Explainable Prediction Technique for Analyzing Information Disorder Using Fuzzy Rough Sets

Udhayashankar S[1], Muniasamy P[2]

[1,2]*Department of Computer Science*
Vysya College
Ayodhyappattanam (P.O, Ramakrishnapuram, Salem,
Masinaickenpatti, Tamil Nadu 636103
[1]*udhaya1707@gmail.com*
[2]*profmuniasamy@gmail.com*

ABSTRACT

A persistent challenge in the domain of social media pertains to the comprehension and representation of information disorder incidents, which include but are not limited to erroneous information, theories of conspiracy, and racial discrimination. These issues frequently become apparent in the written content of posts that are shared on networking platforms and multiple sites. The complexity of information disorder research stems from the wide range of internet epidemic genres and the diverse effects that such information can have on people. Furthermore, the difficulty is exacerbated by the emergence of generative AI models that are capable of generating information that is difficult to identify as synthetic. The findings can assist analysts and decision-makers in acquiring a more in-depth comprehension of the phenomenon occurring in the field of information disorder. Hence, Information Disorder Prediction using Fuzzy Rough Set (IDP-FRS) has been proposed to predict the disinformation in the collected domain accurately and identify fake and real information. Through the incorporation of fuzzy sets of data into modelling and prediction techniques, the research improves the accuracy and interpretability of models, hence facilitating the discovery of disinformation in a more trustworthy manner. Moreover, the introduction of explainable prediction approaches gives users the ability to make educated choices concerning the appropriateness and dissemination of information, which in turn promotes openness and confidence in predictive models.

Keywords: Fuzzy Rough Set; Information Disorder; Prediction Analysis; Generative AI.

1. Introduction

The incorporation of the Explainable Artificial Intelligence (XAI) technique[1] with fuzzy rough set theory in order to tackle the interpretability difficulties encountered in machine learning, specifically in the context of deep learning models to interpret the complexity of information disorder. A novel information fusion approach is suggested for handling multimodal data, with a specific focus on establishing a multiview multimodal integration framework[2] that preserves a greater amount of dynamic content. A collection of features that are related to the actions that individuals take on social media platforms like Facebook or other internet platforms with the purpose of identifying possible targets of disinformation[3]. Rough set discretization has performed well in many situations because it uses the existing knowledge base without previous information. However, the rough set equivalence class is an ordinary set, making it difficult to express fuzzy data set components[4], and reliability is low for certain complicated data kinds in the large-scale ordering of information environments.

Except for a few earlier works by similar authors that concentrated on the research of societies and the psychological implications of social content, this was the first utilization of oppositional hexagons in the study of information disorders[5]. The hexagon is constructed using set-theoretic criteria based on a fuzzy set[6]; its analysis enables us to discern whether the discourse has a tendency toward or away from empathy. A specific framework of opposition is employed using set measures such as Jaccard and symmetrical difference, the graduated hexagon[7], which computes the individuality, diversity(variation), similarities, and divergence of two separate items portrayed as sets. Studies include fuzzy entropy, which is achieved by employing thresholds for feature selection in a variety of medical datasets, as well as fuzzy rough sets, which are utilized for complexity reduction in feature space in order to prevent samples from being misclassified[8]. In situations where there is a gradual idea of indiscernibility between items, fuzzy rough set theory can be utilized as a technique for dealing with inconsistent data. It accomplishes this by supplying approximations of concepts at both the lower and upper levels. When dealing with traditional fuzzy rough sets, both the minimum and maximum functions are utilized to ascertain the lower and upper approximations[9], respectively.

Numerous works have integrated the conventional TOPSIS method with fuzzy rough sets in order to address fuzzy environments, facilitate the advancement of decision-theoretic domains in Decisions Theories Rough Sets, and encompass a wider range of problems[10,11]. The investigation was conducted as part of a real-life analysis utilizing authentic data and pertained to phenomena associated with oppositional information disorder. Social media platforms employ AI functions as a means to combat the proliferation of propagation of hatred and misinformation. Natural language processing (NLP), an area of AI, employs machine learning (ML) and deep learning (DL) algorithms to acquire data representations for detection, and prediction techniques process vast quantities of online content[12]. A fuzzy inference system[13] is applied for extracting information from emotion classifications like anger and happiness through a machine interface and yields the characteristics of document categories.

