Automated Essay Scoring using Word2vec and Support Vector Machine

A. E. E. Elalfi
Computer Science Department
Faculty of Specific Education
Mansoura University

A. F. Elgamal
Computer Science Department
Faculty of Specific Education
Mansoura University

N. A. Amasha
Computer Science Department
Faculty of Specific Education
Mansoura University

ABSTRACT
Essay scoring is one of the most important tools for evaluating and assessing the level of achievement of educational goals. It aims to innovate performance, arrange, integrate ideas, and connect them by using the vocabulary of the particular subjects. Human essay scoring consumes a lot of time and effort, this leads to mistakes. Automated Essay Scoring (AES) solve to great extent problems. A new approach for AES is presented. It is based on Natural Language Processing (NLP) which is used to unify linguistic answers, word2vec model which converts words into features and synonyms in semantic space, Support Vector Machine(SVM) is used to classify students answers and estimate score levels. The system stages consist of preprocessing, feature extraction, classification and similarity algorithm. The results of proposed method reaches high precision (94%) relative to human resident scores.

Keywords
Automated Essay Scoring, Word2vec, Support Vector Machine, Natural Language Processing

1. INTRODUCTION
Assessment is an integral part of education as it determines whether or not the educational goals are being met. It is a necessary part of the learning process. Today's students need to be able to think critically, analyze and make inferences.
Essays are considered as the most useful tool for assessing learning outcomes, meaning the ability to recall, organize incorporate concepts, express oneself in writing, provide more than the analysis and use of data[1].

Additionally it trains students on the consistency of thinking supporting poor students in their language and producing non-expressions of their level of knowledge to reach acceptable degrees. It tests the level of retrieving, understanding, applying assess, thinking and evaluating.

Automated Essay Scoring is a measuring software in which computers measure student answer[2] and educational assessment process. AES systems are mainly used in writing assessment to save time, expense, accuracy, and generalizability issues. AES continues to attract public schools, universities, consulting companies, scholars and educator’s attention[3]. AES would certainly also help teachers and administrators in education. If large numbers of student responses are submitted at once, teachers are trapped in their attempt to provide students with clear evaluations and high-quality feedback in as short a time[4].

Obviously, it would be highly desirable to attach an automated system to the educational tool kit, in general it can provide less costly and more effective results[4]. Automated scoring has the ability to overcome some of the obvious weaknesses in human essay scoring. Computers are not influenced by external factors (e.g. deadlines) or personally attached to an essay. Computers are not influenced by their community of examiner’s stereotypes or preconceptions. Consequently, automated scoring can achieve more objectivity than human scoring[5].

However, researchers argue that a machine can review and evaluate essays in much more detail than a human rater, since it is completely free of any kind of assumptions, misconceptions, false beliefs and biases in value[6]. AES can be seen as a useful alternative, as well as a support tool[7].

In this paper, English is used to put and answer questions in the information systems domain, software engineering, etc., which has a set of targeted answers where students will not be able to manipulate and write imaginary answers.

The aim of this research is to ensure that some of the corrections are not influenced by self-identity, resulting in inaccuracy of the student's degree. It was therefore necessary to find a mechanism for correcting essay grading.

AES system generally consists of software features. These functions are pre-processing the essays, extracting from the essays the necessary features, and finally performing the classification task to determine the score to assign to an essay[3,8].

Pre-processing consists of stop word extraction, phrase-stemming and lingual error handling in the simple model[9]. Automated essay scoring uses NLP techniques to automatically score essays written in an educational setting for given prompts, namely essay topics[10].

NLP had to write a large set of rules manually. Machine learning algorithms learn these rules automatically by training on large corpora of real world examples. NLP is used in many Big Data issues, such as automated summarization, analysis of sentiments, answering questions, identification of anomalies in text data and other such applications[11].

Feature is synonymous of input variable or attribute. Finding a good representation of data is very unique to the field and linked to the available measurements. Human experience, often needed to transform data into a collection of useful features, can be complemented by automated methods of building features[12].

Word2vec is a neural network proposed by Google that processes the text data. Word2Vec is not a single algorithm, but it contains two learning models, Continuous Bag of Words (CBOW) and Skip-gram. CBOW predicts the word given its context, but Skip-gram predicts the context provided by a word [13-15]. Finally Word2Vec produces the word vectors by feeding the text corpus into one learning model. In this process, Word2Vec first builds a vocabulary from the text corpus of training and learns the vector representations of each word[16].

The classification is based on supervised machine learning where the classes are the scores and each essay is defined by a set of features[3,8].

There are many machine learning techniques such as decision trees (DT), naive-bayes (NB), rule induction, artificial neural networks (ANN), K- nearest neighbors (KNN), support vector machines (SVM)[17], and word2vec[18] are used in many text classification techniques.

With Google's introduction of word2vec, a new approach to document representation is emerged. This brings additional semantic features that help in text classification[18].

The proposed system for AES presents a new approach to reduce human resident's subjectivity and eliminate discrimination toward weak responses. Therefore, this research does not find the grammar or spelling errors to be significant. It is based on testing the semantic similarity between the synonyms, as well as predicting the corresponding terms in the sentences. It provides high speed and accuracy in the grades. Word2vec was also used to learn word vector representations, called "word embedding". Then used the support vector machine to help classify them for estimating score.

The paper is organized as follows; section 2 illustrates the proposed system of AES. The system application and results are introduced in section 3. Finally the conclusions are presented in section 4.

