# REAL –TIME AUTOMATED ESSAY EVALUATION SYSTEM FOR MULTIPLE USER

Prof. Afroze Ansari[1], Sana Naaz[2]

[1]Asst Professor, Dept. Of CSE, K.B.N College of Engineering, Kalaburagi, VTU Belagavi
[2]M.Tech (Student) Dept. Of CSE, K.B.N College of Engineering, Kalaburagi,VTU Belgavi,

*Abstract: Descriptive sort of answers or essay are the most ideal path for of surveying the student's brilliance alongside connecting the various thoughts with the capacity to review, however are tedious when they are evaluated physically. Yet, one of the obstacles in essay assessment contrasted with various decision assessments is exertion and sufficient measure of time in assessment which is totally the matter of movement to achieve. Automated scoring when got tied up with use won't just diminish the season of scoring the article yet will likewise make the score sensible when contrasted with the evaluator score. This paper centres around the current mechanized exposition scoring frameworks, what are the innovations behind them and proposes another framework over the current ones with new highlights. Author prepare classifiers on the preparation data set, let it go concluded the downloaded dataset, and after that portion execution our dataset by contrasting the acquired score and the dataset values. Author executed classic utilizing java. Author have the machine learning technique to build the application.*

*Keywords: Machine Learning, Automated scoring, Classifier, training data, descriptive answer.*

## I. INTRODUCTION

Type of test for estimating the student's capacity can be objective or subjective test. The upsides of applying subjective test amid the assessment are having the option to gauge the capacity of understudies in higher request thinking levels. The key objective of learning organizations is to deliver student with the assessment reports concerning their investigation as finest as likely with least errors. If it is matter of evaluating the multiple choice question and answer, then we have many systems already developed in the market which does its job very well. But the main problem is with scoring the answers when it is subjective type. Teacher or evaluator requires lot of time to score thousands of student's answers. Hand scoring the answer is time consuming and a hectic process. Subsequently in this advanced period, steps have been taken to try and create Automated scoring the subjective sort of answer moreover. Mechanized paper reviewing is not more a legend they are reality. As on today, the human written answers are amended not just by inspectors/educators likewise by machines. A framework for computerized evaluation would at any rate be reliable in the manner it marks papers, and expense and time reserve funds could be accomplished if the framework can be appeared to review expositions inside the scope of those granted by social evaluator.

## II. RELATED WORKS

**Project Essay Grade (PEG)** is one of the original and long duration executions of automated essay grading. This system was established by Page and others and initially trusts on style examination of external language features of block of text. Based on the writing quality an essay is predominately graded, without considering of content. Based on the idea of proxies the proposal of technique for PEG was implemented. i. e computer calculations or methods of trins, central variables of interest within the essay to motivate human ratter grading. Proxies content: [1] essay length to denote the trin of effortlessness, counts of arrangements, comparative pronouns and other portions of speech, as a pointer of difficulty of sentence arrangement.

**Intelligent Essay Assessor(IEA**) was established in the early nineties and is based on the Latent Semantic Analysis(LSA) method that was basically "planned for indexing documents and text retrieval. signifies documents and their word contented in a big two-dimensional matrix semantic space". [2] With the help of matrix algebra method recognized as Singular Value Decomposition(SVD), fresh interactions among words and documents are exposed and current connection are improved to extra correctly signify their accurate consequence.

A matrix signifies the words and their circumstances. Respective word being analysed signifies a row of in the matrix during respective column signifies the "sentence, paragraphs and other subdivisions" of the environment in which the word arises. The cells of the matrix comprise the occurrences of the words in respective background. "The initial matrix is then transformed according to an inverse document frequency weighting technique, well known method of the indexing and information recovery domain. The SVD is practical to the matrix to split it into three modules matrices that replicate the unique one if increased collected". [3] An essay to grade, a matrix for the essay document is constructed for the essay subject domain semantic space.

