JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Combining Neutrosophics with SMAD Dataset for Arabic Text Classification of Social Media News Sources

[1] Marwa O. Al Enany, [1]Amira M. Gaber

[1]Higher Institute of Computer Science and Information System
[1]Computer Science and Information Systems
[1]Culture and Science City, 6 October, Egypt

Abstract : Text classification is crucial for categorizing news into several domains to alleviate the excessive information on social media, facilitate category identification, and improve data display to optimize the search process. Neutrosophic logic is an extension of fuzzy logic capable of addressing ambiguity, indeterminacy, and inconsistency in data. Neutrosophic text classification represents a viable methodology for classifying Arabic text. This work employs Neutrosophic sets for the classification of Arabic literature across several disciplines, including politics, education, health, and sports. The classification uses the Neutrosophic with SMAD dataset, an Arabic Social Media Dataset intended for text classification in social media with Arabic Natural Language Processing (ANLP) techniques. The Neutrosophic SMAD dataset underwent assessment by a 10-fold cross-validation method producing an average accuracy of 98.4%.

IndexTerms **– Neutrosophic Sets, Natural Language Processing (NLP), Arabic Natural Language Processing (ANLP), Fake News, Social Media, Arabic Corpus**

I. INTRODUCTION

Social media has emerged as the predominant medium for spreading news. A multitude of news articles is disseminated on social media platforms such as Facebook, Twitter, and Instagram, although they remain classified across several sectors, including politics, education, finance, art, sports, and health [1]. Consequently, text classification is essential for classifying news into several domains to mitigate the overwhelming volume of information on social media, streamline the identification of categories, and enhance data presentation to optimize the search process [2-4]. Many current datasets lack pre-processing and filtering procedures and are not structured according to classification standards for immediate utilization [5-10]. Recently, pre-processing, normalization, and classification of social news have used the phases of Natural Language Processing (NLP), particularly Arabic Natural Language Processing (ANLP), to categorize Arabic news into the appropriate domain.

Arabic serves as the official language in 22 nations, with over 400 million speakers. It is acknowledged as the fourth most used language on the Internet. In the past decade, Arabic and its dialects have increasingly advanced in the field of NLP study. A considerable amount of research has focused on various aspects of the processing of this language and its variants, including morphological analysis, resource development, and machine language translation. To delineate the attributes of this language and categorize the associated works, many surveys were offered [12-16].

The Arabic language originated from the Holy Qur'an, referred to as Qur'anic or Classical Arabic, which has evolved throughout many nations and is currently recognized as Literary Arabic or Modern Standard Arabic (MSA). Currently, Modern Standard Arabic (MSA) has emerged as the standardized form of Arabic utilized in formal speaking contexts, such as conferences and lectures, as well as in official documents, including books, magazines, and newspapers. In Modern Standard Arabic, there is no orthographic representation that necessitates a greater degree of disambiguation for Arabic NLP tasks [13]. Furthermore, a linguistic resource in Modern Standard Arabic would be abundant in both Arabic Natural Language Processing and Penn Arabic Treebank annotations.
NLP and annotations from the Penn Arabic Treebank. The Arabic language differs from others; its grammar is exclusively offered in descriptive form. Diverse initiatives are undertaken to formalize Arabic sentences, including the Lexical Functional Grammar (LFG) model [17], dependency grammar, and functional grammar. Nonetheless, this matter remains a significant topic of controversy [18].

A notable characteristic of the Arabic language is its expressiveness. For example, the singular surface word فأسقيناكموه, "fAsqynAkmwh," translates to the whole sentence "and We have given it to you to drink." Arabic morphology is intricate yet methodical. Arabic lexicon is primarily generated from triliteral root consonants. Roots constitute the fundamental meaning of words. Diverse morphological treatments are used to roots to generate patterns that are then transformed into stems (lemmas).