2. PROPOSED SYSTEM FOR AES

The proposed system is based mainly on two stages. These stages are explained as follows:

- Model answer patterns generation.
- Student answer evaluation.

The block diagrams for the two stages are depicted in figures 1 and 2.

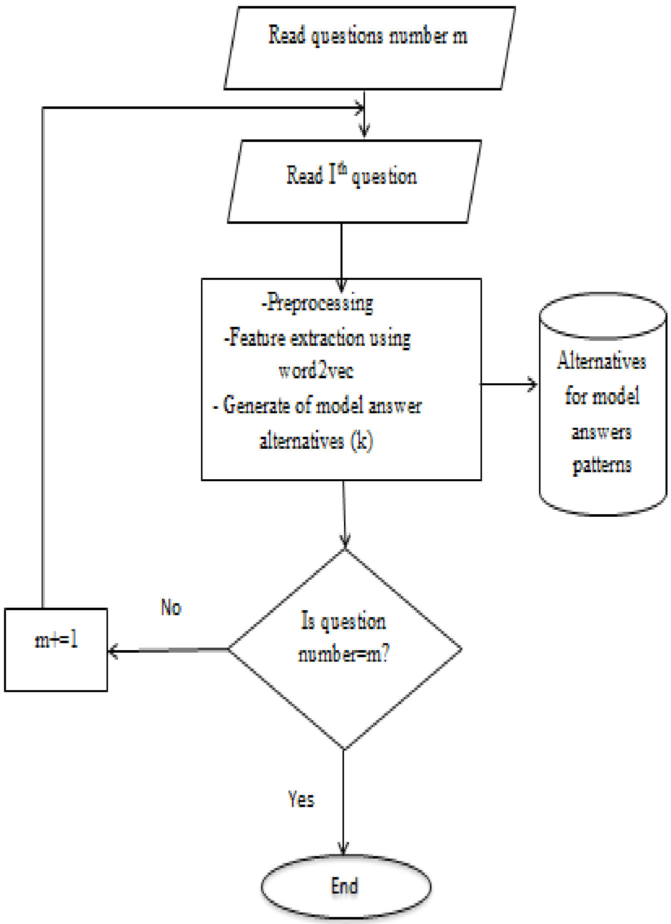


Fig1: Block diagram for model answer patterns generation

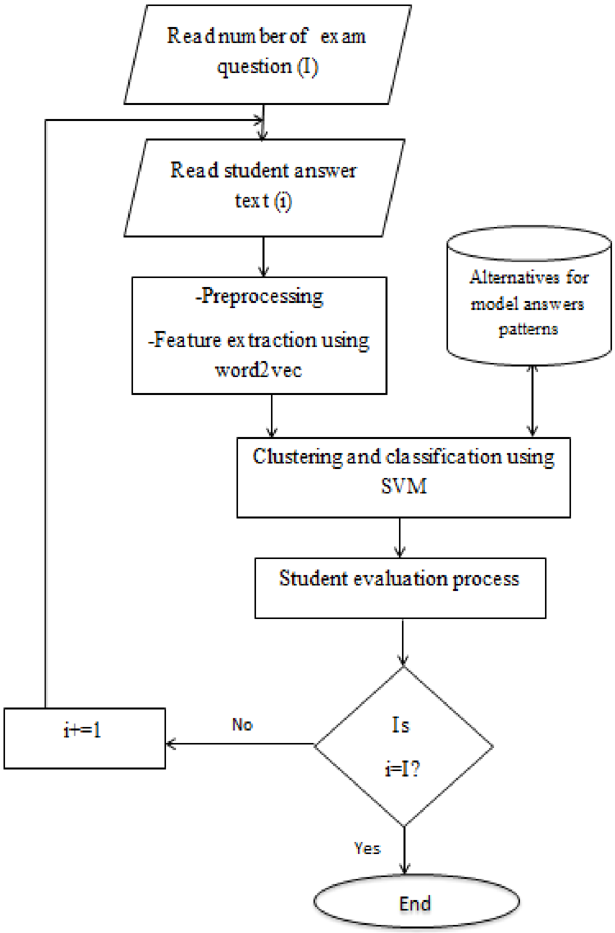


Fig2: Block diagram for student's evaluation

2.1 Model answer patterns generation

Facing the differences in the answers of the students and eliminating the deviations in the answers of the individuals, this stage passes through the following steps:

2.1.1 Model answer preprocessing

Preprocessing processes for model answers are based on NLP techniques. These can be explained in the following parts.

2.1.1.1 Sentence Detection

Text is divided into sentences. Typically, the punctuation character marks the end of a sentence. But not all characters of punctuation end a sentence. Potential sentence boundaries are taken including including ("!", "?", "."). However, full stop "." itself may not mean the end of sentences. If full stop is a member of set1{"Mr.", "Gen.","Prof."," Corp.", "Inc.","S.p.A",…..} or set2{www.wedsite name, 255.0.0.0 (IP addresses)} it does not end the sentence[19,20]. The CoreNLP toolkit is used here to detect the end of the sentence considering the previous exceptions [21].

2.1.1.2 Tokenization

Electronic text is a linear symbol sequence (characters or words or sentences). Obviously, it is important to segment text into linguistic units such as words, punctuation, numbers, alpha-numerics, etc. before any actual text processing is to be performed. This is called tokenization process.

Terms are often separated by blanks (white space) in English, but not all white space is equivalent. Both "Los Angeles" even though they contain multiple words and spaces, are individual thoughts[22]. A tokenizer separates text into a set of tokens that correspond roughly to "words"[21,23]. The CoreNLP toolkit is used to implement category tokenizer.

2.1.1.3 Abbreviations rule

The text may include some abbreviations like "It's = It is", "app=application". Such words need to be returned to their full spilling. Use abbreviations rule, this can be done. This rule is stored in the system knowledge base (KB).

2.1.1.4 Check spelling

This step is aimed to check spelling mistakes, with a high degree of accuracy and speed. In addition to improving written English for student answer. Text automatically checked for correct spelling of the words entered by the student. It is tested against the words in the installed dictionary as soon as students finish typing a word. If the word is not found in the dictionary, it will be underlined in red.

2.1.1.5 Part of Speech

Part of Speech (POS) is used to classify types of tokens according to their use in a sentence such as nouns, pronouns, adjectives, verbs, adverbs etc.. This will be used in the next steps. The category POS tagger is used in CoreNLP toolkit [21].

2.1.1.6 Co-reference resolution

The meaning of any single sentence depends on the preceding sentences and invokes the meaning of the following sentences as well. The word "it" in the sentence "she wanted it" depends upon the prior discourse

context[24]. Opennlp tools provide a deterministic co-reference module, that can handle most complex sentences in co-reference resolution[7].

2.1.1.7 Lemmatization

The key terms of a query or document are represented by stems rather than by the original words. It implies A query or document's main terms are represented by stems rather than the original words. This means that different variants of a word can be conflated into a single representative form; it also reduces the dictionary's size[25]. It involves reducing word forms to their root form after understanding the POS and the word context in the sentence given[26]. This done by predefined rules such as : the words, "identifying"," identified" become "identify".

2.1.1.8 Stop words removal

Stop-words are primarily categorized as conjunctions, prepositions, adverbs and symbols[27]. While processing documents and queries all pre-processing text applications remove stopwords. This improves the efficiency of the system. Stop words of conjunctions[27] are included in set1: {As, Because, But, For, Just, as, or, ……. }. Preposition are included in set2: { before, after, during……..}. Adverbs are included in set3: {almost, also, only,……}[28]. And symbol are included in set4: {(,) , " , . , #, , ,: , ………}. Stop words are removed to save storage space and processing time. This can be done by predefined rules which included in the proposed system. A sample of these rules is illustrated as follows:

#R
 if word ϵ {set1 or set2 or set3 or set4} then remove tag.

2.1.2 Features extractions using word2vec

Once the pre-processing tasks are completed, the student answer is prepared to be compared with the model answer. The comparison requires numerical representations of the essays. This step can be represented as follows:

2.1.2.1 Model answer features extraction

Word2vec is an algorithm that takes a word as the input and produces its vector representation equivalent to the output[29]. More Specifically, it first generates a vocabulary consisting of unique words from the data on training text and then produces vector representation of words in vocabulary. Two models Continuous-Bag-of-Words (CBOW) and Skip-Gram (SG), are used to transform internal vectors[11,30].

Using the Google dataset[6,31], Word2vec is used here to transform model answer to features. In course contents text, word2vec first builds context by linearly parsing the input text from start to end. Each word is represented by unique vector. Sample of the word vectors is presented in table1, where n represents the count of the word w in the text.

Table1: Features of some words

Word(w)	Feature1	Feature2	Feature3	…	Feature n
Word 1	-0.056	-0.099	0.078		
Word 2	0.014	0.035			
. . .	. . .	. . .	. . .	. . .	. . .
Word i	0.066	-0.059	0.040	…	-0.030

2.1.2.2 Generate of model answer alternatives

Words with similar meaning are mapped to a similar position in the vector space consequently; it is represented in space model to appear semantic relations of various words to achieve useful results. For example, synonyms of some words as shown in table2.

Table2: Synonyms of some words

Word(w)	Synonyms1	Synonyms2	…..	Synonyms n
Word 1 method	technique	tool	…..	Procedure formula
Word 2 information	Data	contact	…..	Details
. . .	. . .	. . .	. . .	. . .
Word j System	program	mechanism	…..	method

From the previous table, alternatives model answer will be generated to confront differences in students answers.

2.2 Student answer evaluation

Similarity value or distance between the model answer and student answer is calculated, then student degree is determined. This depends on some steps as follows:

2.2.1 Student answer preprocessing

Before comparing the model answer with the student answer, it is essential to prepare the students answer such as the model answer. The steps are explained in section 2.1.1.

2.2.2 Student answer feature extraction

Student answer features are extracted by using Google dataset. It is explained in section 2.1.2.1. Then comparing the model answer with the student answer will be illustrated in the next section.

2.2.3 Clustering and classification using SVM

A support vector machine is a term used for classification and regression analysis in statistics and computer science for a series of related supervised learning methods to analyze data and recognize patterns. The standard SVM takes a set of input data and predicts which consist of two possible classes forms the input for each given input, making the SVM a non-probabilistic linear binary classifier[32,33].

SVM was trained on model answer and their alternatives for learning a subset of all possible right answer in such a way to classify for answers.

Training process

Input: Question number(i), Model answer(M)

Output: Question(i) model answer alternatives

-Apply word2vec model on M to generate model answer alternatives Mi;

- Store in Question(i) model answer;

-Train SVM model on right answers(Mi) which stored in Question(i) model answer;

Testing process

Input: Question number(i), Right answers(Mi), Student answer(S)

Output: score of student

-Check if S is related to mi then continue else end;

-Determine nearest predict between right answers(Mi) and S to get the score (cosine similarity algorithm is used);

-End

Fig3: SVM classifier model

Consequently, SVM is used of linear mapping which considers away to classify students answers into two classes related to course contents or not related. Figure3 illustrates the model.

Then, cosine similarity algorithm is used to determine the similarity between model and students answers. The value of similarity ratio is ranging from 0 to 1. The following equation identifies cosine similarity[34].

$$os(\Theta)= \frac{\sum_{i=1}^{n} Mi\ Si}{\sum_{i=1}^{n} Mi\ \ \sum_{i=1}^{n} Si} \quad (1)$$

Where ;

M is a vector of model answer alternatives ,

S is a vector of student answer and

n is count of words in answer.

2.2.4 Student evaluation process

The system determines the final score of the students by measuring the similarity between students answer and model answer. Rescaling of the similarity ratio into more human convert scale is done.

3. APPLICATIONS AND RESULTS

AES was developed by web application in C# which used to implement the described algorithms in previous section. Html & css (bootstrap framework) were used for structuring each page, and also styling of each page. Additionally, JavaScript (angularJs framework) was used to send, get, post, put and delete requests to perform the required actions. Finally, Asp.net mvc 5 was used for the backend as well as linking to the database , some preprocessing and machine learning. Entity framework database were used to link between the previous programs with the database. AES was applied into a Mid Term test of course named "System Analysis" in faculty of Computers and Information Sciences, Mansoura University. The test contains 8 essay questions, 120 undergraduate students answers are checked. Table 3 presents a sample of essay scoring given by the proposed system and through two human experts. Appendix A represents sample of the proposed software screen.

Table3 Sample of the proposed system and human for students scoring

St. no.	Proposed Sys. Scoring	Human1 scoring	Human2 Scoring
1	47	50	40
2	54	58	55
3	14	19	16
4	50	40	45
………	……………	…………….	…………
117	36	37	38
118	50	62	55
119	0	0	20
120	39	42	40

Figure 4 illustrates the comparison between AES and two human experts' evaluation for students' answers.

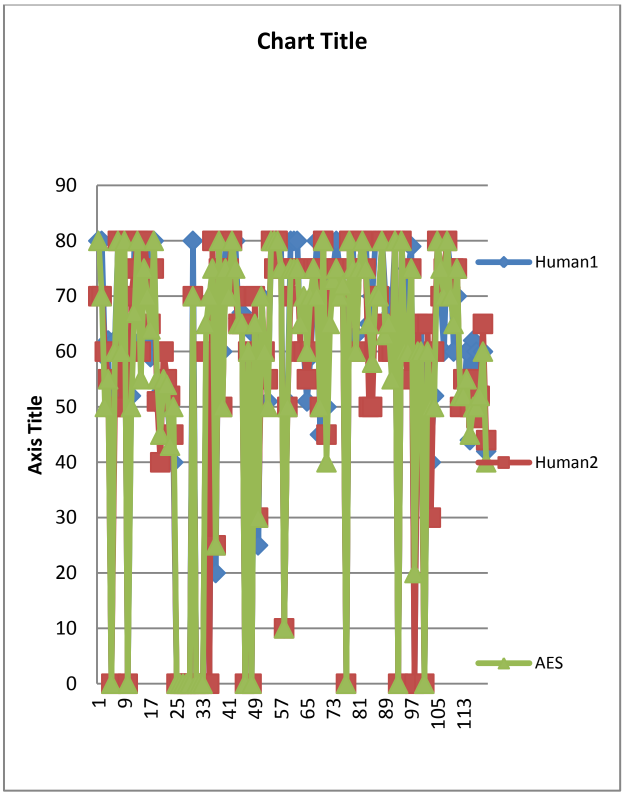


Fig4:Students scoring by AES and two human

The overall performance of the system is measured through the confusion matrix. It is a table that is often used to define a classification model's performance (or "classifier") on a set of test data for which the true values are known[35]. It depends on grades from human expert positive, true negative, false positive and false negative have been denoted as the entries are in the confusion matrix. The confusion matrix provides accuracy, error rate, precision, recall and f- measure. These (target class) and the AES proposed system (predicted class). True values is given by following equations[36]:

$$\text{Accuracy} = \frac{TP+FP}{TP+FP+TN+FN} \qquad (2)$$

$$\text{Error rate} = \frac{TN+FN}{TP+FP+TN+FN} \qquad (3)$$

$$\text{Precision} = \frac{\text{Number of TP}}{\text{Number of TP+Number of FP}} \qquad (4)$$

$$\text{Recall} = \frac{\text{Number of TP}}{\text{Number of TP+Number of FN}} \qquad (5)$$

$$F1 = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \qquad (6)$$

AES measurements can be estimated using these possible types of confusion matrix evaluation parameters in table4 , table5 and table6.

Table4 Confusion matrices among the proposed system and two human

Proposed system												
Degree	**Excellent**		**Very Good**		**Good**		**Pass**		**Fail**		**Sum**	
	Human1	Human 2	Human 1	Human 2	Human1	Human2	Human1	Human2	Human1	Human2	Human1	Human2
Excel-lent	50	50	2	1	0	0	0	0	0	0	52	51
Very Good	0	0	22	22	1	0	0	0	0	0	23	22
Good	0	0	0	1	10	9	0	0	0	0	10	10
Pass	0	0	0	0	0	1	15	15	0	0	15	16
Fail	0	0	0	0	0	1	1	1	19	19	20	21
Total	50	50	24	24	11	11	16	16	19	19	120	120

Table 5: Statistical measures for each category and for all categories (average) according to two human

Proposed system												
Degree	Excellent		Very Good		Good		Pass		Fail		Average	
	Human1	**Human2**	**Human1**	**Human 2**	**Human 1**	**Human 2**	**Human 1**	**Human 2**	**Human 1**	**Human 2**	**Human 1**	**Human2**
Accuracy	0.983333	0.99167	0.975	0.98333	0.99167	0.975	0.99167	0.98333	0.99167	0.98333	0.98667	0.98333
Error Rate	0.016667	0.00833	0.025	0.01667	0.00833	0.025	0.00833	0.01667	0.00833	0.01667	0.01333	0.01667
Recall	0.961538	0.98039	0.95652	1	1	0.9	1	0.9375	0.95	0.90476	0.97361	0.94453
Precision	1	1	0.91667	0.91667	0.90909	0.81818	0.9375	0.9375	1	1	0.95265	0.93447
F-Measure	0.980392	0.99009	0.93617	0.95652	0.95238	0.85714	0.96774	0.9375	0.97435	0.95	0.96220	0.93825

Figure5 shows confusion matrix of human1.

Table6 Confusion matrix entries by proposed system and two human

Calculated TP,TN,FP and FN For each category									
Human1 and proposed system					**Human2 and proposed system**				
	TP	**TN**	**FP**	**FN**		**TP**	**TN**	**FP**	**FN**
Excellent	50	68	0	2	**Excellent**	50	69	0	1
Very Good	22	95	2	1	**Very Good**	22	96	2	0
Good	10	109	1	0	**Good**	9	108	2	1
Pass	15	104	1	0	**Pass**	15	103	1	1
Fail	19	100	0	1	**Fail**	19	99	0	2

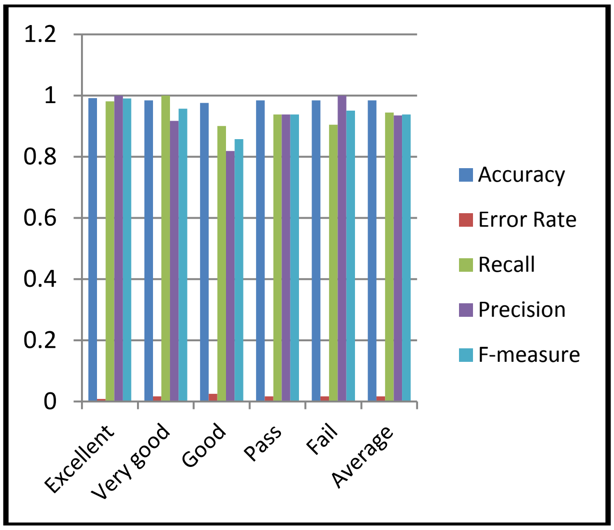


Fig5: Confusion matrix of human1

Figure6 shows confusion matrix of human2.

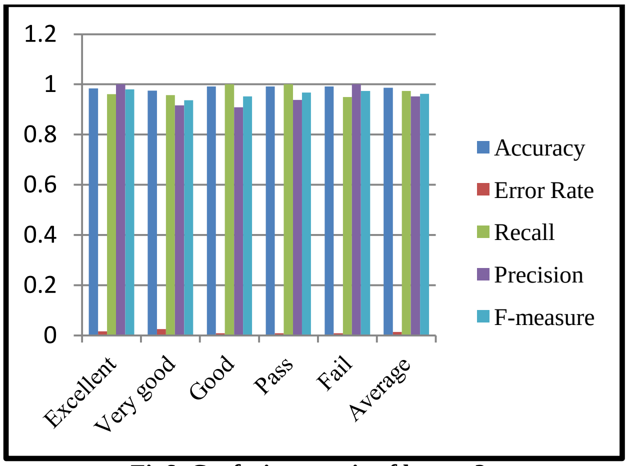


Fig6: Confusion matrix of human2

Finally, table 7 show accuracy, error rate, recall, precision and F- measure for the proposed system. Figure 7 shows comparison between proposed system and two human.

Table7 Overall statistical measures for the proposed system

	Human1	**Human2**	**Average**
Accuracy	0.986667	0.983333	0.985
Error Rate	0.013333	0.016667	0.015
Recall	0.973612	0.944531	0.959071
Precision	0.952652	0.93447	0.943561
F-measure	0.962209	0.938253	0.950231

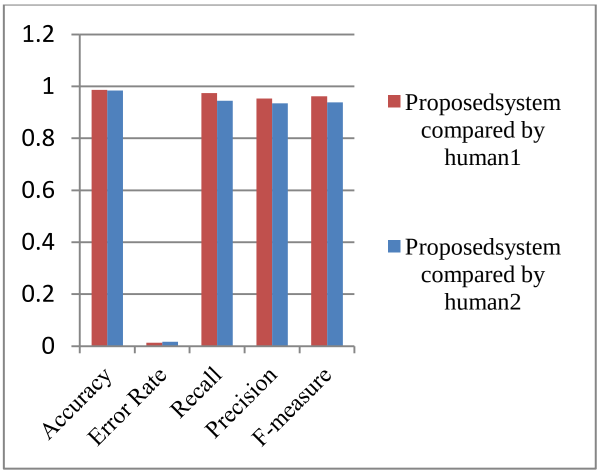


Fig7: Comparison between proposed system and two human

Figure8 shows overall statistical measures for the proposed system.

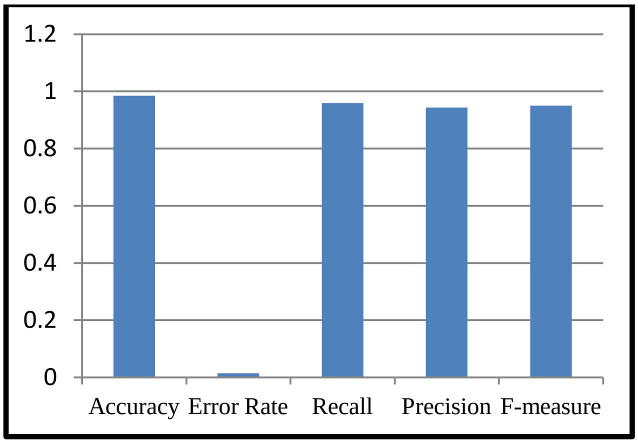


Fig8: Overall statistical measures for the proposed system

4. CONCLUSION

Automated Essay Scoring (AES) in the learning process is a very important research field. In order to advance the educational process, it aims to reduce the teacher's time and effort and save his experience. In this paper, proposed system for automating essay scoring is introduced. The system uses Natural Language Processing, word2vec and support vector machine. It has advantages over previous systems where it considered the word synonyms to generate model answer alternatives. So, students can write answer in different ways. Word2vec used here to extract semantics of the words and take into consideration the word synonyms. The scores were determined by comparing the similarity between the model answer and the student answers. SVM classifier was used, and finally an accurate and consistent score is provided. Experimental results show that the proposed AEG system achieves higher level of accuracy.

The proposed work will be developed in the future towards the scoring of essays that include text, tables, mathematical equation etc.

5. REFERENCES

[1] Salvatore Valenti, Francesca Neri and Alessandro Cucchiarelli, "An Overview of Current Research on Automated Essay Grading", Journal of Information Technology Education, 2003

[2] Mark D. Shermis, Jill Burstein and el, "Automated Essay Scoring: Writing Assessment and Instruction,chapter Automated Essay Scoring".

[3] Semire Dikli," An Overview of Automated Scoring of Essays", The Journal of Technology, Learning, and Assessment, Volume 5, Number1, August 2006.

[4] John Palmer, Robert Williams & Heinz Dreher," Automated Essay Grading System Applied to a First Year University Subject – How Can We do it Better?", InSITE - "Where Parallels Intersect", 2002.

[5] Arshad Arafat, Mohammed Raihanuzzaman," Automated Essay Grading with Recommendation", Bachelor of Science In Computer Science And Engineering, Department of Computer Science & Engineering, BRAC University, 2016.

[6] L. Streeter, J. Pstoka, D. Laham and D. MacCuish, "The credible grading machine: Automated essay scoring in the dod", 2003.

[7] Govinnage R. Perera, Deenuka N. Perera, A. R. Weerasinghe, "A Dynamic Semantic Space Modelling Approach for Short Essay Grading", International Conference on Advances in ICT for Emerging Regions (ICTer), 2015.

[8] André Smolentzov, " Automated Essay Scoring", Scoring Essays in Swedish, 2012.

[9] Parag A. Guruji and el, " Evaluation Of Subjective Answers using Glsa Enhanced with Contextual Synonymy", International Journal on Natural Language Computing (IJNLC) Vol. 4, No.1, February 2015.

[10] Hongbo Chen and Ben He," Automated Essay Scoring by Maximizing Human-machine greement", Conference on Empirical Methods in Natural Language Processing, Seattle, Washington,201

[11] Ketul Barot, Jialing Zhang, Seung Woo Son, "Using Natural Language Processing Modelsfor Understanding Network Anomalies", IEEE, 2016.

[12] Isabelle Guyon and Andr´e Elisseeff, "An Introduction to Feature Extraction", 2006.

[13] T. Mikolov, K. Chen, G. Corrado, J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. Workshop at ICLR, 2013.

[14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Proc. NIPS, 2013.

[15] T. Mikolov, W.Yih, G. Zweig, "Linguistic Regularities in Continuous Space Word Representations," Proc. NAACL HLT, 2013.

[16] Long Ma, Yanqing Zhang, "Using Word2Vec to Process Big Text Data", IEEE International Conference on Big Data, 2015.

[17] Rajni Jindal, Ruchika Malhotra and etc, "Techniques for text classification: Literature review and current trends", Webology, Volume 12, Number 2, December, 2015.

[18] Joseph Lilleberg, Yun Zhu, Yanqing Zhang , "Support Vector Machines and Word2vec for Text

[19] Classification with Semantic Features", Proc. 1IEE 14th Inl'l Coni. on Cognitive Inlormatics & Cognitive Computing, 2015.

[20] Dhanashree K. Thakur, Babasaheb Ambedkar, " Natural Language Processing Basic Tasks for Text Mining ", International Journal of Emerging Technologies in Computational and Applied Sc, 2014.Open NLP (http://opennlp.apache.org).

[21] C. D. Manning, Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, J. Steven, McClosky and David, "The Stanford CoreNLP Natural Language Processing Toolkit.," 2014.

[22] "The Art of Tokenization", available : https://www.ibm.com/developerworks/community/blogs/nlp/entry/tokenization?lang=en, 2013.

[23] https://nlp.stanford.edu/software/tokenizer.shtml.

