TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

Integrating Artificial Neural Network: Model for the Development of SMARTBOT E-learning Platform

[1]Gemma M. Pangilinan, [2]Mary Ann F. Quioc [3]Albert A. Vinluan

[1]Don Mariano Marcos Memorial State University [1], [2], [3]AMA University [1]gpangilinan@dmmmsu.edu.ph
[2]maryannquioc@gmail.com
[3]albert.vinluan@gmail.com

Article Info
Volume 83
Page Number: 1109 - 1116
Publication Issue:
May - June 2020

Article History
Article Received*: 11August 2019*
Revised*: 18November 2019*
Accepted*: 23January 2020*
Publication*: 10 May2020*

Abstract:

E-learning optimization integrating Smartbot is the delivery of learning with recent technologies by allowing a student to learn aside from the traditional face to face classroom instruction. On the other hand, educators and trainers are empowering by allowing the use of technology to communicate with learners anytime in accessing learning resources to plan and deliver the lessons. This study is specifically addressed the (1) Measuring the dataset performance using Decision Tree. (2) Integration of Artificial Neural Network as a model basis for the Development of Smartbot E-learning (3) Development of Model of the SMARTBOT E-learning Platform. This study is conceptualizing the idea to use these gadgets as a form of a learning tool in assessing the predictive readiness of Science, Technology, Engineering and Mathematics (STEM) learners for Work Immersion of Senior High School through an android learning app environment with Artificial Neural Network (ANN) and chatbot integration. This learning tool will further be called SMARTBOT – a short term for smart robots.

Keywords: *ANN, Dataset Performance, SMARTBOT, E-learning, Cosine Similarity*

1. INTRODUCTION

Education Act of 1982 stresses the importance of Science and Technology and its relevance to national growth and development as stipulated in Section 10, to wit: "Science and technology are important to national growth and advancement. The State shall give priority to scientific research, invention, innovation, and production, as well as to science and technology education, training, and services. It shall support aboriginal, sufficient, self-reliant, scientific, and technical capabilities and their application to the competitive system of the nation. [1]".

The provision states the importance of developing science and technological competence as a means of attaining national development goals as well as addressing the interests and demands of growing industrialization and globalization, which is also the concern of the education sector. The learning progress of every student depends on how instruction is being delivered to them by their mentors. Learning pertains to knowledge or skills that are acquiring through appropriate instructions; thus, it is the task of the teachers to look for a better means of delivering lessons to their students. That is, they should provide unique and modern strategies that will suit the needs of the learners, make the teaching-learning process becomes more convenient and comfortable not only to students but to teachers as well. Curriculum integration with the use of technology involves the infusion of technology as a tool to enhance the learning in a content area or multidisciplinary setting. Effective integration of technology is achieving when students can select technology tools to help them obtain information promptly, analyze and synthesize the information, and present it professionally. The technology should become an integral part of how the functions of the classroom – as accessible as all other classroom tools. The focus in each lesson or unit is the curriculum outcome, not the technology. The power, reach, and rapid evolution of technology demand a curriculum that will enable students to become technologically literate. Students need to acquire the technological skills and knowledge that will allow them to participate fully in a competitive global economy

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

and to become responsible citizens in an environmentally vulnerable world.

Artificial Neural Networks (ANNs), on the other hand, is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely basing on the animal neuron. The processing ability of the network is storing in the inter-unit connection strengths, or weights, obtained by the process of adaptation to or learning from a set of training patterns. Neurons communicate via electrical signals that are momentary impulses or "spikes" in the voltage of the cell wall or membrane. The interneuron connections are mediated by electrochemical junctions called synapses, which are located on branches of the cell referred to as dendrites. Each neuron typically receives many thousands of connections from other neurons and is therefore continually receiving a multitude of incoming signals, which eventually reach the cell body. Here, they are integrated or summed together in some way, and, roughly speaking, if the resulting signal exceeds some threshold, then the neuron will "fire" or generate a voltage impulse in response. This is then transmitted to other neurons via a branching fiber known as the axon[2]. Sayed and Baker utilized an artificial neural network model as supervised learning. The network provides an example of the learning input parameters and the optimal optimized and accurate output base on the input [3]. They describe how to use an artificial neural network to produce a converging mathematical model by utilizing e-learning interactions and social analytics. It can predict the performance of the students; there should, therefore, be a reduction in the risk of failure of the enrolled e-course. Furthermore, input parameters have to be investigated in the extensive data generated during the semester and analyzed to improve and correct the process of learning. [4] [5]