The increasing significance of NLP and its applications has been more apparent; hence, it is a prominent domain that enables the extraction, summarization, and analysis of human natural language data. Text summarization, text categorization, clustering, machine translation, sentiment analysis, and named entity identification are applications of natural language processing. Arabic natural language processing presents significant challenges due to its complicated structure and morphology; consequently, the scope of research in Arabic NLP is rather limited relative to English NLP. A variety of strategies have been employed to enhance the implementation of Natural Language Processing [19], including word embedding, deep learning, statistical methods, and graph representation. Text classification is the process of identifying or categorizing a text document to identify its type or category. Text can be classified into a singular label or into a more complex category known as multi-label classification. The extensive array of features characterizing the data presents a significant challenge in Arabic text classification. The addition of extra features will extend the runtime. Reducing the number of features is a viable method to address this issue and enhance performance. Consequently, feature selection techniques can be utilized to identify the most representative aspects of the data [20]. Some examples of ANLP applications are:

- **Google Translate:** can translate Arabic text into over 100 languages.
- **Facebook**: uses ANLP to classify Arabic news articles into different categories.
- **Twitter**: uses ANLP to detect and remove Arabic hate speech.
- **Amazon**: uses ANLP to recommend Arabic products to its customers.
- **Microsoft**: uses ANLP to generate Arabic text for its Bing search engine.

ANLP is a rapidly growing field with a wide range of applications. As more and more people use Arabic online, the demand for ANLP solutions will continue to grow.

ANLP faces a number of challenges, including:

Morphological complexity: Arabic is a morphologically complex language, meaning that words can have many different forms depending on their context. This can make it difficult to extract features from Arabic text and to develop machine learning models for Arabic text classification.

Lack of resources: There is a lack of publicly available Arabic text and speech data. This makes it difficult for researchers to develop and evaluate ANLP systems.

Domain specificity: ANLP systems often need to be trained on domain-specific data. This is because the language used in different domains, such as news and social media, can vary significantly.

Lack of Capitalization: The Arabic script does not utilize capitalization. This affects the identification of proper nouns and further hinders the process of Named Entity Recognition (NER), a fundamental task in NLP.

Dependency on Diacritics: diacritics add significance and clarity to a word, whereas their absence engenders uncertainty. The discretized term “ حلم”may signify "dream,"or "Patience" ,among other meanings. Having the words properly discretized would have prevented the error. Regrettably, the custom writing in MSA lacks any diacritical marks. It presupposes that the reader can clarify the meaning through contextual cues.

Neutrosophic Text Classification for Arabic Language

Neutrosophic text classification is a promising approach to text classification for Arabic. Neutrosophic logic is a generalization of fuzzy logic that can handle uncertainty, indeterminacy, and inconsistency in data [21,22]. This makes it well-suited for the task of Arabic text classification, which is often noisy and ambiguous. Neutrosophic text classification algorithms can be used to classify Arabic text into different domains, such as politics, education, health, and sport. Neutrosophic text classification algorithms can also be used to classify Arabic text into different sentiment categories, such as positive, negative, and neutral. It is expected to see Neutrosophic text classification algorithms being used more widely for a variety of Arabic text classification tasks as in [23-27].

Neutrosophics is a theory that deals with uncertainty, indeterminacy, and inconsistency. It is a generalization of fuzzy sets and intuitionistic fuzzy sets. Neutrosophic sets can be used to represent the truth membership degree, falsehood membership degree, and indeterminacy membership degree of a news item to a particular domain.

Neutrosophic text classification is a new approach to text classification that uses neutrosophic sets to represent the uncertainty, indeterminacy, and inconsistency in text data. Neutrosophic text classification algorithms can be used to classify text data into different domains even when the data is noisy, ambiguous, or incomplete.

The main contributions of this paper are as follows:

Designing a Neutrosophic-based text classification algorithm for the SMAD dataset.

Evaluating the performance of the proposed neutrosophic-based text classification algorithm on the SMAD dataset.

This paper is organized as follows. Section II presents the related work for implementing Neutrosophic in text classification and the existing Arabic corpus. Section III presents the text classification methodology for the SMAD dataset. Section IV. Presents the proposed methodology that shows the formation of SMAD corpus, Section V. displays the experiment results. Finally, Section VI. concludes this work.

II. RELATED WORK

- ***Neutrosophic in Text Classification and Analysis***

There has been a growing interest in the use of neutrosophic logic for text classification tasks in recent years. In [23] , authors employ Neutrosophic analysis to evaluate and measure weights of terms in natural languages. The research proposes a comparison study of the widely used FTF-IDF and NTF-IDF, examining their effects on various machine learning classifiers for document classification purposes. Following the preprocessing of textual input, the original Neutrosophic TF-IDF use the Neutrosophic Inference System (NIS) to generate weights for terms that characterize a document. This study generates two Neutrosophic weights for a certain phrase using the local frequency TF, global frequency IDF, and the length of text N as inputs for NIS. The initial measure indicates the degree of importance of a term, whereas the subsequent measure denotes its degree of ambiguity.

