Received 4 July 2024, accepted 24 July 2024, date of publication 5 August 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3438434

RESEARCH ARTICLE

FTLM: A Fuzzy TOPSIS Language Modeling Approach for Plagiarism Severity Assessment

P. SHARMILA[1], KALAIARASI SONAI MUTHU ANBANANTHEN[2], NITHYAKALA GUNASEKARAN[1], BAARATHI BALASUBRAMANIAM[2], AND DEISY CHELLIAH[1]

[1]Thiagarajar College of Engineering, Madurai, Tamil Nadu 625015, India
[2]Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia

Corresponding author: Kalaiarasi Sonai Muthu Anbananthen (kalaiarasi@mmu.edu.my)

This work was supported by the Multimedia University.

ABSTRACT Detecting plagiarism poses a significant challenge for academic institutions, research centers, and content-centric organizations, especially in cases involving subtle paraphrasing and content manipulation where conventional methods often prove inadequate. Our paper proposes FTLM (Fuzzy TOPSIS Language Modeling), a novel method for detecting plagiarism within decision science. FTLM integrates language models with fuzzy sorting techniques to assess plagiarism severity by evaluating the similarity of potential solutions to a reference. The method involves two stages: leveraging language modeling to define criteria and alternatives and implementing enhanced fuzzy TOPSIS. Word usage patterns, grammatical structures, and semantic coherence represent fuzzy membership functions. Moreover, pre-trained language models enhance semantic similarity analysis. This approach highlights the benefits of combining fuzzy logic's tolerance for imprecision with the semantic evaluation capabilities of advanced language models, thereby offering a comprehensive and contextually aware method for analyzing plagiarism severity. The experimental results on the benchmark dataset demonstrate effective features that enhance performance on the user-defined severity ranking order.

INDEX TERMS Plagiarism detection, semantic analysis, natural language processing, language modelling, fuzzy TOPSIS.

I. INTRODUCTION

Plagiarism has become a significant ethical problem in academic and professional environments, as it can take various forms that are difficult to detect. Despite the development of numerous methods for detecting plagiarism, not all cases can be accurately identified. One common type of plagiarism is near-copying, which involves copying content from a source without proper citation. Disguised plagiarism occurs when individuals rephrase sentences and replace words with synonyms to make the content appear original. Translated plagiarism happens when a source document is translated into another language without correctly citing the source. Idea plagiarism involves changing the structure and wording of a document while covering the same topic as the original source.

Research shows that while near-copying can be effectively identified by plagiarism detection methods, detecting covert, translated, and idea plagiarism remains challenging [1], [2].

Let us examine the following sentence: ''X and Y read the story and the newspaper.'' and ''X read the story, and Y read the newspaper.'' These two sentences use the same words but convey entirely different meanings. Consider these sentences as well: ''The weather is pleasant today.'' and ''Today, the climate is pleasant.'' Although they use different terms, both sentences convey the same general concept. The terminology is different, but the context connects them. Understanding the context and considering both the syntactic (sentence structure) and semantic (meaning) aspects are essential for correctly interpreting and comparing texts. The examples above demonstrate how context influences plagiarism and similarity in addition to the text or the occurrence of words. Therefore, lexical context is crucial for this work, and we should consider the syntactic and semantic aspects of the text.

The associate editor coordinating the review of this manuscript and approving it for publication was Giambattista Gruosso.

TABLE 1. Characteristics of different criteria.

Decision Model	Advantages	Disadvantages
WSM [15]	Robust while dealing with single-dimensional issues Simple to employ Computation and calculation are straightforward. Intelligible and unambiguous	Multidimensional problems are difficult. Occasionally unrealistic The outcome may not be rational.
ELECTRE [16]	Consider uncertainty and employ fuzzy logic. Maintain coherence with established rules. Balancing these ensures effective decision-making in complex situations.	Explaining the process and outcomes can pose challenges. Outranking may obscure strengths and weaknesses. Time consumption is a factor.
PROMETHEE [17]	User-friendliness and transparency Does not require the assumption that the criteria are proportionate	Does not provide a clear method on which to assign weights
AHP [18]	Flexible, understandable model ensuring consistency with interdependent criteria. Hierarchical, accommodating both quantitative and qualitative factors, refined judgments, adaptable across contexts, and fuzzy numbers.	Challenges include rank reversal, potential information loss, managing uncertainty, computational demands, and reliance on decision-maker preferences.
TOPSIS [19]	Easy implementation Suitable for large-scale data Simplicity Constant number of steps in the process	Euclidean distance disregards criteria correlation. Ensuring consistency. Vector normalization relies on criterion function units. Risk of rank reversal. Identifying maximum and minimum values is necessary.

The direction of this work is towards context analysis based on natural language processing (NLP) for content similarity. The NLP-based plagiarism detection method uses syntactic parsing to analyze suspicious content and discover terms with the same meaning.

Latent Semantic Analysis (LSA) measures word similarity by calculating the cosine value of two vectors reflected by the words. The lower the value, the more similar the words are [3]. Lexical analysis plagiarism detection [4] methods for character n-grams often detect plagiarism based on writing style but fail to detect plagiarism in short lines. Plagiarism detection [5] is done by analyzing the similarity of individual words using syntactic dependency trees.

Recently, neural network-based content similarity evaluations have become more accurate but computationally expensive [6]. Advanced approaches, such as transformer architectures, are used to predict similarity and are classified as end-to-end approaches [7]. NLP-based paraphrase detection benefits from highly parameterized, pre-trained models to detect plagiarism [8], [9]. By considering the part-of-speech (POS) element in NLP-based syntactic parsing, we can analyze suspicious content and detect plagiarism terms with the same meaning [10], [11]. The degree of plagiarism and its penalty are discussed using the Multiple Criteria Decision Making (MCDM) method by selecting the best alternative from several criteria or factors [12], [13], [14]. This method is particularly useful when there is no clear criterion for evaluating and comparing alternatives, and decision-makers must weigh competing goals or preferences. It is important to note that MCDM is not related to anti-plagiarism software that detects and prevents plagiarism in written content. Instead, MCDM is a decision-making approach used in complex multi-criteria decisions. Some common MCDM methods are explained in Table 1. We used the TOPSIS [19] distance-based method for easier implementation.

