



AUTOMATED ESSAY — SCORING —

MAIN
PROJECT

VARSHA U
ROLL NO :28

GUIDED BY
SONIA ABRAHAM



Introduction

Automated Essay Scoring (AES) is a tool for evaluating and scoring of essays written in response to specific prompts using computer programs. The process of automating the assessment process could be useful for both educators and learners since it encourages the iterative improvements of students' writings.

Reference papers

PAPER 1

Sharma, Shakshi, and Anjali Goyal. "Automated essay grading: An empirical analysis of ensemble learning techniques." *Computational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 2*. Singapore: Springer Singapore, 2020. 343-362

- **Objective:** Evaluate traditional machine learning and ensemble learning for essay grading.
- **Algorithms:** Logistic Regression, kNN, Naive Bayes, Decision Tree, SVM, Random Forest, AdaBoost, Gradient Boosting, XGBoost
- **Dataset:** ASAP dataset from Kaggle
- **Accuracy-** Logistic Regression(85%), kNN (k=3) (80.93%), kNN(k=5) (81.07%), Naive Bayes (75.77%), Decision Tree (72.89%), SVM (64.07%), Random Forest (82.93%), AdaBoost (70.93%), Gradient Boosting (86.53%), XGBoost (85%)
- **Conclusion:** Ensemble learning enhances automated essay grading

Reference papers

PAPER 2

Ramesh, Dadi, and Suresh Kumar Sanampudi.
"An Improved Approach for Automated Essay Scoring with LSTM and Word Embedding."
Evolution in Computational Intelligence: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021). Singapore: Springer Nature Singapore, 2022.

- **Objective:** Improve Automated Essay Scoring (AES) efficiency and reliability.
- **Algorithms:** Single-dimensional LSTM with Word2Vec.
- **Dataset:** Uses ASAP dataset from Kaggle with eight prompts and human-rated essay scores
- Achieves a mean QWK score of 85.35%.
- **Conclusion:** Outperforms other models on Kaggle dataset, showcasing the effectiveness of proposed LSTM with K-Fold cross-validation.

Reference papers

PAPER 3

Chimingyang, Huang. "An automatic system for essay questions scoring based on LSTM and word embedding." 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT). IEEE, 2020.

- **Objective:** Develop an Automated Essay Scoring (AES) model for efficient evaluation of student responses.
- **Algorithms:**
Logistic Regression: Utilized count vectors and manually crafted features.
LSTM: Employed two LSTM layers with dropout for processing Word2Vec-generated word vectors.
- **Dataset:** Kaggle's ASAP dataset, eight prompts, human-rated scores.
- Logistic Regression (QWK: 0.65 ± 0.03)
LSTM Model (QWK: 0.95 ± 0.01)
- **Conclusion** LSTM outperformed Logistic Regression.

Insights from three studies

	Advantage	Limitation	Algorithms
Paper 1 2020 Springer	Ensemble learning improves automated essay grading accuracy and efficiency, surpassing traditional methods.	Still faces challenges related to subjectivity and requires ongoing research for further improvement.	Traditional ML: logistic regression-85% Ensemble technique : gradient boosting – 86.53%
Paper 2 2022 Springer	Outperform other neural network models on the Kaggle dataset.	The model extracts word-level features, potentially missing the semantic nuances of essays.	LSTM (QWK:85.35%)
Paper 3 2020 IEEE	Demonstrate the effectiveness of deep learning for accurate essay scoring.	The model's interpretability might be challenging due to the complexity of LSTM architectures, making it less straightforward to understand the specific factors contributing to essay scores.	Logistic Regression(QWK: 65%) LSTM(QWK: 95%)

Proposed Architecture

Objective: The scope of this project is to grade essays based on language fluency, vocabulary, structure, organization and content.

Motivation: To facilitate and simplify traditional essay grading which requires a significant amount of time and effort, and to reduce subjectivity and human errors that could possibly lead to an unfair grade.

2 Stages:

Data Transformation using CountVectorizer and word2vec

Models Implementation

We are looking at both traditional machine learning methods and neural networks and compare the results. The models that will be using are:

- Logistic regression

- Gradient boosting

- LSTM

Dataset: Automated Student Assessment Prize (ASAP) on Kaggle

<https://www.kaggle.com/c/asap-aes/data>

Essay Set	Type of Essay	Grade Level	Training Set Size	Valid Set Size	Min Score	Max Score
1	Persuasive / narrative / expository	8	1783	589	2	12
2	persuasive / narrative / expository	10	1800	600	1	6
3	source dependent responses	10	1726	568	0	3
4	source dependent responses	10	1772	586	0	3
5	source dependent responses	8	1805	601	0	4
6	source dependent responses	10	1800	600	0	4
7	persuasive / narrative / expository	7	1569	441	0	30
8	persuasive / narrative / expository	10	723	233	0	60

Dataset Description

8 essay sets

Essays span 150–550 words

Authored by Grade 7 to Grade 10 students

hand-graded, and double-scored

12976 entries, 28 columns



size:120.15 MB

- **essay_id**: A unique identifier for each individual student essay
- **essay_set**: 1-8, an id for each set of essays
- **essay**: The ascii text of a student's response
- **rater1_domain1**: Rater 1's domain 1 score; all essays have this
- **rater2_domain1**: Rater 2's domain 1 score; all essays have this
- **rater3_domain1**: Rater 3's domain 1 score; only some essays in set 8 have this.
- **domain1_score**: Resolved score between the raters; all essays have this
- **rater1_domain2**: Rater 1's domain 2 score; only essays in set 2 have this
- **rater2_domain2**: Rater 2's domain 2 score; only essays in set 2 have this
- **domain2_score**: Resolved score between the raters; only essays in set 2 have this
- **rater1_trait1 score - rater3_trait6 score**: trait scores for sets 7-8

Conclusion

In summary, automating essay grading using machine learning and neural networks not only saves time but also ensures realistic scores when compared to human assessments. This project's focus on classifying textual entities into discrete grade categories signifies a meaningful advancement in improving efficiency and objectivity in the evaluation process.

Resources

- <https://www.kaggle.com/c/asap-aes/data>
- <https://ieeexplore.ieee.org/abstract/document/9363782/>
- https://link.springer.com/chapter/10.1007/978-981-16-6616-2_4
- https://link.springer.com/chapter/10.1007/978-981-15-7907-3_26

Thank you