**Electronic Essay Rater(E-Rater)** was implemented by Burstein and others." E-Rater uses the MsNLP tool for parsing all sentences in essay-Rater uses a both statistical and NLP approach to retrieve linguistic features from the essays to be graded. A standard usual of human graded essays evaluated against the Essays.E-Rater is strong, well-documented argument structure and shows a variability of word use a syntactic structure will accept a score at the advanced end of a six-point scale. E-Rater features contains the examination of the dissertation structure, of the syntactic structure and the vocabulary practice". [4] To analysis of words of essay document, the autonomous module built to evaluate suing a technique called corpus-based approach. Presently unruffled of five modules to identify the features in the text was developed application and text merits stated in human reader scoring standards. Among five modules, three modules recognize features that may be recycled as scoring guide standards for the syntactic variability, the organization of concepts and the vocabulary practice of an essay. A fourth independent module is recycled to choice and weigh projecting features for essay scoring. Final, module is recycled to compute the ultimate score.

### III. DATASET

The dataset applied in this project has been removed from kaggle.com, it comprises of the information from the antagonism directed by the Hewlett Foundation. There are complete 13000 exposition set among which 80% is utilized for preparing and 20% for testing. Each paper is around 150 to 550 words long. All the answers are human reviewed and since the quantity of dataset is very huge so they have been isolated into 8 sets of expositions dependent on the kind of essay. We prepared our machine for one of the set which is essay on Computers.

### IV. METHODLOGY

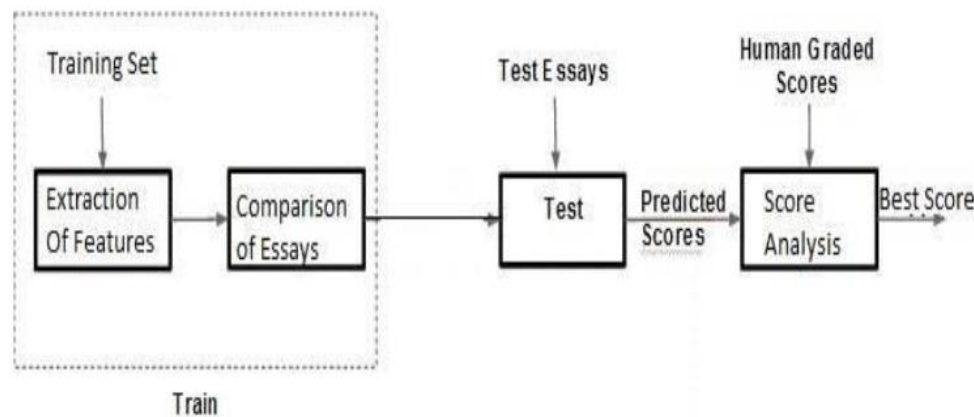The block diagram of the projected scheme is presented in below figure 1:



Figure 1: Architecture of proposed system

We have used Naive Bayes algorithm to train our machine for predicting the scores. To test the algorithm we have trained the machine with the training data taken from the kaggle.com which had over 13000 essay set on 8 different topics. We have used 1 set of essay that consisted of nearly 8000 essay and the trained machine. We have used Java programming language as the backend technology to build our application. We have expanded the classifier with three labels, for example 'extensive', 'partial', and 'unsatisfactory' to predict the score of the written answer. The substance in Bayesian paper scoring is fundamentally highlights of exposition, for example, (explicit words, phrases) and different qualities of essay like the request where definite noun verb word pair shows up or the request of the ideas clarified. We have also incorporated Word count and Sentence count into our system to optimize the predictions.

To achieve Real-time Answer scoring and to show the prediction of many test takers to the examiner in real-time we have used AngulaJS $http directive to send the essay in real-time to server for evaluation. This process is repeated every time when the test taken either hits enter key or completes a sentence using a '.' full stop.

The results in real-time are also shown in a very attractive way, test taker who has got 'unsatisfactory' result prediction is shown in Red colour, 'partial' as prediction is shown in 'gray' colour and 'extensive' as predicted result is shown in green colour.

