ENHANCING SUBJECTIVE ANSWER EVALUATION THROUGH MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

[1]G.RAKESH REDDY, [2]K.S.YASHWANTH, [3]G.BALA PARDHA SARADHI, [4]U.LATHA MAHESWARI

[1,2,3] IV Year B. Tech CSE Students, Dept of Computer Science and Engineering, DR MGR EDUCATIONAL AND RESEARCH INSTITUTE, Maduravoyal, Chennai-95, Tamil Nadu, India

[4]Assistant Professor, Dept of Computer Science and Engineering, DR MGR EDUCATIONAL AND RESEARCH INSTITUTE, Maduravoyal, Chennai-95, Tamil Nadu, India

Abstract-- This research introduces a pioneering framework that harnesses machine learning and natural language processing (NLP) to revolutionize the evaluation of subjective answers in educational contexts. Traditional methods of assessing essays and open-ended responses have been characterized by their labour-intensive nature and subjectivity. Our approach streamlines this process by employing NLP techniques for preprocessing, tokenization, and advanced feature extraction, followed by training machine learning algorithms on diverse datasets of annotated answers. The result is a system capable of providing automated scores and feedback that closely align with human evaluators' judgments, demonstrating effectiveness and reliability across a spectrum of educational domains. Importantly, this automation not only enhances scalability and consistency but also lightens the workload on educators, allowing them to focus on more nuanced aspects of teaching.

Beyond its technical contributions, our research addresses ethical considerations and challenges associated with the deployment of automated evaluation systems in educational settings. This comprehensive exploration encompasses concerns related to bias, transparency, and the overall impact on the learning experience. By navigating these ethical dimensions, our study not only advances the technological aspects of automated evaluation but also underscores the importance of responsible implementation within the educational landscape. This dual emphasis on technical innovation and ethical considerations positions our framework as a promising solution for achieving efficient and objective subjective answer assessment in educational contexts.

Keywords: Machine learning, NLP, Subjective answer assessment, automatic scoring, feature extraction, consistency, feedback, teaching work load reduction, transparent evaluation

1.INTRODUCTION

Navigating the landscape of educational assessments reveals the intricate challenges posed by evaluating subjective answers. In contrast to their objective counterparts, subjective responses are unbounded, allowing students the freedom to express themselves openly. However, this freedom introduces complexities, with subjective answers being not only longer but also requiring more time and cognitive effort to produce. The richness of context in these responses demands a heightened level of concentration and objectivity from teachers during the evaluation process.

Subjective exams are perceived as more intricate and daunting, attributed primarily to their fundamental characteristic—context. The need for evaluators to scrutinize each word actively adds to the complexity. Furthermore, the mental state, fatigue, and objectivity of the evaluator play a substantial role in shaping the overall assessment outcome. Recognizing the challenges embedded in manual evaluation, there is a compelling argument for the adoption of automated systems, particularly in the contemporary context dominated by virtual work environments, accentuated by the Covid-19 pandemic.

While machines have proven adept at swiftly evaluating objective responses, the narrative pivots to the complexities inherent in handling subjective answers. Varied in length and enriched with vocabulary, these responses introduce an additional layer of difficulty, compounded by the use of synonyms and abbreviations. Despite existing efforts in the field, such as text similarity measurement and keyword matching, persistent challenges like semantic context loss and the need for robust datasets persist.

Within this landscape, the project takes center stage, aiming to automate the evaluation of subjective answers through the integration of ML And NLP. This initiative responds directly to the labour- intensive and time-consuming nature of manual evaluation, a concern amplified in the current virtual work environment shaped by the ongoing pandemic.

2. OVERVIEW

This project centers on the automation of the subjective answer evaluation process in educational settings through the utilization of advanced technologies, specifically ML and NLP. Subjective answers, characterized by their open-ended nature, have traditionally posed challenges in terms of varied lengths, diverse vocabulary, and the nuanced contextual understanding required for accurate assessment.