[24] Abhimanyu Chopra, Abhinav Prashar, Chandresh Sain, " Natural Language Processing", International Journal of Technology Enhancements and Emerging Engineering Research, Vol 1, Issue 4, 2013.

[25] Eduard Heindl, Ajay Singh, " Topic- Stemming Technology", " Course- Master Of Sciences- Business Consulting".

[26] Anjali Ganesh Jivani," A Comparative Study of Stemming Algorithms", IJCTA ,NOV-DEC 2011.

[27] Jaideepsinh K. Raulji , Jatinderkumar R. Saini, "Stop-Word Removal Algorithm and its Implementation for Sanskrit Language ", International Journal of Computer Applications , 2016.

[28] http://examples.yourdictionary.com/examples-of-adverbs.html.

[29] https://en.wikipedia.org/wiki/Word2vec.

[30] “What are the continuous bag of words and skip-gram architectures?” [Online]. Available: https://www.quora.com/What-are-the-continuous-bag-of-words-and-skip-gram-architectures.

[31] Alabhya Farkiya, Prashant Saini and etc. , “Natural Language Processing using NLTK and WordNet” , International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5465-5469.

[32] C. Chang and C.-J. Lin, \LIBSVM: A library for support vector machines,"ACM Transactions on Intelligent Systems and Technology, vol. 2, pp. 27:1{27:27, 2011.

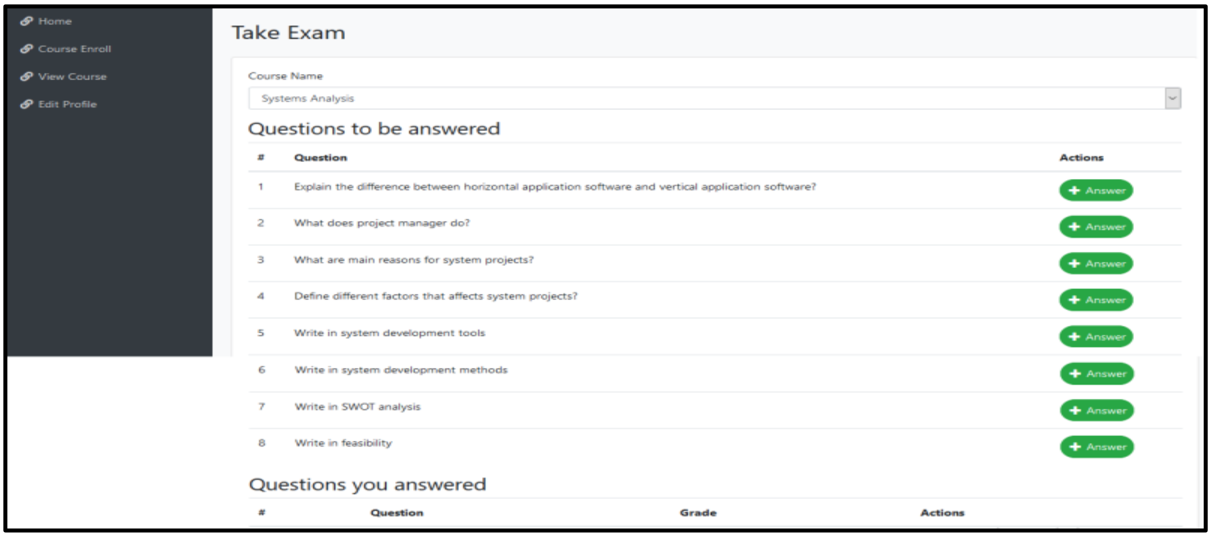
[33] Suresha M, Shilpa N.A, “ Apples Grading based on SVM Classifier”, International Journal of Computer Applications (0975 – 8878) on National Conference on Advanced Computing and Communications - NCACC, April 2012.

[34] Grigori Sidorov, Alexander Gelbukh, Helena G´omez-Adorno, “Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model”, Computación y Sistemas Vol. 18, No. 3, 2014 pp. 491–504.

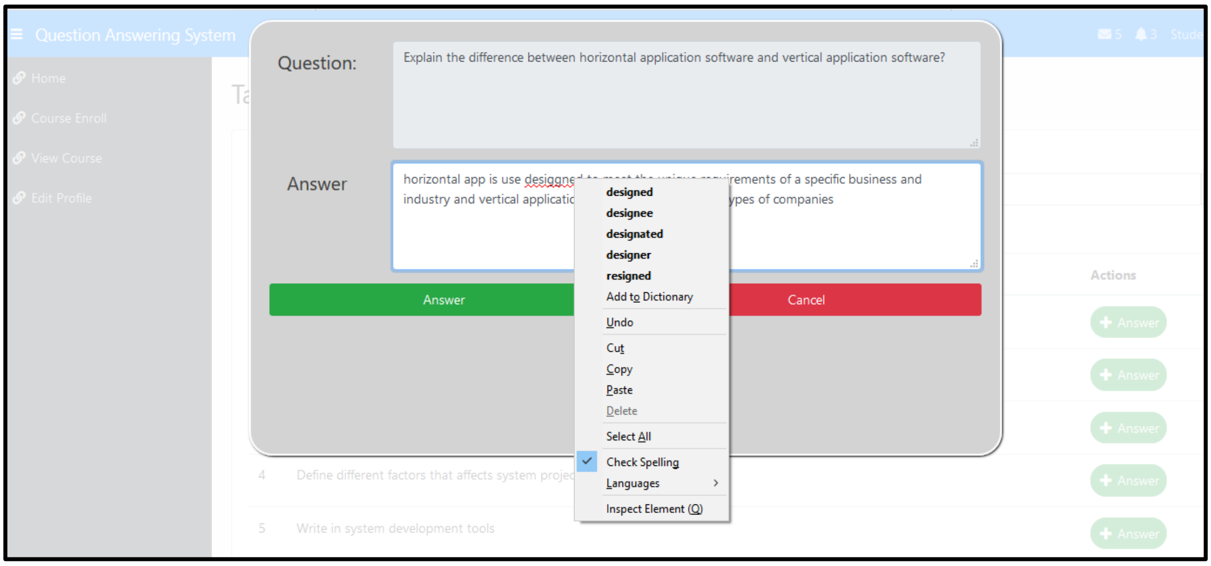
[35] https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/