The research has the following contributions:

✓ To introduce a novel methodology for information disorder prediction (IDP) using FRS combine fuzzy logic and rough set theory to address uncertainty and ambiguity in text information analysis
✓ To apply fuzzy rough sets to educational systems, particularly in decision-making related to recommended information prediction.
✓ To predict the explainable techniques and make informed decisions regarding information dissemination evidence-based decision-making.

An outline of the research paper is given below. In Section 2, the recent articles using various fuzzy set theory concepts related to information disorder are predicted thoroughly. The implemented explainable techniques and research focus on FRS theory factors with lower and upper approximation are discussed in Section 3. Forecasting information analysis about fake and real information from the news prediction results and improved performance metrics is covered in Section 5, which explores the summarized key findings and scope for future work.

2. Existing Research Methodology

For predicting verbal aggressive information in real-time, the information dynamism was identified by Campos et al. [14], and they developed a fuzzy model to identify instances of verbal aggression in interactions that take place on social media platforms. The model in consideration is a Mamdami Fuzzy Inference Approach(MFIA), and the authors claim that it is dynamic due to the fact that it is able to modify the configuration of its parameters from trapezoidal to triangular based on the requirements. The model helps authorities to substitute the information tracking model and make decisions.

A method for analyzing phenomena associated with information disorder that is based on Rough Set Theory(RST) was handled by Gaeta et al. [15] with certain modifications and extensions. RST is utilized to model reasoning on groups of social networking individuals and sets of data that circulate in social media. Lower and higher representations of a target set, indiscernibility, and neighbourhood binary relations.An increase in the ability to comprehend the consequences of information disorder, which is caused by the circulation of news. This is because the ratio that exists between the cardinality of lower and higher approximate values of a rough set, cardinality deviations of components, and an improvement in their fragmentation or cohesiveness are all factors that contribute to the increased interpretation of these effects.

The Fuzzy Rough Purity Approach (FRPA) is a new information theoretic-based method proposed by Uddin et al. [16] aimed to address the issues that are associated with the conventional category clustering algorithms that are based on RST. It is important to note that the RPA takes into account the informational-theoretic attribute purity of categorical-valued information systems. FRPA is shown to perform better than the baseline algorithms, according to the experiments' findings. is suited for use in real life, as evidenced by the large percentage improvement in terms of time taken at 66.7%, iterations at 83.13%, purity at 10.53%, entropy at 14%, and accuracy at 12.15%, as well as the Rough Reliability of clusters.

Weijian and Yongbin [17] developed an algorithm for Detecting Disinformation(DD) by employing a Multidimensional(M) analysis of information. By integrating the opinions of users and psychological considerations into text-based interactive media, this algorithm generates six dimensions that facilitate thorough content analysis. A precise evaluation method that corresponds to six dimensions has been proposed for the purpose of detecting misinformation—an average accuracy of 95.28%. The results demonstrate that this algorithm is capable of predicting misinformation with utmost importance to promptly identify and combat misinformation in order to mitigate its negative effects. There is a need for additional investigation into the outcomes that are more user-centric through communication.

Han et al. [18] introduced incoherent bipolarity into the crisp proximity space for the first time, proposing the generalized notion of a bipolar-valued fuzzy rough set in two different dimensions, which are considered to reflect the interrelationships among incoherent bipolarity information from various information sources. Despite this, the bipolar-valued ragged fuzzy set in a single universe remains inadequate for solving complex practical problems. The comparative analysis with other established methods pertaining to rough sets underscores the importance of the research, which provided an entirely new viewpoint on inconsistent fuzzy bipolarity for rough set theory and its associated implementations.

3. Proposed methodology

The dataset has been taken from the open-access content provided by the Kaggle website in order to assist researchers in recognizing bogus content and combating the spread of disinformation around COVID-19 healthcare [19]. 35 CSV files make up these files, containing

both real and fraudulent content that has been shared on Twitter under the subject of COVID-19. Through the process of screening and amending the initial data, the dataset that will be used for the experiment has been generated. In particular, the following criteria have been applied. All news information that did not have a date of release has been eliminated because it wasn't possible to organize them in chronological order according to the needs, primarily information that was circulated by a minimum of three individuals and individuals who shared at least three articles have been brought into consideration. The ratio of disinformation to true information has been brought into proportion. In the end, Twitter programming interfaces13 served to extract the data linked to comments related to the information that had been shared on Twitter by the users that had been taken into consideration. This occurred for each news item that has been included in the dataset.

