardly_RB, measure_NNS, wash_VBP, mask_NNS, nose_JJ, hand_NN, government_NN
7	delta_NN, measure_NNS, safety_NN, covid_NN, water_NN, student_NNS, mask_NNS, nose_JJ, hardly_RB, covid_JJ
8	delta_NN, measure_NNS, water_NN, student_NNS, safety_NN, mask_NNS, covid_NN, wash_VBP, nose_JJ, hand_NN
9	government_NN, hardly_RB, wash_VBP, wash_NN, college_NNS, measure_NNS, important_JJ, hand_NN, water_NN, student_NNS

Topic	Topic keywords
1	nose_VBP, covid_JJ, delta_NN, government_NN, covid_NN, student_NNS, mask_NNS, nose_NN, water_NN, wash_NN
2	covid_NN, delta_NN, nose_VBP, covid_JJ, student_NNS, government_NN, nose_NN, mask_NNS, water_NN, safety_NN
3	government_NN, wash_NN, college_NNS, water_NN, nose_VBP, covid_JJ, hand_NN, delta_NN, hardly_RB, covid_NN
4	nose_VBP, covid_JJ, delta_NN, student_NNS, covid_NN, government_NN, mask_NNS, nose_NN, water_NN, wash_NN
5	nose_VBP, covid_JJ, delta_NN, government_NN, covid_NN, student_NNS, mask_NNS, nose_NN, water_NN, wash_NN
6	nose_VBP, covid_JJ, delta_NN, government_NN, covid_NN, student_NNS, mask_NNS, nose_NN, water_NN, wash_NN
7	nose_VBP, covid_JJ, delta_NN, government_NN, covid_NN, student_NNS, mask_NNS, nose_NN, water_NN, wash_NN
8	nose_VBP, covid_JJ, delta_NN, government_NN, covid_NN, student_NNS, mask_NNS, nose_NN, water_NN, wash_NN
9	hardly_RB, wash_VBP, covid_JJ, nose_VBP, college_NNS, student_NNS, mask_NNS, nose_JJ, delta_NN, hand_NN
10	student_NNS, mask_NNS, hand_NN, nose_JJ, covid_JJ, nose_VBP, safety_NN, water_NN, washing_VBG, important_JJ