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.

III. 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 citation-based 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

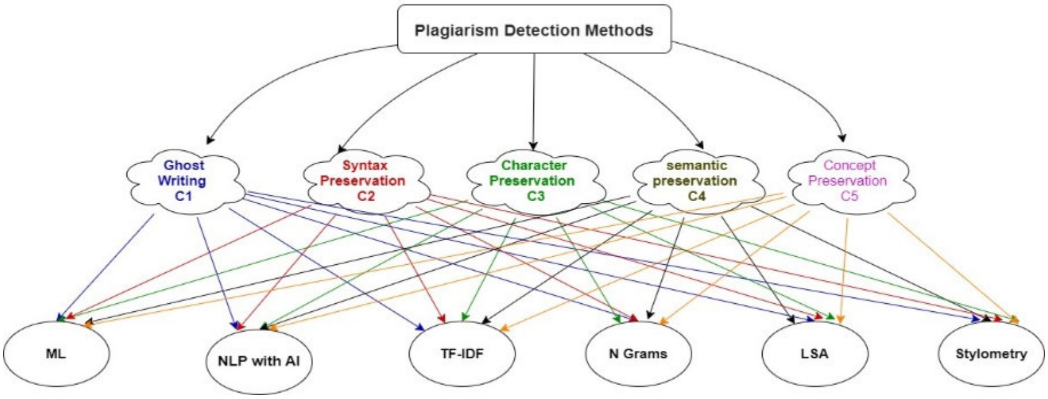


FIGURE 1. Criteria and alternatives for language modeling phase.

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.

IV. PROPOSED FTLM DETECTION MODEL

The FTLM 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 1.

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.

Context and structure: Semantic-preserving recognition considers not only individual words but also the context and structure of the text. It examines how words are used in sentences and how the sentences are structured in a paragraph or document. In this way, cases can be identified in which the text is structurally similar, even if words are rearranged or synonyms are used.

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.

Step 1: Generate a matrix of alternatives.

The fuzzy TOPSIS approach is used in this study to analyze five criteria (variables) and six alternatives. The category represents the different types of criteria. Suppose there are γ members of the decision team. With respect to the α-th alternative to the β-th criterion, if the fuzzy score and priority weighting of the γ-th evaluation expert are: $\dddot{R}^{\gamma}_{\alpha\beta} = \left(x^{\gamma}_{\alpha\beta}, y^{\gamma}_{\alpha\beta}, z^{\gamma}_{\alpha\beta}\right)$ and $\dddot{W}^{\gamma}_{\beta} = \left(w^{\gamma}_{\beta1}, w^{\gamma}_{\beta2}, w^{\gamma}_{\beta3}\right)$ correspondingly where if $\alpha = 1, 2, \ldots$, m, and $\beta = 1, 2, \ldots$, n,

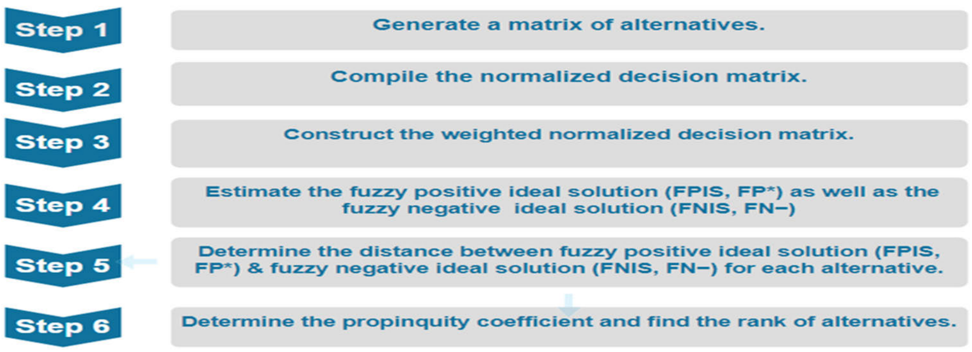

Step 1
Generate a matrix of alternatives.
Step 2
Compile the normalized decision matrix.
Step 3
Construct the weighted normalized decision matrix.
Step 4
Estimate the fuzzy positive ideal solution (FPIS, FP*) as well as the fuzzy negative ideal solution (FNIS, FN−)
Step 5
Determine the distance between fuzzy positive ideal solution (FPIS, FP*) & fuzzy negative ideal solution (FNIS, FN−) for each alternative.
Step 6
Determine the propinquity coefficient and find the rank of alternatives.

FIGURE 2. Workflow of fuzzy TOPSIS phase.

then the combined fuzzy ratings R ···_αβ^γ of the alternatives (α) concerning every criterion (β) are identified by $\dddot{R}_{\alpha\beta}^{\gamma} = \left(x_{\alpha\beta}^{\gamma}, y_{\alpha\beta}^{\gamma}, z_{\alpha\beta}^{\gamma}\right)$

TABLE 2. Characteristics of different criteria.

	Criteria	Category
1	C1	+
2	C2	+
3	C3	+
4	C4	+
5	C5	+

TABLE 3. Fuzzy scale.

Code	Linguistic Terms	T	M	H
1	Very low	1	1	3
2	Low	1	3	5
3	Medium	3	5	7
4	High	5	7	9
5	Very high	7	9	9

Table 2 shows the category of the criterion and the weighting of the individual criteria. H, M, & T represent a triangular fuzzy number (TFN). H, M and T represent the highest possible, most probable, and least probable values, respectively. The fuzzy scale applied to the model is shown in Table 3.

Step 2: Compute the normalized decision matrix.