[36] Md. Monjurul Islam, A. S. M. Latiful Hoque, "Automated Essay Scoring Using Generalized Latent Semantic Analysis", Journal of Computers, Vol. 7, No. 3, March 2012.

6. APPENDIX A

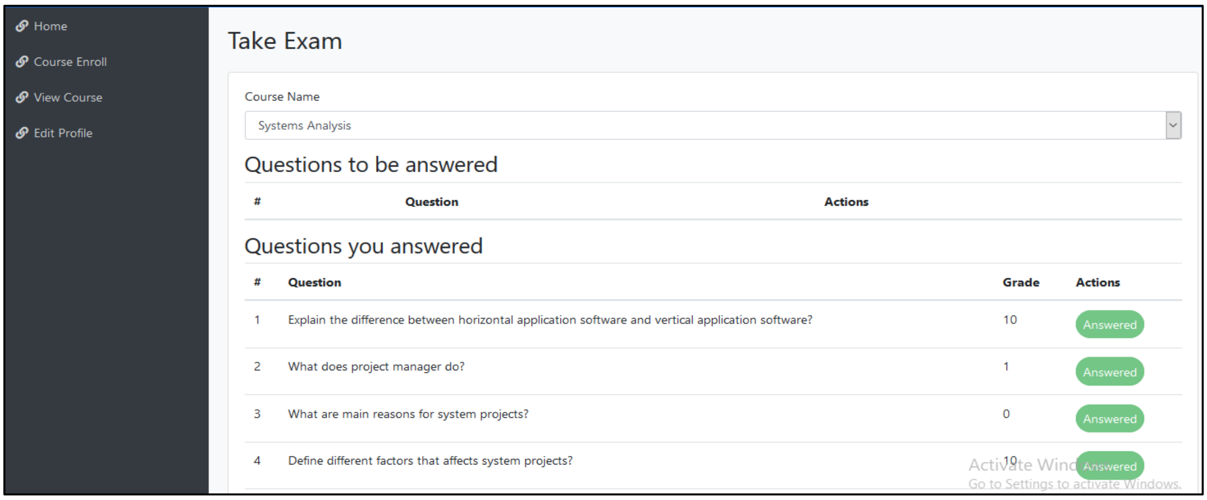


Screen of test questions to answer by student

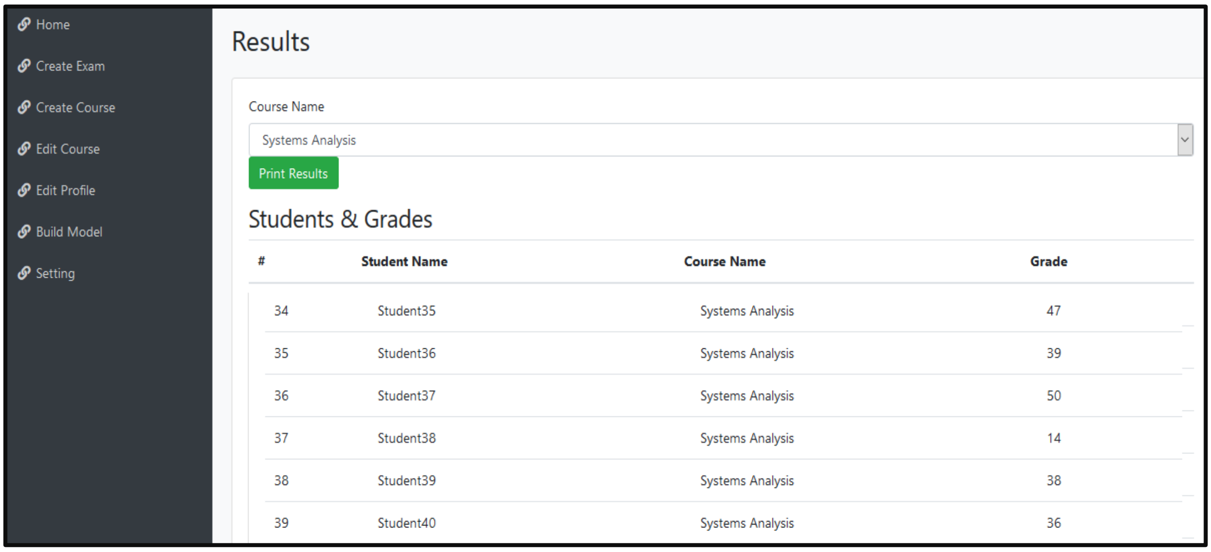


Screen of questions after answer

Screen of entered student answer



Screen of result for student



Screen of results for all students