Fig. 7. LSI and LDA topic modelling with generated keywords respectively

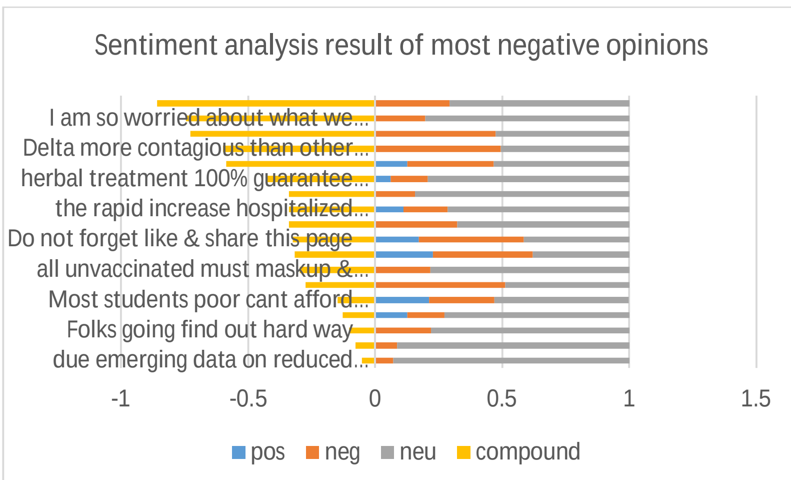


Fig. 8. Result of sentiment analysis

From the REPTree experimental result, handwashing garnered highest information gain hence at the root of the tree, being the deciding factor of the predictive model. A critical study of the decision tree reveals that respondents who wash their hands '*always*' *(W-A)* regularly, from the south-west, south-south and north-east zones of Nigeria, are yet Omicron-variant fearful while counterparts from the south-east, north-central and north-west does not dread the variant. The decision tree shows those who never wash their hands are understandably fearful of the variant just as Junior staff, who '*often*' wash their hands but opines students' compliance with safety measures are poor with those who feel students' compliance are good, does not in any way exhibit COVID-19 fears. Similarly, senior staff who '*often*' wash their hands does not exhibit fears while discharging their daily duties; whereas, respondents who '*rarely*' and '*sometimes*' wash their hands still does not exhibit Omicron fear according to the machine learning modelled outcome. Result further implies that issues relating to face mask wearing, individual COVID-19 test status, COVID-19 case count, and staff awareness of COVID-19 updates does not have significant information gain to stimulate the Omicron-variant fears in the academic communities. Regular wearing of nose masks likewise does not necessarily implies fear of the variant while experimental result suggest that issues about college precautionary measures and students' safety measures prominently matters to staff of TEIs in Nigeria, and could stimulate Omicron fear.

The sentiment analysis result shows a higher neutral sentiments as expressed by the entire academic communities such that most of their item 15th opinions are hardly of optimism but more of pessimism as regards the subject of the Omicron-variant and school reopening. Their negative sentiments is almost eroded by the neutrality inherent of their opinions. This is a pointer to their constructive criticism of government's decision for school opening in the face of Omicron-variant prevalence, which echoes their concern and yearn for non-clinical preventive measures other than criticizing the decision to open schools. As may be observed from the topic modelling and word cloud results, compliance with non-clinical safety measures remains topical in the minds of the TEI stakeholders and in relation to the output of the supervised REPTree decision tree prediction, issues surrounding handwashing, provision of water, and student's compliance stimulates the Omicron-variant fear.