According to the positive and negative ideal alternatives, a normalized decision matrix may also be found using the following relation:

Positive ideal solution

$$P_{\alpha\beta} = \left(\frac{x_{\alpha\beta}}{z_{\beta}^{+}}, \frac{y_{\alpha\beta}}{z_{\beta}^{+}}, \frac{z_{\alpha\beta}}{z_{\beta}^{+}}\right); z_{\beta}^{+} = \max_{\beta} z_{\alpha\beta}; \tag{1}$$

Negative ideal solution

$$N_{\alpha\beta} = \left(\frac{x_{\beta}^{-}}{x_{\alpha\beta}}, \frac{x_{\beta}^{-}}{y_{\alpha\beta}}, \frac{x_{\beta}^{-}}{z_{\alpha\beta}}\right); x_{\beta}^{-} = \min_{\beta} x_{\alpha\beta}; \tag{2}$$

Step 3: Construct the weighted normalized decision matrix.

Given the different weights assigned to each criterion in the normalized fuzzy decision matrix, the weighted normalized decision matrix may be produced by multiplying the weights of each criterion by the following Eqn.3.

$$\underset{\alpha\beta}{\dddot{V}} = \underset{\alpha\beta}{\dddot{R}} * \underset{\alpha\beta}{\dddot{W}} \tag{3}$$

where $\dddot{W}_{\alpha\beta}$ denotes the weight of the criterion Z_{β}

Step 4: Estimate the fuzzy positive ideal solution (FPIS, FP^{+}) as well as the fuzzy negative ideal solution (FNIS, FN^{-})

The Fuzzy Positive Ideal Solution (FPIS) and Fuzzy Negative Ideal Solution (FNIS) of the alternatives may be well-defined as Eqn.4 and 5, respectively:

$$\begin{aligned} FP^{+} &= \left\{\dddot{v}_{1}^{+}, \dddot{v}_{2}^{+}, \ldots., \dddot{v}_{n}^{+}\right\} \\ &= \left\{\left(\max_{\beta} v_{\alpha\beta} | \alpha\epsilon P\right), \left(\min_{\beta} v_{\alpha\beta} | \alpha\epsilon N\right)\right\} \end{aligned} \tag{4}$$

$$\begin{aligned} FN^{-} &= \left\{\dddot{v}_{1}^{-}, \dddot{v}_{2}^{-}, \ldots., \dddot{v}_{n}^{-}\right\} \\ &= \left\{\left(\max_{\beta} v_{\alpha\beta} | \alpha\epsilon P\right), \left(\min_{\beta} v_{\alpha\beta} | \alpha\epsilon N\right)\right\} \end{aligned} \tag{5}$$

where FP^{+} is the maximal value of α for all the alternatives and FP^{-} is the minimal value of α for all the alternatives. P and N symbolize the corresponding positive and negative ideal solutions, respectively.

Step 5: Estimate The Amount Of Space Between Each Interim Solution And The Fuzzy Negative Ideal Solution FN^{-} And Between Each Option And The Fuzzy Positive Ideal Solution FP^{+}.

To calculate the distance between each option and FPIS and between every alternative and FNIS, Eqn. 6 and 7, respectively, are used.

$$DP_{\alpha}^{+} = \sum_{\beta=1}^{n} d_{e}\left(\dddot{v}_{\alpha\beta}, \dddot{v}_{\beta}^{+}\right); \alpha = 1, 2, \ldots.m \tag{6}$$

$$DN_{\alpha}^{-} = \sum_{\beta=1}^{n} d_{e}\left(\dddot{v}_{\alpha\beta}, \dddot{v}_{\beta}^{-}\right); \alpha = 1, 2, \ldots.m \tag{7}$$

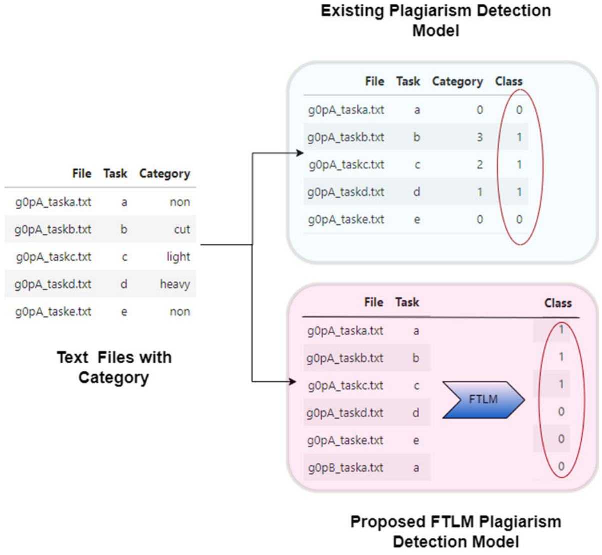
$$d_e\left(\ddot{\mathrm{P}}_1, \ddot{P}_2\right) = \sqrt{\frac{1}{3}\left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right]} \tag{8}$$

Remember that both $d_e\left(v_{\alpha\beta}, \dddot{v}_{\alpha\beta}^{+}\right)$ and $d_e\left(v_{\alpha\beta}, \dddot{v}_{\alpha\beta}^{-}\right)$ are distinct numbers.

Step 6: Along with ranking the selections, find the closeness coefficient.

The following formula could be used to get the closeness coefficient of each option:

$$CC_\alpha = \frac{DN_\alpha^-}{DN_\alpha^- + DP_\alpha^+} \tag{9}$$


Existing Plagiarism Detection Model
Text Files with Category
FTLM
Proposed FTLM Plagiarism Detection Model

File	Task	Category
g0pA_taska.txt	a	non
g0pA_taskb.txt	b	cut
g0pA_taskc.txt	c	light
g0pA_taskd.txt	d	heavy
g0pA_taske.txt	e	non

File	Task	Category	Class
g0pA_taska.txt	a	0	0
g0pA_taskb.txt	b	3	1
g0pA_taskc.txt	c	2	1
g0pA_taskd.txt	d	1	1
g0pA_taske.txt	e	0	0

File	Task	Class
g0pA_taska.txt	a	1
g0pA_taskb.txt	b	1
g0pA_taskc.txt	c	1
g0pA_taskd.txt	d	0
g0pA_taske.txt	e	0
g0pB_taska.txt	a	0

FIGURE 3. Sample output.