Even a Real-time Line chart is available for examiner to view the test writing trend of each test taker. We have also considered test cases were in test taker might repeat same sentences multiple times to increase the length of essay and fool the machine. For every repeated sentence 1 mark is deducted from the predicted score.
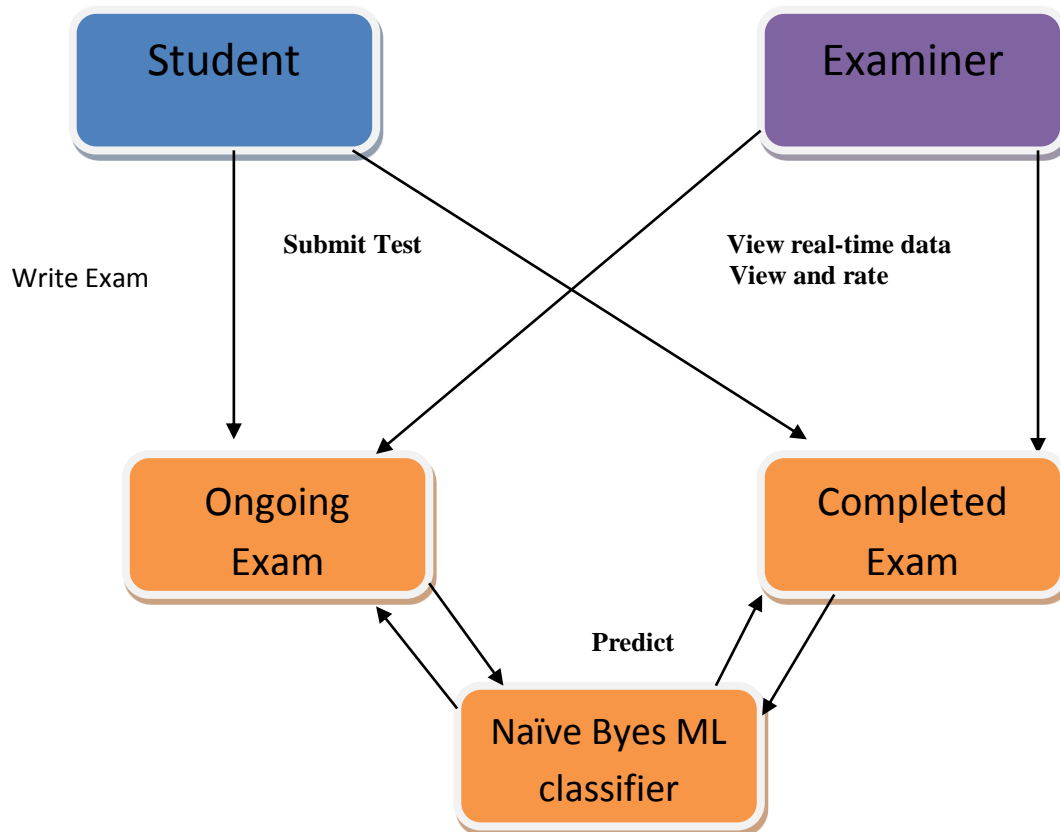
## V. IMPLEMENTATION



**Figure 2: Dataflow diagram of proposed system.**

The implementation of the proposed scheme is distributed into below mentioned three modules:
MODULES:
➢ Student module
➢ Examiner module
➢ Prediction module(Naive bayes algorithm)

These modules are implemented as follows:
➢ **Student Module:**
To get the students registered with application, a responsive registration and login form is designed. Once login, the student can access the real-time automated system. Database used to store the record is "MySQL". After login the student enters into the profile page where he can choose the exam to write the text. After choosing the exam he enters into the text area provided to write the answer. One feature added to the design of text area page is that as the student choose the exam and enters into text area this page is made automatically full screen so that student cannot navigate from that page anywhere and copy the answers. While writing the essay the student record is created in the "ongoing exam table" created in MySQL. Once the student submits the answer he is navigated back to his profile page and his record from "ongoing exam table" is deleted and created in the "completed exam table". After this student is navigated to his profile page and select the option of view result and can view the result in real time.

➢ **Examiner module:**
Admin of the application creates the record of examiner and provide the credentials to them so that no other party can claim as the examiner as the application is web based. Examiner now login and enter into the dashboard where he can view multiple students taking exams. Examiner now can choose the exam for which they want to view real-time data and scoring of all the students taking the test. To make application friendlier we have added a live chart. Examiner View a line-chart which shows the progress of student in real-time. The chart contains the text written by the student on the x-axis and the predicted score on the y-axis.
We cannot complete rely on the machine for prediction because machine is not always 100% accurate. So to be fair enough with score for students answers we have provided one more option here that is if at all after the machine giving the score, the examiner finds that the student deserves more marks, then examiner can provide his own marks too.

