International Journal of Electrical and Electronics Engineers
Volume 18, Issue No. 01, Jan-June 2026
ISSN (0) 2321-2055 ISSN (P) 2321-2045

A Hybrid Model for Intelligent Plagiarism Detection

Mrs. L. Saritha, M.Tech. G. Saraswathi, J. Navya, K. Sharoon, A. Tejaswini

Department of Electronics and communication Engineering

Tirumala Engineering College, Andhra Pradesh, India

***Abstract* :Plagiarism detection has become a significant concern in academic, research, and industrial domains due to the rapid expansion of digital content and easy accessibility of information. However, distinguishing between human-written and AI-generated content has become a significant challenge in academic, research, and industrial domains. Traditional plagiarism detection systems mainly rely on exact string matching and surface-level similarity measures, which are not effective in identifying AI-generated or semantically altered content. This paper presents a hybrid model for intelligent human versus AI content detection that integrates machine learning techniques with semantic analysis and deep learning-based natural language processing. The proposed system combines traditional similarity metrics with advanced embedding methods, including word vectors and contextual representations, to capture both syntactic and semantic characteristics of text. Furthermore, supervised learning algorithms and transformer-based models are used to analyze writing patterns and classify content as either human-written or AI-generated. The system also examines contextual coherence, sentence structure, and paraphrasing patterns to improve detection accuracy. In addition to classification, the model provides a percentage-based output indicating the likelihood of content being human or AI-generated, enabling a clear and interpretable final decision. Overall, the proposed hybrid approach enhances accuracy, scalability, and reliability, offering an effective solution for modern content authenticity verification.**

Keywords— Plagiarism Detection, Machine Learning, Natural Language Processing, Semantic Analysis, Hybrid Model, Text Similarity, Deep Learning, Severity Classification

I. INTRODUCTION

Plagiarism detection has become an essential requirement in academic, research, and industrial environments due to the exponential growth of digital content and the ease of accessing online information. With the increasing availability of text data, individuals can easily copy, modify, or paraphrase existing content, making it difficult to ensure originality and maintain academic integrity. Despite the availability of various plagiarism detection tools, many existing systems rely heavily on exact string matching and surface-level similarity measures, which are often insufficient for identifying semantically altered or paraphrased content [1]. These limitations create challenges in accurately detecting more sophisticated forms of plagiarism, thereby necessitating the development of more advanced and intelligent detection systems.

Traditional plagiarism detection approaches typically utilize techniques such as keyword matching, n-gram comparison, and basic statistical similarity metrics to identify overlapping text segments. While these methods are effective in detecting direct copying, they often fail when the content is restructured, paraphrased, or contextually modified [2]. Furthermore, such systems lack the ability to understand the deeper meaning and contextual relationships with in the text, resulting in reduced accuracy and reliability. As digital content continues to expand, there is a growing need for systems that can analyze both syntactic and semantic similarities to provide a more comprehensive assessment of plagiarism [3]. Recent advancements in machine learning (ML) and natural language processing (NLP) have opened new possibilities for improving plagiarism detection systems. Techniques such as word embeddings, semantic similarity measures, and transformer-based models enable the analysis of contextual meaning and relationships between words and sentences [4]. These approaches allow the detection of paraphrased and semantically similar content that traditional methods often overlook. Additionally, supervised learning algorithms can be trained to classify text based on patterns of similarity and transformation, further enhancing detection capabilities [5].

This paper proposes a hybrid model for intelligent plagiarism severity detection that integrates traditional similarity measures with advanced ML and deep learning-based NLP techniques. The system combines multiple models to capture both lexical and semantic similarities and classifies plagiarism into different severity levels such as low, moderate, and high. By analyzing contextual alignment, sentence restructuring, and paraphrasing patterns, the proposed framework aims to improve detection accuracy and provide a more detailed evaluation of plagiarism. The use of hybrid methodologies not only enhances performance but also addresses the limitations of existing systems, offering a scalable and reliable solution for modern plagiarism detection and academic integrity enforcement [6].

