

# Neural-Network Architecture Approach: An Automated Essay Scoring Using Bayesian Linear Ridge Regression Algorithm

Jeff Jojer Jones E. Catulay  
*Eulogio "Amang" Rodriguez Institute of Science and Technology*  
 Manila, Philippines  
 catulay.jjj.bscs@gmail.com

Danlord O. Ancheta  
*Eulogio "Amang" Rodriguez Institute of Science and Technology*  
 Manila, Philippines  
 ancheta.d.bscs@gmail.com

Maco E. Magsael  
*Eulogio "Amang" Rodriguez Institute of Science and Technology*  
 Manila, Philippines  
 magsael.m.bscs@gmail.com

Jefferson A. Costales  
*Eulogio "Amang" Rodriguez Institute of Science and Technology*  
 Manila, Philippines  
 jacostales@earist.ph.education

**Abstract**—Most people are relying on technology in various industries, such as education, where people are more likely to use technology as a platform for knowledge acquisition. In this study, the researchers proposed an Automated Essay Scoring or AES to help the teachers minimize the time in grading essay work of the student, prevent biased ratings and provide a feedback mechanism for the students. The study proposed a Neural-Network Approach Architecture and Bayesian Linear Ridge Regression Algorithm to improve the correctness and reliability of the AES system and dimensions we are going to use. We also use the new Hybrid Model or Hybrid Agile to have a detailed approach in developing our study. We used the dataset from the Hewlett Foundation, the Automated Student Assessment Prize (ASAP) from Kaggle. We also gather information in reliable sources and collect data to test the accuracy and reliability of our AES system.

**Keywords**—automated essay scoring, bayesian linear ridge regression algorithm, neural-network approach architecture

## I. INTRODUCTION

Educational institutions are challenged to preserve academic integrity through objective evaluation of students' performance [1], [2]. As such, technologies that support the improvement of the assessment process are vital in today's connected learning environment [3]. Prior research explored the impact of technologies in the learning process and recent advances [4], various platforms [5] and diverse users [6], [7] point to the need for innovative approach in assessing learners' comprehension of a subject matter [1], [2], [7A].

Examinations are vital components in our learning journey which may come in various forms. To write a good essay not only requires sustained practice but also demands instructional feedback from teachers [9]. The student's essay will vary in different answers and format, reading and checking the essay is a hard task for the teacher and it will consume a lot of time. Also, the Human rater of an essay can be biased.

Automated Essay Scoring or AES is a computerized way to evaluate and rate the students' essay without the help of a human rater which will decrease the bias ratings. In the pandemic situations, student's education relies on online education, student test occurs in different education platforms such as Google Classroom, Google Forms, and others. For a multiple-choice question, the answer will be specific but in an essay part of the test the teacher needs to rate the work manually.

The process of assigning a score to a piece of writing using computation is known as AES. A successful AES system would support the field of education in a number of ways including preventing teacher bias towards students, saving them time spent when conducting grading papers, as well as giving students immediate feedback on their written work [10].

AES should have different dimensions in grading; Grammatically, Usage (User of a preposition, word usage), Mechanics (Spelling, punctuation, capitalization), Style, Relevance (relevance to the topic), Organization, Development, Thesis Clarity , Persuasiveness, Cohesion and Coherence, [11]. Efficient cross-instruction AES systems are critical for applications in the real world because it is normal to have an inadequate number of essays labelled for the target [10].

In recent years, neural networks have been widely used to grade student essays automatically and achieve an output that is state-of-the-art. In particular, a distributed representation is learned for an essay with variant neural networks, and a linear layer is then used to produce the final score [12].

This study uses a Neural Network Architecture and the Bayesian Linear Ridge Regression Algorithm for the AES. This study will be beneficial for the students and the teachers in the field of Essay writing.

## II. RELATED WORKS

Neural Network is a model based on a human brain, or pertaining to the nerves or nervous system. Neural Network

Architecture is one of the leading approaches in the development of an AES. It focuses on developing an AES using the recurrent neural network approach. It used a deep learning for the AES because it is more reliable than a feature-based system. Using a Deep Learning model on an AES is more reliable rather than using a traditional feature selection method [13].

Ontology in Computer Science is a set of categories and concepts in subject area that shows the relationship between them. An approach that presents extracting information using an ontology that uses basic NLP for word tagging, tokenization, frequency distribution, counting the characters, word tagging, semantic matching, and frequency distribution of text in essay scoring. Its aim is to see how useful the ontology is for teachers to calculate essay scores. and extracting features for grading essays by experimenting with various language tools. However, it has lacked of recommenders that may help the students and they need to explore other different natural language tools that may help in the extraction of different traits in grading an essay [14].