To address these challenges, the project employs ML algorithms and NLP techniques. The process involves preprocessing and tokenization of textual responses, followed by feature extraction using NLP methodologies. The ML models are then trained on annotated datasets, allowing them to discern the subtleties of human assessment. The ultimate objective is to develop a system capable of providing automated scores and feedback that align closely with human evaluators' judgments, thereby enhancing scalability, consistency, and objectivity in the subjective answer assessment process.

A key focus of the project is on mitigating the burden on educators associated with manually evaluating subjective answers. The inefficiencies and subjectivity inherent in traditional evaluation methods are addressed through the introduction of a more streamlined and automated approach. The system is designed not only to handle the challenges posed by varied answer lengths and rich vocabulary but also to navigate the complexities introduced by synonyms and abbreviations frequently used by students.

Moreover, the project takes into account ethical considerations associated with deploying +automated evaluation systems in educational settings. It strives to ensure transparency, fairness, and responsible use of technology in the assessment process. The overarching goal is to revolutionize the way subjective answers are evaluated, providing a more efficient, consistent, and objective framework that aligns with the demands of contemporary educational environments.

3. METHODOLOGY

The system architecture designed for subjective answer evaluation seamlessly integrates machine learning and natural language processing (NLP) techniques to ensure precision in assessments. The process initiates with raw input data, encompassing both the main answer and the student's response. To enhance consistency and eliminate unnecessary noise, the input data undergoes preprocessing, involving case folding for uniformity and the removal of special characters, non-alphanumeric elements, and punctuation marks.

The architecture uses the Glove approach for word embedding after preprocessing. By examining the co-occurrence statistics of words across large text corpora, glove vectors are able to accurately represent the contextual meaning of individual words. This strategic approach allows the system to encode semantic information, transforming the answers into a vector space for meaningful comparison and evaluation.

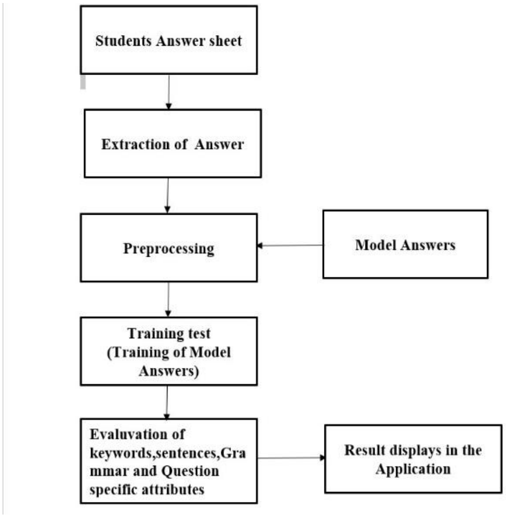

Students Answer sheet
Extraction of Answer
Preprocessing
Model Answers
Training test (Training of Model Answers)
Evaluvation of keywords,sentences,Grammar and Question specific attributes
Result displays in the Application

Fig 3.1 Workflow diagram

The pre-processed and word-embedded data is efficiently stored for retrieval and comparison during the evaluation process. When assessing the student's answer, the architecture utilizes cosine similarity, a metric gauging the likeness between two vectors. In this context, the vectors correspond to the student's response and the main answer. A higher cosine similarity score indicates a more significant match between the two answers. The architecture generates an evaluation output based on this score, offering either a numerical value or a qualitative assessment that reflects the correctness of the student's answer. In essence, this architecture combines preprocessing, word embedding, and cosine similarity metrics to provide a robust and accurate framework for evaluating subjective answers.