Moreover, this can also be a Learning Management System. The Learning Management System (LMS) is a software application for the implementation, recording, tracking, reporting, and distribution of training courses, training programs, and learning and development programs. The concept of the learning management system emerged directly from e-learning. Although the first LMS has appeared in the higher education sector, currently, most LMSs focus on the corporate market. Learning Management Systems is the largest segment of the learning system market. The introduction of LMS was in the late 1990s. Learning management systems have been designed to identify differences in training and learning, to use analytical data and to document. LMSs focus on online learning delivery but promote a variety of uses, serving as a portal for online content, including courses, both asynchronous-based and synchronous-based. An LMS may offer classroom management for instructor-led training or a flipped classroom, used in higher education, but not in the corporate space[6].

This ultimate goal of the study is to Measure and validate the dataset performance using Decision Tree and Integrate Artificial Neural Network as a model basis for the Development of SMARTBOT E-learning platform.

2.METHODOLOGY

Research design is a critical and exhaustive investigation or experimentation to discover and gather facts that are significant for analyzing data in the development and implementation of the study. In this study, the researcher will use the following methods in creating a model basis for the Development of SMARTBOT E-learning that integrates Artificial Neural Network (ANN) and Smarbot.

A. Measuring the dataset performance using Decision Tree

Many decision tree induction algorithms follow a top-down approach, starting with a pair of training sets and associated class labels. When the tree builds, the training set is again split into smaller subsets[7][8]. A decision tree is a tree-like flow-diagram [9]. The node in the tree denotes a research attribute. Every branch is the test output, and each leaf is a distribution of class or class. The top node serves as the root and can be a cluster from the root node to a single leaf. The application of the classification rules in the tree of decisions is straightforward. It has several algorithms, but the main idea is to use the induction approach from top to bottom [10].

WEKA or Information Analysis Waikato Environment is a workspace array of trendsetting machine learning algorithms and pre-processing

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

software [11], [12]. A relatively large data set is required to create a decision rule[13], [14]. One significant variable of this study is the dataset. For this study, the dataset was collated through data mining activities from past school records. The dataset was divided into two parts – the learners' profile and the lessons. For the learners' profile, it is subdivided into eight, namely ID number, gender, JHS academic profile, SHS academic profile, hobbies, skills, ambition five years ago, and five years from now. The second significant part of the dataset is the lesson. To measure the dataset performance, the researcher considered two important factors – the highest rate correctly classified instances and information gain. For the highest rate correctly classified instances, the researcher considered six (6) dataset algorithms under Decision Tree, and these are the REP Tree, Random Tree, Random Forest, J48, C4.5, and the Classification and Regression Tree (CART). The results of the six (6) decision tree datasets were compared by the researcher to find out which of these six has the highest rate of correctly classified instances. Meanwhile, to cross-validate the result, the researcher will further consider the F-score and the information gain. The F-score was even computed using the formula.

$$F_1 = 2 \text{ x } \frac{(\text{Precision})(\text{Recall})}{\text{Precision} + \text{Recall}} \qquad (1)$$

The researcher also used F-score online calculator, which was developed by FX Solver Education. The calculator is designed to input data analytics for precision and recall. As to the reduction in entropy or surprise of transforming the data set, the researcher used information gain, and the formula is illustrated below:

$$H = -\sum_{i=1}^{K} p_K \log_2 p_K \qquad (2)$$

After identification of the best Decision Tree algorithm to use in the study, Classification and Regression Tree (CART) has the highest rate for correctly classified instances. The researcher even secured data classifying and cleansing. Moreover, the CART was utilized to map learners' datasets as to what lesson or group of lessons they should be reviewing. On the other hand, the researcher also ensured that the lessons are aligned in the instructional/curricular activities of the school by considering the approved Work Immersion Learning Plans of DMMMSU – Mid La Union Campus. The said set of learning plans is divided into ten (10) significant lessons. The said lessons are Lesson 1 – Expected Behavior, Lesson 2 – Work Immersion Rules and Regulations, Lesson 3 – Writing a Resume, Lesson 4 – Filling-out An Application Form, Lesson 5 – Job Interview and Skills Training, Lesson 6 – Portfolio Instructions, Content Packaging, and Presentation, Lesson 7 – Appreciating Management Processes, Lesson 8 – Appreciating Business Processes, Lesson 9 – Applying Skills Learned and Proper Values Acquired from School, and Lesson 10 – Work Immersion Experience. CART was adequately supported and optimized by the researcher through entropy – a process that was even included under the concept of Information Gain [15]. For smooth mapping of the lessons to be suggested to the learner, the researcher used one significant attribute in the learners' dataset profile. This attribute was the learner's ID number. This attribute helps the learner that once he/she entered his/her ID number in the system, the output was a suggested lesson or a group of lessons to review. To secure that there was a standardized generated model for each learner, the researcher considered the models generated by WEKA.

B. Creation of Artificial Neural Network Model using Cosine Similarity

Merging the applied and developmental research design helps the researcher to assess the outcome in determining the algorithm's performance to achieve the desired output. The applied research design [16] involves seeking new applications of scientific knowledge to the solution of a problem such as the development of a new system or procedure or new method to solve the problem. The developmental research design [16] is decision-oriented research involving the application of the steps of the scientific method in response to an immediate need to improve existing practices. Concerning the above discussion, the researcher was able to create or develop an Artificial Neural Network Model by considering inputs and nodes. The inputs of the

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

study are the lessons 1 to 10 of the Work Immersion of DMMMSU-MLUC, which was optimized by the Cosine Similarity Algorithm. The underlying procedures for the said algorithm may include the Text Conversion and Extraction, Pre-processing, and Bag of Words. Meanwhile, there are hidden layers for the nodes, and some of these concepts are presented in the next discussions.

Natural Language Processing in Php is a library for natural language processing written in PHP. The own needs drive its development for text classification, clustering, tokenizing, stemming from others. The natural language processing chatbots (NLP) find a way to translate the speech or text of the user into structured data following to selection of a corresponding response. The measures include the natural language process of tokenization in which the method will divide the set of words into linguistically descriptive tokens or pieces taken from the lesson plan and the expected response from the student with the terms having a different value[17]. The next step is the normalization of data responses where the text to find out the typographical errors and common spelling mistakes that might alter the intended meaning of the user's request. It will the go-to name entity recognition wherein the program model of a chatbot looks for different categories of words, similar to the topic, lesson, or response from both teacher or student, whichever information is required[18]. It will then go to dependency parsing: The Chatbot searches for

the subjects, verbs, objects, everyday phrases, and nouns in the user's text to discover related phrases that what users want to convey—the chatbot's linked to the database for many applications. The database is used to manage the chatbot and to provide appropriate responses to each user. NLP can use a mixture of text and patterns to turn human language into data information that can be useful in finding appropriate responses.

Term Frequency-Inverse Document Frequency, or TF-IDF for the brief, where Term Frequency is a measure of the word frequency in this document whereas Inverse Document Frequency is an indicator of the occurrence of the word in documents at times[19].

$$TF = \frac{\text{(The number of times the term } \mathbf{t} \text{ is shown in the document)}}{\text{Number of terms in the document}} \quad (3)$$

$$IDF = 1+\log(N/n), \quad (4)$$

Where N is the number of documents, and n is the number of documents in which t appears. TF-IDF weight is a weight usually used in the recovery of data and text mining. This weight is a statistical assessment utilized to measure the significance of a word to a document in a corpus or collection.

