

# Intelligent Essay Grading System using Hybrid Text Processing Techniques

Kuyoro S.O., Eluwa J. M., Akinsola J.E.T, Ayankoya F.Y., Omotunde A.A., Adegbenjo A.A.

Department of Computer Science, Babcock University, Ilishan-Remo, Nigeria

## ABSTRACT

### Article Info

Volume 6, Issue 5

Page Number: 264-270

Publication Issue :

September-October-2020

Educational Institutions are facing enormous tasks of marking and grading students at the end of every examination within the shortest possible time. Marking theoretical essay questions which involves thousands of examinees can be biased, subjective and time-consuming, leading to variation in grades awarded by different human assessors. This study presents an Essay Grading System called Intelligent Natural Language Processing Essay Grading System (iNLPEGS) with high accuracy percentage and minimal loss function for scoring assessment that can accommodate more robust questions. Secondary dataset collected from Kaggle provided by The Hewlett Foundation was used to aid semantic analysis and Part of Speech tagging. Assemblage of Computer Science questions and answers were collected from Babcock University Computer Science Department to create a more robust dataset to ensure high reliability. An Intelligent Natural Language Processing Essay Grading Model was designed based on Enhanced Latent Semantic Analysis using Part of Speech n-gram Inverse Document Frequency. Web based application was developed using Django, Gensim, Jupyter Notebook and Anaconda as the development tools due to availability of several Python libraries with SQLite as the database. Results of performance evaluation of iNLPEGS showed accuracy of 89.03% and error of 10.97% connoting that there is very little difference between scores from the developed intelligent essay grading system and a human grader. Also, the loss function from Root Mean Square Error (RSME) showed value of 0.620 which is very small and thus signifies closeness to the line of best fit from the regression equation.

### Article History

Accepted : 05 Sep 2020

Published : 15 Oct 2020

**Keywords :** Enhanced Latent Semantic Analysis (ELSA), Essay Grading System (EGS), Inverse Document Frequency (IDF), Natural Language Processing (NLP), Part of Speech (POS) Tagging.

## I. INTRODUCTION

NLP is the capacity of computer programs to grasp human language the way it is spoken. It is a technology frequently applied for data analysis

which are in textual form. The technology is used to break the texts into tokens that can be further used for calculations. This approach is used for easier alteration during data analysis. These tokens are correlated and analyzed with the adjacent

tokens to confirm that the sentences are grammatically as well as accurately correct. The several techniques used in NLP are stemming, lemmatization, discourse, etc. [9].

Natural Language Processing (NLP) has demonstrated immeasurable value in textual data mining difficulties especially with regards to huge databases that contain valuable information that can only be revealed automatically. Deeper knowledge of programming and semantics analysis often need to be discovered so as to process the ambiguous and linguistic pattern essential in any language. In a study by [11], Natural Language Processing is described to be useful in grammar checking, language translation, development of efficient human-computer interface and alteration of human language into text format; and Sentiment Analysis, Chatbots, Client Service, Management of Advertisement Funnel and Market Intelligence [7].

Grading theoretical questions can be a constantly challenging issue for instructors but it is a very necessary aspect of teaching as it helps to determine students' level of understanding, provide feedback on the effectiveness of teaching staff, to determine if the course objectives were achieved and also to find out effectiveness of learning environment among other things. Grading theoretical questions is sometimes supported by the lecturer's individual insight, consideration, as well as explanations of a specified solution. Therefore, the grading falls on the needed typical expected words highlighted within the answer structure prepared by the teacher in the marking guide. It may be a hard method particularly during most important assessment as examiners are burdened by many solutions written system. If manually written, they need to evaluate different kinds of writing styles.

An Intelligent Grading System (IGS) is the utilization of expert software to assign scores to compositions written in natural language. It is a technique of

educational evaluation and an application of intelligent natural language processing. Grading of theoretical questions, essays, articles and other textual information especially when it involves very large number of students is a major problem in the educational Institutions. This problem requires thorough investigation, consideration and analysis. This is due largely to the fact that the manual grading can be biased, prone to human error and usage of inefficient scoring attributes. One of the difficulties of grading essays is the subjectivity, or at least the perceived subjectivity, of the grading process. The manual grading system has a difficulty on getting report cards especially when it concerns several number of students which is also occasioned by many computation errors. Also, there is general problems of slow retrieval of data with manual system. This challenge has made the evaluation of student performance after the examination a herculean task.

Assistive technology like Intelligent Grading System (IGS) is a tool that can be used to solve this problem. The development of intelligent grading system using hybrid text processing techniques involve text preprocessing and text classification, feature extraction and keywords segmentation and pattern recognition. Each of these stages comes with limitations and require high computation and feature extraction without losing the lexicon in the texts. Hence, the need to develop a generic, robust and more systematic hybrid intelligent grading system which combines the use of syntactic features with an Enhanced Latent Semantic Analysis (ELSA) that can easily grade students' theoretical essay.

The remaining part of this work is arranged as follows: Section 2 presents the literature review; section 3 presents the methodology used, section 4 discusses the results of the work while section 5 gives the conclusion and recommendation for further works.

## II. LITERATURE REVIEW

Several studies on the feasibility of Intelligent Grading System (IGS) for learners' evaluations have been in existence for decades, in addition to many methods that have been utilized. The foremost grading system is referred to as Project Essay Grade (PEG and it is still being in use as of today. This grading system was produced by Page and other researchers [5,8] and principally make use of style examination of superficial semantic highlights of a square of content. An article is overwhelmingly reviewed based on composing quality, failing to assess its content. PEG simply depends on a factual methodology dependent on the suspicion that the nature of expositions is revealed by the quantifiable proxes.

The perception of an Intelligent Natural Language Processing Grading System (INLPGS) falls on the objectivity of a grader and the idea of evaluation. Appraisal scores might be utilized for very diverse purposes [1]. In a few circumstances, NLP is utilized to offer opinion based response for text-based teaching in the schoolroom. In different circumstances, the solutions are utilized to measure pass or fail in a subject or to gain admission to advanced studying or schooling programme.

Textual information Preprocessing is one of the key parts in numerous text mining procedures. For instance, a conventional text classification system contains preprocessing, attribute or characteristics extraction, attribute selection and classification steps. The preprocessing steps more often than not comprises of the following tasks briefly described below.

---

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

### ***Tokenization***

This is the process of separating a character sequence into pieces (words/phrases) called tokens, and perhaps in the meantime throw away certain characters, for example, punctuation marks. The list of tokens at that point is utilized for additional transformation [6].

### ***Filtering***

In other to remove some words or some parts of the words from a document, filtering is employed. A typical filtering is a stop-words removal. These are the words that consistently occur in a text with no much content data. Examples are conjunctions (such as and, or) and prepositions [2,4].

### ***Lemmatization***

This involves the gathering together of various inflected forms of a word in order to be evaluated solely. It shows the morphological analysis of the words. This method assists in mapping forms of verbs to infinite tenses while nouns are mapped to a single form. So as to lemmatize the documents, we initially should hypothesize the part of speech (POS) of each expression of the documents and on the grounds that POS consumes a lot of time and is prone to error, practically speaking stemming techniques are preferred.

### ***Stemming***

Stemming methods deals with acquiring stem (root) of resultant words. Stemming techniques depends on language. It is a method employed in Natural Language Processing which helps change words to their base or stem [3].

## III. Methodology

An enhanced Latent semantic analysis matrix, based on POS n-gram Inverse Document Frequency (IDF) (POS IDF) was built to improve the syntactic feature

of the data and then applied. In order to find various category of the parts of speech existing in the text, python NLTK was used to list the verbs, nouns and adjectives in parted vectors in respect to the responses. A cosine similarity was calculated among every pair of answers so as to measure the degree of similarity among the answers. Finally, a Euclidian distance was utilized to calculate the root mean square error so as to compare the accuracy of the scores generated by the system and the manual scores. The proposed system incorporates a penalty module for student cheat and irregularity check to enhance the full proof feature of the system as shown in figure 1.

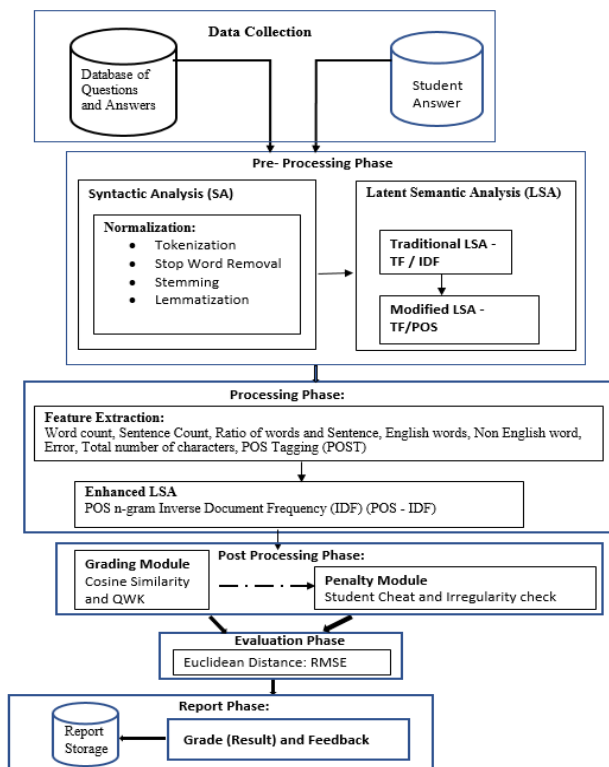


Figure 1. Intelligent Natural Language Processing Essay Grading Model

### Data Collection

Data collected from William and Flora Hewlett Foundation (Hewlett, 2012) on Automated Student Assessment Prize (ASAP) taken during the summer academic year 2011/2012. The essays were collected in a project ran by the William and Flora Hewlett Foundation [10]. The sample was randomly selected

to represent the population of students that took the test. Also, assemblage of Computer Science questions and answers were collected from Babcock University Computer Science department to create a more robust dataset to ensure high similarity index. A total of 6 Attributes used are as follows:

- **Essay\_id:** This is the primary key with unique identification for differentiating each students essay from another.
- **Essay\_set:** The essays diverges in 8 dissimilar essay set number. Each essay set comprises of related topic of the essay.
- **Essay:** This is the attribute that contains the essay. This column is made up of the ASCII value of the essays or the essay texts are written by the students.
- **Rater1\_domain1:** this column encompasses the essay score graded by first teacher.
- **Rater2\_domain1:** this column consists of the essay score evaluated by second teacher.
- **Domain1\_score:** this is used to compare the similarity between the two teachers that graded the essays.

### Data Preprocessing

Data preprocessing is the leading step when it concerns any data mining perspective. Data preprocessing is required to translate the raw unstructured inoperative data to structured functional pattern. Also dataset is inherent with many outliers and additional noises that could have negative outcome on the result obtained from the research. Nevertheless, the dataset used for this research was well fine-tuned at the pre-processing stage. Marginally variation also carried out to reduce the effect of outliers and other noises such as removal of specific missing values and so on.

### *Development Tools*

The research tools used for this study are Porter algorithm, Latent semantic analysis algorithm and WordNet for English Language. They were used in the stemming process, development of the model for intelligent grading system and creation of a domain specific dictionary respectively.

- **Porter Algorithm** used in the stemming process of the data used.
- **Latent Semantic Analysis Algorithm** used alongside part of speech tagging POS to create a modified LSA, which serves as the base model for the proposed system.
- **WordNet** is another great tool that can be used in Natural language processing. It has a huge lexical database of English Language comprising of Nouns, verbs, adjectives and adverbs which are arranged into collection of cognitive synonyms, with each one of them expressing a distinct concept. For the purpose of this study, the WordNet was imported into the development environment so as to be able to create a synonym of each word in the data set collected.
- **SQLite** is an end program embedded relational database management system. This software was used for database storage.
- **Django** is a python web framework based for web application development. This was utilized for web interface design. It supports quite a lot of database applications. It facilitates the development of pragmatic and clean design.
- **Apache server**, an open source cross platform web server solution was utilized for creating a local web server for testing and deployment purposes.
- **Anaconda**, used for package management and deployment based distribution of the Python and R programming languages for scientific computing.

### *Accuracy*

It can be defined as the extent of doubtfulness in a measurement in relation to an absolute standard. Accuracy qualifications usually contain the consequence of error due to gain and offset parameters.

$$\text{Accuracy\%} = 100 - \text{Error\%} \quad (1)$$

### *Error*

Altogether measurements are liable to error, which gives rise to the doubtfulness of the results.

$$\text{Error\%} = \left( \frac{|\text{Experimental} - \text{Manual}|}{\text{Manual}} \right) * 100 \quad (2)$$

## IV. Result and Discussion

In this study data were collected from twenty-four students six out of which were considered as test data set. A total of twelve questions were used as sample questions. The scoring results from developed intelligent essay grading system and human grader are compared for the purpose of performance analysis. This is to ensure the accuracy of the proposed solution. The study measures the variations from the expected score and actual score visually. The metrics considered are accuracy and error. Also, Mean Square Error (MSE) was calculated to measure the loss function. It is used to cross-check different estimators by matching the value of their MSE. It is also referred to as variance, which gives insight of the best fit estimate. The more reduced the MSE, the lesser the error and the more reliable the estimator. Table 1 and Figure 2 shows the comparison of scores between the proposed Intelligent Essay Grading System and Human Grader while Figure 3 shows the error percentage and accuracy percentage based on the results from Table 1.

TABLE I. STUDENTS SCORES

X	Y	
<b>Intelligent Grader</b>	<b>Human Grader</b>	<b>NEW Y'</b>
4	5	4.999
6	6	6.383
5	6	5.691
7	6	7.075
7	8	7.075
8	8	7.767

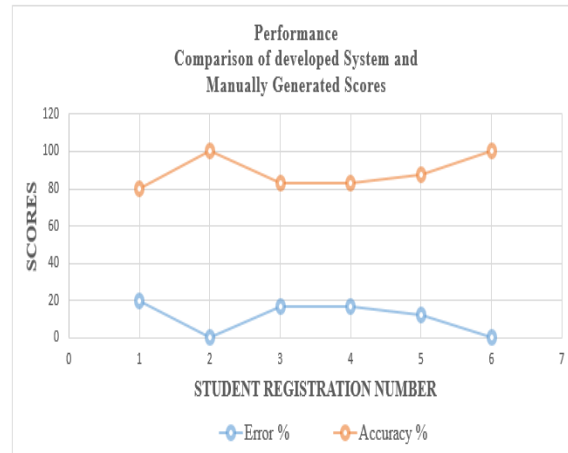


Figure 3. Performance Comparison

It can be interpreted from the graph above that there is very less difference or no difference between both the scores from the intelligent essay grading system and human grader. Hence, the developed intelligent essay grading system is valid and can be used for student examination grading. Also, the RMSE value of 0.620 is very small which connotes closeness to the line of best fit from the regression equation.

### V. Conclusion and Recommendation for Further Works

The study addressed the problem associated with the manual grading which can be biased, prone to human error and usage of inefficient scoring attributes as well as subjectivity in essays grading. This research has been able to demonstrate the efficacy of Natural Language processing (NLP) in developing an Intelligent Essay Grading System (IEGS). The model developed in this research has proven to be very efficient in electronic textual essay grading. The study was able to meet its set objectives of designing a model for efficient essay grading with high accuracy and minimal root mean square error. The result showed that Natural Language Took Kit (NLTK) is capable of playing a central role in the multilingual information society. Therefore, NLP can be used to develop unbiased, non-subjective and efficient grading system.

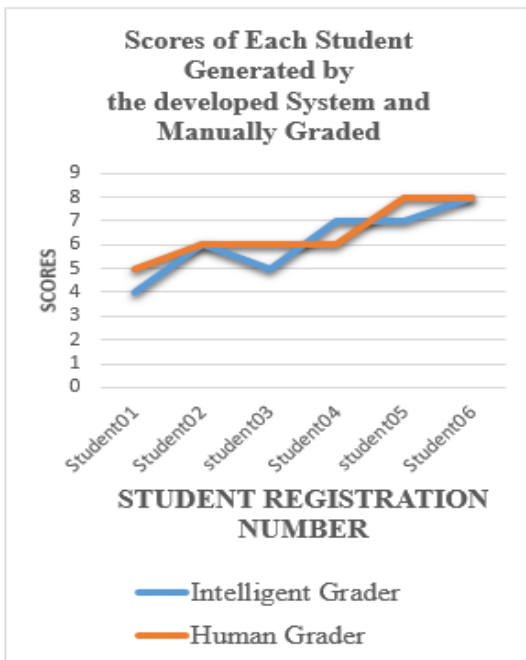


Figure 2. Students Score

The study recommends that grammatical error checker should be introduced to enhance improvement especially with the complexity for languages around the world. In information retrieval problem especially textual data mining, more accuracy is recommended where other performance calculation models other than Quadratic Weighted Kappa (QWK) or combination of several accuracy calculation models could be employed. This is for ensuring higher agreement between two raters as opposed to the QWK score of 0.961.

Furthermore, larger datasets could be considered to enhance the strength of the performance of the grading system. Natural Language Processing can be employed with machine learning especially deep learning for developing highly intelligent solutions.

## VI. REFERENCES

- [1]. Y. Attali and J. C. Burstein (2006), "Automated Essay Scoring with e-rater V.2", *The Journal of Technology, Learning and Assessment (JTLA)*, 4(3).
- [2]. S. Catarina and R. Bernardete (2003), "The importance of stop word removal on recall values in text categorization", In *Neural Networks, 2003. Proceedings of the International Joint Conference on, IEEE 3*, 1661–1666.
- [3]. Y. Harika, , I. S. Latha, Sai V.L., P.S. Krishna and M. Suneetha (2017), "Automated Essay Grading using Feature Selection", *International Research Journal of Engineering and Technology (IRJET)* 4(3).
- [4]. S. Hassan, F. Miriam, H. Yulan and A. Harith, (2014), "On stop words, filtering and data sparsity for sentiment analysis of twitter". In *Proceeding Irec conf 2014*.
- [5]. M. Hearst (2000), "The debate on Automated Essay Grading", *IEEE Intelligent Systems*, 15(5), 2237, IEEE CS Press.
- [6]. J. W. Jonathan and K. Chunyu (1992), "Tokenization as the initial phase in NLP", In *Proceedings of the 14th conference on Computational Linguistics-Volume 4*. Association for Computational Linguistics, 1106–1110.
- [7]. J. Khare, (2018), "Applications of Natural Language Processing for Businesses in 2018", Available at: <https://www.upgrad.com/blog/5-applications-of-natural-language-processing-forbusinesses-in-2018/>
- [8]. E. B. Page, (1966), "The Imminence of Grading Essays by Computer", *Phi Delta Kappan*, 48:238-243.
- [9]. M. Rouse, (2017), "Natural Language Processing (NLP)? Search Business Analytics", Available: <https://searchbusinessanalytics.techtarget.com/definition/natural-languageprocessing-NLP>.
- [10]. The Hewlett Foundation (2012): Automated Essay Scoring. Available at: <https://www.kaggle.com/c/asap-aes>
- [11]. TwinWord (2018). 4 Applications of Natural Language Processing. Available at: <https://www.twinword.com/blog/some-common-applications-of-natural-languageprocessing/>

### Cite this article as :

Kuyoro S.O., Eluwa J. M., Akinsola J.E.T, Ayankoya F.Y., Omotunde A.A., Adegbenjo A.A., "Intelligent Essay Grading System using Hybrid Text Processing Techniques ", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 6 Issue 5, pp. 229-235, September-October 2020. Available at doi : <https://doi.org/10.32628/CSEIT206547>  
Journal URL : <http://ijsrcseit.com/CSEIT206547>