Different sets of criteria can be used to evaluate and rank alternates using various methods. Every strategy has advantages over the others as well as disadvantages. The fuzzy TOPSIS technique has the benefits of being transparent in its mathematical expression, easily illustrating human priorities, and facilitating direct and explicit trade-offs between various criteria [30]. The strategy is also classified as a compromising notion, meaning that even though there is never an ideal circumstance, finding a solution with optimal values for every criterion is still possible. Consequently, this research study uses fuzzy TOPSIS with a triangular membership value to assess various plagiarism detection techniques.

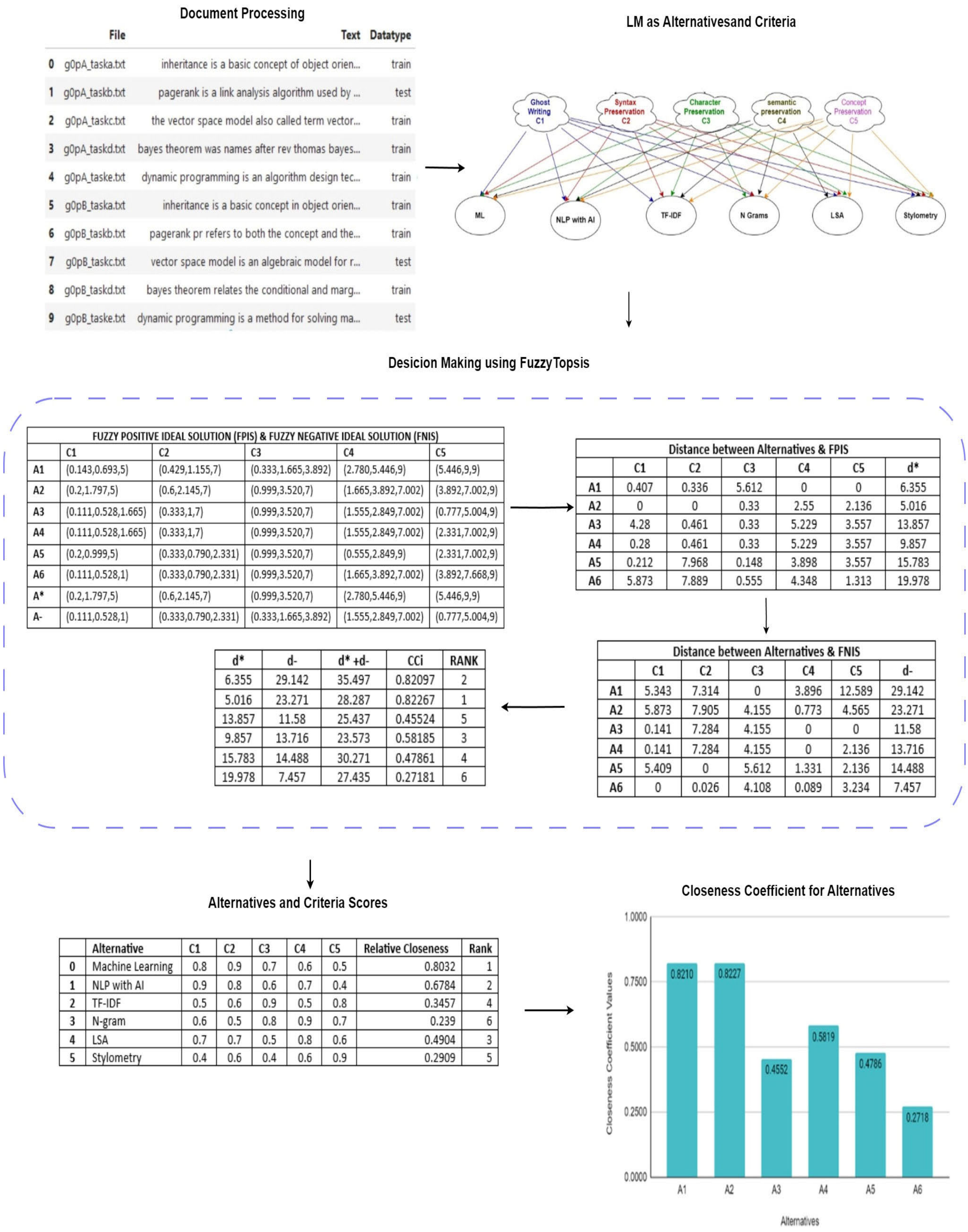
V. DATASET AND IMPLEMENTATION

The dataset is made of multiple text files; each text file is associated with one **task (tasks A-E) and one category** of plagiarism. The dataset categorizes text files into five labels: "cut" for direct copying, "light" for paraphrasing, "heavy" for challenging plagiarism, "non" for non-plagiarized, and "orig" for original source text. The "non" category indicates no plagiarism, while "orig" serves for comparison purposes. Sample output data is shown in Figure.3