Following the completion of the phase of collecting data, the news in the data has been divided into three distinct rough sets, one per time window, according to the month in which the information first appeared. The users have been provided with three decision-making tables, one for each time instant that was taken into consideration, which were constructed using identical groups of fuzzy rough sets. A variant of set theory that is designed to account for inaccurate information is known as Rough Sets (RS), where the formal approximation of a typical set that is crisp is referred to as a rough set. This approximation is expressed as a combination of a pair of sets that provide the lower and higher equivalents of the original set. The indiscernibility relation is an important notion in the field of RS. This is an example of a binary relation, which illustrates the notion that two items cannot be differentiated from one another based on their distinct characteristics. The properties of polarity and subjectivity result in the generation of average information values. This is because each news item is associated with the average polarity significance, which is determined by combining all the polarity values within the bounds of [-1, 1] of textual information remarks that are connected with the shares of the news item in consideration. The same method is utilized to calculate the mean subjectivity values.

a. Defining Fuzzy Set Theory

This research extends the existing fuzzy set operations where F_S denotes the fuzzy set expressed using the equation (1) below.

$$F_S = \left\{ \frac{\mu(f_i)}{f_i} \right\} \tag{1}$$

Where f_i in Equation (1) denotes the ith element of the set and $\mu(f_i)\epsilon[0,1]$ is its membership degree. Fuzzy logic operators like t-norm are a $fn(T): [0,1]^2$ implies $[0,1]$ that monotonic, commutative and associative properties that satisfy the fuzzy boundaries of information $\min T(a,b) = min\{a,b\}$. The crisp indiscernibility relationship of RS can be transformed into a fuzzy indiscernibility relation, which is one of the possibilities that are offered. This method is commonly referred to as Fuzzy Rough Sets (FRS). Any fuzzy relation typically indicates the degree to which two items are indistinguishable from one another, and a fuzzy indiscernibility correlation is the relationship that is employed. The FRS relation between a and b is the fuzzy tolerance relation.

$$FRS_a(i,j) = \frac{|a(i)-a(j)|}{|a_{max}-a_{min}|} \tag{2}$$

Where a_{max} in Equation (2) defines the minimum and maximum values provided for the attribute a. $FRS_a(i,j)$ defines the properties of symmetric and reflexivity. Let T be the T-norm operator with the fuzzy rough set model for information disorder, which can be given as lower and upper approximations using FRS along with decision attribute

$$FRS_b(i,j) = T(FRS_a(i,j)) \tag{3}$$

The circumstance in Equation (3) is shown to be more realistic as a result of the implementation of a fuzzy indiscernibility relationship, which gives $FRS_b(i,j)$, and it has been observed that the boundary region for both groups appears to be more disordered. That is, different behaviours are taken into consideration, but it consistently tends to reduce its cardinality without decreasing to zero.
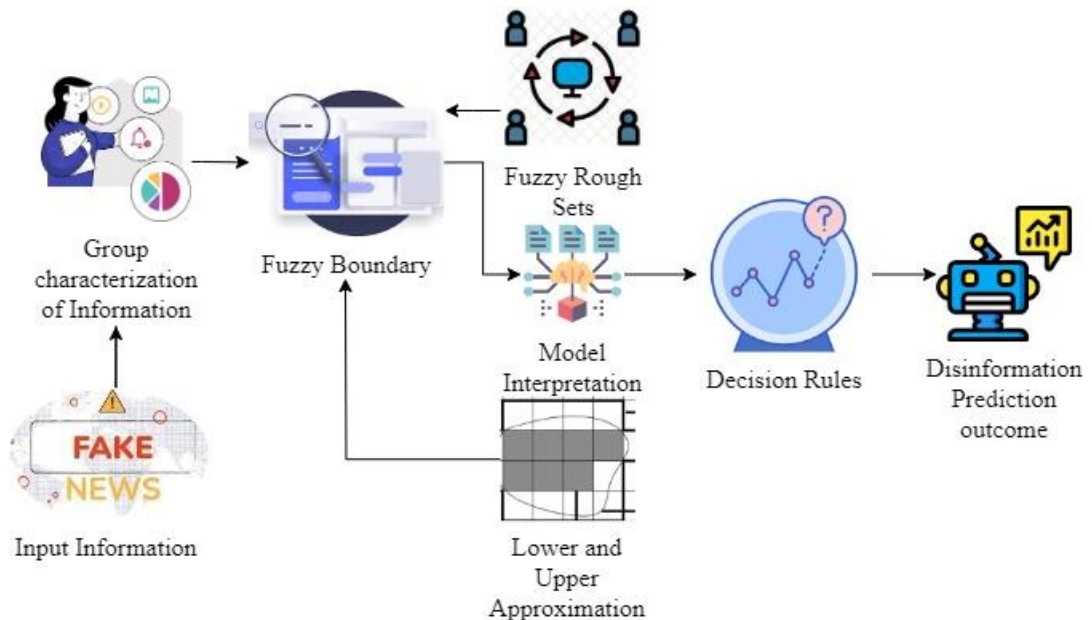