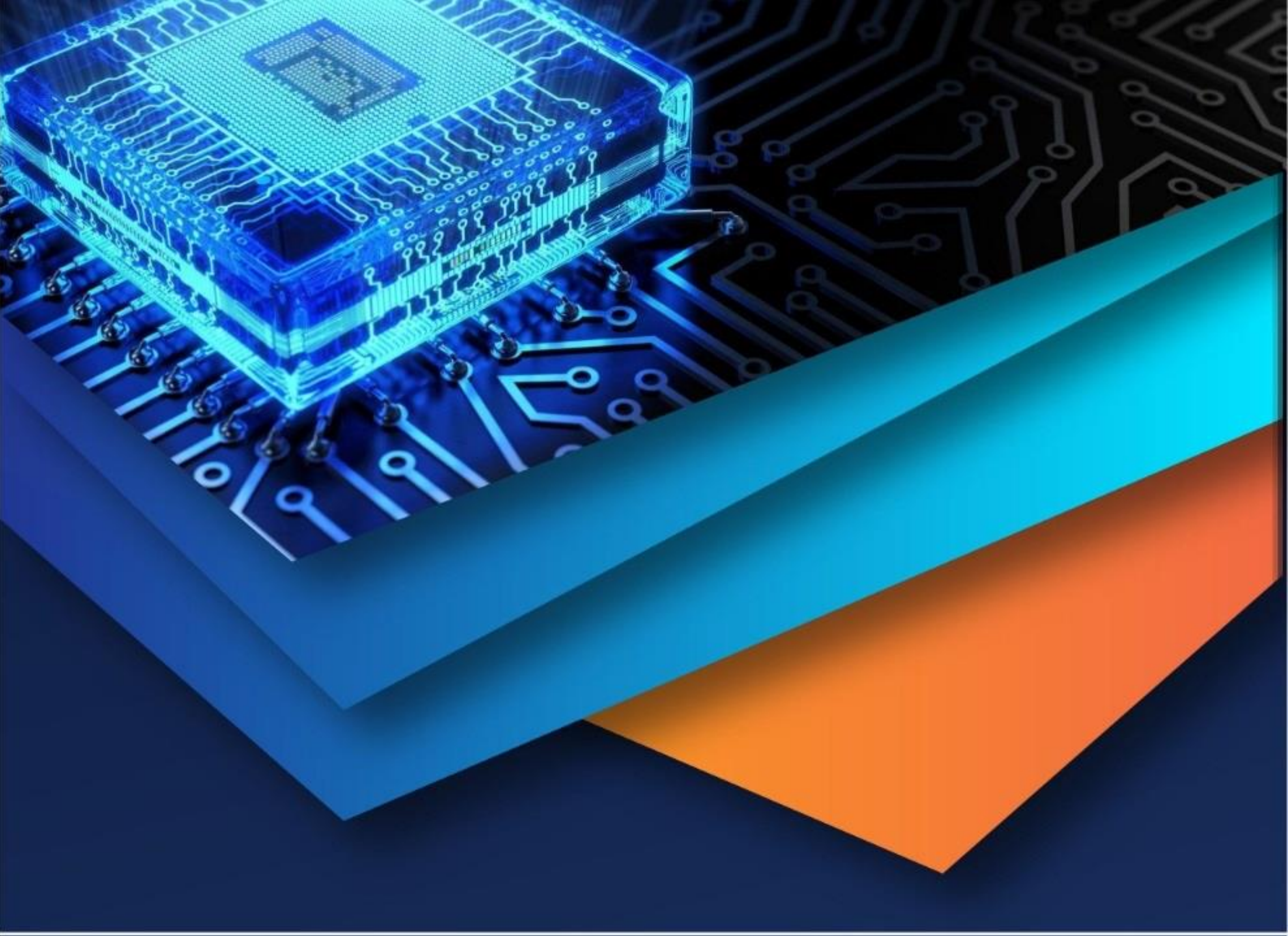
V. CONCLUSION & RECOMMENDATIONS

COVID-19 related personal opinions, especially over the Omicron-variant, acquired from Tertiary Educational Institutions is captured for a predictive modelling experiment. REPTree machine learning algorithm, VADER-based sentiment analysis, and Topic modelling techniques were executed to uncover stimulus of COVID-19 fears in educational communities across Nigeria in the face of the Omicron invasion and government's decision to open schools. The resulting decision tree model is validated with Precision, Recall and ROC area performance metrics yielding a state-of-the-earth prediction accuracy of 94.68% and similarly to the topic modelling of textual corpus, the experimental model returns the attribute *hand-wash* as the most significant possessing the highest information gain towards determining the Omicron-variant fear-stimuli in college communities. The sentiment analysis of the opinions expressed likewise shows more of neutral sentiments, lower negative and positive sentiments concerning government decision to open schools in the face of the new wave. Consequently, the fear rate of the virus amongst staff of TEIs is a function of issues associated with compliance of stakeholders with non-clinical precautionary measure of regular hand washing and other issues associated therewith. The palpable fret being exhibited by staff of TEIs therefore need to be addressed by school operators and regulators with the mind view of ensuring adequate compliance with regulations put in place by government for tertiary institutions which will in no small way tame the fear-trend of staff towards the third wave of the virus. Awareness of COVID-19 case-count across the country, the attribute *casePositive*, which checks if respondents know of any COVID-19 positive acquaintance, surprisingly does not stimulate fear in staff. A research with wider dataset scope is recommended for future work while ensuring a balanced class representation in the training set.