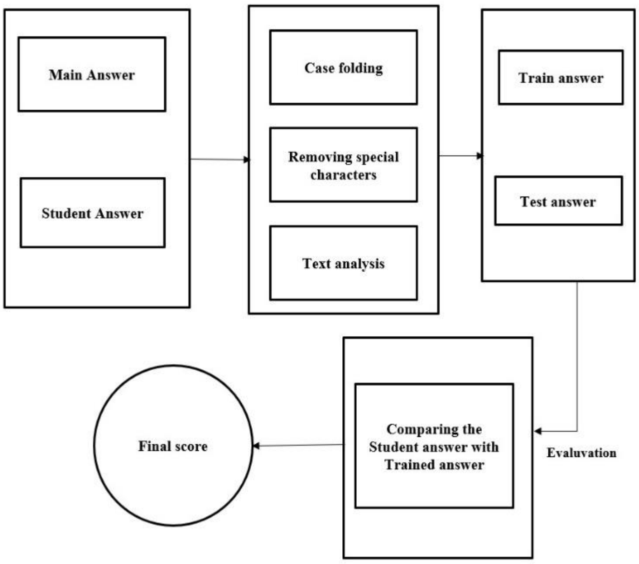


Fig 3.2 System architecture

3.1 DATA PRE-PROCESSING

Data pre-processing is a critical phase in data analysis that involves cleaning and transforming raw data to ensure its quality and reliability. This includes tasks such as handling duplicates, addressing missing values, scaling numerical data, and encoding categorical variables. Removing duplicates ensures that each data point is unique, preventing distortions in the analysis caused by redundant information. Handling missing values is essential to avoid biased or inaccurate analyses, and scaling numerical data ensures that variables with different units or scales contribute equally to the analysis. Encoding categorical variables converts non-numeric data into a format suitable for analysis, allowing algorithms to process and derive insights from this information. Exploratory Data Analysis (EDA) complements data pre-processing by providing a deeper understanding of the dataset. It involves visually exploring data distributions, identifying patterns, and detecting outliers. EDA helps in making informed decisions about data transformation and preprocessing techniques by revealing underlying structures or anomalies in the dataset. The insights gained from EDA guide subsequent analysis and model building.

3.2 TEXT CLEARING

Text clearing, a crucial facet of NLP and text analysis, is a systematic process aimed at enhancing the quality and reliability of textual data. This essential step involves the methodical removal of irrelevant or unwanted elements from the text to standardize and refine its form for accurate analysis. Punctuation marks are eliminated to simplify the text, and a uniform lowercase representation is adopted to ensure consistency. Stop words, common words that may not contribute significant meaning, are often removed to focus on more meaningful terms. Special characters and non-alphanumeric elements are also excluded to eliminate potential sources of noise. Additionally, normalization techniques, such as reducing words to their base or root form, contribute to a standardized representation of the text. The process may also involve spell checking to rectify any inaccuracies in word representation. Abbreviations are addressed to ensure clarity and consistency. Overall, text clearing plays a pivotal role in preparing textual data for NLP tasks, providing a refined and standardized foundation for subsequent analyses, such as sentiment analysis, text classification, and information retrieval.

3.3 WORD EMBEDDING

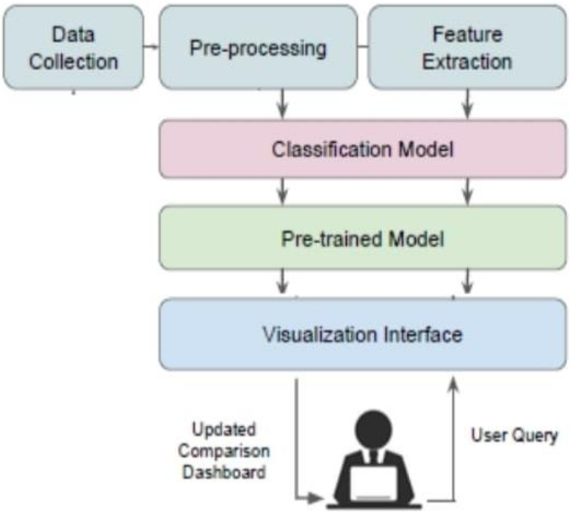
Words are converted into numerical vectors using word embedding, a complex natural language processing (NLP) approach, allowing computers to understand and process language in a mathematical and computational framework. The fundamental idea behind word embedding is to represent words as continuous, dense vectors in a high-dimensional space, where the geometric relationships between these vectors capture semantic meanings and relationships between words.

One of the key benefits of word embedding is its ability to capture contextual information and semantic relationships between words. Unlike traditional methods that represent words as discrete symbols or indices in a sparse matrix, word embeddings leverage dense vector representations that encapsulate the meaning of a word based on its context within a given dataset. For applications like sentiment analysis, text categorization, and machine translation, this contextual awareness is essential.

Neural networks, specifically shallow or deep learning models, are commonly employed in word embedding algorithms. These models learn to represent words as vectors by analyzing vast amounts of text data during a training phase. The vectors are positioned in such a way that words with similar meanings or contexts are closer to each other in the vector space. A popular word embedding model is Word2Vec, which uses a shallow neural network to predict the context of words.

GloVe is another notable word embedding model. It considers the global context of word co-occurrences across an entire corpus and captures semantic relationships between words by analyzing their distributional patterns. GloVe has gained widespread adoption due to its ability to generate meaningful and contextually rich word representations.

Word embeddings have revolutionized NLP tasks by providing a continuous and dense representation of words that preserves semantic relationships. This enables algorithms to better capture the nuances of language, understand word similarities, and generalize across different contexts. Word embedding models not only enhance the efficiency of various NLP applications but also contribute to the advancement of machine learning models in understanding and processing natural language more effectively.



3.4 COSINE SIMILARITY

Cosine similarity is a mathematical measure employed in NLP and text analysis to quantify the similarity between two vectors. Within the field of Natural Language Processing, vectors are frequently used to depict words or documents in a high-dimensional space, with each dimension denoting a distinct term or characteristic. The measure, which is based on the cosine of the angle that separates these vectors, sheds light on the direction of similarity between them instead of just magnitudes.

The computation of cosine similarity involves the dot product of the vectors and their magnitudes. When vectors are identical or point in the same direction, the cosine similarity is 1. If vectors are perpendicular, the similarity is 0, and if they point in opposite directions, the similarity is -1. The score ranges from -1 to 1, with 1 indicating complete similarity, 0 denoting no similarity, and -1 suggesting complete dissimilarity.

Cosine similarity is particularly useful for comparing text documents or word embeddings. It considers the direction of vectors, making it robust to differences in vector lengths. Its insensitivity to vector magnitudes is advantageous, allowing for effective comparison even when document lengths or word frequencies differ. This metric finds applications in various NLP tasks, including document similarity analysis, clustering, and information retrieval systems.

3.5 SIMILARITY SCORE

A similarity score in natural language processing (NLP) is a quantitative measure used to assess the degree of similarity between two entities, such as text or vectors. Several mathematical metrics capture this concept, each serving specific purposes. One prominent metric is the cosine similarity score, expressed by the equation:

$$\text{Cosine Similarity (A,B)} = \frac{A.B}{||A|| .||B||}$$

This formula calculates the cosine of the angle between vectors A and B, providing a measure of their directional agreement. A higher cosine similarity suggests a closer match and greater resemblance.

Jaccard similarity is another metric, particularly useful for set-based comparisons, and is calculated as:

$$\text{Jaccard Similarity (A,B)} = \frac{|A\cap B|}{|A\cup B|}$$

It measures the intersection over the union of elements in sets A and B, offering insights into their shared elements. Euclidean similarity, capturing the proximity between vectors A and B, is expressed as:

$$\text{Euclidean Similarity (A, B)} = \frac{1}{1+\sqrt{\sum_{i=1}^{n}(A_i-B_i)^2}}$$

A higher Euclidean similarity indicates closer proximity between vectors, emphasizing their overall similarity.

Pearson similarity, assessing linear relationships between variables, is given by the absolute value of the Pearson correlation:

Pearson Similarity (A,B) =
|Pearson Correlation (A,B)|

The Pearson correlation coefficient measures how strongly two variables are related, and whether that relationship is positive or negative.

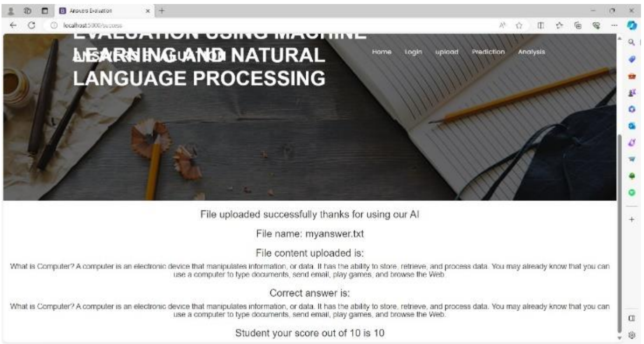
These equations represent the mathematical foundations of similarity scores, which play a crucial role in various NLP tasks, including text matching, clustering, and information retrieval. The choice of a specific metric depends on the nature of the data and the objectives of the analysis, highlighting the versatility and applicability of these mathematical measures in NLP.

5.RESULT

The project resulted in the development of an advanced system for evaluating responses through a comprehensive process involving feature extraction, sentiment analysis, semantic analysis, and other relevant techniques. This system offers a quantitative assessment of subjective qualities, enhancing its utility across various domains such as educational assessment, customer feedback analysis, and automated essay grading.

One significant aspect of the project involved updating keywords, sentences, and words for future evaluation challenges. This process included identifying and incorporating previously non-existent keywords into the system, ensuring its adaptability and relevance in evolving contexts.

The implementation of these updates has led to a notable enhancement in the quality and efficiency of response evaluation compared to existing machine learning models. By integrating cutting-edge techniques and continuously refining its evaluation criteria, the system demonstrates improved accuracy and efficacy in assessing the quality, coherence, relevance, and fluency of responses.



File uploaded successfully thanks for using our AI

File name: myanswer.txt

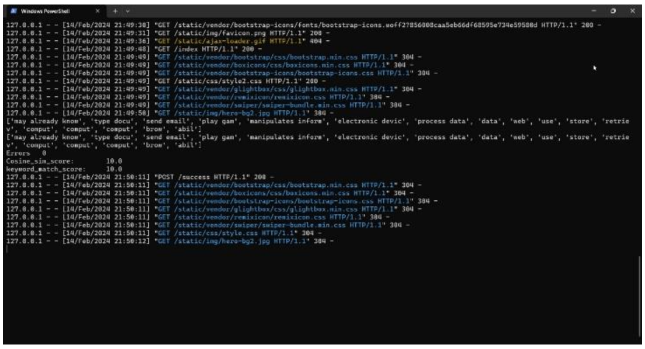
File content uploaded is:

What is Computer? A computer is an electronic device that manipulates information, or data. It has the ability to store, retrieve, and process data. You may already know that you can use a computer to type documents, send email, play games, and browse the Web.

Correct answer is:

What is Computer? A computer is an electronic device that manipulates information, or data. It has the ability to store, retrieve, and process data. You may already know that you can use a computer to type documents, send email, play games, and browse the Web.

Student your score out of 10 is 10



```
127.0.0.1 - - [14/Feb/2024 21:49:30] "GET /static/vendor/bootstrap-icons/fonts/bootstrap-icons.woff2?856008caa5eb66df68595e734e59580d HTTP/1.1" 200 -
127.0.0.1 - - [14/Feb/2024 21:49:31] "GET /static/img/favicon.png HTTP/1.1" 200 -
127.0.0.1 - - [14/Feb/2024 21:49:36] "GET /static/ajax-loader.gif HTTP/1.1" 404 -
127.0.0.1 - - [14/Feb/2024 21:49:48] "GET /index HTTP/1.1" 200 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/vendor/bootstrap/css/bootstrap.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/vendor/boxicons/css/boxicons.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/vendor/bootstrap-icons/bootstrap-icons.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/css/style2.css HTTP/1.1" 200 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/vendor/glightbox/css/glightbox.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/vendor/remixicon/remixicon.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:49:49] "GET /static/vendor/swiper/swiper-bundle.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:49:50] "GET /static/img/hero-bg2.jpg HTTP/1.1" 304 -
['may already know', 'type docu', 'send email', 'play gam', 'manipulates inform', 'electronic devic', 'process data', 'data', 'web', 'use', 'store', 'retrie
v', 'comput', 'comput', 'comput', 'brow', 'abil']
['may already know', 'type docu', 'send email', 'play gam', 'manipulates inform', 'electronic devic', 'process data', 'data', 'web', 'use', 'store', 'retrie
v', 'comput', 'comput', 'comput', 'brow', 'abil']
Errors   0
Cosine_sim_score:        10.0
keyword_match_score:     10.0
127.0.0.1 - - [14/Feb/2024 21:50:11] "POST /success HTTP/1.1" 200 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/vendor/bootstrap/css/bootstrap.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/vendor/boxicons/css/boxicons.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/vendor/bootstrap-icons/bootstrap-icons.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/vendor/glightbox/css/glightbox.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/vendor/remixicon/remixicon.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/vendor/swiper/swiper-bundle.min.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:11] "GET /static/css/style.css HTTP/1.1" 304 -
127.0.0.1 - - [14/Feb/2024 21:50:12] "GET /static/img/hero-bg2.jpg HTTP/1.1" 304 -
```

5. CONCLUSION

The assessment of subjective answers is tackled using machine learning and natural language processing techniques. Two score prediction algorithms are introduced, achieving up to 88% accuracy. To address semantically loose answers, various similarity and dissimilarity thresholds are explored, incorporating measures like keyword presence and sentence percentage mapping. Experimental results indicate the superiority of the word2vec approach over traditional word embedding techniques, maintaining semantic integrity. Word Mover's Distance outperforms Cosine Similarity in most cases, expediting machine learning model training. With adequate training, the model becomes proficient in predicting scores independently, eliminating the need for semantic checking. Tailoring word2vec models for specific domains and increasing the number of classes with large datasets are viable strategies. Data preprocessing tasks, such as handling duplicates and encoding variables, play a crucial role. Text cleaning ensures consistency and removes irrelevant information, while word embedding techniques like GLOVE enhance data quality, proving valuable in subjective paper evaluation and other text-based tasks.

References

[1] J. Wang and Y. Dong, "Measurement of text similarity: A survey," Information, vol. 11, no. 9, p. 421, Aug. 2020.

[2] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, "A survey on the techniques, applications, and performance of short text semantic similarity," Concurrency Comput., Pract. Exper., vol. 33, no. 5, Mar. 2021.

[3] M. S. M. Patil and M. S. Patil, "Evaluating Student descriptive answers using natural language processing," Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 1716–1718, 2014.

[4] P. Patil, S. Patil, V. Miniyar, and A. Bandal, "Subjective answer evaluation using machine learning," Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 1–13, 2018.

[5] J. Muangprathub, S. Kajornkasirat, and A. Wanichsombat, "Document plagiarism detection using a new concept similarity in formal concept analysis," J. Appl. Math., vol. 2021, pp. 1–10, Mar. 2021.

[6] X. Hu and H. Xia, "Automated assessment system for subjective questions based on LSI," in Proc. 3rd Int. Symp. Intell. Inf. Technol. Secur. Informat., Apr. 2010, pp. 250– 254.

[7] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger, "From word embeddings to document distances," in Proc. Int. Conf. Mach. Learn., 2015, pp. 957–966.

[8] C. Xia, T. He, W. Li, Z. Qin, and Z. Zou, "Similarity analysis of law documents based on Word2vec," in Proc. IEEE 19th Int. Conf. Softw. Qual., Rel. Secur. Companion (QRSC), Jul. 2019, pp. 354–357.

[9] H. Mittal and M. S. Devi, "Subjective evaluation: A comparison of several statistical techniques," Appl. Artif. Intell., vol. 32, no. 1, pp. 85–95, Jan. 2018.

[10] L. A. Cutrone and M. Chang, "Automarking: Automatic assessment of open questions," in Proc. 10th IEEE Int. Conf. Adv. Learn. Technol., Sousse, Tunisia, Jul. 2010, pp. 143– 147.

[11] G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, "A two-stage text feature selection algorithm for improving text classification," Tech. Rep., 2021.

[12] H. Mangassarian and H. Artail, "A general framework for subjective information extraction from unstructured English text," Data Knowl. Eng., vol. 62, no. 2, pp. 352– 367, Aug. 2007.

[13] B. Oral, E. Emekligil, S. Arslan, and G. Eryigit, "Information extractionˇ from text intensive and visually rich banking documents," Inf. Process. Manage., vol. 57, no. 6, Nov. 2020, Art. no. 102361. 23

[14] H. Khan, M. U. Asghar, M. Z. Asghar, G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, "Fake review classification using supervised machine learning," in Proc. Pattern Recognit. Int. Workshops Challenges (ICPR). New York, NY, USA: Springer, 2021, pp. 269–288.

[15] S. Afzal, M. Asim, A. R. Javed, M. O. Beg, and T. Baker, "URLdeepDetect: A deep learning approach for detecting malicious URLs using semantic vector models," J. Netw. Syst. Manage., vol. 29, no. 3, pp. 1–27, Mar. 2021.

[16] N. Madnani and A. Cahill, "Automated scoring: Beyond natural language processing," in Proc. 27th Int. Conf. Comput. Linguistics (COLING), E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Santa Fe, NM, USA: Association for Computational Linguistics, Aug. 2018, pp. 1099–1109.

[17] Z. Lin, H. Wang, and S. I. McClean, "Measuring tree similarity for natural language processing based information retrieval," in Proc. Int. Conf. Appl. Natural Lang. Inf. Syst. (NLDB) (Lecture Notes in Computer Science), vol. 6177, C. J. Hopfe, Y. Rezgui, E. Métais, A. D. Preece, and H. Li, Eds. Cardiff, U.K.: Springer, 2010, pp. 13–23.

[18] G. Grefenstette, "Tokenization," in Syntactic Wordclass Tagging. Springer, 1999, pp. 117–133.

[19] K. Sirts and K. Peekman, "Evaluating sentence segmentation and word Tokenization systems on Estonian web texts," in Proc. 9th Int. Conf. Baltic (HLT) (Frontiers in Artificial Intelligence and Applications) vol. 328, U. Andrius, V. Jurgita, K. Jolantai, and K. Danguole, Eds. Kaunas, Lithuania: IOS Press, Sep. 2020, pp. 174–181.

[20] A. Schofield, M. Magnusson, and D. M. Mimno, "Pulling out the stops: Rethinking stopword removal for topic models," in Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics (EACL) vol. 2, M. Lapata, P. Blunsom, and A. Koller, Eds. Valencia, Spain: Association for Computational Linguistics, 2017, pp. 432–436.

[21] M. Çagatayli and E. Çelebi, "The effect of stemming and stop-wordremoval on automatic text classification in Turkish

language," in Proc. 22nd Int. Conf. Neural Inf. Process. (ICONIP) (Lecture Notes in Computer Science), vol. 9489, S. Arik, T. Huang, W. K. Lai, and Q. Liu, Eds. Istanbul, Turkey: Springer, 2015, pp. 168–176.

[22] M. Divyapushpalakshmi and R. Ramalakshmi, "An efficient sentimental analysis using hybrid deep learning and optimization technique for Twitter using parts of speech (POS) tagging," Int. J. Speech Technol., vol. 24, no. 2, pp. 329–339, Jun. 2021.