There are lots of open-source library that may help in the development of the AES. XGBoost is one of them. It is an open-source library that can provide a framework for different programming languages like C++, Python, java, and many more. Using XGBoost as the classifier, they conducted a narrative and argumentative essay in junior high school students to make as their dataset. Furthermore, based on their study they concluded that the accuracy does not depend on a huge number of features because not all patterns and features presented in an essay can be made by a little quantity of training data [15].

Text analysis is an important feature in the development of an AES. Coh-Metrix is a tool that can provide a representation of text, it can also analyze text with many different features. It proposes new semantic steps, such as estimating the overlap between two essays' topics work. Using neural networks as prototype implementation was used to test the effectiveness of the presented AES system. However, they need future directions to consider the examination of making rubrics for grading, domain-based semantic features, as well as extending the training and research datasets. Moreover, there are essays connected to the person who writes the essay and it needs more investigation about academic levels of work models [16].

### III. METHODOLOGY

The research attempts to create a model using the Bayesian Linear Ridge Regression Algorithm as presented in Figure 3 – Automated Procedures in essay assessment. The first part of the study is to load the dataset needed, afterwards the dataset will undergo to cleaning, removing any null values and unnecessary symbols. Second is to identify different features to extract in the essay. Third, once the features are successfully extracted, the correlation value of the extracted features and the score will be computed using Spearman's Correlation. Lastly, the features with correlation with the score will be set up to the regression analysis. Through this the model will be created using Bayesian Ridge, and will analyze its performance.

#### A. Bayesian Linear Ridge Regression Algorithm

Bayesian Ridge model is one of the most know Bayesian Regression, it measures the probabilistic model of a regression problem. The Bayesian ridge algorithm is shown as follows:

$$p(w | \lambda) = N(w | 0, \lambda^{-1} I_p)$$

#### B. Features Used

TABLE I. FEATURE USED

Features	Definition
Word Count	Depicts the number of words in the essay
Sentence Count	Shows the number of sentences in the essay
Part of Speech Count	The number of nouns, adjective, verb and adverb
Grammar Errors	Includes the number of misspelled words, wrong grammar and capitalization
Average Word Length	Shows the average length of words of the essay
Unique Words	Word count without the stop words
Original Sentences	Shows the originality of the essay

Table 1 depicts the different features to be extract in the essay, such as word count, part of speech count, sentence count, grammar errors, average word length, unique word, original sentences.

#### C. Design

First step of the study is gathering dataset using csv file. The researchers used the dataset from Hewlett foundation called Automated Student Assessment Prize or ASAP dataset. The dataset includes different essay set as well as the different essay. The dataset also includes the score of two human rater and the overall score of the essay.

Once the dataset is loaded the researchers then cleaned the dataset, removing empty values that may cause any irregularity. The researchers also removed any emoji, Unicode, random symbols present in the essay. This character shows no relevance in an essay work and may cause inconsistency in the dataset.

Once the researchers cleaned the dataset the essay will then be processed to the automated procedures in essay assessment.

Figure 1 shows the Automated procedures in essay assessment. The first one is the feature extraction of the essay. There are variety of dimensions for a good quality essay. The first step in the procedure is to identify the features to extract in an essay. Once the researchers identify the different features, it will undergo to correlation computation with the score to identify what features is correlated to the score. Once the researchers successfully identified correlated features, this features will be set up in the regression analysis. Setting up regression analysis will identify the different parameters in the regression model. Once the parameters is identified the researcherss will compare different models. Comparing is

beneficial in cross validation between different models. The researchers used the pycaret library in the model creation. Once the different model was compared the Bayesian Ridge model will be used. One of the good practices is to tune the model for better performance. The tuned model will now undergo to the evaluation, measuring prediction error and features of importance. Once the researchers successfully evaluated the model the training and test dataset will be entered in the model for prediction result. The prediction result will be the basis of the researchers to evaluate the performance of the model. The last step would be the finalization before saving and exporting the model. The model will now be used in the development of the Automated Essay Scoring.