II. PLAGIARISM TYPES

The context is mainly categorized under the following types of plagiarism: Unintentional plagiarism, intentional plagiarism, and self-plagiarism. There are three main points to consider when combating these types of plagiarism:

- Similarity detection approaches aim to identify likely source documents in a large database for a particular problematic document.
- Text-matching systems track possible sources using a variety of detection techniques and provide an interface for users.
- Policies that establish institutional guidelines and procedures for preventing plagiarism or dealing with detected cases.

The main contributions of the proposed FTLM detection model are as follows:

- Selection of criteria and alternatives based on different language modeling
- Identification of plagiarism based on the criteria and alternatives chosen above
- Use the proposed TOPSIS model to rank them.

This paper is organized as follows: Section II discusses the background and related work, followed by the proposed plagiarism detection model in Section III; Section IV describes the dataset, results, and analysis of the experiments; and the conclusion is presented in Section V.

I. RELATED WORK

TOPSIS [19] is one of the best-known MCDM techniques. Distance calculations are a useful and practical method for evaluating and selecting a plausible choice. It is based on the idea that the best outcome must be the furthest distance from the negative ideal solution (NIS), i.e., the solution that maximizes the cost criteria while minimizing the benefit criteria and the solution that minimizes the cost criteria while maximizing the benefit criteria. This is referred to as the positive ideal solution (PIS), which maximizes the benefit criteria and minimizes the cost criteria.

Based on the weighted Euclidean distance, public decision-making is modeled using WEDTOPSIS [29]. A comparison between TOPSIS and Modified TOPSIS is made by [20] using simulation and mathematical analysis. To reduce the complexity of the original MCDM problem, [21] developed a new ranking index by assigning different weights to the criteria "cost" and "benefit." Additionally, [22] Kuo [22] presented an improved fuzzy semantics-dependent plagiarism detection scheme for analyzing and matching texts using the WordNet lexical dataset. This scheme includes a pre-processing stage to identify plagiarism in texts translated from or into Arabic, facilitating the application of the fuzzy method with the available information.

Further work on plagiarism detection is based on stylometry N-grams, and the Vector Space Model focuses only on text similarities. To circumvent these limitations, plagiarism detection systems often use additional techniques and strategies besides context representation. These can include:

Semantic analysis [23]: The use of NLP techniques to analyze the semantic content of a text, which can help detect paraphrased or rephrased content. Syntactic analysis [24]: Examining the grammatical structure of sentences to recognize structural similarities between documents. Machine learning [25]: Using machine learning algorithms to learn patterns of plagiarism from large data sets. These models can be trained to recognize different forms of plagiarism, including paraphrasing and patchwork plagiarism.

Analyzing citations [26]: Check for proper attribution and compare citation patterns to detect improper citations or citation plagiarism. Review by human experts: Human experts often review suspicious cases to determine plagiarism.

Existing work in plagiarism detection spans various techniques, from classical text similarity measures to advanced methods involving semantic and syntactic analyses, machine learning, and human review. However, significant gaps remain, particularly in detecting paraphrasing, handling structural and semantic variations, and leveraging citationbased techniques. Classical TOPSIS might not be ideal for plagiarism detection because it assumes definite scores and equal importance for all criteria. Fuzzy TOPSIS addresses this limitation by allowing fuzzy scores, which account for vagueness, and fuzzy weights, which prioritize important criteria like word order similarity. The proposed FTLM (Fuzzy TOPSIS Language Modeling) aims to address these gaps by combining fuzzy logic with advanced language modeling to provide a more comprehensive and context-aware approach to plagiarism detection.

A. RESEARCH GAP

Many research gaps can be derived from the literature reviews, which are listed as follows:

- Improvement of detection techniques with an emphasis on the detection of paraphrases and intelligent manipulations.

- The available tools cover only structural and semantic variations or manipulations. Therefore, the efficiency of the algorithms should be improved in this respect.
- Focus on plagiarism using idea adoptions, i.e., summarizing obfuscations that are difficult to combat.

Computational intelligence, soft computing, and advanced NLP techniques can be used in these aspects. The literature shows that most of the work has been done with N-gram models, VSM, etc. Very few works were found with semantic and intelligent implementations. Citation-based techniques are still under-researched and offer good opportunities to improve recognition efficiency when properly combined with text-based techniques. Focus on techniques for searching for candidates, especially in the context of online resources. Search query formulation and keyword extraction techniques must be explored to regulate and improve the performance of a PDS.