Figure 1. IDP-FRS Illustration of Proposed Model

b. *Fuzzy Knowledge Explainability*

Rough set theory is concerned with the approximation and uncertainty that are present in data. To establish boundary regions of rough sets, lower and higher approximations are taken into consideration. The objective of fuzzy rough set reduction is to simplify the fuzzy rough set by deleting qualities that are redundant or irrelevant while maintaining the attributes' ability to be distinguished from one another. Methods for explaining model inference are reviewed in construction approaches. This category has subcategories: data-driven approaches that extract fresh information from trained models without considering domain experts' experience. Expert knowledge and logic are captured through knowledge-driven techniques, frequently based on the concept of an agent. The explanations are presented in the form of an interactive interface, in which the user, upon selecting a point, the tool displays the characteristics that had the most impact on the classification of that point, as well as the boundary value that is closest to that point.

$$R_r = \frac{|reduct|}{|att|} \tag{4}$$

Hence $reduct$ in Equation (4) represents the reduced set of attributes, and att gives the total number of polarity and subjectivity of the fake and real information order gathered from the dataset.
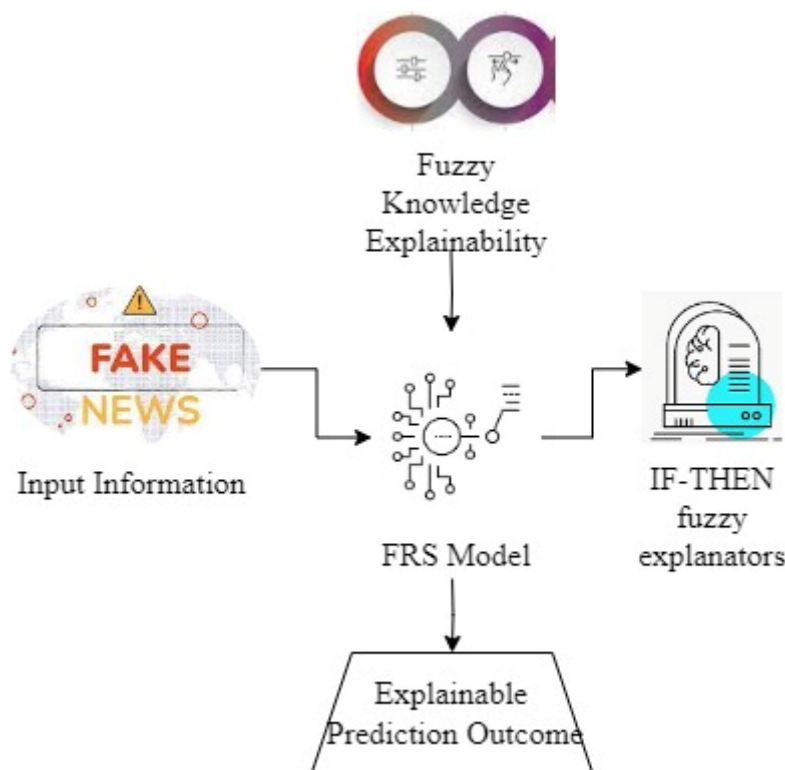
Figure 2. Prediction of Information Disorder using FRS

Figure 2 discusses the gathered dataset with the combination of both fake and real information news articles with the knowledge explainability using ML techniques with FRS approximations followed using IF-THEN fuzzy explanators in the design model. The applied fuzzy boundary range follows the composition of polarity and subjectivity measures to make a perfect prediction about misinformation.