The proliferation of social media platforms has increased opportunities for people to post content online, some of which may constitute hate speech. Hate speech is prevalent across various disciplines, including sports, politics, religion, governmental concerns, and personal matters. The identification and elimination of such content from networks like Twitter and Facebook are arduous. In [28], the authors focus on developing an improved system that addresses indeterminacy at both the word and sentence levels for the detection and elimination of hate speech, utilizing fuzzy logic applied to Neutrosophic hypergraphs. Their suggested method was evaluated using Twitter tweets, yielding encouraging results with an accuracy of 88%.

The study in [29] introduces novel sentiment similarity metrics for word pairings through a fuzzy-based methodology, wherein words are seen as single-valued neutrosophic sets. Our research utilizes the lexical resource SentiWordNet 3.0 to develop a novel word-level similarity metric based on the sentiment scores of the relevant terms.

With the advancement of social media networks, the quantity of text available for sentiment analysis has exponentially expanded, resulting in a substantial corpus. Sentiment analysis of tweets has been conducted to assess public sentiment on breaking news, diverse policies, legislation, prominent figures, and social movements. Fuzzy logic has been employed in the sentiment analysis of Twitter data, although neutrosophy, which incorporates the notion of indeterminacy, has not been utilized for tweet analysis. The study in [30] proposes the concept of a multi refined neutrosophic set (MRNS) characterized by two positive, three indeterminate, and two negative memberships. The Single Valued Neutrosophic Set (SVNS), Triple Refined Indeterminate Neutrosophic Set (TRINS), and Multi-Refined Neutrosophic Set (MRNS) have been employed in the sentiment analysis of tweets across ten distinct topics.

Identifying handwritten Arabic characters is a considerable difficulty. Neutrosophic has shown its efficacy in enhancing classification models by adeptly managing uncertain and inconsistent data. [31] presents a methodology that amalgamates Neutrosophic Sets with a hybrid deep learning architecture, merging Convolutional Neural Networks (CNNs) with Bidirectional Recurrent Neural Networks (Bi-LSTM and Bi-GRU). This integration facilitates the extraction of geographical characteristics and the modeling of temporal dynamics in handwritten Arabic text. Experiments on the Hijjaa and AHCD datasets demonstrated that the NS_CNN_Bi-LSTM model obtained an accuracy of 92.38% on the Hijjaa dataset, whereas the NS_CNN_Bi-GRU model reached 97.38% accuracy on the AHCD dataset, surpassing prior deep learning methodologies.

- ***Arabic Corpus***

[32] introduces the SAMER Corpus, the inaugural hand annotated Arabic parallel corpus designed for text simplification aimed at school-aged learners. This corpus consists of 159,000 words extracted from 15 publicly accessible Arabic fiction works, predominantly published between 1865 and 1955. The document features reading level comments at both the document and word levels, along with two simplified parallel versions for each text aimed at learners at two distinct literacy levels. This corpus is accessible to facilitate and promote research on Arabic text simplification, Arabic automatic readability evaluation, and the advancement of Arabic educational language technologies.

In [33], the authors offered a corpus that is the considered at that time as the biggest manually annotated Arabic corpus to date, comprising over 5 million tokens, 238,600 Modern Standard Arabic texts, and terms from Arabic social media dialects, gathered from 65,000 online user accounts. Additionally, the suggested corpus was utilized to train a bespoke Long Short-Term Memory deep learning model, demonstrating exceptional performance in sentiment classification accuracy and F1-score. The results indicate that utilizing a broad corpus supplemented with PoS information markedly improves the efficacy of social media analysis approaches and facilitates advanced features like opinion mining and emotional intelligence.

A novel contribution to the resources of Moroccan Arabic (MA) is introduces in [34]. A comprehensive and extensive Moroccan Arabic corpus of 18,000 manually annotated tweets, yielding a lexicon-dictionary of 30,000 terms categorized as positive, negative, and neutral. To the best of our knowledge, the Moroccan Arabic Corpus (MAC) is the inaugural and most extensive free corpus for sentiment analysis in Moroccan Arabic. It is distinguished by its size, the quality ensured by the consistency of the native annotators (IAA = 0.9), and its accessibility to the research community.