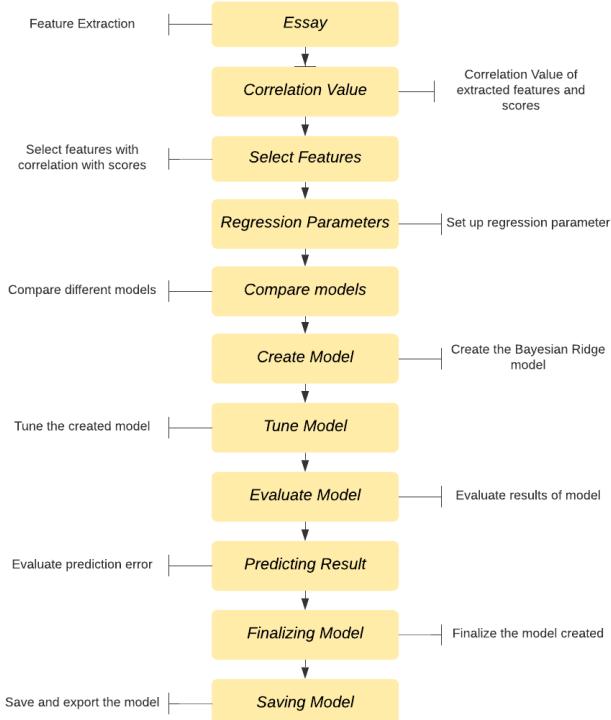


Fig. 1. Automated procedures in essay assessment.

#### IV. RESULT AND DISCUSSION

To analyze the data, the researchers created a data frame from the rows and columns of the dataset.

	essay_id	essay_set	essay	rater1_domain1	rater2_domain1	domain1_score
0	1	1	Dear local newspaper. I think effects computer...	4	4	8
1	2	1	Dear @CAPS1 @CAPS2. I believe that using compu...	5	4	9
2	3	1	Dear, @CAPS1 @CAPS2 @CAPS3 More and more peopl...	4	3	7
3	4	1	Dear Local Newspaper, @CAPS1 I have found that...	5	5	10
4	5	1	Dear @LOCATION1, I know having computers has a...	4	4	8
5	6	1	Dear @LOCATION1, I think that computers have a...	4	4	8
6	7	1	Did you know that more and more people these d...	5	5	10

Fig. 2. Essay dataset.

The researchers will now extract different features in the essay to be used in the regression analysis. Once the researchers analyzed the different features to be extract in the essay the extracted features will be measured based on its correlation to the essay score. The below code will

plot the graph of the feature, measure its correlation and interpretation and the linear regression result. The researchers will then plot the correlation of the feature to the score of the essay

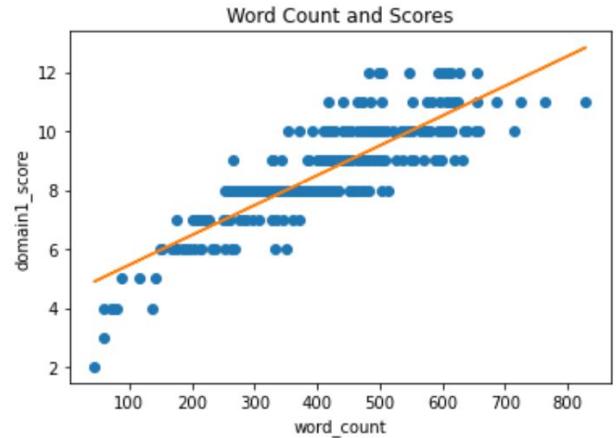


Fig. 3. Result of word count and scores.

The graph shows that the higher the word count of the essay is its more likely to have a higher score.

**Correlation Value**  
85% - Very Strong

Fig. 4. Correlation value of word count and scores.

The correlation value was computed using the Spearman's correlation by the SciPy library, the formula is shown as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

The researchers calculate the correlation of the word count and the score of the essay. The value of word count and the score is graph to help the researchers visualize and identify Linear regression parameters. Using the `scipy.stats.linregress`, the researchers calculates the Linear Regression instances. Result is shown below:

TABLE II. LINEAR LEAST-SQUARE REGRESSION

slope	0.01012003772696644
intercept	4.443612559776924
rvalue	0.8502658686481264
pvalue	1.713751351088606e-05
stderr	0.0003641321308342928

The table 2 shows the Linear Least-Square Regression of the word count and the score. After the extracted features measures its correlation with the score, the researchers set up the regression analysis parameters.

	Description	Value
0	session_id	100
1	Target	domain1_score
2	Original Data	(298, 8)
3	Missing Values	False
4	Numeric Features	7
5	Categorical Features	0
6	Ordinal Features	False
7	High Cardinality Features	False
8	High Cardinality Method	None
9	Transformed Train Set	(208, 6)
10	Transformed Test Set	(90, 6)

Fig. 5. Result of the regression analysis.

The researchers will now compare different models based on the regression parameter setup to evaluate metrics using cross-validation.

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
ada	0.6430	0.6649	0.8097	0.7015	0.0000	0.0792	0.0490
lar	0.6472	0.6870	0.8197	0.6970	0.0000	0.0795	0.0140
huber	0.6516	0.6952	0.8247	0.6941	0.0000	0.0802	0.0240
ridge	0.6555	0.7009	0.8277	0.6900	0.0000	0.0803	0.0140
lr	0.6553	0.7008	0.8278	0.6906	0.0000	0.0804	1.4120
lasso	0.6556	0.7038	0.8297	0.6898	0.0000	0.0793	0.0170
en	0.6582	0.7106	0.8339	0.6865	0.0000	0.0796	0.0120
br	0.6632	0.7217	0.8401	0.6824	0.0000	0.0801	0.0220
omp	0.6652	0.7219	0.8416	0.6786	0.0000	0.0811	0.0090
rf	0.6662	0.7220	0.8437	0.6750	0.0000	0.0796	0.1510

Fig. 6. Result of the model comparison.

After the comparison of the model, the researchers developed a Bayesian Ridge model using pycaret python library. Once the model is created it will undergo to tuning for better result of the model. Once the model is tuned the researchers will evaluate the model. Running the model evaluation code will output the prediction error graph and the features of importance in the model.

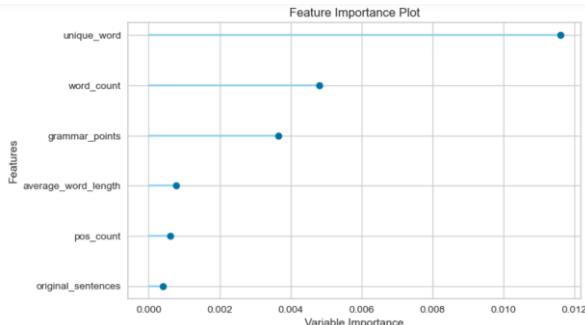


Fig. 7. The model's important features.

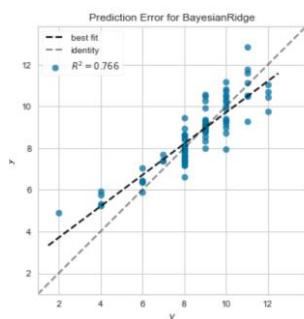


Fig. 8. The model's prediction error.

Figure 7 shows the different features of importance in the model. Figure 8 shows that the r2 is greater than 0.7 and closer to one. It shows that the dots are not far in the hyperplane of the linear line, which indicates that the regression model is good. Once the model is analyzed, it will measure the prediction result. By default, training data set is seventy percent and test dataset are thirty percent based on the pycaret library.

Neural Network is trained to produce the final score of the essay. The Neural Network have a structure of 200 input with two hidden layers with 200 neurons with an activation of rectified linear unit. The output contains a one neuron with linear activation which shows the final score of the essay. The Neural Network was trained with 200 epochs.

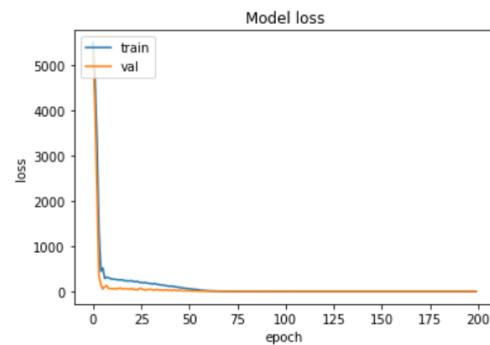


Fig. 9. Neural network model loss.

The Figure 9 shows that the increasing number of the epoch training of the neural network the lower the loss value is. This shows a great performance for the Neural Network. Once the Bayesian model produces the different parameters of the essay the output will be passed on the neural network which is in charge of producing the final essay score.

## V. CONCLUSION

The study was conducted to demonstrate the different features in an essay and how it will affect the score of the work. The different features extracted in the essay shows different correlation value with the score. Measuring the correlation of the features and scores shows that the feature with the highest correlation is the numbers of unique word in the essay. Similarly, in the model analyzation the most important features that affects the Bayesian ridge model is the unique word of the essay. Setting up the regression parameters and creating the Bayesian Ridge model, shows that it is effective in the data. Therefore, Bayesian Linear Ridge Regression Algorithm proves that it fits in the Automated Essay Scoring.

The study provided a feedback mechanism based on the features for automated scoring of essay assessment but results should be interpreted in light of our limitations. The study doesn't include feedback analysis however future initiatives may extend this study by integrating feedback which is vital in self-regulated learning. The study also doesn't include sentiment analysis which may provide a better understanding of essays in the future. The study will be the foundation for future studies as it provides different dimensions on the automated essay scoring process. The study will be a milestone

in the field of education especially in essay writing as it will support better education through increased productivity among educators and faster assessment feedback on the part of the learners.