c. Accurate Detection of Information Disorder

Target participants of users referring to the set of news information Inf need to be approximated using the neighbour correlation $\{X/Y\}$ to obtain a lower approximation with a fuzzy boundary predicted with accuracy of each specified information class specified as

$$Inf = < \frac{|lower(x_1)|}{|upper(x_1)|}, ..., \frac{|lower(x_n)|}{|upper(x_n)|} > \tag{5}$$

Where n in Equation (5) gives the number of unique combinations of the information classified on decision features $D = \{d_1, ..., d_n\}$ Along with S tells the sentimental conditions belong to the user's view on specified news topics and nature. Each piece of information is categorized on the news $i \epsilon Inf$.

$$s = \nabla_{ur \epsilon U} f(x, ur) \tag{6}$$

Where ∇ in Equation (6) indicates the aggregation depends on the resultant value of $f(x, ur)$ is a function that is used to predict s from the content of an ur in response to the Inf. Different types of attributes like opinions, emotional cases, and polarity segments are defined by f. The theory of fuzzy sets facilitates a progressive ascertainment of the membership of each element within an FRS, denoted by a membership function whose value falls within the real unit range $[0,1]$. Due to the fact that the role of membership is frequently a single-valued functioning, it is frequently unsuitable for representing supportive and challenge evidence.

d. Time Window Information Analysis Evaluation

The process of analyzing time windows in information disorder prediction entails partitioning the total time frame of interest into discrete intervals. This allows for a more detailed assessment of the prevalence and features of disinformation, fake news, and other related phenomena with time. Depending on the level of temporal resolution that is necessary for the study, each time frame represents a different temporal segment. These temporal segments can be defined as per-hour, weekly, or monthly intervals. A number of pertinent measures, such as the false news score, are generated inside each frame in order to quantify the degree to which information disorder has occurred. On the basis of the features that have been retrieved from the data, predictive modelling techniques, which may include algorithms for machine learning, are then utilized in order to predict the occurrence of the data disorder throughout each time window.

In general, fuzzy rough sets provide a flexible framework for evaluating textual data and forecasting information disorder. This framework allows for the handling of unpredictability, uncertainty, and complexity in the data analysis process by using the attributes of fuzzy logic combined with rough set theory. Users are able to make more educated judgments on the moderation and dissemination of material with the assistance of the rules that are developed through analysis. These rules offer insights into the elements that influence information disorder. For information disorder prediction, fuzzy rough sets offer models that are both visible and interpretable. This makes it simpler for users to comprehend and have confidence in the predictions obtained.

4. Results and Discussion

A library written in Python that is dedicated solely to rough set analysis and could include representations of fuzzy rough set methods is employed in this research. The existing algorithms, such as Mamdami Fuzzy Inference Approach (MFIA) [14], Fuzzy Rough Purity Approach (FRPA) [16] and Multidimensional Detection of Disinformation (MDD) [17], are taken for comparison study with the following metrics: prediction accuracy, FRS approximation, interpretability ratio and fake news score for predicting the impact of explainable technique using FRS. The investigation was conducted as part of a case study that utilized real data and pertained to phenomena associated with antagonistic information disorder.

a. Prediction Accuracy

For the purpose of validating the accuracy of generalization of predictive models spanning a variety of datasets or time periods, prediction accuracy is utilized. The durability and reliability of models in forecasting information disorder across a variety of contexts can be demonstrated by models that consistently attain high levels of prediction accuracy. The practical ramifications of the deployment of predictive models for disorders of information identification are impacted by the accuracy of the predictions made when using these models. In applications that take place in the real world, such as moderation of content and fact-checking, models that have a good ability to predict are more likely to produce insights that can be put into action and make it easier to make educated decisions.
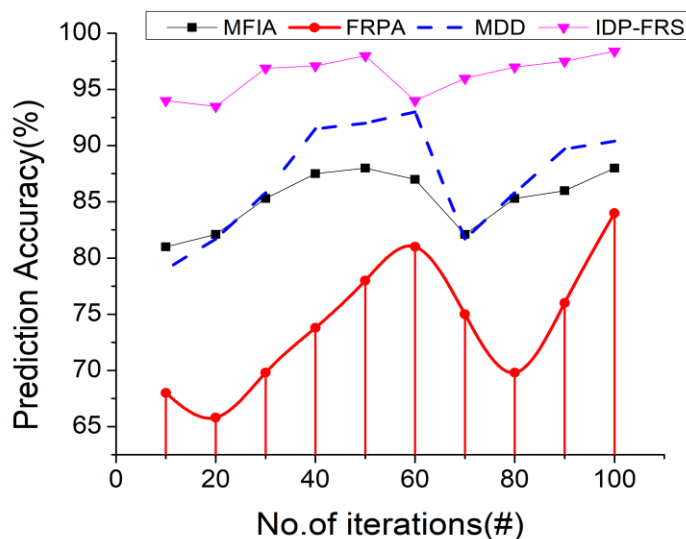
Figure 3. Explainable Prediction Information Accuracy Analysis