Algorithm 1 FTLM for Plagiarism Detection

```
Input:
Initialization:
  For each document in Documents:
    Documents: Set of documents including reference and candidate documents.
    Criteria: Defined based on language modeling features
      (word usage patterns, grammatical structures, semantic coherence).
    Weights: Importance assigned to each criterion.
Output:
Ranked List: Documents ranked by their similarity and severity of plagiarism compared to the reference.
Pre-processing and Feature Extraction:
  For each document in Documents:
    Tokenize and normalize the document.
    Extract language modeling features (word usage patterns, grammatical structures, semantic coherence).
    Store the extracted features in data structures (e.g., arrays, matrices).
Fuzzy Logic Integration and Criteria Definition:
  For each criterion in Criteria:
    Define fuzzy membership functions to handle linguistic uncertainty.
    Calculate fuzzy membership values for each document based on the defined criteria.
Construct Decision Matrix:
Initialize a decision matrix with dimensions (number of documents) x (number of criteria).
  For each document and criterion:
    Populate the decision matrix with the calculated fuzzy membership values.
Normalization:
  For each criterion:
    Normalize the values in the respective column of the decision matrix to ensure comparability.
Weighted Decision Matrix:
  For each criterion and document:
    Apply weights to the normalized values in the decision matrix according to the predefined weights.
Identify Ideal Solutions:
Initialize variables to store the Positive Ideal Solution (PIS) and Negative Ideal Solution (NIS).
  For each criterion:
      Determine the maximum (PIS) and minimum (NIS) values across all documents for the weighted decision matrix.
Calculate Distance to Ideal Solutions:
Initialize arrays to store distances from each document to the PIS and NIS.
  For each document:
    Calculate distances to the PIS and NIS using a suitable distance metric (e.g., Euclidean distance).
Relative Closeness Calculation:
Initialize an array to store relative closeness coefficients for each document.
  For each document:
    Calculate the relative closeness coefficient based on distances to the PIS and NIS.
Ranking and Output:
Sort documents based on their relative closeness coefficients in descending order.
Output a ranked list where higher coefficients indicate higher severity of plagiarism.
```

Data Source link: https://s3.amazonaws.com/video.udacity-data.com/topher/2019/January/5c4147f9_data/data.zip



Document Processing

	File	Text	Datatype
0	g0pA_taska.txt	inheritance is a basic concept of object orien...	train
1	g0pA_taskb.txt	pagerank is a link analysis algorithm used by ...	test
2	g0pA_taskc.txt	the vector space model also called term vector...	train
3	g0pA_taskd.txt	bayes theorem was names after rev thomas bayes...	train
4	g0pA_taske.txt	dynamic programming is an algorithm design tec...	train
5	g0pB_taska.txt	inheritance is a basic concept in object orien...	train
6	g0pB_taskb.txt	pagerank pr refers to both the concept and the...	train
7	g0pB_taskc.txt	vector space model is an algebraic model for r...	test
8	g0pB_taskd.txt	bayes theorem relates the conditional and marg...	train
9	g0pB_taske.txt	dynamic programming is a method for solving ma...	test

LM as Alternativesand Criteria

Desicion Making using FuzzyTopsis

FUZZY POSITIVE IDEAL SOLUTION (FPIS) & FUZZY NEGATIVE IDEAL SOLUTION (FNIS)					
	C1	C2	C3	C4	C5
A1	(0.143,0.693,5)	(0.429,1.155,7)	(0.333,1.665,3.892)	(2.780,5.446,9)	(5.446,9,9)
A2	(0.2,1.797,5)	(0.6,2.145,7)	(0.999,3.520,7)	(1.665,3.892,7.002)	(3.892,7.002,9)
A3	(0.111,0.528,1.665)	(0.333,1,7)	(0.999,3.520,7)	(1.555,2.849,7.002)	(0.777,5.004,9)
A4	(0.111,0.528,1.665)	(0.333,1,7)	(0.999,3.520,7)	(1.555,2.849,7.002)	(2.331,7.002,9)
A5	(0.2,0.999,5)	(0.333,0.790,2.331)	(0.999,3.520,7)	(0.555,2.849,9)	(2.331,7.002,9)
A6	(0.111,0.528,1)	(0.333,0.790,2.331)	(0.999,3.520,7)	(1.665,3.892,7.002)	(3.892,7.668,9)
A*	(0.2,1.797,5)	(0.6,2.145,7)	(0.999,3.520,7)	(2.780,5.446,9)	(5.446,9,9)
A-	(0.111,0.528,1)	(0.333,0.790,2.331)	(0.333,1.665,3.892)	(1.555,2.849,7.002)	(0.777,5.004,9)

Distance between Alternatives & FPIS						
	C1	C2	C3	C4	C5	d*
A1	0.407	0.336	5.612	0	0	6.355
A2	0	0	0.33	2.55	2.136	5.016
A3	4.28	0.461	0.33	5.229	3.557	13.857
A4	0.28	0.461	0.33	5.229	3.557	9.857
A5	0.212	7.968	0.148	3.898	3.557	15.783
A6	5.873	7.889	0.555	4.348	1.313	19.978

Distance between Alternatives & FNIS						
	C1	C2	C3	C4	C5	d-
A1	5.343	7.314	0	3.896	12.589	29.142
A2	5.873	7.905	4.155	0.773	4.565	23.271
A3	0.141	7.284	4.155	0	0	11.58
A4	0.141	7.284	4.155	0	2.136	13.716
A5	5.409	0	5.612	1.331	2.136	14.488
A6	0	0.026	4.108	0.089	3.234	7.457

d*	d-	d* +d-	CCi	RANK
6.355	29.142	35.497	0.82097	2
5.016	23.271	28.287	0.82267	1
13.857	11.58	25.437	0.45524	5
9.857	13.716	23.573	0.58185	3
15.783	14.488	30.271	0.47861	4
19.978	7.457	27.435	0.27181	6

Alternatives and Criteria Scores

	Alternative	C1	C2	C3	C4	C5	Relative Closeness	Rank
0	Machine Learning	0.8	0.9	0.7	0.6	0.5	0.8032	1
1	NLP with AI	0.9	0.8	0.6	0.7	0.4	0.6784	2
2	TF-IDF	0.5	0.6	0.9	0.5	0.8	0.3457	4
3	N-gram	0.6	0.5	0.8	0.9	0.7	0.239	6
4	LSA	0.7	0.7	0.5	0.8	0.6	0.4904	3
5	Stylometry	0.4	0.6	0.4	0.6	0.9	0.2909	5

Closeness Coefficient for Alternatives

FIGURE 4. Visual illustration of proposed FTLM model.