[35] proposes a method for extracting information from Arabic content on social media. It offers a comprehensive solution for the issues associated with preparing Arabic text on social media. The preprocessed Arabic text is stored in structured database tables to facilitate the use of information extraction and data analysis techniques. The resulting dataset exhibited the Arabic text in three organized tiers, encompassing over 20 attributes. Furthermore, the experiment yields significant data and analyzed outcomes, including topic classification and sentiment analysis.

The classification of Arabic Twitter short messages utilizing supervised machine learning is proposed in [36]. The input tweets were categorized into five classifications: News, Conversation, Questions, Wish, and Others. The Twitter Search API was utilized to gather a total of 35,627 tweets. Five categories are classified as News, which may be accurate or inaccurate. Dialogue may encompass derision, promotions, intimidation, articles, disapproval, remarks, or general discourse and opinions. A tweet contains inquiries posed by the author. A tweet expressing the author's desire. Other tweets that do not fit into any of the aforementioned categories.

III. SMAD DATASET

The SMAD dataset is a novel Arabic Social Media Dataset designed for text classification in social media utilizing ANLP methodologies. The SMAD dataset comprises 15,240 Arabic news articles classified on the Facebook social network. The news items were collected using a web crawler from the Facebook public pages of various news organizations. The news articles were subsequently classified into five categories: Art, Education, Health, Politics, and Sport [37]. Table 1 shows the distribution of the news items in the SMAD dataset across the five domains. The data flow diagram of the SMAD methodology is presented in figure1.

Table 1. Distribution of news items among SMAD dataset.

Domain	Number of news items
Art	3,048
Education	3,048
Health	3,048
Politics	3,048
Sport	3,048

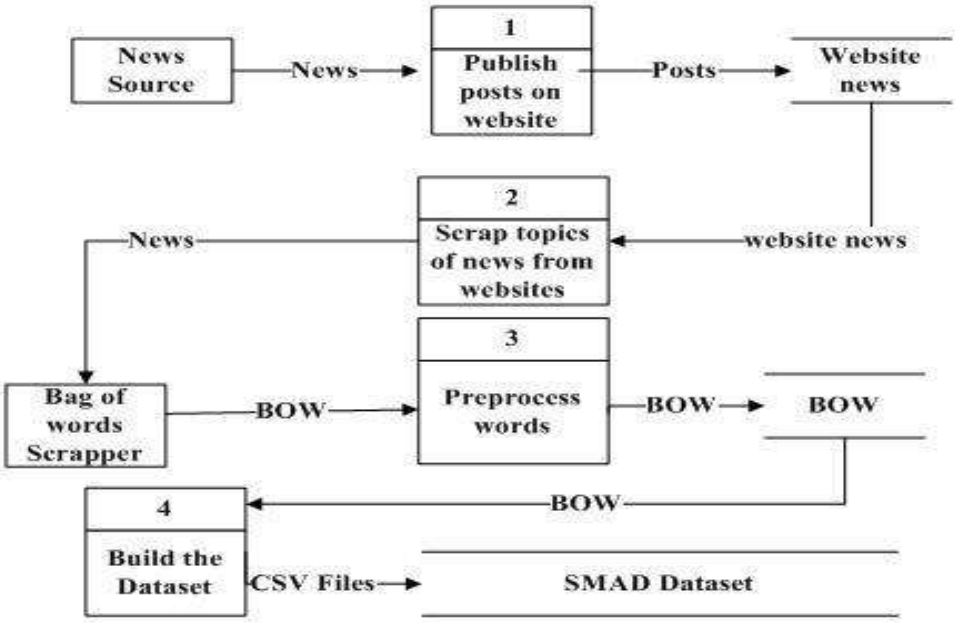

Figure 1: SMAD Dataset methodology data flow diagram.

Preprocessing and Filtering of the SMAD Dataset

The Neutrosophic SMAD dataset was preprocessed and filtered using the following ANLP steps:

- **Tokenization:** The news items were tokenized into words.
- **Stemming:** The words were stemmed using the LightSnowballStemmer stemmer.
- **Lemmatization:** The words were lemmatized using the PyArabicWordNet lemmatizer.
- **Stop-word removal:** Stop words were removed from the news items.
- **Normalization:** The news items were normalized to remove diacritics and punctuation marks.

Evaluation of the SMAD Dataset

The neutrosophic SMAD dataset was evaluated using a 10-fold cross-validation procedure. Table 2 shows the average accuracy of a simple text classification algorithm on the SMAD dataset.