REFERENCES

[1] P. C. Addo, F. Jiaming, N. B. Kulbo and L. Liangqiang, "COVID-19: fear appeal favoring purchase behavior towards personal protective equipment," The Service Industries Journal, vol. 40, no. 7, pp. 471-490, 2020.
[2] WHO, "Coronavirus disease 2019 (COVID-19) situation report," World Health Organization, 8 April 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. [Accessed 8 April 2020].
[3] H. Plecher, "Age structure in Nigeria," Statista, 28 July 2021. [Online]. Available: https://.statista.com/statistics/382296/age-structure-in-nigeria/. [Accessed 1 December 2020].
[4] NCDC, "COVID-19 daily Statistics," Nigerian Centre for Disease Control, 12 January 2022. [Online]. Available: https://covid19.ncdc.gov.ng/. [Accessed 12 January 2022].
[5] L. Wen, H. Wang, L. Yuxing and L. Li, "Psychological status of medical workforce during the COVID-19 pandemic: A cross-sectional study," Psychiatry research, vol. 288, no. 112936, 2020.
[6] S. Taylor, L. Caeleigh, M. Paluszek, T. Fergus, D. McKay and G. Asmundson, "Development and initial validation of the COVID Stress Scales," Journal of Anxiety Disoders, vol. 72, no. 102232, 2020.
[7] T. O. Olaleye and O. R. Vincent, "A Predictive Model for Students' Performance and Risk Level Indicators Using Machine Learning," in International Conference in Mathematics, Computer Engineering and Computer Science, Lagos, 2020.
[8] O. Arogundade, C. Atasie, S. Misra, A. Sakpere, O. Abayomi-Alli and K. Adesemowo, "Improved Predictive System for Soil Test Fertility Performance Using Fuzzy Rule Approach," in 2nd International Conference on Soft Computing and its Engineering Application, India, 2021.
[9] F. Ornell, J. Schuch, A. Sordi and F. Kessler, "Pandemic fear" and COVID-19: mental health burden and strategies," Brazilian Journal of Psychiatry, vol. 42, no. 3, pp. 232-235, 2020.
[10] S. J. Eder, D. Steyrl, M. M. Stefanczyk, M. Pieniak, J. M. Molina, P. Ondra, B. Jakub, S. Patrick, F. Scharnowski and A. A. Nicholson, "Predicting fear and perceived health during the COVID-19 pandemic using machine learning: A cross-national longitudinal study," Plus one, vol. 16, no. 3, 2021.
[11] E. C. Green and W. Kim, "Fear arousal, sexual behavior change and AIDS prevention," Journal of Health Communication, vol. 11, no. 3, pp. 245-259, 2006.
[12] S. Taylor, "The psychology of pandemics: Preparing for the next global outbreak of infectious disease," Cambridge Scholars Publishing, 2019.
[13] G. J. Asmundson and S. Taylor, "Coronaphobia: Fear and the 2019-nCoV outbreak," Journal of anxiety disorders, vol. 70, no. 102196, 2020.
[14] V. Giallonardo, S. Gaia, V. D. Vecchio, M. Luciano, U. Albert, C. Carmassi, G. Carra, F. Cirulli, B. Dell'Osso, M. Nanni and M. Pompili, "The impact of quarantine and physical distancing following COVID-19 on mental health: study protocol of a multicentric Italian population tria," Frontiers in Psychiatry, vol. 11, p. 533, 2020.
[15] S. A. Lee, "Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety," Death studies, vol. 44, no. 7, pp. 393-401, 2020.

[16] N. O. Okoi and J. E. John, "Nosophobia, hypochondriasis, and willingness of people to seek healthcare amidst the COVID-19 pandemic in Calabar Metropolis of Cross River State, Nigeria," Open Journal of Psychiatry & Allied Sciences, vol. 12, no. 1, pp. 36-42, 2020.
[17] S. Paul, "Nosophobia of corona: A loss or gain for India?," International Journal of Health & Allied Sciences, vol. 9, no. 4, pp. 367-367, 2020.
[18] G. Mertens, L. Gerritsen, S. Duijndam, E. Salemink and M. Iris, "Fear of the coronavirus (COVID-19): Predictors in an online study conducted," Journal of Anxiety Disorders, vol. 74, no. 102258, 2020
[19] P. Kaila, Dr. and D. A. Prasad, "Informational Flow on Twitter - Corona Virus Outbreak – Topic Modelling Approach," International Journal of Advanced Research in Engineering and Technology, vol. 11, no. 3, pp. 128-134, 2020.
[20] F. Martin and M. Johnson, "More Efficient Topic Modelling Through a Noun Only Approach," in Proceedings of Australasian Language Technology Association Workshop, 2015.
[21] T. Olaleye, T. Arogundade, A. Abayomi-Alli and K. Adesemowo, "An Ensemble Predictive Analytics of COVID-19 Infodemic Tweets using bag of words," in Data Science for COVID-19, Singapore, Elsevier, 2021.
[22] T. Olaleye, O. Arogundade, C. Adenusi and S. Misra, "Evaluation of Image Filtering Parameters for Plant Biometrics Improvement Using Machine Learning," in 2nd International Conference on Soft Computing, India, 2021.
[23] T. Sherstinova, O. Mitrofanova, T. Skrebtsova, E. Zamiraylova and M. Kirina, "Topic Modelling with NMF vs. Expert Topic Annotation: The Case Study of Russian Fiction," in Mexican International Confeence on Artificial Intelligence, 2020.
[24] H. Hota, D. Sharma and N. Verma, "Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad," in Data Science for COVID-19, London, Elsevier, 2021, pp. 275-293.
[25] NCCE, "List of colleges of education in Nigeria.," Nigerian Commission for Colleges of Education, 28 July 2021. [Online]. Available: http://www.ncceonline.edu.ng/colleges.php. [Accessed 28 July 2021].



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY