iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45755

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 VII July 2022- Available at www.ijraset.com

IBM Watson Studio for Building an Automated Essay Grading System

Santosh K C[1], Vinayak Kyatnatti[2], Ankith G D[3], Jagadish[4], Nikhil H N[5]
[1, 2, 3, 4, 5]*Department of Computer Science and Engineering, BIET Davanagere*

***Abstract*: Essays are one of the most important methods for assessing learning and intelligence of a student. Manual essay grading is a time-consuming process for the evaluator, a solution to such problem is to make evaluation through computers. Aim of this paper is to understand and analyze current essay grading systems and compare them primarily focusing on technique used, performance and focused attributes.**

***Keywords*: LSTM, IBM Watson and RNN**

I. INTRODUCTION

Essays are crucial testing tools for assessing academic achievement, integration of ideas and ability to recall, but are expensive and time-consuming to grade manually. Manual grading of essays takes up a number of instructors' valuable time and hence is an expensive process. Automated grading, if proven to match or exceed the reliability of human graders, will reduce costs. Currently, automated grading is used instead of second graders in some high-stakes applications, and as the only grading scheme in low stakes evaluation. This application can have a high utility in many places. For instance, currently, evaluation of essay writing sections in exams like GRE, GMAT, and TOEFL is done manually. And, so automating such a system may prove to be highly useful.

An automated grading system is built with the magical powers of neural networks. Using automation reduces time and effort in evaluation. NLTK libraries for feature extraction and LSTM are used for the learning process.

The Automated Essay Grading System is a computer-based evaluation of student responses. Now a day's education system is shifted to the online based education system and assessment also automatically. Till now more than 15 essay grading systems were there and their working on two categories like content-based and style-based assessment few are on both by considering features like number of sentences, number of words, parts of speech and grammar, etc. And remaining is considering features like semantic analysis, using natural language tools grading the essays.

A. Problem Statement

The AI graphical tool in Watson Studio automatically analyzes data and generates candidate model pipelines customized for predictive modeling problem. These model pipelines are created iteratively as AI analyzes dataset and discovers data transformations, algorithms, and parameter settings that work best for problem setting. Results are displayed on a leader board, showing the automatically generated model pipelines ranked according to problem optimization objective.

B. Objectives

The Objectives of our project are:

1) You'll be able to understand the text and process it.
2) You will be able to extract important features from the text.
3) You will be able to understand RNN and LSTM working principles.
4) You will be able to know how to pre-process/clean the text using different data pre-processing techniques.
5) You will be able to know how to find the accuracy of the model.
6) You will be able to build web applications using the Flask framework.

C. Proposed System

The system will have the capability of identifying the local languages (Indian) present in the submitted essay and it will also find out how much effect is there for these words. It will also help the students to resubmit the essay with corrections where the students will be asked to re-enter the similar words for those local languages. It will also give a proper scorecard mentioning that still how much the essay is influenced by the local languages and the no of local words present, number of times corrections made by students.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

1) Poornima M, Mrs Shibily Joseph "A Survey on Automated Essay Scoring (AES)" (2019). Automated Essay Scoring (AES) is a challenging task which assigns grades to essays written in an educational institute. It reduces human errors, inequality problem, time consuming and so on. There are diverse approaches for this task like natural language processing, machine learning, deep learning etc. The overall performance of such systems is tightly bound to high-quality features. The essential purpose of this paper is to evaluate these strategies of AES in both in-domain and cross domain settings.

2) Kshitiz Srivastava, Namrata Dhanda and Anurag Shrivastava "An Analysis of Automated Essay Grading Systems" (2020). Essays are one of the most important methods for assessing learning and intelligence of a student. Manual essay grading is a time-consuming process for the evaluator, a solution to such problem is to make evaluation through computers. Many systems were proposed over past few decades. Each system works on different approach having focus on different attributes. Aim of this paper is to understand and analyze current essay grading systems and compare them primarily focusing on technique used, performance and focused attributes.

3) Mohamed Abdellatif Hussein, Hesham Hassan2 and Mohammad Nassef "Automated language essay scoring systems: a literature review" (2019). Writing composition is a significant factor for measuring test-takers' ability in any language exam. However, the assessment (scoring) of these writing compositions or essays is a very challenging process in terms of reliability and time. The need for objective and quick scores has raised the need for a computer system that can automatically grade essay questions targeting specific prompts. Automated Essay Scoring (AES) systems are used to overcome the challenges of scoring writing tasks by using Natural Language Processing (NLP) and machine learning techniques. The purpose of this paper is to review the literature for the AES systems used for grading the essay questions.

4) Dadi Ramesh, Suresh Kumar Sanampudi "An automated essay scoring systems: a systematic literature review" (2021). Assessment in the Education system plays a signifcant role in judging student performance. The present evaluation system is through human assessment. As the number of teachers' student ratio is gradually increasing, the manual evaluation process becomes complicated. The drawback of manual evaluation is that it is time-consuming, lack's reliability, and many more. This connection online examination system evolved as an alternative tool for pen and paper-based methods. Present Computer-based evaluation system works only for multiple-choice questions, but there is no proper evaluation system for grading essays and short answers. Many researchers are working on automated essay grading and short answer scoring for the last few decades, but assessing an essay by considering all parameters like the relevance of the content to the prompt, development of ideas, Cohesion, and Coherence is a big challenge till now. Few researchers focused on Content-based evaluation, while many of them addressed style-based assessment. This paper provides a systematic literature review on automated essay scoring systems. We studied the Artifcial Intelligence and Machine Learning techniques used to evaluate automatic essay scoring and analyzed the limitations of the current studies and research trends. We observed that the essay evaluation is not done based on the relevance of the content and coherence.

5) Hesham Ahmed Hassan, Mohamed Abdel-latif Hussein, Mohammad Nassef "Automated language essay scoring systems: a literature review" (2019). Background Writing composition is a significant factor for measuring test-takers' ability in any language exam. However, the assessment (scoring) of these writing compositions or essays is a very challenging process in terms of reliability and time. The need for objective and quick scores has raised the need for a computer system that can automatically grade essay questions targeting specific prompts. Automated Essay Scoring (AES) systems are used to overcome the challenges of scoring writing tasks by using Natural Language Processing (NLP) and machine learning techniques. The purpose of this paper is to review the literature for the AES systems used for grading the essay questions. Methodology We have reviewed the existing literature using Google Scholar, EBSCO and ERIC to search for the terms "AES", "Automated Essay Scoring", "Automated Essay Grading", or "Automatic Essay" for essays written in English language. Two categories have been identified: handcrafted features and automatically featured AES systems. The systems of the former category are closely bonded to the quality of the designed features. On the other hand, the systems of the latter category are based on the automatic learning of the features and relations between an essay and its score without any handcrafted features. We reviewed the systems of the two categories in terms of system primary focus, technique(s) used in the system, the need for training data, instructional application (feedback system), and the correlation between e-scores and human scores.

6) Mohamed Abdellatif Hussein,corresponding author Hesham Hassan and Mohammad Nassef "Automated language essay scoring systems: a literature review" (2019). Writing composition is a significant factor for measuring test-takers' ability in any language exam. However, the assessment (scoring) of these writing compositions or essays is a very challenging process in terms of reliability and time. The need for objective and quick scores has raised the need for a computer system that can

automatically grade essay questions targeting specific prompts. Automated Essay Scoring (AES) systems are used to overcome the challenges of scoring writing tasks by using Natural Language Processing (NLP) and machine learning techniques. The purpose of this paper is to review the literature for the AES systems used for grading the essay questions.

7) Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., & Harris, M. "Automated Scoring Using A Hybrid Feature Identification Technique" (1998). This study exploits statistical redundancy inherent in natural language to automatically predict scores for essays. We use a hybrid feature identification method, including syntactic structure analysis, rhetorical structure analysis, and topical analysis, to score essay responses from test-takers of the Graduate Management Admissions Test (GMAT) and the Test of Written English (TWE). For each essay question, a stepwise linear regression analysis is run on a training set (sample of human scored essay responses) to extract a weighted set of predictive features for each test question. Score prediction for cross-validation sets is calculated from the set of predictive features. Exact or adjacent agreement between the Electronic Essay Rater (e-rater) score predictions and human rater scores ranged from 87% to 94% across the 15 test questions.

8) *Dadi Ramesh and Suresh Kumar Sanampudi* "An automated essay scoring systems: a systematic literature review" (2022). Assessment in the Education system plays a significant role in judging student performance. The present evaluation system is through human assessment. As the number of teachers' student ratio is gradually increasing, the manual evaluation process becomes complicated. The drawback of manual evaluation is that it is time-consuming, lacks reliability, and many more. This connection online examination system evolved as an alternative tool for pen and paper-based methods. Present Computer-based evaluation system works only for multiple-choice questions, but there is no proper evaluation system for grading essays and short answers. Many researchers are working on automated essay grading and short answer scoring for the last few decades, but assessing an essay by considering all parameters like the relevance of the content to the prompt, development of ideas, Cohesion, and Coherence is a big challenge till now. Few researchers focused on Content-based evaluation, while many of them addressed style-based assessment. This paper provides a systematic literature review on automated essay scoring systems. We studied the Artificial Intelligence and Machine Learning techniques used to evaluate automatic essay scoring and analysed the limitations of the current studies and research trends. We observed that the essay evaluation is not done based on the relevance of the content and coherence.

III. SYSTEM DESIGN

A. *System Architecture*

The figure 1 depicts the architectural diagram of proposed system. System designs the main aim of this structure incorporated in study can fetch out data from economic news and propose this sets into prognosticate model. Major phases in formulated system include data collection and pre-processing, feature and factor selection and price appraisal and prediction. In the initial hand, news, financial and market data are gathered and processed. In Further aspect, unstructured documents are modified into structured extract by classification. Data retrieval and pre-processing in data retrieval, datasets can be fetched such of news data, black gold price data and market data. Dataset from news can be retrieved through headlines as it is easier to obtain and justifies in one line. Factors that affect the reduction are expert business, stock market and later business. Sentimental Analysis In this era of modernization, big data is also assisting through study of sentiment analysis which focuses on retrieving data through news and proposing prediction model. In this kind of analysis dictionary-based approach is accounted to gather the data regarding markets and essential factors affecting it. In case of trend prediction, the sentiment and prediction models are considered as variables. Back Propagation Back-propagation is considered as an algorithm which can be used for the purpose of training feed forward neural networks for prognosticate learning model. This leads to the attainably use gradient methods to teach multi – layer networks, by modifying weights to minimum loss. The process fetches the inputs and outputs and modify its inner state that will be capable enough to calculate the output that will be very precise to the expected output. Back propagation can also be described as "backward propagation of errors." It is a natural function to teach artificial neural networks. The forecasting can be done with TensorFlow which is followed by getting the data, generating features, generating ML model, training the ML model and testing and predicting with ml model. RNN (Recurrent Neural Network) RNN are not same as feed-forward networks. To predict things, it uses internal memory. It is capable of doing things which humans can't do such as handwriting, speech recognition. Sequential information is added as input to these networks. We assume inputs are independent of each other which is a false assumption. We should know the previous words so that next words can be predicted without any trouble.

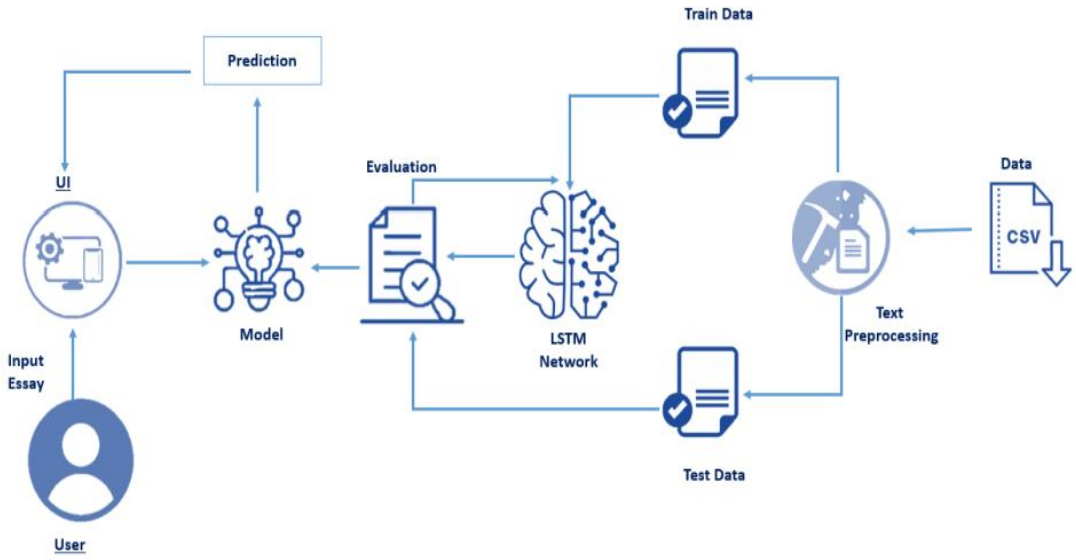


Fig 1 Architecture of proposed System

B. Flowchart

1) Data Collection - The process of gathering data depends on the type of project, for an ML project, Realtime data is used. The data set can be collected from various sources such as a file, database, sensor and other sources and some free data sets from internet can be used. Here we are working with the Box and Jenkins (1976) Airline Passengers dataset, which contains time series data on the monthly number of airline passengers between 1949 and 1960.

2) Data pre-processing - Data pre-processing is a process of cleaning the raw data i.e., the data is collected in the real world and is converted to a clean data set. There are certain steps executed to convert the data into a small clean data set and make it feasible for analysis, this part of the process is called as data pre-processing.

Most of the real-world data is messy, like:

- Missing Data
- Noisy Data
- Inconsistent Data

Some of the basic pre-processing techniques that can be used to convert raw data are:

- Conversion of Data
- Ignoring the missing values
- Filling the missing values
- Detection of outliers

3) Model Selection - Model selection is the process of selecting one final machine learning model from among a collection of candidate machine learning models for a training dataset. Model selection is a process that can be applied both across different types of models and across models of the same type configured with different model hyper parameters.

The types of classification models are:

- K-Nearest Neighbour
- Naive Bayes
- Decision Trees/Random Forest
- Support Vector Machine
- Logistic Regression

4) *Train and Test Data* - For training a model we initially split the model into 2 sections which are 'Training data' and 'Testing data'. The classifier is trained using 'training data set', and then tests the performance of classifier on unseen 'test data set'. Training set: The training set is the material through which the computer learns how to process information. Machine learning uses algorithms to perform the training part. Training data set is used for learning and to fit the parameters of the classifier. Test set: A set of unseen data used only to assess the performance of a fully-specified classifier.
5) *Evaluation* - Model Evaluation is an integral part of the model development process. It helps to find the best model that represents the data and how well the chosen model will work in the future. To improve the model hyper-parameters of the model can be tuned and the accuracy can be improved. Confusion matrix can be used to improve by increasing the number of true positives and true negatives. The output is predicted by analysing the test data as input along with test data output and then the output is displayed.
6) *Interface* - A web interface is built to take input and display an output. Flask language is used to build a web interface and pickle library is used to integrate both model and web page.

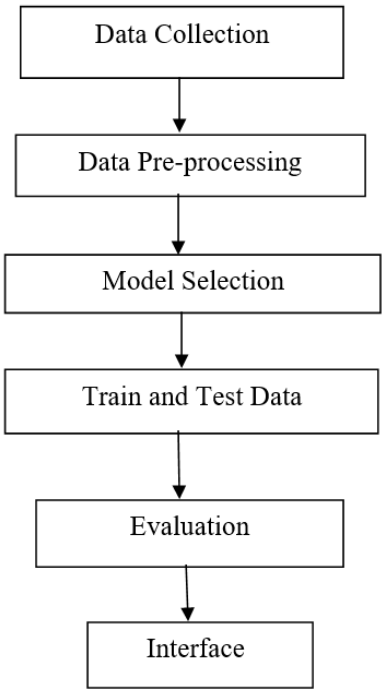


Fig. 2 Flowchart of Automated Essay Grading System

IV. IMPLEMENTATION

A. Algorithms

1) *Recurrent Neural Network (RNN):* Recurrent neural network are a type of Neural Network where the output from previous step is fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. In Recurrent Neural networks, the information cycles through a loop to the middle-hidden layer. The input layer 'x' takes in the input to the neural network and processes it and passes it onto the middle layer. The middle layer 'h' can consist of multiple hidden layers, each with its own activation functions and weights and biases. If you have a neural network where the various parameters of different hidden layers are not affected by the previous layer, the neural network does not have memory, then you can use a recurrent neural network. The Recurrent Neural Network will standardize the different activation functions and weights and biases so that each hidden layer has the same parameters. Then, instead of creating multiple hidden layers, it will create one and loop over it as many times as required. Figure 5 shows the working of RNN.

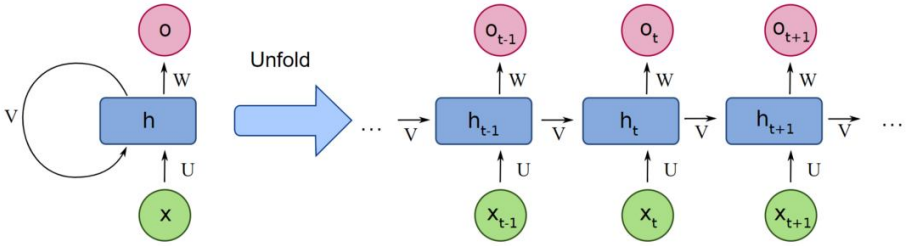
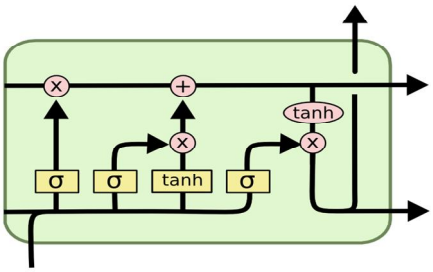


Fig. 3 Working of RNN

2) *Long short-term memory (LSTM):* LSTMs have chain like structure with the repeating module having a different structure. There are four neural network layers which are interacting to each other in a special way. Initially LSTM decides what information we're going to throw away from the cell state. This decision is made by a sigmoid layer called the "forget gate layer". Next it decides what new information we're going to store in the cell state as shown in fig 7. This has two parts. First, a sigmoid layer called the "input gate layer" decides which values are to be updated. Next, a tanh layer creates a vector of new candidate values, C̃t, that could be added to the state. Then it combines these two to create an update to the state. The old cell state, Ct−1, updates into the new cell state Ct. The old state is multiplied by ft, forgetting the things which were decided to forget earlier. Then we add it xC̃t. This is the new candidate values, scaled by how much it is decided to update each state value. This output will be based on the cell state, but will be a filtered version. First, a sigmoid layer runs which decides what parts of the cell state will be output. Then, the cell state through tanh (to push the values to be between −1 and 1) and multiply it by the output of the sigmoid gate to get desired output.



x_t	LSTM input
h_t	Current LSTM output
h_{t-1}	Previous LSTM output
C_t	Current candidate value
C_{t-1}	Previous candidate value
→	Vector Transfer
→	Concatenate
→	Copy
○	Point-wise Operation
▭	Neural Network Layer
σ	Sigmoid layer
tanh	tanh layer to create a vector of new candidate values

Fig. 4 RNN Single LSTM

V. RESULTS AND SCREENSHOTS

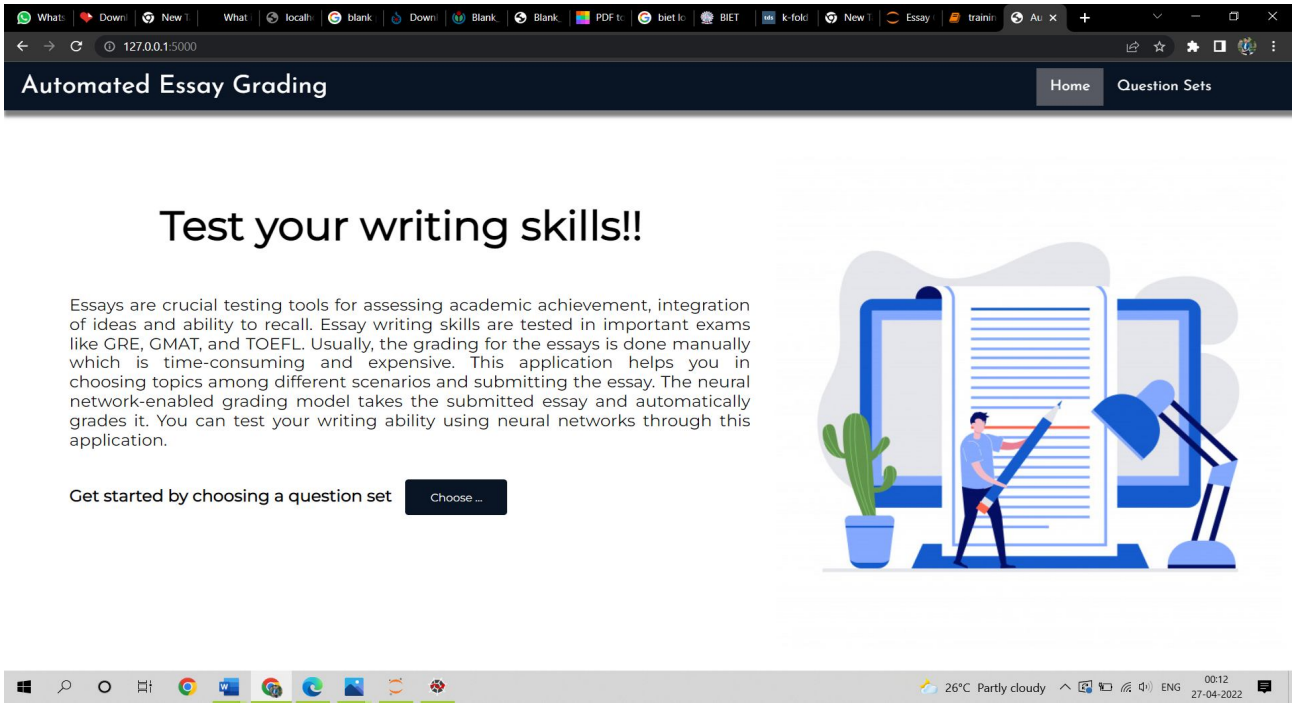


Fig 5 Home page

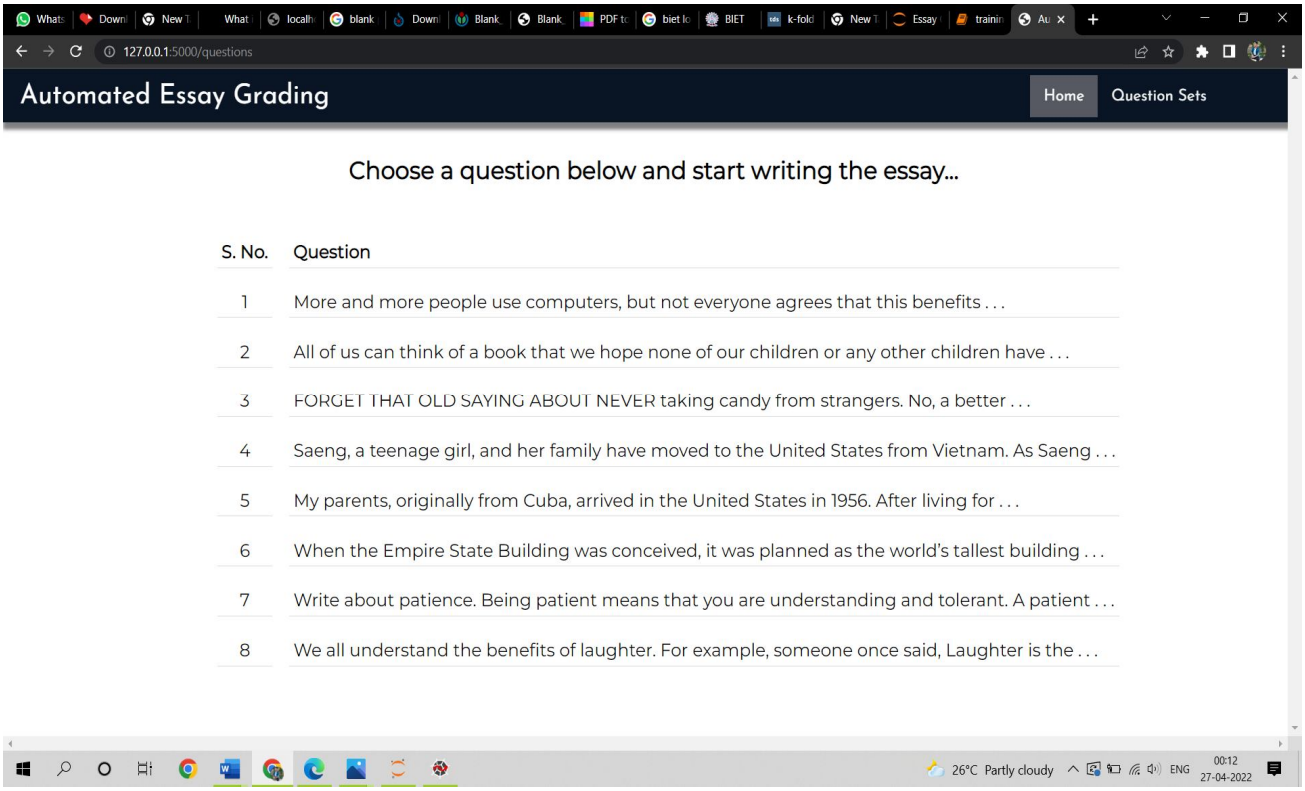


Fig 6 Set of questions

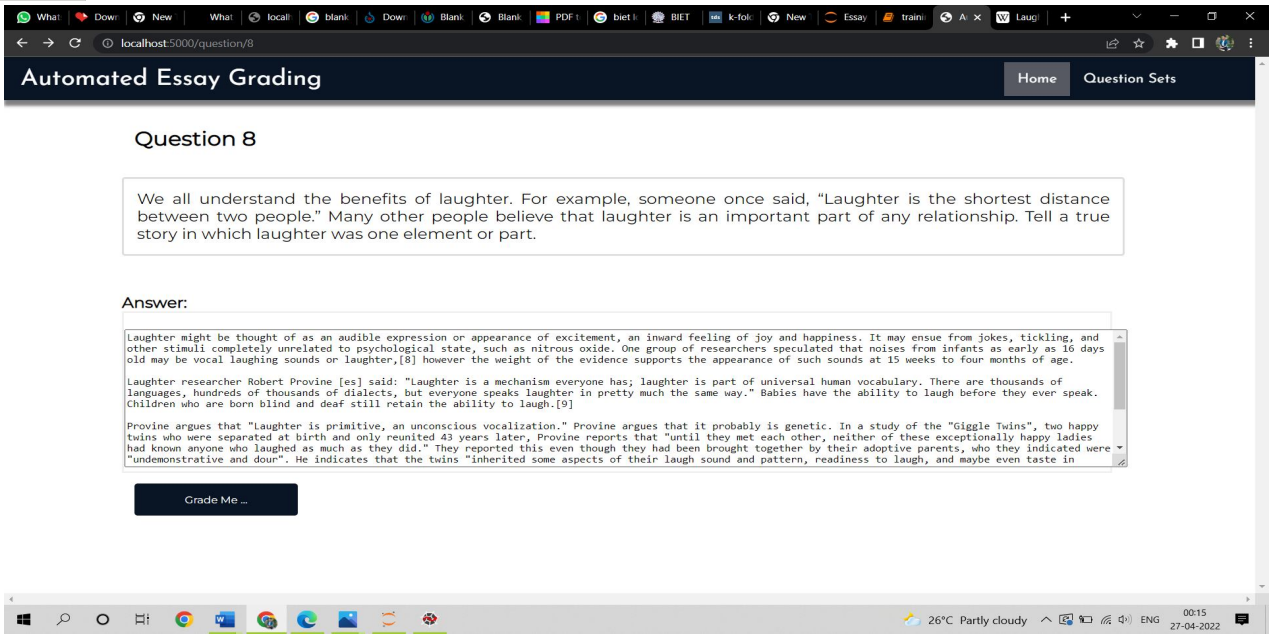


Fig 7 Essay description for the given question

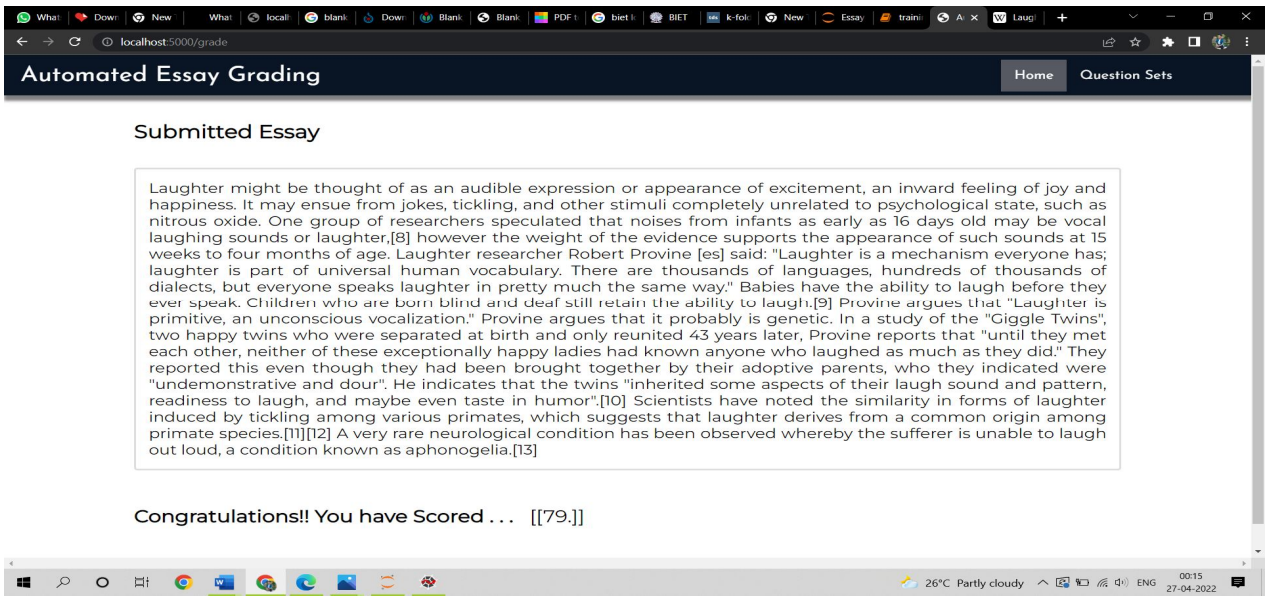


Fig 8 Grade for the submitted essay

VI. CONCLUSIONS

Automated essay scoring (AES) is a compelling topic in Learning Analytics for the primary reason that recent advances in AI find it as a good testbed to explore artificial supplementation of human creativity. However, a vast swath of research tackles AES only holistically; few have even developed AES models at the rubric level, the very first layer of explanation underlying the prediction of holistic scores. Consequently, the AES black box has remained impenetrable. Although several algorithms from Explainable Artificial Intelligence have recently been published, no research has yet investigated the role that these explanation models can play. Automated essay scoring systems basically find features in the essay and grade for them. We can divide these features in terms of three attributes namely: style, content and semantic. Most of the system proposed works on either style, content or both. In future we require systems that primarily focus on semantic attributes for essay grading, we can also have new semantic attributes that can help in evaluating essay more accurately. Current grading systems cannot detect correctness of the essay we can also have different ways to check the consistency of the essay.

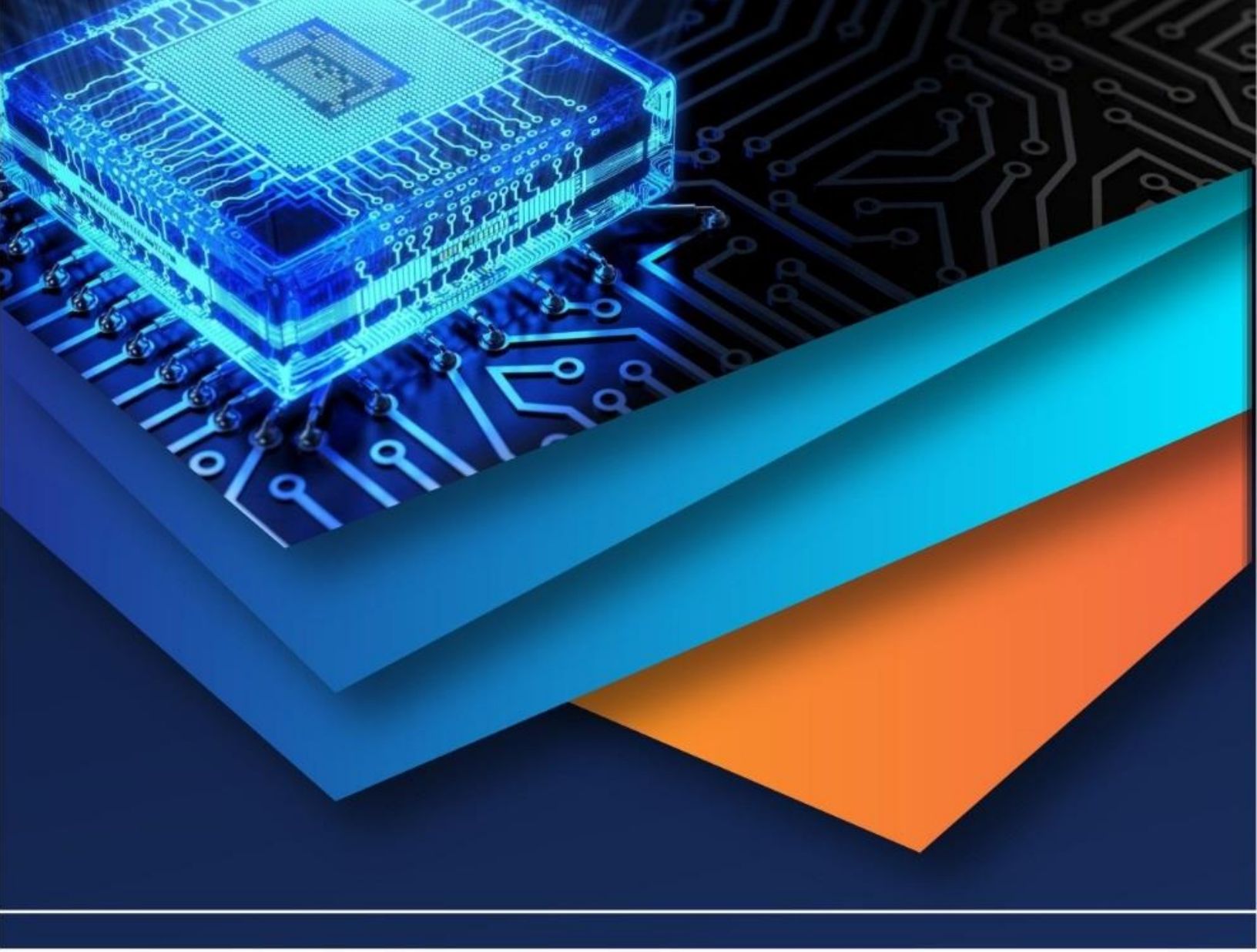
VII. FUTURE SCOPE

The future scope of the given problem can extend in various dimensions. One such area is to search and model good semantic and syntactic features. For this, various semantic parsers etc. can be used. Other area of focus can be to come up with a better tool than linear regression with polynomial basis functions like neural networks etc.

REFERENCES

[1] Poornima M, Mrs Shibily Joseph "A Survey on Automated Essay Scoring (AES)" (2019).
[2] Kshitiz Srivastava, Namrata Dhanda and Anurag Shrivastava "An Analysis of Automated Essay Grading Systems" (2020).
[3] Mohamed Abdellatif Hussein, Hesham Hassan2 and Mohammad Nassef "Automated language essay scoring systems: a literature review" (2019).
[4] Dadi Ramesh, Suresh Kumar Sanampudi "An automated essay scoring systems: a systematic literature review" (2021).
[5] Hesham Ahmed Hassan, Mohamed Abdel-latif Hussein, Mohammad Nassef "Automated language essay scoring systems: a literature review" (2019).
[6] Mohamed Abdellatif Hussein,corresponding author Hesham Hassan and Mohammad Nassef "Automated language essay scoring systems: a literature review" (2019).
[7] Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., & Harris, M. "Automated Scoring Using A Hybrid Feature Identification Technique" (1998).
[8] Dadi Ramesh and Suresh Kumar Sanampudi "An automated essay scoring systems: a systematic literature review" (2022).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)