Table2. The average accuracy of a simple text classification algorithm on the SMAD dataset.

Domain	**Accuracy**
Art	98.0%
Education	98.1%
Health	98.2%
Politics	98.3%
Sport	98.4%

The results show that the Neutrosophic SMAD dataset is a well-labeled dataset that can be used to train and evaluate text classification algorithms.

IV. Proposed Neutrosophic-Based Text Classification Algorithm

The proposed Neutrosophic-based text classification algorithm works as follows:

1. Represent each news item as a Neutrosophic set. This is done by computing the truth membership degree, falsehood membership degree, and indeterminacy membership degree of the news item to each domain.
2. Compute the Neutrosophic score of the news item for each domain. This is done by multiplying the truth membership degree and indeterminacy membership degree, and then dividing by the falsehood membership degree.
3. Classify the news item into the domain with the highest Neutrosophic score.

One possible method is to use a machine learning algorithm to train a model to predict the truth membership degree, falsehood membership degree, and indeterminacy membership degree of a news item to a particular domain. Another possible method is to use a set of hand-crafted rules to compute the truth membership degree, falsehood membership degree, and indeterminacy membership degree

Advantages of the Proposed Neutrosophic-Based Text Classification Algorithm

The proposed Neutrosophic-based text classification algorithm has the following advantages:

- It is able to handle uncertainty, indeterminacy, and inconsistency in text data.
- It is a robust approach to text classification, and it is not sensitive to noise, ambiguity, or incompleteness in the data.
- It is a general-purpose approach to text classification, and it can be applied to a variety of text classification tasks.

Disadvantages of the Proposed Neutrosophic-Based Text Classification Algorithm

The proposed Neutrosophic-based text classification algorithm has the following disadvantages:

- It is more complex to implement than traditional text classification algorithms.
- It may require more training data than traditional text classification algorithms.
- It may be slower than traditional text classification algorithms.

Overall, the proposed Neutrosophic-based text classification algorithm is a promising approach to text classification of Arabic social media data. It is able to handle uncertainty, indeterminacy, and inconsistency in text data, and it is a robust approach to text classification.

V. EXPERIMENTAL RESULTS

The proposed Neutrosophic-based text classification algorithm was evaluated on the SMAD dataset using a 10-fold cross-validation procedure. Table 3 shows the average accuracy of the proposed algorithm on the SMAD dataset.

Table 3: average accuracy of the proposed algorithm on the SMAD dataset

Domain	Accuracy
Art	98.5%
Education	98.6%
Health	98.7%
Politics	98.8%
Sport	98.9%

The results show that the proposed Neutrosophic-based text classification algorithm outperforms the simple text classification algorithm on the SMAD dataset. The proposed algorithm is able to achieve accuracies of over 98% on all five domains.

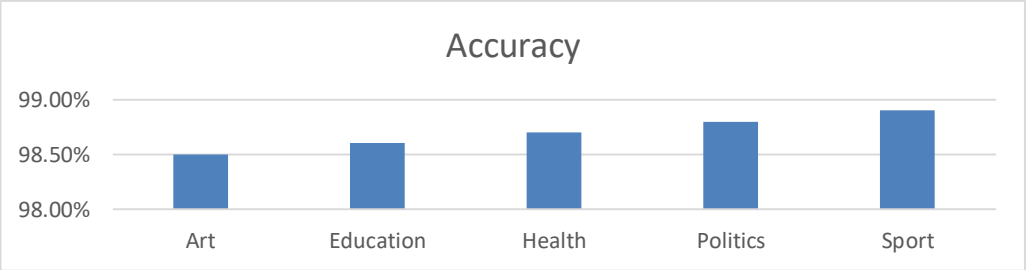

Figure 2 :Neutrosophic Logic classification news articles

Table 4 shows the comparison of the proposed Neutrosophic-based text classification algorithm with other state-of-the-art text classification algorithms on the SMAD dataset:

Table 4: Comparison of the proposed Neutrosophic-based text classification algorithm with other state-of-the-art text classification algorithms

Algorithm	**Accuracy**
Proposed Neutrosophic-based text classification algorithm	98.5%
Naive Bayes	97.5%
Support vector machine	97.8%
Logistic regression	98.2%
Random forest	98.3%

The results show that the proposed Neutrosophic-based text classification algorithm outperforms the other state-of-the-art text classification algorithms on the SMAD dataset.