Finally, the examiner can view all the completed exams and view the predicted score and complete essay written by student and score accordingly.

➢ **Prediction module ( Naive bayes algorithm)**

This module is responsible for prediction the test scores using the naïve bayes classifier. In our Auto Score dataset which comprises of essays on computer science. In datasets, we test a hypothesis given multiple evidence(feature) The Naïve bayes classifier considers the already written essays and their respective scores as evidence and the test that we write is considered as hypothesis.

The essays in dataset were rated on the scale from 1-6. We have divided the score into 3 different classes.

   Rating 1-2 = "unsatisfactory".
   Rating 3-4 = "Partial"
   Rating 5-6 = "Extensive"

Given an essay the Naïve bayes classifier calculates the probability of rating that can be given to this hypothesis considering the evidences the classifier has already been trained with. A closest or more likely class index of amongst "unsatisfactory, partial or extensive" is returned by the Naïve bayes classifier.

Classifier as mentioned does not account for repeated sentences or word count. For this we have improvised our algorithm to work on the rating provided by naïve bayes and account for number of words written by a student and whether or not sentences have been repeated by the student.

Two conditions have been given here to predict the score for essay

▪ If the word count of the essay is less than 20 words, then the prediction for that essay will be 1. This condition is given because the student just cannot give a short essay without concentrating on length which also matters in answer.
▪ If the any sentence in the answer is repeated more than once, then minus one marks from the predicted score i.e.: Prediction= Prediction-1.

## VI. RESULTS

The final results of the planned scheme are presented in following screenshots:
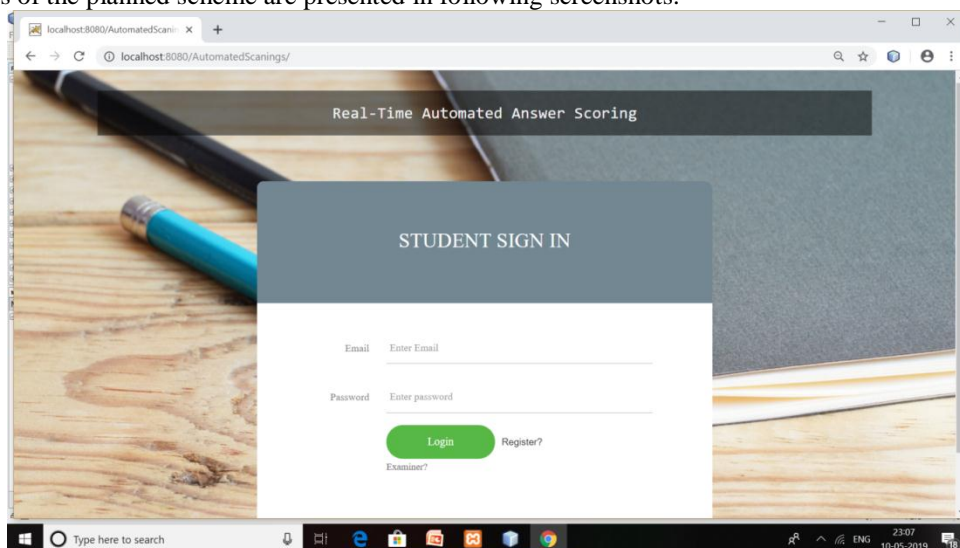