In Figure 3, An enormous quantity of varied information about disorders is gathered in this research study. These data include fake and genuine data, which provides for fuzzy boundary cases reported and analyzed in the FRS domain. A measure of how well predictive algorithms are able to accurately classify occurrences of information disorder, such as fake news and conspiracy ideas, based on the attributes derived from textual data, is referred to as prediction accuracy. In doing so, it offers insights into the general effectiveness of the models in distinguishing between instances of information disorder and cases that do not exhibit those characteristics. There is a correlation between high prediction accuracy and increased interpretability and transparency of predictive models. This is because high prediction accuracy promotes trust in the reliability of model predictions. Users can have faith in and rely on models that have a high prediction accuracy in order to assist them in efficiently identifying and responding to instances of information disorder.

b. FRS Approximation

Frequently, the FRS approximation process entails extracting decision criteria using the lower estimate in order to make the analysis and interpretation of the data more straightforward. The interpretability and effectiveness of the derived rules for decision-making are evaluated using metrics that pertain to the quality of the rules, their simplicity, their coverage, and their accuracy, among other things. Academics can measure the information effectiveness and quality of fuzzy rough sets approximation techniques in addressing uncertainty and ambiguity in data processing, particularly in predicting disinformation prediction. The size and composition of reducts, as well as the presence of a core (that is, qualities that are essential for discernibility), are crucial metrics that are utilized in the process of analyzing the efficiency and effectiveness of attribute reduction approaches.

In Figure 4, the computation of experiments utilizing the precise indiscernibility relationship was conducted with a single equivalence class containing all users in the boundary region between the $t0$ and $t1$.This depletion of the sole user block within the boundary area signifies that the behaviour of each of these users has changed. This seemingly implausible reality results from the challenge posed by the precise indiscernibility association when attempting to model minute fluctuations in the actions of users.
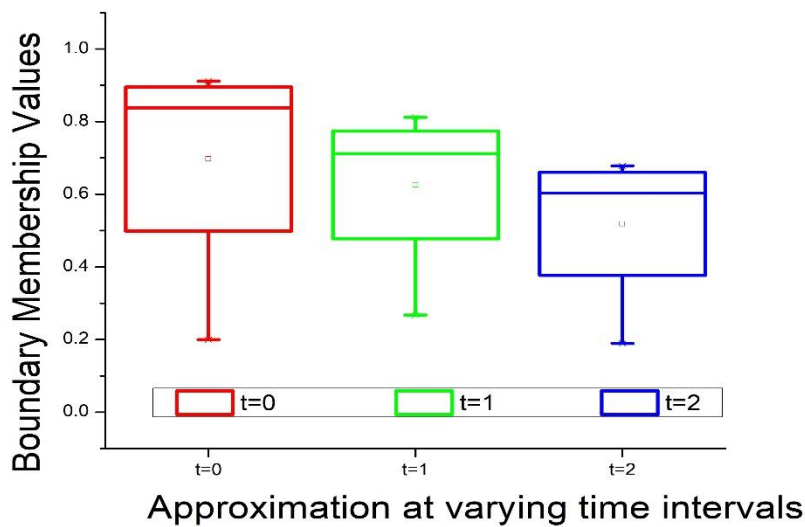
Figure 4. Boundary Approximation on Varying used actions at intervals