TF-IDF is a numerical statistic that shows the relevance of keywords to some specific documents, or it can be said that it provides those keywords, using which some specific documents can be identified or categorized [20]. For example, a blogger is running a blog with hundreds of contributors, and he just hired an internee whose main task is to add new blog posts daily. It has been observed that most of the time, internees do not take care of tags due to which many blog posts are not categorized. This is one of the ideal conditions for applying the TF-IDF algorithm, which can identify the tags automatically for the bloggers. It will save plenty of time for bloggers and internees, as they would not need to take care of tags.

TF-IDF is a conversion used for text to obtain two vectors of actual value in vector space. Then, by trying to take their dot product and divide it by the product of their norms, we can acquire the Cosine similarity of any combination of vectors. It provides the angle cosine among the vectors. A similarity measure between two non-zero vectors is the cosine similarity. Use this formula to determine the similarity between any two documents d1 and d2.

$$Cosine\ Similarity\ (d1, d2) = Dot\ product(d1, d2) / ||d1|| \quad (5)$$

where d1,d2 is two non-zero vectors.

After dealing with these underlying processes, mapping the needs of the learners and address it with an output which is the chatbot. To secure that this chatbot is accessible to the needs of the learners, there will be two types of output, and these types can run using a web-based platform and android phone.

C. Model prototype of the SMARTBOT E-learning Platform

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

The text Conversion and Extraction is the principal activity in the development of the SMARTBOT e-Learning platform. The primary source of data, which is the lesson plan for the Senior High School of Don Mariano Marcos Memorial State University, will be converted into basic text pre-processing with Natural Language Toolkit (NLTK). NLTK is a powerful package that provides a set of various natural language algorithms [21]. NLTK consists of the most common algorithms, such as tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition. NLTK helps the computer to analyze, preprocess, and understand the written text [22]. Unstructured data is generated and collected in a wide range of forms; expressly, the data set properties, which includes the profile of the students and lessons gathered.

The pre-processing stage involves a four-step principle. The following are Tokenization, Removing Stop Words and Stemming, and Lemmatization and Bag of Words. Tokenization is the first step in text analytics. The process of breaking down a text paragraph into smaller chunks such as words or sentences is called Tokenization. A token is a single entity that is building blocks for sentences or paragraphs. Sentence tokenizer breaks text paragraph into sentences . Word tokenizer breaks text paragraph into words. As the text is cleansed, each token was tagged with their part-of-speech labels. Removing Stop Words is considered as noise in the text. In NLTK for removing stopwords, we need to create a list of stopwords and filter out your list of tokens from these words. Ubiquitous words that would appear to be of little value in helping select documents matching a user need are excluded from the vocabulary entirely. These words are called stop words.

Stemming is a process of linguistic normalization, which reduces words to their word root word or chops off the derivational affixes. For example, connection, connected, connecting word reduce to a common word "connect." Lemmatization reduces words to their base word, which is linguistically correct lemmas. It transforms the root word with the use of vocabulary and morphological analysis. Lemmatization is usually more sophisticated than stemming. Stemmer works on an individual word without knowledge of the context. For example, The word "better" has "good" as its lemma. This thing will miss by stemming because it requires a dictionary look-up. The Bag-of-words model (BoW) is the simplest way of extracting features from the text. BoW converts text into the matrix of the occurrence of words within a document. This model concerns about whether given words occurred or not in the document.

Example: There are three documents:
Doc 1: I love dogs. Doc 2: I hate dogs and knitting.
Doc 3: Knitting is my hobby and passion.

Now, we can create a matrix of documents and words by counting the occurrence of words in the given document. This matrix is known as Document-Term Matrix (DTM), as shown in Figure 1.

	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	1	1	1							
Doc 2	1		1	1	1	1				
Doc 3					1	1	1	2	1	1

Figure 1 – Document-Term Matrix (DTM)

This matrix is using a single word.

3.RESULTS AND DISCUSSION

1. Measuring the dataset performance using Decision Tree

Waikato Environment Knowledge Analysis (WEKA) was administered to measure the dataset's accuracy. Based on the generated results, it was found out that out of six (6) decision tree algorithms for dataset performance, it was disclosed that Classification and Regression Tree (CART) has the highest dataset performance with 97.3826 percent for correctly classified instances. The next two highest after CART are the Random tree with 96.8347 percent and the RepTree with 96.2279 percent. Despite such comparisons above, all dataset performances for each of the six algorithms are possible to consider. However, the researcher considered CART for it entails the lowest incorrectly classified instances and has an excellent model to be generated that is fit on the system.