Figure 3: Home page of the real- time automated answer scoring system and the login registration page of student.
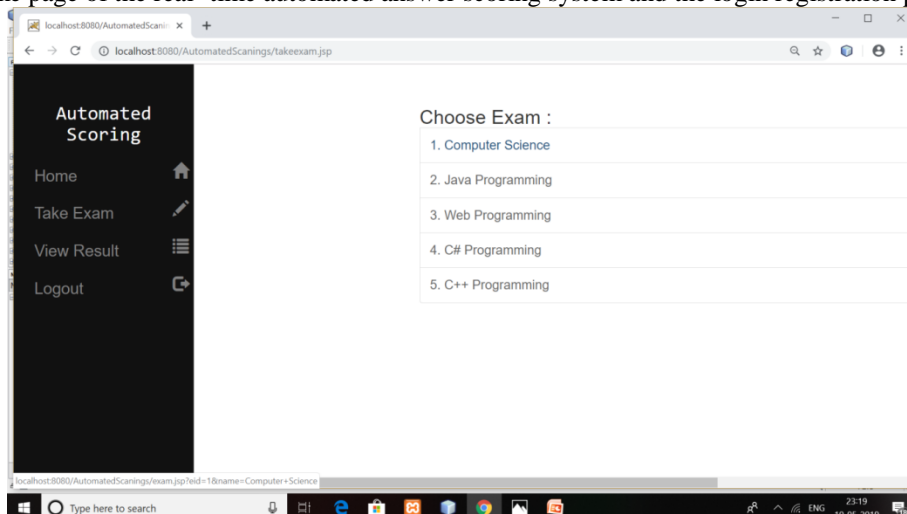


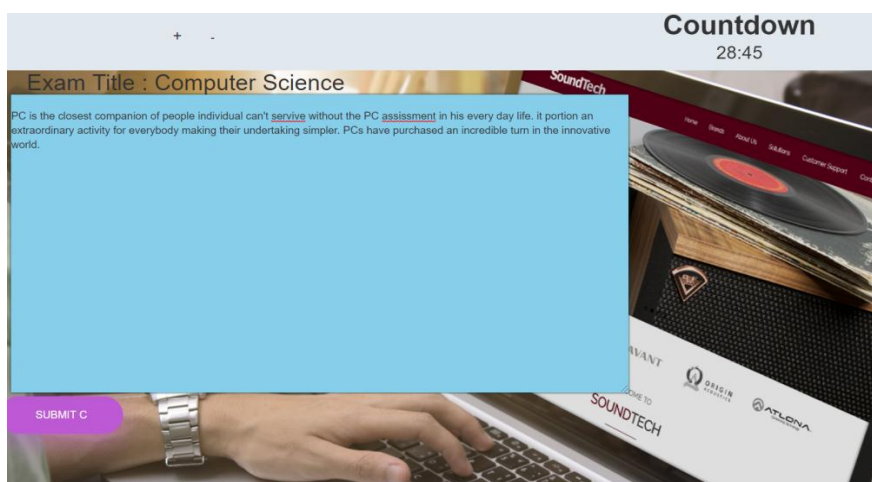Figure 4:  Student selects the topic on which he wants to write the essay.
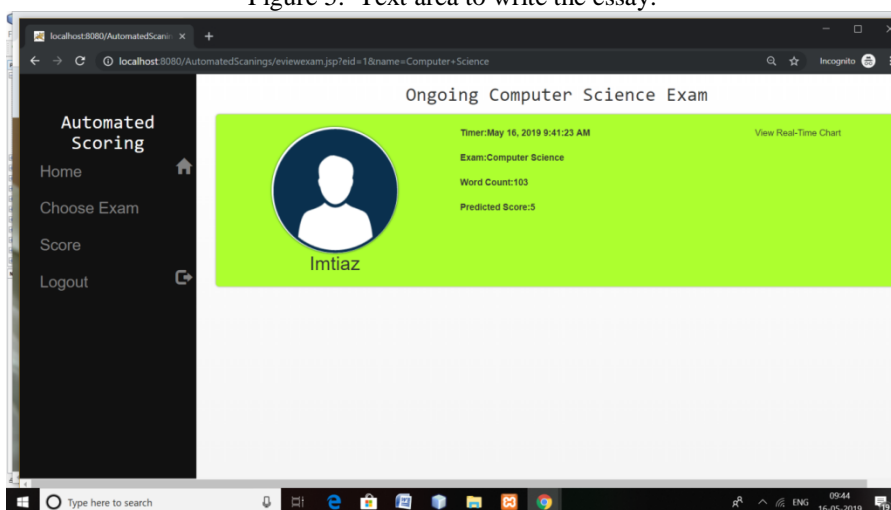
Figure 5: Text area to write the essay.