c. *Interpretability (%)*

Evaluation factors such as the number of rules, the degree of complexity of rule structures, and the readability of rule formulations are taken into consideration. After that, the score for interpretability is determined for every FRS model or set of rules by adding up the scores that were allocated to the various interpretability variables. For the purpose of making comparisons between other models or datasets easier, this score is normalized to a scale that has been in place for some time. Lastly, the interpretability ratio is calculated by multiplying the standardized interpretability rating by the total accuracy of predictions or performance indicators of the model. This method is used to determine the interpretability ratio. A larger interpretability ratio implies that the framework is more interpretable in comparison to its performance. This provides important insights into the relationship that exists between the interpretability of the model and its prediction accuracy. Researchers can make educated judgments on model selection and deployment by utilizing FRS theory to compute the interpretability ratio. This helps to ensure that the predictive models successfully balance comprehension with prediction effectiveness in the setting of information disorder detection interpretability, which is illustrated in Figure 5.
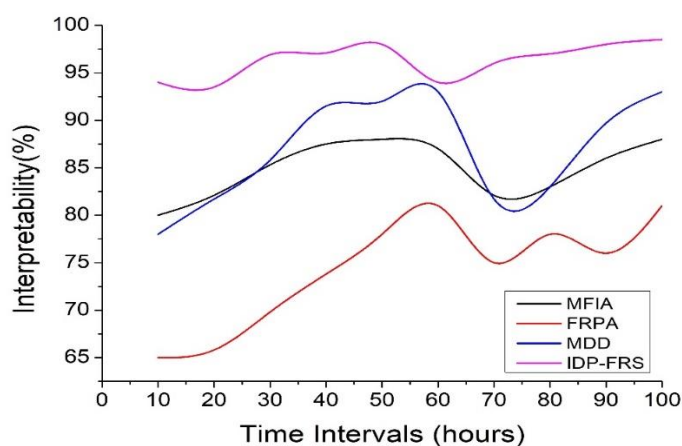

Figure 5. Analyzing Interpretability Ratio
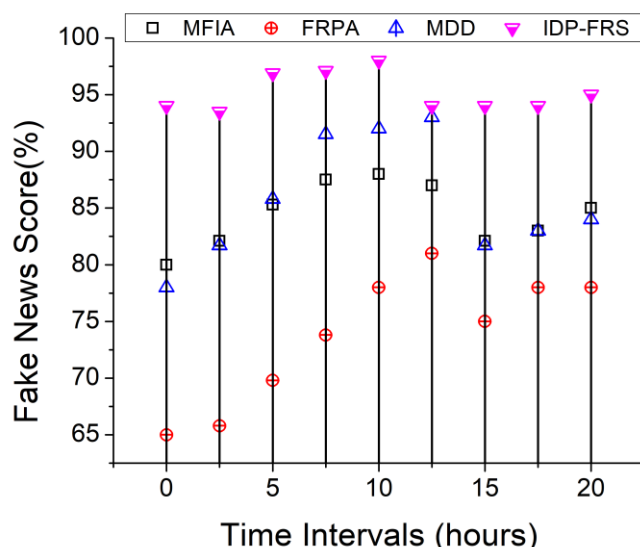
d. *Fake News Score Calculation*

Figure 6. Calculating Fake News Score

Figure 6 shows the illustration to delegate membership levels to every attribute based on how relevant it is to fake news, fuzzy membership functions need to be clearly defined. The degree of relationship between each characteristic and the idea of fake news is captured by these membership functions, which take into account a variety of characteristics, including the prominence of the context and the frequency with which the characteristic occurs. In order to enable understanding and comparison across a variety of cases or datasets, it is recommended that the fake news scores be normalized against a preset scale or range for instance, between 0 and 1. Because of this, the scores for fake news are guaranteed to be standardized and comparable, independent of the particular characteristics or traits that were utilized in the analysis. It is possible to designate the boundaries of regions of fake news inside the dataset by employing techniques that are based on fuzzy rough set approximation. Calculating lower and higher approximations is required in order to determine objects or occurrences that are absolutely or maybe related to disinformation according to their feature values and respective membership degrees and done in order to identify instances or objects.

5. Conclusion and Future Work

The proposed research presents a novel approach to predicting. In summary, this study has provided evidence for the effectiveness of fuzzy rough sets as a method for examining information disorder, specifically in the identification of misinformation and false news. The methodology's integration of fuzzy logic and rough set theory enhances decision-making capabilities, thereby improving the effectiveness of educational systems in recommending groups within e-learning environments. The integration of explainable prediction methods provides users with additional authority to form well-informed assessments, thereby promoting openness and confidence in predictive models. Subsequent investigations in this field will be indispensable in formulating resilient approaches to counter misinformation and advance a more knowledgeable digital environment. The research improves the accuracy and interpretability of models by incorporating fuzzy raw sets into predictive modelling techniques. This enables more dependable identification of misinformation. Furthermore, the integration of explainable prediction methodologies enables users to exercise informed discernment with

respect to the regulation and distribution of information, thereby fostering openness and confidence in predictive models.