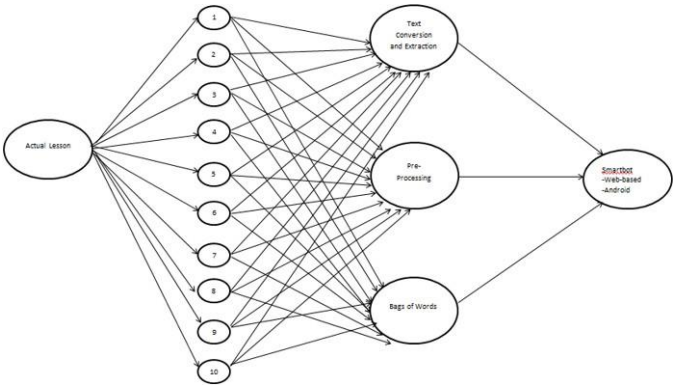
TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

Ensuring that all-important details for CART are adequately presented and validated, the generated result for CART via WEKA is further presented in Figure 2.

```
Correctly Classified Instances        1265               97.3826 %
Incorrectly Classified Instances        34                2.6174 %
Kappa statistic                          0.805
Mean absolute error                      0.0504
Root mean squared error                  0.1542
Relative absolute error                 38.1005 %
Root relative squared error             60.1271 %
Total Number of Instances             1299

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.984     0.163     0.988      0.984    0.986      0.805    0.978     0.998     Passed
               0.837     0.016     0.802      0.837    0.819      0.805    0.978     0.788     Failed
Weighted Avg.  0.974     0.153     0.974      0.974    0.974      0.805    0.978     0.983

=== Confusion Matrix ===

    a    b   <-- classified as
 1188   19 |    a = Passed
   15   77 |    b = Failed
```

Figure 2 - CART Result for Dataset Performance



As to the cross-validation phase of the computed dataset performance, the researcher obtained a similar F-score or F-measure performance of .986 using the F-score online calculator. The said calculator required the researcher to appropriately fill in the precision (p) and recall (r) values in the workable cell to generate the actual computed value. Illustration of the online F-score calculator is presented in Figure 3.

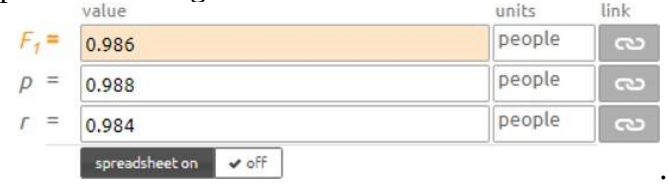


Figure 3 - Computation of F-score using an online calculator

With the conducted cross-validation, it was found out that CART, as the classification algorithm, is applicable and possible for this study. This further connotes that CART can be possibly used in analyzing the dataset and create a possible illustration and model that is fit for the study.

2. Creation of Artificial Neural Network Model using Cosine Similarity

At the onset of such development of the program, the researcher conceptualized and created an Artificial Neural Network Model using Cosine Similarity. In the created model, it can be gleaned that it is divided into three layers. The first layer is the input; the second layer is the hidden layer, which refers to the actual nodes processes, which are aligned and anchored to the concept of the Cosine Similarity Algorithm. For the first layer, the input refers to the set of lessons for Work Immersion. Each of the lessons is attached to the three significant nodes found in the hidden layer. The three significant nodes for this layer are the text conversion and extraction, followed by pre-processing. In the pre-processing stage, it includes three latent functions, and these are: lemmatization, reducing noise, and stemming. For the last primary node, it is the bag of words. The bag of words served as the pool of mapping words for the developed system or the output. For the output of the study, the researcher came up with a chatbot. This chatbot can run into two possible environments. The first environment is via the webpage, and the second environment is via android. These two environments secure that the output could be accessible to learners with two types of environment preference. Visualization of the ANN Model using Cosine Similarity is illustrated in Figure 4.