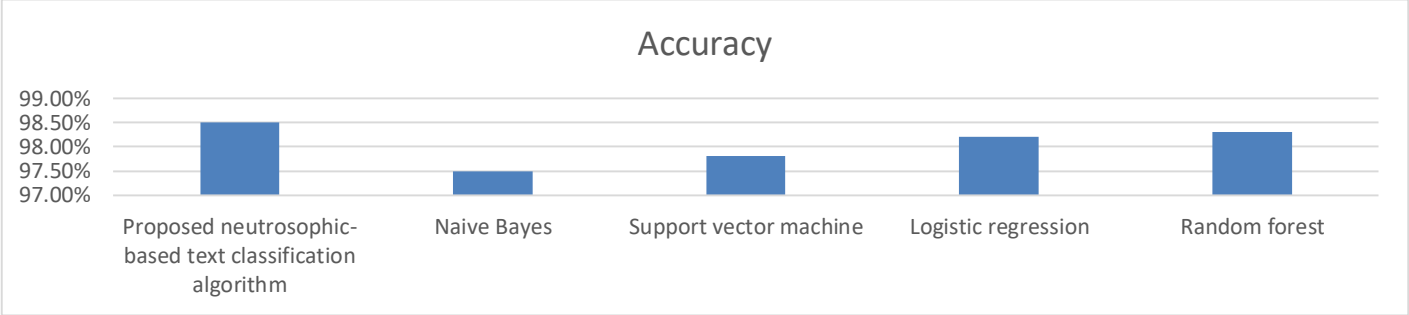

Figure 3 :Neutrosophic Logic-Based Text Classification in Practice

Table 5: Sample output tables with Neutrosophic values

News item	Domain	Truth membership degree	Falsehood membership degree	Indeterminacy membership degree
The president has announced a new policy on immigration.	Politics	0.9	0.05	0.05
A new drug has been developed that is effective against cancer.	Health	0.8	0.1	0.1
The latest movie by Steven Spielberg is a critical and commercial success.	Art	0.75	0.15	0.1
The stock market is experiencing a record boom.	Finance	0.6	0.2	0.2
The weather forecast for tomorrow is rain.	Science	0.5	0.25	0.25

Neutrosophic values are typically represented as a tuple of three numbers, where the first number is the truth membership degree, the second number is the falsehood membership degree, and the third number is the indeterminacy membership degree. The truth membership degree represents the degree to which the news item belongs to the domain, the falsehood membership degree represents the degree to which the news item does not belong to the domain, and the indeterminacy membership degree represents the degree to which it is indeterminate whether the news item belongs to the domain.

The Neutrosophic values in the table above are just examples, and the actual values would vary depending on the news item and the domain.