TABLE 4. Combined decision matrix.

weights	(1,3,5)	(1,1,3)	(3,5,7)	(5,7,9)	(7,9,9)
	C1	C2	C3	C4	C5
A1	(1,4.33,7)	(1,4.33,7)	(1,3,7)	(5,7,9)	(7,9,9)
A2	(1,1.667,5)	(1,2.33,5)	(3,6.33,9)	(3,5,7)	(5,7,9)
A3	(3,5.667,9)	(1,5,9)	(3,6.33,9)	(1,3.667,7)	(1,5,9)
A4	(5,8.33,9)	(3,5,7)	(5,7.669,9)	(3,6.33,9)	(3,7,9)
A5	(1,3,5)	(3,6.33,9)	(3,7.669,9)	(1,3.667,9)	(3,7,9)
A6	(5,8.33,9)	(3,8.33,9)	(3,5.667,9)	(1,4.33,7)	(5,7.669,9)

TABLE 5. Normalized decision matrix.

weights	(1,3,5)	(3,5,7)	(3,5,7)	(5,7,9)	(7,9,9)
	C1	C2	C3	C4	C5
A1	(0.143,0.231,1)	(0.143,0.231,1)	(0.111,0.333,0.556)	(0.556,0.778,1)	(0.778,1,1)
A2	(0.2,0.599,1)	(0.2,0.429,1)	(0.333,0.704,1)	(0.333,0.556,0.778)	(0.556,0.778,1)
A3	(0.111,0.176,0.333)	(0.111,0.2,1)	(0.333,0.704,1)	(0.111,0.407,0.778)	(0.111,0.556,1)
A4	(0.111,0.176,0.333)	(0.111,0.2,1)	(0.333,0.704,1)	(0.111,0.407,0.778)	(0.333,0.778,1)
A5	(0.2,0.333,1)	(0.111,0.158,0.333)	(0.333,0.852,1)	(0.111,0.407,1)	(0.333,0.778,1)
A6	(0.111,0.176,0.2)	(0.111,0.176,0.333)	(0.333,0.631,1)	(0.111,0.481,0.778)	(0.556,0.852,1)

TABLE 6. Weighted normalized decision matrix.

	C1	C2	C3	C4	C5
A1	(0.143,0.693,5)	(0.429,1.155,7)	(0.333,1.665,3.892)	(2.780,5.446,9)	(5.446,9,9)
A2	(0.2,1.797,5)	(0.6,2.145,7)	(0.999,3.520,7)	(1.665,3.892,7.002)	(3.892,7.002,9)
A3	(0.111,0.528,1.665)	(0.333,1,7)	(0.999,3.520,7)	(0.555,2.849,7.002)	(0.777,5.004,9)
A4	(0.111,0.528,1.665)	(0.333,1,7)	(0.999,3.520,7)	(0.555,2.849,7.002)	(2.331,7.002,9)
A5	(0.2,0.999,5)	(0.333,0.790,2.331)	(0.999,4.260,7)	(0.555,2.849,9)	(2.331,7.002,9)
A6	(0.111,0.528,1)	(0.333,0.880,2.331)	(0.999,3.155,7)	(0.555,3.367,7.002)	(3.892,7.668,9)

Also visual illustrations of our proposed FTLM model shown in Figure.4

VI. RESULT AND DISCUSSION

The investigators obtained the values using the regular fuzzy scale displayed in Table 3 and Equations (1) – (9). The answers are evaluated according to several criteria, and the results of the decision matrix are shown below. All expert opinions are provided in Table 4 as the combined decision matrix. The Normalized Decision Matrix with weights for criteria C1 through C5 and options A1 through A6 is shown in Table 5.

The values represent the normalized scores for each alternative in relation to the corresponding criteria inside the matrix. The weights for each criterion (in brackets) indicate its importance.

TABLE 7. Fuzzy positive ideal solution (FPIS) & fuzzy negative ideal solution (FNIS).

	FP^{+}	FN^{-}
C1	(0.2,1.797,5)	(0.111,0.528,1)
C2	(0.6,2.145,7)	(0.333,0.790,2.331)
C3	(0.999,4.260,7)	(0.333,1.665,3.892)
C4	(2.780,5.446,9)	(0.555,2.849,7.002)
C5	(5.446,9,9)	(0.777,5.004,9)

For instance, the numbers in columns C1 through C5 of the first row (A1) represent the normalized scores after considering the allocated weights (1,3,5), (3,5,7), (5,7,9), and (7,9,9) for the corresponding criterion. The values in the

TABLE 8. (a). Distance between alternatives & FPIS, (b). Distance between alternatives & FNIS.

(a)

	C1	C2	C3	C4	C5	DP_{α}^{+}
A1	0.407	0.336	5.612	0	0	6.355
A2	0	0	0.33	2.55	2.136	5.016
A3	4.28	0.461	0.33	5.229	3.557	13.857
A4	0.28	0.461	0.33	5.229	3.557	9.857
A5	0.212	7.968	0.148	3.898	3.557	15.783
A6	5.873	7.889	0.555	4.348	1.313	19.978

(b)

	C1	C2	C3	C4	C5	DN_{α}^{-}
A1	5.343	7.314	0	3.896	12.589	29.142
A2	5.873	7.905	4.155	0.773	4.565	23.271
A3	0.141	7.284	4.155	0	0	11.58
A4	0.141	7.284	4.155	0	2.136	13.716
A5	5.409	0	5.612	1.331	2.136	14.488
A6	0	0.026	4.108	0.089	3.234	7.457

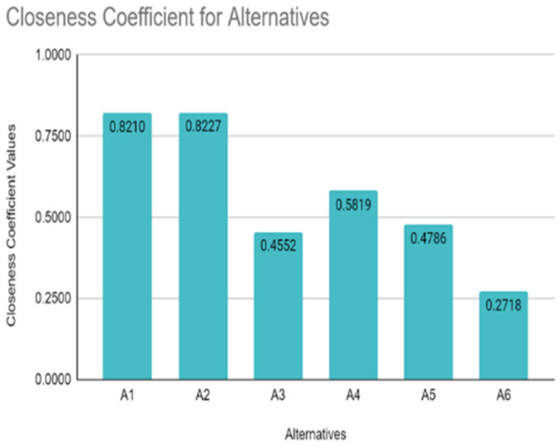


FIGURE 5. Closeness coefficient graph.