3. Development Model of the SMARTBOT E-learning Platform

The prototype system architecture for the SMARTBOT E-learning platform is a by-product of recommendations from experts in order to meet and sustain such technical and educational standards. Concretizing these inputs and suggestions through documentary reviews and analyses, a series of the dataset of STEM learners were considered. The said dataset was composed of their profile and the school's official guided lessons for Work Immersion. The said dataset was subjected to the

Figure 4 - ANN Model using Cosine Similarity

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

initial phase of system development, and that is text conversion and extraction, which yield probable text inputs. These text inputs are subjected to the text pre-processing stage. In the pre-processing stage, translation, tokenization, data cleansing are the expected processes, further optimized by the process of stemming and lemmatization. After such thorough procedures, the text inputs are now ready for final dataset alignment for topic detection and transformation or the so-called Bag of Words (BoW) with the use of cosine similarity algorithm.

Moreover, to meet the demands of learners as to possible platform environment, the prototype of the SMARTBOT E-learning platform was made ready for android-based and web-based environments. This means that STEM learners can be opened with the use of their mobile phones with Android as the operating system, and even with their laptops, netbooks, and tablets. The said prototype was even subjected to ISO 25010 in terms of functional suitability, performance efficiency, usability, security, reliability, and portability.

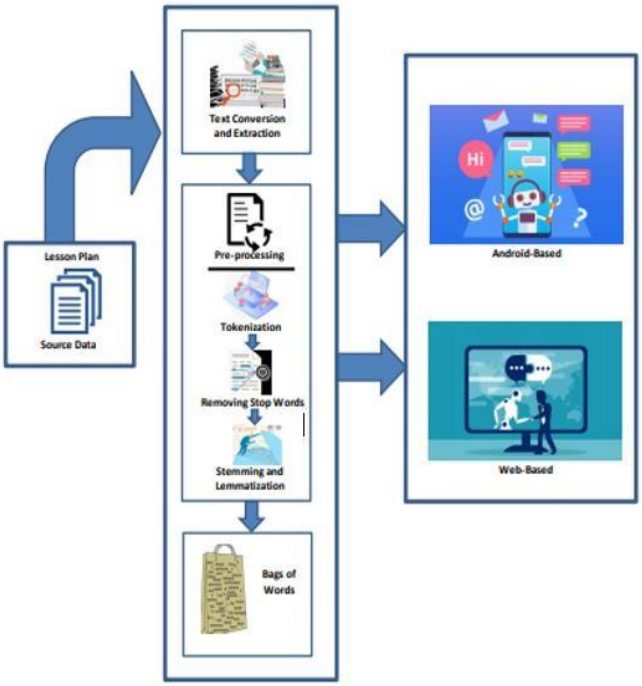


Figure 5 – Prototype Architecture of SMARTBOT E-learning System

4. CONCLUSION

In the field of advanced technological concepts, particularly in learning management systems, dataset properties and attribution should be checked with its dataset performance and accuracy appropriately. In this study, out of six accuracy tests based on the correctly classified instances, CART has an enormous dataset accuracy performance of 97.3826%. F-score was even administered and determined to cross-validate the results, and it is found out to be the perfect match. These concepts of CART was used to illustrate the process of assigning for possible suggested lessons to every enrolled learner. The student's ID number was used as an attribution for the CART. Further, an example of a 162-parameter size visualizes decision tree was even generated, and it is said to have at least 97.0612 correctly classified instances.

Among the different types of algorithms associated with Artificial Neural Network, cosine similarity was the perfectly matched algorithm for this study. In developing a chatbot, the researcher discovered that a Hypertext Preprocessor (PHP) language could be integrated into Natural Language Processing, although NLTK is the main program to develop Python programs to operate with the information about the human language. The use of the data mining approach by algorithms is a methodology used to extract and classify the information according to the requirements of measures and thresholds. The diversification of e-learning systems has resulted in practical methods for managing the complex and dynamic functions of the e-learning environment. Learning Management Systems (LMS) that deliver instructional content at the right time and in the correct format to ensure contact between the user and the learning

REFERENCES

[1] Senate Bill 920, “Education Act of 1982,” *Senate Bill 920*, no. Fifteen Congress. 2010.