RQ1. What common elements are responsible for the occurrence of plagiarism?

RQ2. How can one determine the criteria and alternatives for the multi-decision method?

RQ3. How can the options be evaluated and a decision reached on which one is better using the fuzzy approach?

B. PROBLEM STATEMENT

Most work on plagiarism detection models is based on text matching. Overlapping texts and similarities are not always an indication of plagiarism. Therefore, one should never rely on a percentage of semantic similarity when assessing whether plagiarism is present. Estimating the semantic similarity of text data is one of the most challenging and open research tasks in the field of NLP. The diversity of natural language makes it difficult to develop rule-based systems for determining semantic similarity measures [27]. Various semantic similarity algorithms have been presented together with FUZZY TOPSIS to solve this problem. Hence, Fuzzy concepts can be applied in aggregating multiple sources of evidence, considering various factors like writing style, vocabulary, and structural similarities to provide a more comprehensive view.

C. CHOSEN ALTERNATIVES AND CRITERIA

The MCDM method for decision-making is based on criteria and alternatives. To select effective strategies for

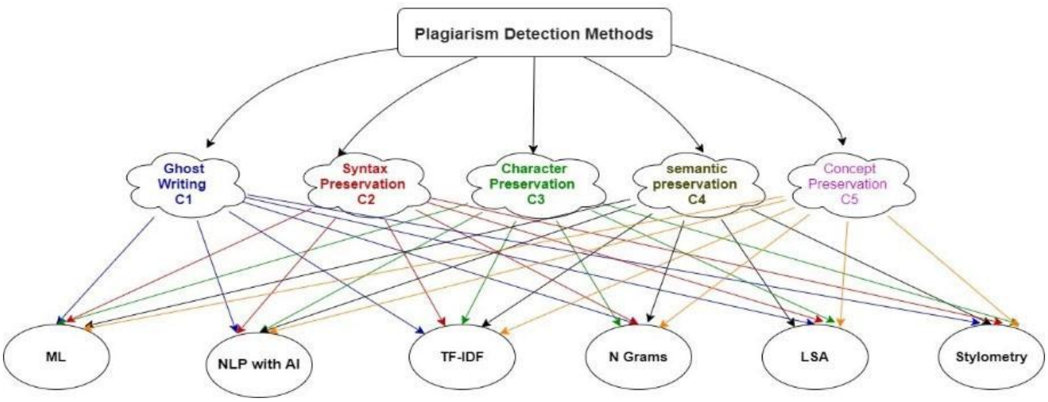

Fig1. Criteria and alternatives for language modeling phase.

detecting academic plagiarism from a wide range of options, a hierarchy was created based on five groups of criteria: [28] Ghostwriting (C1), syntax preservation (C2), character preservation (C3), semantic preservation (C4) and concept preservation (C5).

Alternatives A1 to A6 are machine learning, NLP with AI, TF-IDF, N-gram, LSA, and stylometry.

C1, Ghostwriting is ubiquitous in the world of the written word. Famous people often use ghostwriters to publish their work, as this is common practice in the publishing industry. The terms "ghostwriting" and "plagiarism" will forever be intertwined. A paraphrased or copied text that is almost identical to the original in its syntactic structure, sentence structure, and style is plagiarism. Syntax-preserving plagiarism avoids changing words or phrases to hide the similarity by keeping the sentence structure as close as possible to the source. In "character-preserving" plagiarism, also known as "character-level plagiarism", a source is copied or paraphrased while preserving the original text's letters, numbers, punctuation, and spaces. This plagiarism is difficult to detect because the characters are very similar.

The term "semantic-preserving" refers to the idea that the system attempts to detect plagiarism while preserving the underlying meaning or semantics of the text in the context of detecting plagiarism and preserving ideas.

III. PROPOSED SYSTEM

The model we propose must focus on detecting plagiarism, and it is based on two phases. The first phase is the similarity language modeling phase, which detects plagiarism through similarity computations based on different language modeling. The second phase is the decision phase, the fuzzy TOPSIS phase, in which an improved fuzzy TOPSIS is applied to select alternatives and criteria. The chosen criteria and alternatives for Language modeling are shown in Figure

A. CRITERIA AND ALTERNATIVES WITH LANGUAGE MODELING PHASE

Verbatim Copying: Detecting exact text matches is a simple task in plagiarism detection. It usually involves identifying cases where a particular part of the text in one document is identical to another.

Paraphrasing and rewriting: Detecting plagiarism while preserving semantics goes beyond verbatim copying. It also involves recognizing cases in which the content has been rephrased or paraphrased without changing the original meaning. In such cases, the system must analyze the semantic equivalence between the text in different documents.

Semantic Similarity: NLP techniques are often used to evaluate the semantic similarity between different documents to detect plagiarism while preserving semantics. This can include methods such as vector embedding (e.g., Word2Vec, Glove), which represent words and phrases in a way that captures their semantic relationships.

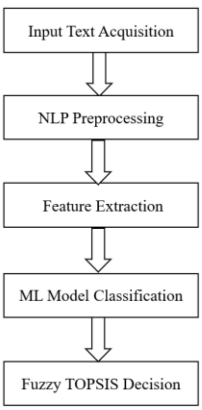

Fig 2 :proposed System

Machine learning and AI: Many plagiarism detection systems use ML with AI to automatically identify and categorize plagiarism, including those that use semantics-preserving techniques. These systems can be trained on large datasets of known plagiarism examples.

B. FUZZY TOPSIS PHASE

The fuzzy TOPSIS is a step-by-step sequential method for weight calculation and significance rating, and Figure 2 shows its workflow.

In the first step, a decision matrix is created by listing all the alternatives and criteria. In your project, the alternatives represent different machine learning models such as Logistic Regression, SVM. The criteria represent the features used to evaluate these models, such as accuracy, semantic similarity, perplexity, and sentence structure. This matrix shows how each model performs with respect to each feature when detecting whether the content is human-written or AI-generated.

In the second step, the decision matrix is normalized so that all the values are brought into a common scale, usually between 0 and 1. This is important because different features may have different ranges, and normalization ensures that no single feature dominates the others. In your project, this helps in fairly comparing features like accuracy and perplexity while analyzing human versus AI content.

In the third step, weights are assigned to each criterion based on their importance. For example, accuracy and semantic similarity may be given higher importance compared to other features. The normalized values are then multiplied by these weights to form the weighted normalized decision matrix. This step ensures that more important features have a greater influence on the final decision in identifying whether the text is human or AI-generated.

In the fourth step, the fuzzy positive ideal solution (FPIS) and fuzzy negative ideal solution (FNIS) are determined. The FPIS represents the best possible values for all criteria, meaning the ideal case where the model performs perfectly in identifying human-written content. The FNIS represents the worst possible values, indicating poor performance or strong AI-like characteristics. These two solutions act as reference points for comparison.

A model that is closer to the FPIS and farther from the FNIS is considered better at detecting human content, while a model closer to FNIS indicates poorer performance or bias toward AI-generated patterns.

In the final step, the closeness coefficient is calculated for each alternative. This value indicates how close a model is to the ideal solution. A higher closeness coefficient means the model is more reliable and performs better in distinguishing human-written text from AI-generated text. Based on this value, all models are ranked, and the best model is selected to make the final decision in your system.

Overall, in your Human vs AI detection project, Fuzzy TOPSIS helps in combining multiple model outputs and features into a single decision-making process. It ensures that the final result is accurate, balanced, and based on both performance metrics and semantic understanding, making the system more effective in identifying whether the content is written by a human or generated by AI.

IV. DATASET AND IMPLEMENTATION

The dataset used in this project consists of a large collection of textual data stored in CSV format. It contains approximately 50,000 text samples, where each sample represents either human-written content or AI-generated content.

Each text sample in the dataset is labeled into two categories:

0 – Human-written content

1 – AI-generated content

The dataset includes diverse text data such as articles, paragraphs, and short content written in different styles. This variation helps the system learn differences in writing patterns, sentence structure, and semantic meaning between human and AI-generated text.

The dataset is used for both training and testing the model to ensure that the system can accurately classify unseen data. Sample output results are shown in Figure 2.

Algorithm used to detect the plagiarism

Input:

Initialization:

For each document in Documents:

- Documents: Set of text samples containing both human-written and AI-generated content.
- Criteria: Defined based on hybrid features:
  - Word usage patterns (TF-IDF)
  - Grammatical structures
  - Semantic coherence
  - Outputs from machine learning models (Logistic Regression, SVM, Naïve Bayes)
- Weights: Importance assigned to each criterion based on contribution to classification.

Output:

- Classification Result:
  - Human Content (0)
  - AI Content (1)
- Percentage Output:
  - Human Content (%)
  - AI Content (%)
- Final Decision:
  - Based on higher percentage value

Pre-processing and Feature Extraction:

For each document in Documents:

- Tokenize and normalize the document
- Convert text to lowercase and remove stop words and punctuation
- Extract features using:
  - TF-IDF (word importance)
  - Writing style patterns
  - Sentence structure
- Store the extracted features in data structures (arrays or matrices)

Hybrid Model Integration and Criteria Definition:

For each criterion in Criteria:

- Apply machine learning models:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Naïve Bayes
- Combine outputs from all models
- Define fuzzy membership functions to handle uncertainty
- Calculate membership values for each document based on:
  - Probability scores
  - Semantic similarity
  - Pattern recognition

Construct Decision Matrix:

- Initialize a decision matrix with dimensions: (number of documents) × (number of criteria)

For each document and criterion:

- Populate the matrix with feature values and model outputs

Normalization:

For each criterion:

- Normalize values in the decision matrix to ensure all features are comparable

Weighted Decision Matrix:

For each criterion and document:

- Apply weights to normalized values based on importance
- This improves the influence of important features

Identify Ideal Solutions:

- Initialize:
  - Positive Ideal Solution (PIS): Represents AI-generated content (label = 1)
  - Negative Ideal Solution (NIS): Represents Human-written content (label = 0)

For each criterion:

- Determine maximum values → PIS
- Determine minimum values → NIS

Calculate Distance to Ideal Solutions:

For each document:

- Calculate distance to PIS (AI similarity)
- Calculate distance to NIS (Human similarity)
- Use distance measures such as Euclidean distance

Relative Closeness Calculation:

For each document:

- Calculate closeness coefficient
- Higher value → AI-generated content

- Lower value → Human-written content
- Convert closeness values into percentage form:

V. RESULT AND DISCUSSION

The Results and Discussion section plays a crucial role in evaluating the performance, effectiveness, and practical applicability of the proposed Human vs AI Content Classification System. This system is developed using advanced Natural Language Processing (NLP) techniques combined with a Fuzzy TOPSIS-based decision-making model, which enables accurate classification of textual data into human-written or AI-generated categories. After the successful design and implementation of the system, a series of experiments were conducted to assess its ability to analyze, interpret, and classify large volumes of text data efficiently.

The primary objective of this system is not only to detect plagiarism but also to differentiate between human-authored content and AI-generated content, which has become increasingly important due to the rapid growth of AI writing tools. Traditional plagiarism detection systems mainly rely on surface-level similarity measures, which are often ineffective in detecting paraphrased or AI-generated text. In contrast, the proposed system focuses on deeper linguistic and semantic analysis, allowing it to capture subtle patterns in writing style, structure, and contextual meaning.

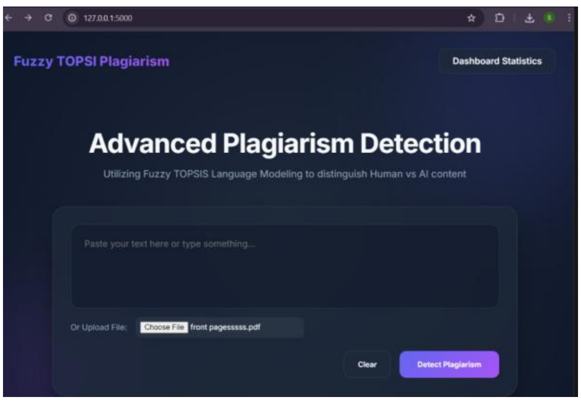

Fig 3: User Input Page

The dataset used in this project consists of approximately 500,000 text samples, stored in CSV format. Each sample is labeled based on its origin, where human-written content is assigned a label of 0, and AI-generated content is assigned a label of 1. The dataset includes a wide variety of textual data from different domains, ensuring diversity in writing styles, vocabulary usage, and contextual complexity. This diversity is essential for training a robust model capable of handling real-world scenarios.

The fig3 shows the how the user input page is working feeding the data into the model, several preprocessing steps are applied to ensure data quality and consistency. These steps include tokenization, stop-word removal, text normalization, and feature extraction using techniques such as TF-IDF (Term Frequency–Inverse Document Frequency). These preprocessing techniques help convert raw textual data into a structured numerical format that can be effectively processed by the model. By extracting meaningful features, the system can better identify patterns that distinguish human-written text from AI-generated content.

The classification process is carried out in two major phases. In the first phase, NLP techniques analyze the input text and extract important features related to syntax, semantics, and writing style. In the second phase, the Fuzzy TOPSIS decision model is applied to evaluate the extracted features based on multiple criteria. This approach allows the system to consider uncertainty and imprecision in textual data, leading to more accurate and reliable classification results.

One of the key strengths of the proposed system is its ability to provide percentage-based output rather than just a binary classification. Instead of simply labeling content as "Human" or "AI," the system generates a probability-like score that indicates the degree to which the content belongs to each category. For example:

Human Content – 72%

AI Content – 28%

These percentage values are derived from the closeness coefficient calculated using the Fuzzy TOPSIS method. This coefficient represents how close the given text is to the ideal human-written or AI-generated characteristics. The final decision is made by comparing these percentages, where the higher value determines the classification result. For instance, if the human percentage is higher, the content is classified as human-written; otherwise, it is classified as AI-generated.

This percentage-based approach provides greater transparency and interpretability compared to traditional classification methods. Users can clearly understand not only the final decision but also the confidence level of the system in its prediction. This is particularly useful in real-world applications such as academic integrity checking, content verification, and plagiarism detection, where decision transparency is essential.

To evaluate the performance of the system, several standard classification metrics are used, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the system, indicating the percentage of correctly classified samples. Precision evaluates how many of the texts predicted as AI-generated are actually correct, which helps reduce false positives. Recall measures the system's ability to identify all AI-generated content, ensuring that such content is not missed. The F1-score provides a balanced evaluation by combining both precision and recall.

In addition to numerical metrics, visualization techniques are used to better understand the system's performance. A confusion matrix is used to compare actual labels with predicted outputs, showing the number of true positives, true negatives, false positives, and false negatives. This helps in identifying the types of errors made by the system and provides insights into areas for improvement. Performance comparison graphs further illustrate the effectiveness of the model across different metrics.

Fig 4 : Result Display Page

The Fig4 shows experimental results demonstrate that the proposed system is highly effective in analyzing textual data and accurately distinguishing between human-written and AI-generated content. The integration of NLP techniques with the Fuzzy TOPSIS decision model allows the system to capture both linguistic patterns and semantic relationships, which are critical for reliable classification. The system performs consistently across different types of text inputs, including formal, informal, and technical content, indicating its robustness and adaptability.

Furthermore, the system is designed to handle large-scale datasets efficiently, making it suitable for real-world applications. Its ability to process and analyze text in real time, combined with its percentage-based output, enhances its usability in practical scenarios. The web-based implementation using frameworks such as Flask enables users to input text or upload documents and receive instant classification results.

Overall, the Results and Discussion section highlights the effectiveness of the proposed Human vs AI Content Classification System in achieving its objectives. The system not only improves the accuracy of plagiarism detection but also addresses the growing challenge of identifying AI-generated content. By leveraging advanced machine learning techniques and fuzzy decision-making, the model provides a reliable, interpretable, and scalable solution for content verification.

In conclusion, the experimental analysis confirms that the proposed system successfully combines semantic understanding, statistical analysis, and fuzzy logic to deliver accurate and meaningful results. This makes it a valuable tool for ensuring authenticity, maintaining academic integrity, and supporting content verification in the modern digital environment.

VI. CONCLUSION

The use of the model represents a significant advancement in addressing the challenges of plagiarism detection, particularly in cases involving subtle paraphrasing and AI-generated content. The two-stage procedure, which combines language modeling with fuzzy TOPSIS decision-making, effectively captures fine-grained linguistic patterns, semantic relationships, and contextual coherence. By utilizing pre-trained language models, the system enhances its ability to measure semantic similarity while also identifying distinctive patterns that differentiate human-written text from AI-generated content. This integration of semantic analysis with the imprecision-handling capability of fuzzy logic enables the model to provide a more robust and reliable evaluation.

Furthermore, the model calculates closeness coefficients (D^+ and D^-) to determine the similarity of the input text to both human and AI reference patterns. Based on these values, a percentage score is generated, indicating the likelihood of the content belonging to either category. This allows the system to produce clear and interpretable results, such as identifying content as "Human-written – 75%" or "AI-generated – 82%," thereby improving transparency and decision-making.

In decision science, FTLM stands out as a context-aware and systematic approach for evaluating both plagiarism and content originality. In the future, this model can be extended to handle multilingual and unstructured data, as well as adapted for broader real-world applications such as academic verification systems, content moderation platforms, and intelligent authorship analysis tools.

VII. REFERENCES

[1] F. Ullah, S. Jabbar, and L. Mostarda, "An intelligent decision support system for software plagiarism detection in academia," *Int. J. Intell. Syst.*, vol. 36, no. 6, pp. 2730–2752, Jun. 2021.

[2] M. N. Mansoor and M. S. H. Al-Tamimi, "Computer-based plagiarism detection techniques: A comparative study," *Int. J. Nonlinear Anal. Appl.*, vol. 13, no. 1, pp. 3599–3611, 2022.

[3] A. K. Dipongkor, R. Islam, M. Shafiuzzaman, M. A. Nashiry, S. M. Galib, and K. M. Mazumder, "AcPgChecker: Detection of plagiarism among academic and scientific writings," in *Proc. Joint 10th Int. Conf. Informat., Electron. Vis. (ICIEV) 5th Int. Conf. Imag., Vis. Pattern Recognit. (icIVPR)*, Aug. 2021, pp. 1–6.

[4] A. S. Altheneyan and M. E. B. Menai, "Automatic plagiarism detection in obfuscated text," *Pattern Anal. Appl.*, vol. 23, no. 4, pp. 1627–1650, Nov. 2020.

[5] H.Arabiand M.Akbari,"Improving plagiarism detectio n in text document using hybrid weighted similarity," *Expert Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 118034.

[6] S. Vamosi, T. Reutterer, and M. Platzer, "A deep recurrent neural network approach to learn sequence similarities for user-identification," *Decis. Support Syst.*, vol. 155, Apr. 2022, Art. no. 113718.

[7] X. Yang, X. He, H. Zhang, Y. Ma, J. Bian, and Y. Wu, "Measurement of semantic textual similarity in clinical texts: Comparison of transformerbased models," *JMIR Med. Informat.*, vol. 8, no. 11, Nov. 2020, Art. no. e19735.

[8] H. El Mostafa and F. Benabbou, "A deep learning based technique for plagiarism detection: A comparative study," *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, p. 81, Mar. 2020.

[9] K. Vani and D. Gupta, "Integrating syntax-semantic-based text analysis with structural and citation information for scientific plagiarism detection," *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 11, pp. 1330–1345, Nov. 2018.

[10] K. S. M. Anbananthen and A. M. H. Elyasir, "Evolution of opinion mining," *Austral. J. Basic Appl. Sci.*, vol. 7, no. 6, pp. 359–370, 2013.

[11] K. Yalcin, I. Cicekli, and G. Ercan, "An external plagiarism detection system based on part-of-speech (POS) tag n-grams and word embedding," *Expert Syst. Appl.*, vol. 197, Jul. 2022, Art. no. 116677.

[12] K. M. Jambi, I. H. Khan, and M. A. Siddiqui, "Evaluation of different plagiarism detection methods: A fuzzy MCDM perspective," *Appl. Sci.*, vol. 12, no. 9, p. 4580, Apr. 2022.

[13] J. Gyani, A. Ahmed, and M. A. Haq, "MCDM and various prioritization methods in AHP for CSS: A comprehensive review," *IEEE Access*, vol. 10, pp. 33492–33511, 2022.

[14] P. Karande, E. K. Zavadskas, and S. Chakraborty, "A study on the ranking performance of some MCDM methods for industrial robot selection problems," *Int. J. Ind. Eng. Comput.*, vol. 7, no. 3, pp. 399–422, 2016.

[15] E. Mulliner, N. Malys, and V. Maliene, "Comparative analysis of MCDM methods for the assessment of sustainable housing affordability," *Omega*, vol. 59, pp. 146–156, Mar. 2016.

[16] J. R. Figueira, S. Greco, B. Roy, and R. Słowiński, "An overview of ELECTRE methods and their recent extensions," *J. Multi-Criteria Decis. Anal.*, vol. 20, nos. 1–2, pp. 61–85, Jan. 2013.

[17] M. Behzadian, R. B. Kazemzadeh, A. Albadvi, and M. Aghdasi, "PROMETHEE: A comprehensive literature review on methodologies and applications," *Eur. J. Oper. Res.*, vol. 200, no. 1, pp. 198–215, Jan. 2010.

[18] W.Ma,X.Luo,andY.Jiang,"Multicriteriadecisionmaking withcognitive limitations: A DS/AHP-based approach," *Int. J. Intell. Syst.*, vol. 32, no. 7, pp. 686–721, Jul. 2017.

[19] A. R. Fayek and M. N. Omar, "A fuzzy topsis method for prioritized aggregation in multi-criteria decision making problems," *J. Multi-Criteria Decis. Anal.*, vol. 23, nos. 5–6, pp. 242–256, Sep. 2016.

[20] S. Chakraborty, "TOPSIS and modified TOPSIS: A comparative analysis," *Decis. Anal. J.*, vol. 2, Mar. 2022, Art. no. 100021.

[21] T. Kuo, "A modified TOPSIS with a different ranking index," *Eur. J. Oper. Res.*, vol. 260, no. 1, pp. 152–160, Jul. 2017.

[22] S. M. Alzahrani, N. Salim, and V. Palade, "Uncovering highly obfuscated plagiarism cases using fuzzy semantic-based similarity model," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 27, no. 3, pp. 248–268, Jul. 2015.

[23] I. Mukherjee, B. Kumar, S. Singh, and K. Sharma, "Plagiarism detection based on semantic analysis," *Int. J. Knowl. Learn.*, vol. 12, no. 3, p. 242, 2018.

[24] W. Massagram, S. Prapanitisatian, and K. Kesorn, "A novel technique for Thai document plagiarism detection using syntactic parse trees," *Eng. Appl. Sci. Res.*, vol. 45, no. 4, pp. 290–300, 2018.

[25] H. Chavan, M. Taufik, R. Kadave, and N. Chandra, "Plagiarism detector using machine learning," *Int. J. Res. Eng., Sci. Manage.*, vol. 4, no. 4, pp. 152–154, 2021.

[26] D. Childers and S. Bruton, "'Should it be considered plagiarism?' Student perceptions of complex citation issues," *J. Academic Ethics*, vol. 14, pp. 1–17, Mar. 2016.

[27] K. S. M. Anbananthen, J. K. Krishnan, M. S. Sayeed, and P. Muniapan, "Comparison of stochastic and rule-based POS tagging on Malay online text," *Amer. J. Appl. Sci.*, vol. 14, no. 9, pp. 843–851, Sep. 2017.

[28] R. Padillah, "Ghostwriting: A reflection of academic dishonesty in the artificial intelligence era," *J. Public Health*, vol. 46, no. 1, pp. e193–e194, Mar. 2024.

[29] M. Singh and P. Gupta, "AI-Driven Plagiarism Detection Framework for Academic Documents," *IEEE Access*, 2023., Art. no. 1750004.

[30] P. Sharmila, "A Fuzzy TOPSIS Language Modeling Approach for Plagiarism Severity Assessment," *IEEE Access*, 2025.vol. 54, no. 6, pp. 928–948, Jun. 2022.