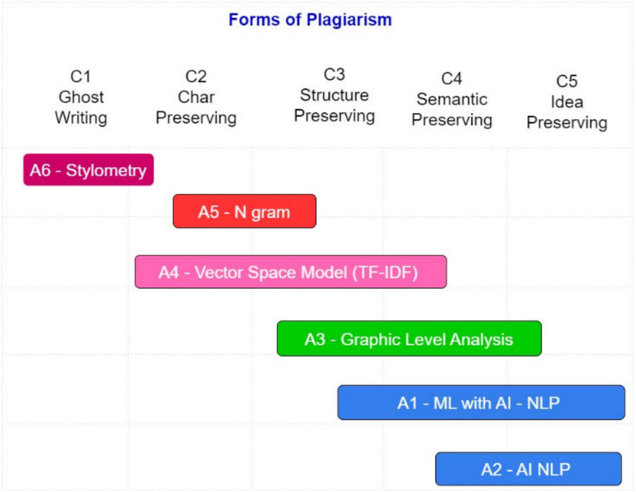


FIGURE 6. Observed plagiarism for chosen alternatives.

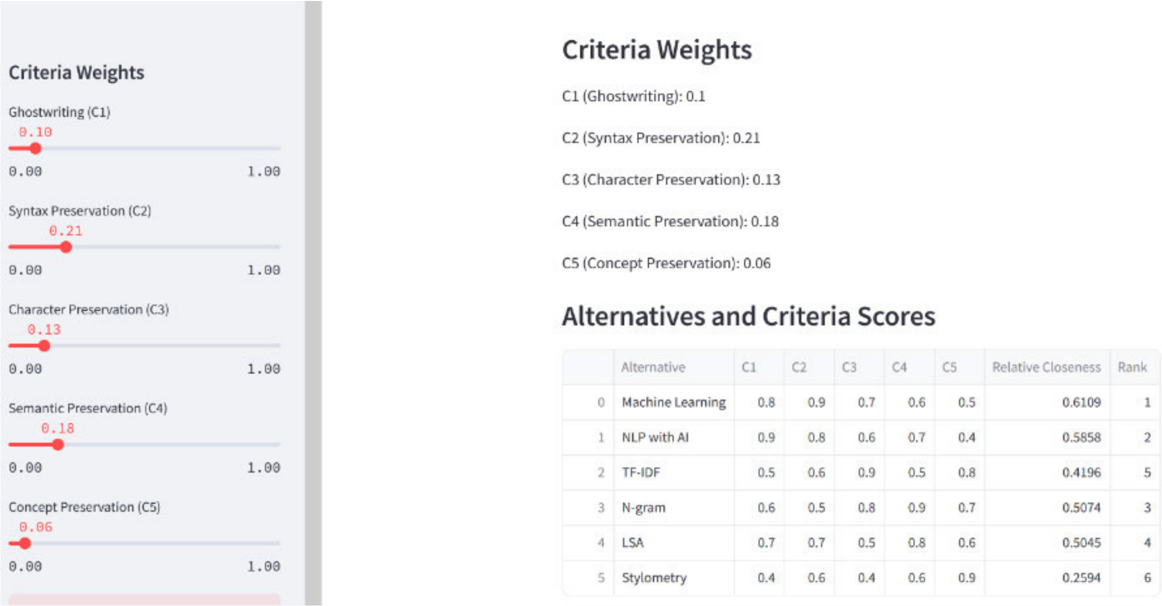
subsequent rows (A2 to A6), which show the normalized scores for each choice and criterion, should be interpreted similarly. The normalization process, which considers the weights provided to each alternative in relation to its relative importance, makes a standardized comparison of alternatives across various criteria possible.

The Weighted Normalized Decision Matrix for criteria C1through C5 and alternatives A1 through A6 is shown in Table 6. Triple values, which include the specified weights, are present in each cell and represent the weighted and normalized scores. This matrix makes it easier to evaluate alternatives thoroughly by taking relative performance and the importance of the criteria into account.

Additionally, Table 7 presents FPIS and FNIS for criteria C1 to C5. These solutions represent the ideal and least desirable values for each criterion.

Tables 8(a) & 8(b) indicate the distances to both positive and negative ideal solutions. Table 8 (a) displays the distances between alternatives A1 to A6 and the Fully Preferable Ideal Solution (FPIS) across criteria C1 to C5. The values in each cell represent the calculated distances, providing insights into the relative performance of each alternative with respect to the ideal solution.

From Tables 8(a) & 8(b), we observed the distance between each alternative's Fuzzy Positive and Fuzzy Negative Ideal



Criteria Weights

Ghostwriting (C1) 0.10 (0.00 – 1.00)

Syntax Preservation (C2) 0.21 (0.00 – 1.00)

Character Preservation (C3) 0.13 (0.00 – 1.00)

Semantic Preservation (C4) 0.18 (0.00 – 1.00)

Concept Preservation (C5) 0.06 (0.00 – 1.00)

Criteria Weights

C1 (Ghostwriting): 0.1

C2 (Syntax Preservation): 0.21

C3 (Character Preservation): 0.13

C4 (Semantic Preservation): 0.18

C5 (Concept Preservation): 0.06

Alternatives and Criteria Scores

	Alternative	C1	C2	C3	C4	C5	Relative Closeness	Rank
0	Machine Learning	0.8	0.9	0.7	0.6	0.5	0.6109	1
1	NLP with AI	0.9	0.8	0.6	0.7	0.4	0.5858	2
2	TF-IDF	0.5	0.6	0.9	0.5	0.8	0.4196	5
3	N-gram	0.6	0.5	0.8	0.9	0.7	0.5074	3
4	LSA	0.7	0.7	0.5	0.8	0.6	0.5045	4
5	Stylometry	0.4	0.6	0.4	0.6	0.9	0.2594	6

FIGURE 7. Sample output by varying criteria weights.

Solutions. Larger distances from FPIS signify greater deviation from ideal conditions, while larger distances from FNIS indicate closer proximity to undesirable conditions. A3, A5, and A6 exhibit higher distances from FPIS, suggesting they are relatively further from ideal conditions shown in Table 8 (a). Conversely, A2 and A1 have smaller distances from FNIS, as shown in Table 8 (b), implying closer proximity to less desirable conditions. These distances provide a quantitative basis for decision-making across defined criteria. These distance values are represented as $DP_{\alpha}^{+}\&DN_{\alpha}^{-}$ shown in Table 9.

TABLE 9. Distance from FPIS & FNIS.

	DP_{α}^{+}	DN_{α}^{-}
A1	6.355	29.142
A2	5.016	23.271
A3	13.857	11.58
A4	9.857	13.716
A5	15.783	14.488
A6	19.978	7.457

Table 10 displays the Closeness Coefficient and rankings for alternatives A1 to A6. The coefficient, ranging from 0 to 1, signifies proximity to the ideal solution. A2 and A1 lead with the highest coefficients (0.8227 and 0.8210), while A6 lags with the lowest (0.2718). The rankings reflect the overall performance order of alternatives, offering a concise assessment of their relative closeness to the desired solution across criteria C1 to C5, as shown in Figure 5.

TABLE 10. Closeness coefficient.

	CC_{α}	RANK
A1	0.8210	2
A2	0.8227	1
A3	0.4552	5
A4	0.5819	3
A5	0.4786	4
A6	0.2718	6

A. A OBSERVATION

Applying the fuzzy TOPSIS technique to evaluate methods for detecting academic plagiarism leads to a comprehensive ranking. The analysis in Figure 6 shows that the effectiveness of the methods varies, with A2 (AI with NLP) performing best. This is closely followed by A1 (machine learning), demonstrating its efficiency in detecting plagiarism.

The ranking also includes A4, A5, A3, and A6 in their respective positions in the hierarchy of effectiveness. This detailed assessment provides valuable insights into each method's comparative strengths and weaknesses, helping decision-makers select the most appropriate approach for their specific needs.

Due to its superior performance, A2 (AI with NLP) is the preferred choice for academic plagiarism detection. Its ability to delve deep into the semantics and ideas of textual content sets it apart and makes it a robust solution for accurately and efficiently detecting plagiarism cases.

Figure 7 depicts a sample output screenshot. Assigning variable weights (preference or importance) to each criterion changes the alternatives' rank, as demonstrated by our proposed FTLM model. When weights are altered, the model recalculates the document ranks. This means that documents previously deemed less similar or more plagiarized may rank differently when the criteria weights are adjusted. This dynamic nature makes the concept adaptable to many circumstances and requirements.

VII. CONCLUSION

The use of FTLM represents a significant advancement in addressing the challenges of plagiarism detection, particularly in cases involving subtle paraphrasing. The two-stage procedure, which combines fuzzy sorting algorithms with language models, effectively captures minute linguistic details and semantic coherence. The method's use of pre-trained language models enhances semantic similarity analysis. This innovative approach advances the field of plagiarism detection by integrating the semantic evaluation capabilities of advanced language models with the imprecision tolerance of fuzzy logic. In decision science, FTLM stands out as a rigorous and contextually aware method for determining the degree of plagiarism.

In the future, FTLM could be adapted to apply to different alternatives with conflicting needs. Additionally, the strategy holds potential for broader real-world applications. This detection method may also be extended for unstructured and multilingual paraphrased plagiarism detection.

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. Arabi and M. Akbari, "Improving plagiarism detection 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 transformer-based 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, and Y. Jiang, "Multicriteria decision making with cognitive 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] T. Arslan, "A weighted Euclidean distance based TOPSIS method for modeling public subjective judgments," *Asia–Pacific J. Oper. Res.*, vol. 34, no. 3, Jun. 2017, Art. no. 1750004.

[30] R. A. Liaqait, S. S. Warsi, M. H. Agha, T. Zahid, and T. Becker, "A multi-criteria decision framework for sustainable supplier selection and order allocation using multi-objective optimization and fuzzy approach," *Eng. Optim.*, vol. 54, no. 6, pp. 928–948, Jun. 2022.

P. SHARMILA received the Ph.D. degree in information and communication engineering from Anna University, Chennai, Tamil Nadu, India. She is currently an Assistant Professor with the Department of Applied Mathematics and Computational Science, Thiagarajar College of Engineering, Madurai, Tamil Nadu. Her research interests include natural language processing, machine learning, and deep learning.

KALAIARASI SONAI MUTHU ANBANANTHEN is currently an Associate Professor with the Faculty of Information Science and Technology, Multimedia University (MMU), Malaysia. She was a Program Coordinator for the Masters of Information Technology and the Coordinator for Business Intelligence Analytics. She has published more than 100 papers in journals, conferences, and book chapters. Her current research interests include data mining, sentiment analysis, artificial intelligence, machine learning, deep learning, and text analytics. She is a reviewer in various Scopus and SCI-indexed technical journals.

NITHYAKALA GUNASEKARAN received the Ph.D. degree from Periyar University, Tamil Nadu, India, in 2021. She is currently an Assistant Professor with the Department of Applied Mathematics and Computational Science, Thiagarajar College of Engineering, Tamil Nadu. Her current research interests include graph theory, multi-criteria decision-making techniques, natural language processes, and predictive analytics. She is a member of ORSI and ISDE.

BAARATHI BALASUBRAMANIAM received the bachelor's degree in nutrition and community health from Universiti Putra Malaysia, in 2022, where she is currently pursuing the master's degree in community nutrition with a focus on data analytics. Holding a professional certification in big data, she is a Research Officer with Multimedia University. Her current research interests include applying NLP, machine learning, and deep learning in the fields of healthcare and nutrition.

DEISY CHELLIAH is currently a Professor and the Head of the Information Technology Department, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India. She completed two projects sponsored by Microsoft and AICTE. She published more than 70 papers in journals and conferences. Her research interests include image analysis and text analytics. She is a member of ISTE and CSI. She acts as a reviewer in various SCI and Scopus-indexed technical journals.