[2] S. S. Cross, R. F. Harrison, and R. L. Kennedy, *Introduction to neural networks*, vol. 346, no. 8982. 1995.

[3] M. Sayed and F. Baker, “E-Learning Optimization Using Supervised Artificial Neural-Network,” *J. Softw. Eng. Appl.*, vol. 08, no. 01, pp. 26–34, 2015.

[4] R. Ferguson and S. B. Shum, “Social learning analytics: Five approaches,” *ACM Int. Conf. Proceeding Ser.*, no. May, pp. 23–33, 2012.

[5] J. Yoo and J. Kim, “Can online discussion

May – June 2020
ISSN: 0193-4120 Page No. 1109 - 1116

participation predict group project performance? Investigating the roles of linguistic features and participation patterns," *Int. J. Artif. Intell. Educ.*, vol. 24, no. 1, pp. 8–32, 2014.

[6] G. M. Pangilinan, M. A. F. Quioc, and L. L. Lacatan, "Integrating artificial neural network and smartbot on the development of an E-learning platform," *Test Eng. Manag.*, vol. 82, 2020.

[7] A. A. Elacio, F. G. Balazon, and L. L. Lacatan, "Digital transformation in managing employee retention using agile and C4.5 algorithm," *Test Eng. Manag.*, vol. 82, no. 15217, pp. 15217–15225, 2020.

[8] D. V. Patil and R. S. Bichkar, "A hybrid evolutionary approach to construct optimal decision trees with large data sets," *Proc. IEEE Int. Conf. Ind. Technol.*, pp. 429–433, 2006.

[9] S. Kalmegh, "Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News," *IJISET -International J. Innov. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 438–446, 2015.

[10] A. A. Elacio, L. L. Lacatan, A. A. Vinluan, and F. G. Balazon, "Machine Learning Integration of Herzberg's Theory using C4.5 Algorithm," vol. 9, no. 1, pp. 57–63, 2020.

[11] A. A. Elacio, "Machine Learning Integration of Herzberg ' s Theory using C4 . 5 Algorithm."

[12] A. A. Elacio and F. G. Balazon, "Title: Digital Transformation in Managing Employee Retention using Agile and C4 . 5 Algorithm," p. 25010.

[13] A. Kumar, "Classification and Regression Tree (CART) Techniques," *Data Min. Tech. Tools Knowl. Discov. Agric. Datasets*.

[14] R. J. Lewis, D. Ph, and W. C. Street, "An Introduction to Classification and Regression Tree (CART) Analysis," *2000 Annu. Meet. Soc. Acad. Emerg. Med.*, no. 310, p. 14p, 2000.

[15] M. A. Muharram and G. D. Smith, "Information Gain and Gini Index," pp. 379–388, 2004.

[16] D. Illustration, *Methodology of Scientific Research Illustration of Qualitative Research Degrees Educational & Didactic Communication 2014.* .

[17] J. Cahn, "CHATBOT: Architecture, Design, and Development," p. 46, 2017.

[18] J. Hoare and T. Bock, "A brand's eye view of correspondence analysis," *Int. J. Mark. Res.*, vol. 61, no. 1, pp. 12–21, 2019.

[19] F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic Cosine Similarity," *Semant. Sch.*, vol. 2, no. 4, pp. 4–5, 2012.

[20] E. Garcia, "Cosine Similarity Tutorial," *Inf. Retr. iItelligence*, no. April, pp. 4–10, 2015.

[21] N. Akma, M. Hafiz, A. Zainal, M. Fairuz, and Z. Adnan, "Review of Chatbots Design Techniques," *Int. J. Comput. Appl.*, vol. 181, no. 8, pp. 7–10, 2018.

[22] K. Jayakodi, M. Bandara, I. Perera, and D. Meedeniya, "WordNet and cosine similarity based classifier of exam questions using bloom's taxonomy," *Int. J. Emerg. Technol. Learn.*, vol. 11, no. 4, pp. 142–149, 2016.