By integrating dynamic time windows into the analysis, a more detailed understanding of the temporal structure of information disorder could be obtained. By adjusting time windows in response to shifting patterns and trends, the methodology can more efficiently capture evolving misinformation trends.

REFERENCES

[1]. Vilone, Giulia, and Luca Longo. "Explainable artificial intelligence: a systematic review." arXiv preprint arXiv:2006.00093 (2020).

[2]. Wen, Jintao, et al. "Dynamic interactive multiview memory network for emotion recognition in conversation." Information Fusion 91 (2023): 123-133.

[3]. Vicario, Michela Del, et al. "Polarization and fake news: Early warning of potential misinformation targets." ACM Transactions on the Web (TWEB) 13.2 (2019): 1-22.

[4]. Chen, Qiong, and Mengxing Huang. "Rough fuzzy model based feature discretization in intelligent data preprocess." Journal of Cloud Computing 10.1 (2021): 5.

[5]. Abbruzzese, Roberto, et al. "Detecting influential news in online communities: an approach based on hexagons of opposition generated by three-way decisions and probabilistic rough sets." Information Sciences 578 (2021): 364-377.

[6]. Gaeta, Angelo. "Evaluation of emotional dynamics in social media conversations: an approach based on structures of opposition and set-theoretic measures." Soft Computing 27.15 (2023): 10893-10903.

[7]. Dubois, Didier, Henri Prade, and Agnès Rico. "Structures of opposition and comparisons: Boolean and gradual cases." Logica Universalis 14.1 (2020): 115-149.

[8]. Kokkotis, Christos, et al. "Explainable machine learning for knee osteoarthritis diagnosis based on a novel fuzzy feature selection methodology." Physical and Engineering Sciences in Medicine 45.1 (2022): 219-229.

[9]. Theerens, Adnan, Oliver Urs Lenz, and Chris Cornelis. "Choquet-based fuzzy rough sets." International Journal of Approximate Reasoning 146 (2022): 62-78.

[10]. Zhang, Kai, Jianming Zhan, and Xizhao Wang. "TOPSIS-WAA method based on a covering-based fuzzy rough set: an application to rating problem." Information Sciences 539 (2020): 397-421.

[11]. Zhang, Kai, and Jianhua Dai. "A novel TOPSIS method with decision-theoretic rough fuzzy sets." Information Sciences 608 (2022): 1221-1244.

[12]. Gongane, Vaishali U., Mousami V. Munot, and Alwin D. Anuse. "A survey of explainable AI techniques for detection of fake news and hate speech on social media platforms." Journal of Computational Social Science (2024): 1-37.

[13]. Shakeel, P. Mohamed, and S. Baskar. "Automatic human emotion classification in web document using fuzzy inference system (FIS): human emotion classification." International Journal of Technology and Human Interaction (IJTHI) 16.1 (2020): 94-104.

[14]. Campos, Obed, Pablo Pancardo, and José Adán Hernández-Nolasco. "Dynamic fuzzy model for detecting verbal violence in real time." Computer Science 23 (2022).

[15]. Gaeta, Angelo, et al. "A novel approach based on rough set theory for analyzing information disorder." Applied Intelligence 53.12 (2023): 15993-16014.

[16]. Uddin, Jamal, et al. "Rough set based information theoretic approach for clustering uncertain categorical data." Plos one 17.5 (2022): e0265190.

[17]. Weijian, F. A. N., and W. A. N. G. Yongbin. "Multidisciplinary Fusion Perspective Analysis Method for False Information Recognition." Advances in Electrical & Computer Engineering 24.1 (2024).

[18]. Han, Ying, Sheng Chen, and Xiaoning Shen. "Fuzzy rough set with inconsistent bipolarity information in two universes and its applications." Soft Computing 26.19 (2022): 9775-9784.

[19]. https://www.kaggle.com/code/kerneler/starter-covid-19-fake-news-dataset-19713c8a-1/data