A Study of Automated Evaluation of Student's Examination Paper using Machine Learning Techniques

Ganga Sanuvala
Research Scholar, Department of CSE
UCE (A), Osmania University
Hyderabad, India

Syeda Sameen Fatima
Professor (Retd), Department of CSE
UCE (A),Osmania University
Hyderabad, India

2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 978-1-7281-8529-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCIS51004.2021.9397227

***Abstract*— The written exam is a universal tool for evaluating student performance in the field of education. The written exam provides a mechanism by which instructors and organizations ensure the consistency of the assessment process. Human effort required for the assessment is very high and it depends on several factors such as knowledge of the teacher, application level understanding of the teacher, criteria of the marking and time allotted. However, traditional evaluation processes consume very costly efforts and take huge time for the completion of the complete evaluation, verification and publishing of the result process. This research introduces the design and implementation of Handwritten Answer Evaluation (HAES) system for student exam papers. The HAES is an automatic response assessment system that enables the identification of text in answer sheets and can evaluate the grade of each answer based on the previous knowledge of the model. In this study, Optical Character Recognition (OCR) tool is used to extract the text from human written scanned answer script and machine learning/natural language processing (NLP) techniques are used for grading the answer sheets. The scores are based on cosine set similarity measures, where each sentence in the evaluated answer paper carries their corresponding mark. The developed model can be used to evaluate the marks of the unscored descriptive answers.**

Keywords*— OCR tool; Cosine Similarity; Examination; Handwritten Answer Evaluation System; Natural Language Processing; Verification of Results.

I. Introduction

Writing is precarious to learning and academic victory, and students who cannot write well are deprived of paid jobs and often find it difficult to complete college courses [1]. Students' writing quality or ability can be improved by getting feedback from the teacher. In recent years, automated essay/exam scoring has become a hot issue in the research of natural language processing with the growing need of essay/exam scoring in English writing skill[2]. In many cases, the essay/exam scoring task costs huge human resources but with less efficiency and the score given by human evaluator is mostly determined by his knowledge and energy [2]. Thus, the correctness of essay/exam scoring cannot be guaranteed. So answer paper evaluation and scoring can be tedious and time consuming process for many of evaluators. For many teachers to finish the essay/exam scoring of all student essays/exam papers in a short time interval is the major challenging task. Thus, in order to solve these issues, the researchers have been proposed automated essay/exam scoring techniques [3].

The goal of Machine Learning (ML) is to program computers to use data samples or previous experience to solve specific problems. Image detection, education, computer vision, bioinformatics, etc. are some of the areas in which ML can be applied [4]. This study emphasizes the use of ML, because education is changing every day. For example, ML is used to support teachers, measure student performance, test student's knowledge, and so on. In this study, a scheme to predict the score of the student's paper by using ML techniques is presented. The evaluated answer text files are used during training to develop a model [5]. The human evaluator's answer key text file is also used during training and each sentence in the text has its corresponding mark. The semantic similarity between the sentences can be calculated by a cosine similarity based approach [6]. During testing each unscored answer text files can be used as input to the developed model.

This research paper is organized as follows: Section 2 provides the evaluation system model; Section 3 describes the study of existing techniques for evaluations. The challenges presented in the research study and objective of the study is presented in Section 4 and 5. Finally, the conclusion of the study is illustrated in Section 6.

II. Literature Review

Kashi et al. [7] presented a basic version of an automated rating system that provides recommended ratings for descriptive answers. The assessment recommendation was based on an analysis of the student's response and the answer key, which includes some key phrases and solutions. The similarity points were then combined using ML to get recommended points for the answer. According to the training data, the weights of each model were learned by the system to map the score by comparing the answers of the students with actual answers given by the graders. . The proposed system works well for responses with fewer deviations from the expectations of evaluators, but requires more precise coordination to handle more application-based responses. Another drawback of this approach was that there was less training data in each language model.

Saha et al. [8] proposed a practical system to evaluate long or descriptive answers that can be evaluated in a small class scenario. The proposed system was not based on large training data and it uses a link or answer template instead. The system uses a written response from an expert to calculate the similarity of the student's response. Various similarity measures were used to calculate the similarity at word and phrase level, including TFIDF, hidden semantic indexing,

ISBN: 978-1-7281-8529-3/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on May 31,2021 at 14:40:55 UTC from IEEE Xplore. Restrictions apply.

LDA, TextRank adder and InferSent based on the integration of neuronal phrases. The student's answer may contain some facts that are not included in the model's answer. The system identifies these suggestions, checks their relevance and correctness to accordingly assign additional evaluations. In the final phase, the system uses a cluster-based reliability analysis was tested for the evaluation of textbooks with answers in the social sciences. But, the system generates an accurate estimate only for sure answers.

V. Nandini, and P. U. Maheswari, [9] implemented an automatic evaluation system called syntactical relation-based feature extraction technique for evaluating descriptive type of answers. This model was handled by the supervised learning environment. A new classification algorithm was proposed to classify students' questions and answers in the exam. The classification of answers ensures that students who respond in style fall into the exact category of questions asked. If the expected answer is descriptive, the student must meet the requirements; otherwise, if the answer is contradictory, it is reflected in the assessment, which clearly reflects the knowledge of the students in the context of the particular subject. The implementation showed that the developed feature extraction technique achieved 95% precision, 94% recall and 94.5% sensitivity. The method didn't consider the analysis of grammar, which played a major role in conveying the meaning of a text.

Vij, et al. [10] developed a ML-based approach using Wordnet chart that looks for the similarity between the answers given by the student and the ideal answers provided by the teacher for the ease of automating the answer sheet. This research study was the first work in the field of short answer-based evaluation using WordNet charts. This algorithm takes into account the semantic relationship of the response text. The root mean square error of the method was 0.319, when the experiments were tested on 400 answer sheets. In addition, the developed method generates Wordnet graphics and automatically assigns scores, which helps reduce evaluation time. However, this task is only appropriate in situations where the student enters the spelling of the appropriate words. Because a WordNet number will not be created for the wrong words.

Alrehily et al. [11] proposed a new project for an electronic exam assessment system using the concepts of semantic similarity and document similarity to find the correspondence between the teacher's answer and the student's answer to each question. The evaluation system has four sections: pre-processing, keyword extension, matching, and evaluation. The system retrieved the score based on the parity of the percentages. The electronic grades are correlated with teacher grades using Spearman correlation. Thus the proposed electronic system increased time, cost, resources, efficiency and performance in setting and evaluating the test. However, this method does not eliminate syntactic errors in keywords and does not even study high-quality documentation.

Neethu et al. [12] development of a system of description of examination of answers and certification based on NLP and advanced training. The functions were removed to create a model from a human response script data set. The proposed sequential model consists of a layer of a recurrent long-term neural network (LSTM-RNN), which represents the glove vector in each sentence of the word and transforms it into a representation of the incorporation vector. The successive models consist of an embedding layer, an LSTM layer, a drop layer and a dense layer. The regularization techniques reduce overlap by preventing complex co-adaptations for these workouts. The Softmax activation function in a dense layer (a fully connected neural network layer) gave a coded estimate for each response. The model compares the answer key with the number of descriptive answers. This approach has been very useful for searching essays, descriptive scripts, checking the similarity of documents and search results. The method need to reduce the computation time of the LSTM-RNN model by introducing optimization techniques.

Sijimol and Varghese [13] Introduced project and implementation of automated systems to evaluate short answers using neural networks. The proposed system was an automated system to evaluate short answers, which was able to identify the text in the document containing the answers and evaluate the grade for each short answer based on previous knowledge gained by the model. In the proposed system, optical character recognition tools were used to extract handwritten letters. The NLP and neural networks were used to extract from a set of individual evaluated documents and feedback keys. The proposed model evaluated estimates based on indicators of similarity of cosine sentences. The dataset consists of the answer and its evaluated mark. The developed model was used to evaluate the marks of the unscored short answers. Due to the usage of neural network, the model has high computation time, which was not considered in this study.

Zupanc, and Bosnić, [14] proposes an extension of the automated essay assessment system, including additional semantic stability and consistency. The synthesis of the novel was developed by transforming successive portions of the essay into meaningful spaces and measuring changes in them. Consistent errors in information extraction and logical reasoning identify novel stability information. The resulting system (called SAJZ - Semantic Automatic Grader for Essays) gives the author meaningful feedback and a higher accuracy of evaluation compared to existing automated essay evaluation systems. The transformation of text into attributes were done by TF-IDF approach and WordNet only that didn't considered other approaches for unsupervised taxonomy learning. In addition, the method was unable to predict the implicit errors and facts/relations that are not explicitly written in an essay.

Devaki Priya et al. [15] proposed Optical Character Recognition Tool (OCR) that provides a system that converts a handwritten answer sheet into a text document with an image and stores it directly in the database. This method improves database security, detects word errors, compares sentence values, and evaluates ratings using NLP methods. The length of response in the project, the presence of the keyword, and the context of the keyword are related to the same factors considered by the actual person. The students

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on May 31,2021 at 14:40:55 UTC from IEEE Xplore. Restrictions apply.

will have a certain amount of freedom to write answers when checking the system for keywords, synonyms, word-of-mouth and coverage of all ideas. This procedure is performed only to assess the situation of subjective responses without diagrams or equations.

Shashavali et al. [16] implemented a methodology that derives Sentence Similarity score based on N-gram and Sliding Window and uses the Fast Text Word Embedding technique which outperforms the existing sentence similarity results. The method used the collected data on the shopping domain to build conversational agents. The extensive experiments done on the fetched dataset and achieved better results in accuracy, precision and recall by 98.80%, 92.26% and 94.08% respectively. It also evinces that the developed solution generalizes well on the low corpus and requires no training. The important parameters such as dependency and constituency parsing information to improve the word representation were not considered in this study.

Atoum, and Otoom [17] used the novel hybrid word similarity when measuring text similarity measuring (TSM). The TSM-specific information was based on the relationship between content and order and it includes the exact word match, the length of two sentences in pairs, and the maximum similarity between a word and a comparable sentence. The similarity index shows significant differences between similarity levels, which are significant at 0.05. The reason for the high efficiency of hybrid method was the use of additional information (corpus and information content) and the effectiveness of measuring the similarity of a borrowed word. However, the proposed method has fewer tools and fewer sources of information.

Hassan et al. [18] used two contributions, namely the Babelnet, a new synset-oriented word aligner based on a multilingual semantic network. Second, three unsupervised STS (USTS) approaches have been proposed: string kernel-based (SK), deployment-based, and loaded alignment-based and the USTS includes the advantages of four similar measures: specific alignment-based, surface-based, body-based, extended machining spacing. Experimental results proved that the proposed alignment role was effective in STS. Soft cardinality levels cannot be applied to multilingual STS functions because they depend on the surface coverage between the two texts, and not every text pair with a different language has common characters.

David Becerra-Alonso et al [19] developed EduZinc which is a tool which enables teachers to go through the complete process of creating and evaluating the activities and materials of a course. This model enables managing two teaching related topics-Firstly, the creation of individual products like activities, exams and test. Secondly it does the automatic grading for all the learning products, automatic creation of students, classes, and reports. This model also has the facility of creating automatic warnings, in case the students fall behind and in case students excel, they are also notified the same. However, the developed model was not related to classification technology with the above mentioned student's final marks which lowered the accuracy performance.

III. TAXONOMY FOR EVALUATION OF STUDENT'S EXAM PAPERS

Many schemes and methods are currently available for evaluation of essays [20]. But these methods cannot be adopted for descriptive answers. The proposed assessment system consists of three modules, the first part is the scanning phase. The scanning phase scans the document and identifies and extracts student answers using OCR tool. The created text file (dataset) consists of identified answers and its human evaluated score. The answers are given as input to the pre-processing phase which is the NLP part that enables to filter out the essential required parts from the answer sheets. The second part is the learning part which consists of training and testing. After the creation of the trained model, user of the system is able to provide the unscored answers in the form of text file to grade marks.

Training deals with creating a model by learning knowledge from the scored answers dataset and the answer key. Testing deals with the scoring of unscored answers based on the learned data in the trained model. The evaluation system makes the system capable of efficiently evaluating answer papers based on an answer key and a sample data set of scored answers with reduced human effort and reduced cost in a very short span of time under the right supervision of an evaluator. The proposed system accepts answer paper in pdf format. Then, the text is extracted from the pdf file using OCR tool. Next, the relevant feature is extracted from the answers and the created model is used to evaluate the marks of the unscored answers. The proposed assessment system extracts the semantics to efficiently represent the text in answers. A model is developed from answer key as well as scored answers to grade answers. The figure 1 presents the basic architecture of the planned system.

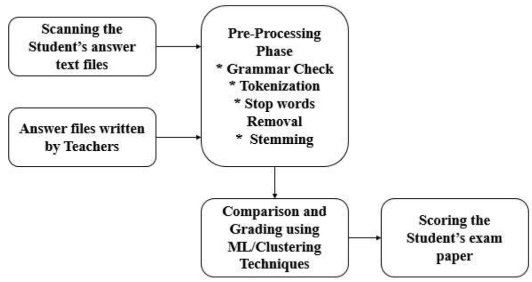

Figure 1: General Workflow to evaluate the Student's Exam paper

The proposed scheme consists of Pre-processing, Comparison and Scoring. The first part of learning is the training which is used to create the trained model. The scored answers and high weightage given answer key is taken as the input for the training phase. The pre-processing extracts important features of the input. Next, each pre-processed answer and key is mapped into a vector space based on term frequency–inverse document frequency (TF_IDF) [21] score and cosine similarity score between the produced word vector in the answer and word vectors in the key is calculated.

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on May 31,2021 at 14:40:55 UTC from IEEE Xplore. Restrictions apply.

The testing deals with the scoring of unscored answers based on the learned data in the trained model. The unscored answers are converted into the TF_IDF vectors and cosine based similarity matching is performed based on the trained model. The scores corresponding to the most similar sentences to the sentences in the unscored answers are further used for grading.

NLP is used to extract the semantics of the answer sheets as a pre-processing step. The output of the pre-processing step is a set of unique words corresponding to each sentence in the answer. NLTK is a NLP Toolkit that has a wide collection of predefined libraries for NLP. Following are the key steps used in the proposed system of text evaluation in digitized descriptive answer. The pre-processing consists of grammar check, tokenization, stop words removal, synonym and antonym checking, stemming. The first step is the tokenization where the process will convert a sentence into a set of words. After that we can ignore some of the commonly used words, the stop words such as is, was and so on. The stop words are words that have no meaning. The text file defines a set of stop words loaded into the "stopwords.txt" program. After loading, it is compared to the words in the list. If a match is found, the word in this list will be deleted. The next step is antonym checking. The student may write answers with sentences that have negative meaning. Those negated words should be identified and can be expanded and added to the set of words. For extracting the antonyms, WordNet can be used and the next step towards pre-processing is stemming.

In order to capture the similarity of words, the researchers normalized to a common root form - the stem and this process is known as stemming. For example: the word "writing", "written" and "written" is stemmed into its basic word called "write". Here, porter stemmer algorithm is the most widely used for stemming process. The duplicate words should be removed from the list in order to get a set of unique words corresponding to a sentence. For getting the vector representation of the sentence, TF_IDF calculation is used. TF_IDF is a numerical statistic that is used to find out the importance of a word to a document in a collection or corpus. The total number of times a term, t occurs in a document, d is called its term frequency IDF measures the information provided by a single word. A cosine based similarity checking algorithm is used to check the sentence similarity. A bag of words model is created based on similarity score between each key and every descriptive answer. During testing the unscored answers are converted to its vector representation and the bag of words model is used to find the most similar sentence vectors with its score. The extracted marks of the most similar sentences are added to get the final score of the answer. Equation 1 is used to measure the final score of the answer.

$$Sim(S_1, S_2) = {(\overrightarrow{S_1}.\overrightarrow{S_2})}\Big/{\|\overrightarrow{S_1}\|\|\overrightarrow{S_2}\|}$$

Where, vector representation of the sentences are defined as $\overrightarrow{S_1}\ and\ \overrightarrow{S_2}$. Similarity cosine measurements are used to find similarities between two non-zero vectors of the interior of the products that measure the cosine of the angle between them. Based on the parity value, the system calculates unpredictable response estimates. Moreover, the ML methods are also used to calculate the marks of the student's exam paper.

2.1 Prediction using Machine Learning Techniques

In recent years, the widespread use of ML to evaluate exam scenarios has been shown to be effective and accurate. Features are extracted from the data and then used as the basis for classification using a computer. The features are usually extracted at three stages: vocabulary, syntactic, and semantic. In this paper, lexical features are considered as sentence keywords, emotion words and all words. The keywords are retrieved using a graph-based ranking model, the Textrank [22] method for processing text and extracting keywords. The ML models are then selected according to the resilient characteristics mentioned above to assess student performance. In this study, students' textual work is divided into two main sections: high grade documents and low grade documents and the widely used models like Naive Bayes (NB) [23], Logistic Regression (LR) [24], Support Vector Machine (SVM) [25], and Gradient Boost Decision Tree (GBDT) [26] are considered to conduct experiment.

NB: This is a classical classification system based on the conditional assumption of the prejudice theorem and the freedom of characteristics. This model is used to discover the maximum power of the lateral potential for a given input x. The NB is easy to implement, because the learning efficiency and prediction rate are high. Shufen Ruan et al [27] developed a Class Specific Deep Feature Weighting for NB text classification. The developed model used the statistical feature weighting technique that used class specific deep feature weighting method for text classification. However, the developed model is to be tested for its effectiveness for structured data.

SVM: This is the binary classification model and its original model feature is a linear classifier with the largest distance defined in space. SVM consists of several kernel techniques, which basically make them a non-linear classification. It has many unique benefits over the problem of identifying small sample, linear, high-dimensional pattern, and other ML problems can also be applied. Ardy Wibowo Haryanto et al [28] developed an influence for Word Normalization and Chi-squared Feature Selection for Text classification using SVM. The developed SVM was used for analyzing text classification that obtained better results for lemmatization that was enhanced by Chi-squared (method 2). However, the developed model would have improved the performance by using other techniques such as data pre-processing, word normalization, feature selection, and other text classification algorithms.

LR: LR is a general linear model driven by a binary classification model based on the Sigmoid function. The ramp and sigmoid functions then convert the value to a floating-point number among 0 and 1. If it is greater than 0.5, it will be classified as a positive class and vice versa. Sujeevan Aseervathama et al [29] developed a Sparse Version of the

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on May 31,2021 at 14:40:55 UTC from IEEE Xplore. Restrictions apply.

Ridge Logistic Regression for Large-Scale Text Categorization. The developed model performed selection method which tried to approach the ridge solution by using solution. However, the developed model was in required of feature selection process as the model was in required of obtaining ridge solution by a sparse model.

GBDT: GBDT is an algorithm that classifies data using a combination model, which reduces the amount of fragmentation created by the training process. The main point of GBDT is that each tree gets to learn the findings of all previous trees. The remainder after adding the approximate values is the sum of the actual values. Kaiwei Yan [30] worked on Analysis and simulation of Multimedia English Auxiliary Handle Based on Decision Tree Algorithm. The developed model used improved decision tree algorithm that handled multimedia English assistance are parsed and simulated. The developed improved decision tree algorithm handled the multimedia English assistance that parsed and simulated. However, the developed model perceived the sense of language in English composition and still needed improvement for the rationality of intelligent evaluation

By using the ML techniques, the student's exam papers are evaluated and final scores are presented according to the similarity of the texts. The measure of correctness and accuracy of a student's response is calculated on the basis of the given categories:

1. Student answer's resemblance with the model answer
2. Answers containing only keywords
3. Answers whose word ordering is changed
4. Vague and contradictory answers

The similarity score is computed by evaluating and comparing the model against the answers fed in the system, namely model answers. The grammatical syntax, synonyms, and the context of the answers are also taken into consideration. Similarity measure is computed using Cosine similarity. The final score reviewed would comprise of all the above stated metrics.

IV. Challenges of the Research Study

There are many challenges acquired in the research study, which are stated as follows:

- Finding a suitable (dis)similarity degree among documents.
- Selecting right features of the documents for similarity measuring.
- Selecting proper ML techniques for utilizing the above (dis)similarity measure.
- Implementing the machine learning using an efficient and effective method.

V. Objective of the Research Study

The present study has the following objectives:

- To analyse the scores of the selected students' answer papers in Education at Higher Secondary, College and University levels.
- Design and implementation of multi-constraint optimization/ML algorithms for evaluation of student's answer paper.
- Development and implementation of grouped algorithms using similarity metrics to evaluate these short descriptive answers.

VI. Conclusion

The exams help to regularly evaluate students' skills, deliver regular feedback to students and regulate the effectiveness of training by monitoring student performance. There are many problems occurred in the research that are related to the manual evaluation scheme. Significant resources are needed and it is also a tedious and time consuming task. The purpose of this study is to appraise the current level of technology in the application of ML in the field of education. This paper presents a study of the existing ML techniques for evaluation of student's examination paper and also proposes an approach for the design and implementation of automated handwritten answer evaluation system using machine learning. The scored answer paper dataset and high weightage given answer sheets are used to train the model. The unscored answer papers are tested by the model. . If the answer written by the student falls into the category of the question asked, it turns out that the student knows little how to express the topic. The future scope in this project is to reduce the computation time by introducing hashing techniques into this system.

References

[1] [1] Wang, E.L., Matsumura, L.C., Correnti, R., Litman, D., Zhang, H., Howe, E., Magooda, A. and Quintana, R., 2020. eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing*, p.100449.

[2] Mohsen, M.A. and Abdulaziz, A., 2019. The Effectiveness of using a Hybrid Mode of Automated Writing Evaluation Systen on EFL Students Writing. *Teaching English with Technology*, *19*(1), pp.118-131.

[3] Galhardi, L.B. and Brancher, J.D., 2018, November. Machine learning approach for automatic short answer grading: A systematic review. In *Ibero-american conference on artificial intelligence* (pp. 380-391). Springer, Cham.

[4] Kučak, D., Juričić, V. and Đambić, G., 2018. Machine Learning In Education -A Survey of Current Research Trends. *Annals of DAAAM & Proceedings*, *29*.

[5] Choudhari, S., Moray, S., Salunke, N. and Dhawale, A., 2019. An Automated Approach to Conduct Pune University's In-sem Examination.

[6] Al-Anzi, F.S. and AbuZeina, D., 2017. Toward an enhanced Arabic text classification using cosine similarity and Latent Semantic Indexing. *Journal of King Saud University-Computer and Information Sciences*, *29*(2), pp.189-195.

[7] Kashi, A., Shastri, S., Deshpande, A.R., Doreswamy, J. and Srinivasa, G., 2016, December. A score recommendation system towards automating assessment in professional courses. In *2016 IEEE Eighth International Conference on Technology for Education (T4E)* (pp. 140-143). IEEE.

[8] Saha, S.K. and Rao CH, D., 2019. Development of a practical system for computerized evaluation of descriptive answers of middle school level students. *Interactive Learning Environments*, pp.1-14.

[9] Nandini, V. and Maheswari, P.U., 2018. Automatic assessment of descriptive answers in online examination system using semantic relational features. *The Journal of Supercomputing*, pp.1-19.

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on May 31,2021 at 14:40:55 UTC from IEEE Xplore. Restrictions apply.

[10] Vij, S., Tayal, D. and Jain, A., 2020. A Machine Learning Approach for Automated Evaluation of Short Answers Using Text Similarity Based on WordNet Graphs. *Wireless Personal Communications*, *111*(2), pp.1271-1282.

[11] Alrehily, A.D., Siddiqui, M.A. and Buhari, S.M., 2018. Intelligent Electronic Assessment for Subjective Exams. *ACSIT, ICITE, SIPM*, pp.47-63.

[12] Neethu George, Sijimol PJ, Surekha Mariam Varghese, "Grading Descriptive Answer Scripts Using Deep Learning", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Volume-8 Issue-5 March, 2019.

[13] Sijimol, P.J. and Varghese, S.M., 2018. Short Answer Scoring System Using Neural Networks. *International Journal of Computer & Mathematical Sciences IJCMS ISSN*, *7*, pp.2347-8527.

[14] Zupanc, K. and Bosnić, Z., 2017. Automated essay evaluation with semantic analysis. *Knowledge-Based Systems*, *120*, pp.118-132.

[15] Devaki Priya, Harini S, Haripriyaa A and Dharaniya R, "A Novel Approach – Automatic paper evaluation system", International Research Journal of Engineering and Technology (IRJET), Vol. 06, no. 03, 2019.

[16] Shashavali, D., Vishwjeet, V., Kumar, R., Mathur, G., Nihal, N., Mukherjee, S. and Venkanagouda Patil, S., 2019. Sentence Similarity Techniques for Short vs Variable Length Text using Word Embeddings. *Computación y Sistemas*, *23*(3).

[17] Atoum, I. and Otoom, A., 2016. Efficient hybrid semantic text similarity using WordNet and a corpus. *Int. J. Adv. Comput. Sci. Appl*, *7*(9), pp.124-130.

[18] Hassan, B., Abdelrahman, S.E., Bahgat, R. and Farag, I., 2019. UESTS: An unsupervised ensemble semantic textual similarity method. *IEEE Access*, *7*, pp.85462-85482.

[19] Becerra-Alonso, D., Lopez-Cobo, I., Gómez-Rey, P., Fernández-Navarro, F. and Barbera, E., 2020. EduZinc: a tool for the creation and assessment of student learning activities in complex open, online, and flexible learning environments. Distance Education, 41(1), pp.86-105.

[20] Wilson, J. and Czik, A., 2016. Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education*, *100*, pp.94-109.

[21] Christian, H., Agus, M.P. and Suhartono, D., 2016. Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, *7*(4), pp.285-294.

[22] Mihalcea, R. and Tarau, P., 2004, July. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411).

[23] Rish, I., 2001, August. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).

[24] Couronné, R., Probst, P. and Boulesteix, A.L., 2018. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC bioinformatics*, *19*(1), p.270.

[25] Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y. and Xu, W., 2018. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics-Proteomics*, *15*(1), pp.41-51.

[26] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp.1189-1232.

[27] Ruan, S., Li, H., Li, C. and Song, K., 2020. Class-Specific Deep Feature Weighting for Naïve Bayes Text Classifiers. *IEEE Access*, *8*, pp.20151-20159.

[28] Haryanto, A.W. and Mawardi, E.K., 2018, September. Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine (SVM) Text Classification. In *2018 International Seminar on Application for Technology of Information and Communication* (pp. 229-233). IEEE.

[29] Aseervatham, S., Antoniadis, A., Gaussier, É., Burlet, M. and Denneulin, Y., 2011. A sparse version of the ridge logistic regression for large-scale text categorization. *Pattern Recognition Letters*, *32*(2), pp.101-106.

[30] Yan, K., 2020. Analysis and Simulation of Multimedia English Auxiliary Handle Based on Decision Tree Algorithm. *Complexity*, *2020*.

Authorized licensed use limited to: University of Prince Edward Island. Downloaded on May 31,2021 at 14:40:55 UTC from IEEE Xplore. Restrictions apply.