Figure 4 : Student profile in green colour depicting result is extensive.
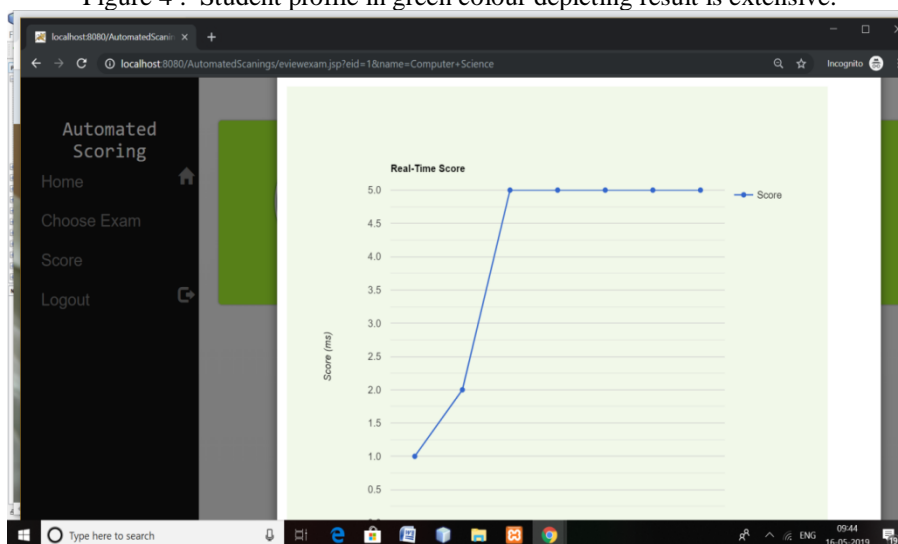


Figure 6: Live chart displaying the progress of student's answer.

## VII. CONCLUSION

Compare to human raters the automated scoring system is highly impartial and reliable. The same activity will perform by automated scoring system repetitively several times with uniformity and diversity of current educational evaluation techniques, this system will concentration on more attention of student's academic routine. With the help of various methods like latent semantic analysis this system planned pretty a number of times. The present technique enervates to model the language features like grammatical accuracy, language articulacy, and domain information comfortable of the essays, word count and sentence count.

### FUTURE WORK:

The upcoming possibility of the assumed difficult can spread in several arenas. One such area is to search and classical decent semantic and syntactic features. For this, several semantic parsers etc., can be used. Other area of focus can be to come up with a better approach which can even check for grammar and organization and styling of answers. If such s system be implemented for any Indian language, it will open the doors for other similar Indian languages.

### REFERENCES

[1] Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An Overview of Current Research on Automated Essay Grading, 2

[2] Manvi Mahana, Mishel Johns, Ashwin Apte CS229 Machine Learning , "Automated Essay Grading Using Machine Learning"-Autumn 2012 Stanford University

[3] Kaggle (2012). The Hewlett Foundation: Automated Essay Scoring. Retrieved 17 October 2012 from Kaggle: http://www.kaggle.com/c/asap-aes

[4] Leacock, C. and Chodorow, M. 2003. C-rater: Automated Scoring of Short-Answer Questions. Computers and Humanities 37:4

[5] Pennington, Jeffrey, Richard Socher, and Christopher Manning, "Glove: Global vectors for word representation," Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 15321543, 2014.

[6] Hongbo Chen and Ben He. 2013. Automated essay scoring by maximizing human-machine agreement. In EMNLP, pages 1741–1752

[7] Peter W Foltz, Darrell Laham, and Thomas K Landauer. 1999. Automated essay scoring: Applications to educational technology. In proceedings of EdMedia, volume 99, pages 40–64

[8] Bennett, R. E. (2011). Automated scoring of constructed response literacy and mathematics items. Retrieved from http://www.ets.org/s/k12/pdf/k12_ commonassess_automated_scoring_math.pdf

[9] Educational Testing Service. (2008). CriterionSM online writing evaluation service.

[10] Junker, M, M. Sintek & M. Rinck 1999. Learning for Text Categorization and Information Extraction with ILP. In: Proceedings of the 1st Workshop on Learning Language in Logic, Bled, Slovenia, 84-93.

[11] Leacock, C. and Chodorow, M. 2003. C-rater: Automated Scoring of Short-Answer Questions. Computers and Humanities 37:4.