## REFERENCES

- [1] M. J. Hazar, Z. H. Toman, and S. H. Toman, "Automated Scoring for Essay Questions in E-learning," *J. Phys. Conf. Ser.*, vol. 1294, no. 4, pp. 0–14, 2019, doi: 10.1088/1742-6596/1294/4/042014.
- [2] R. Ebardo, "Predictors of cyber-plagiarism: The case of Jose Rizal University," ICCE 2018 - 26th Int. Conf. Comput. Educ. Main Conf. Proc., no. December, pp. 778–783, 2018.
- [3] T. M. Tashu and T. Horváth, "Semantic-based feedback recommendation for automatic essay evaluation," *Adv. Intell. Syst. Comput.*, vol. 1038, pp. 334–346, 2020, doi: 10.1007/978-3-030-29513-4\_24.
- [4] H. Trapero et al., "Using augmented reality (AR) in innovating pedagogy: Students and psychologists' perspectives," ICCE 2020 - 28th Int. Conf. Comput. Educ. Proc., vol. 1, no. November, pp. 87–89, 2020.
- [5] R. Ebardo, J. de la Cuesta, J. Catedrilla, and S. Wibowo, "Peer influence, risk propensity and fear of missing out in sharing misinformation on social media during the COVID-19 pandemic," ICCE 2020 - 28th Int. Conf. Comput. Educ. Proc., vol. 1, no. November, pp. 351–359, 2020.
- [6] R. Ebardo, J. B. Tuazon, and M. T. Suarez, "We learn from each other: Informal learning in a facebook community of older adults," ICCE 2020 - 28th Int. Conf. Comput. Educ. Proc., vol. 2, no. November, pp. 594–601, 2020.
- [7] R. Ebardo and M. T. Suarez, "Older Adults and Online Communities: Recent Findings, Gaps and Opportunities," *Int. J. Web Based Communities*, vol. 17, no. 3, p. 200, 2021, doi: 10.1504/ijwbc.2021.10033661.
- [8] A. R. Razon, P. C. Naval, and R. C. L. Guevara, "Automated Essay Content Analysis using Concept Indexing," *Dimens. Contemp. Ger. Arts Lett.*
- [9] C.-T. Tsai, J.-J. Chen, C.-Y. Yang, and J. S. Chang, "LinggleWrite: a Coaching System for Essay Writing," no. 1, pp. 127–133, 2020, doi: 10.18653/v1/2020.acl-demos.17.
- [10] R. Ridley, L. He, X. Dai, S. Huang, and J. Chen, "Prompt Agnostic Essay Scorer: A Domain Generalization Approach to Cross-prompt Automated Essay Scoring," Aug. 2020, [Online]. Available: <http://arxiv.org/abs/2008.01441>.
- [11] Z. Ke and V. Ng, "Automated Essay Scoring: A Survey of the State of the Art," 2019. [Online]. Available: <https://www.kaggle.com/c/asap-aes>.
- [12] Y. Wang, Z. Wei, Y. Zhou, and X. Huang, "Automatic essay scoring incorporating rating schema via reinforcement learning," *Proc. 2018 Conf. Empir. Methods Nat. Lang. Process. EMNLP 2018*, pp. 791–797, 2020, doi: 10.18653/v1/d18-1090.
- [13] C. Cai, "Automatic essay scoring with recurrent neural network," in *ACM International Conference Proceeding Series*, Mar. 2019, pp. 1–7, doi: 10.1145/3318265.3318296.
- [14] J. O. Contreras, S. Hilles, and Z. B. Abubakar, "Automated Essay Scoring with Ontology based on Text Mining and NLTK tools," 2018 Int. Conf. Smart Comput. Electron. Enterp. ICSCEE 2018, pp. 1–6, 2018, doi: 10.1109/ICSCEE.2018.8538399.
- [15] Y. Salim, V. Stevanus, E. Barlian, A. C. Sari, and D. Suhartono, "Automated English Digital Essay Grader Using Machine Learning," 2019.
- [16] A. Ikram and B. Castle, "Automated Essay Scoring (AES); A Semantic Analysis Inspired Machine Learning Approach: An automated essay scoring system using semantic analysis and machine learning is presented in this research," in *ACM International Conference Proceeding Series*, Oct. 2020, pp. 147–151, doi: 10.1145/3436756.3437036.