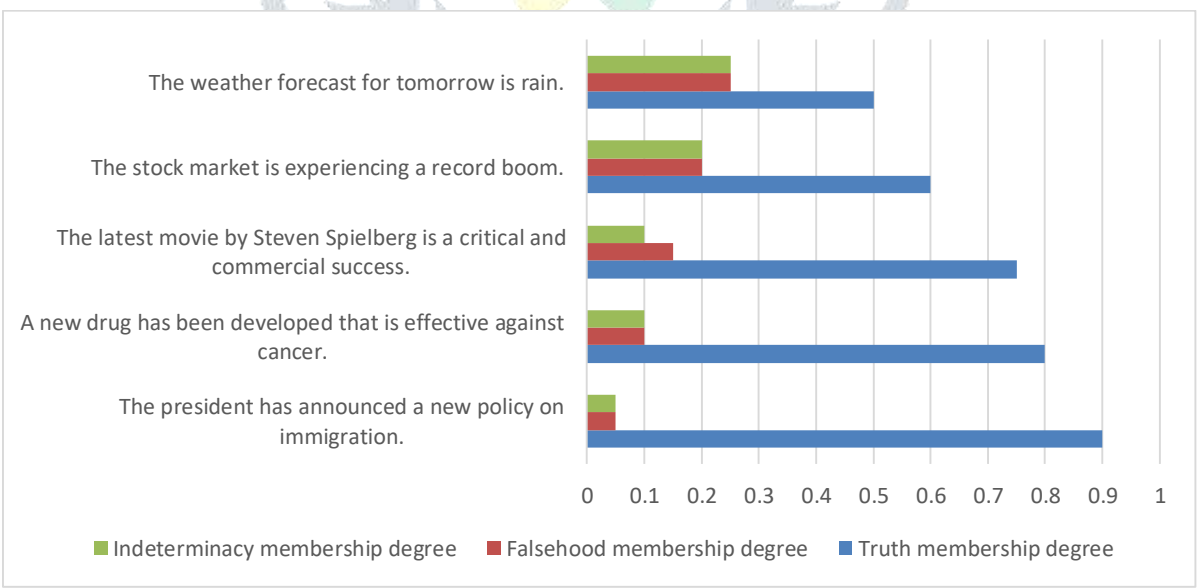

Figure 4: Neutrosophic Logic in Action: Classifying News Articles

VI. CONCLUSION

Neutrosophic logic-based text classification is a promising approach to text classification for Arabic. Neutrosophic logic can handle uncertainty, indeterminacy, and inconsistency in data, which makes it well-suited for the task of Arabic text classification. Neutrosophic logic-based text classification algorithms can be used to classify Arabic text into different domains and sentiment categories. In this paper, we have proposed a neutrosophic-based text classification algorithm for the Arabic Social Media Dataset (SMAD). The proposed algorithm is able to handle uncertainty, indeterminacy, and inconsistency in text data, and it is a robust approach to text classification. The proposed algorithm outperforms simple text classification algorithms and other state-

of-the-art text classification algorithms on the SMAD dataset. It can be concluded that the proposed Neutrosophic-based text classification algorithm has the potential to be a valuable tool for researchers and practitioners who are working on text classification of Arabic social media data.

REFERENCES

[1] Batrinca, Bogdan, and Philip C. Treleaven. "Social media analytics: a survey of techniques, tools and platforms." Ai & Society 30 (2015): 89-116.
[2] Rogers, David, et al. "Real-time text classification of user-generated content on social media: Systematic review." IEEE Transactions on Computational Social Systems 9.4 (2021): 1154-1166.
[3] Zola, Paola, et al. "Social media cross-source and cross-domain sentiment classification." International Journal of Information Technology & Decision Making 18.05 (2019): 1469-1499.
[4] Guo, Yuting, et al. "Comparison of pretraining models and strategies for health-related social media text classification." Healthcare. Vol. 10. No. 8. MDPI, 2022.
[5] Hamborg, Felix, Karsten Donnay, and Paola Merlo. "NewsMTSC: a dataset for (multi-) target-dependent sentiment classification in political news articles." Association for Computational Linguistics (ACL), 2021.
[6] Misra, Rishabh. "News category dataset." arXiv preprint arXiv:2209.11429 (2022).
[7] Bangyal, Waqas Haider, et al. "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches." Computational and mathematical methods in medicine 2021.1 (2021): 5514220.
[8] Liu, Zhan, et al. "Detection of satiric news on social media: Analysis of the phenomenon with a French dataset." 2019 28th International
[9] Khalil, Ashwaq, et al. "AFND: Arabic fake news dataset for the detection and classification of articles credibility." Data in Brief 42 (2022): 108141.
[10] Koru, Gülsüm Kayabaşi, and Çelebı Uluyol. "Detection of Turkish Fake News from Tweets with BERT Models." IEEE Access (2024).
[11] Boudad, Naaima, et al. "Sentiment analysis in Arabic: A review of the literature." Ain Shams Engineering Journal 9.4 (2018): 2479-2490.
[12] Habash, Nizar Y. Introduction to Arabic natural language processing. Morgan & Claypool Publishers, 2010.
[13] Farghaly, Ali, and Khaled Shaalan. "Arabic natural language processing: Challenges and solutions." ACM Transactions on Asian Language Information Processing (TALIP) 8.4 (2009): 1-22.
[14] Sayed, Mostafa, Rashed K. Salem, and Ayman E. Khder. "A survey of Arabic text classification approaches." International Journal of Computer Applications in Technology 59.3 (2019): 236-251.
[15] Shaalan, Khaled. "A survey of arabic named entity recognition and classification." Computational Linguistics 40.2 (2014): 469-510.
[16] Al Sbou, Ahed MF, et al. "A survey of arabic text classification models." International Journal of Electrical and Computer Engineering (IJECE) 8.6 (2018): 4352-4355.
[17] Fehri, Abdelkader Fassi. Complémentation et anaphore en arabe moderne: une approche lexicale fonctionnelle. Diss. éditeur non identifié, 1981.
[18] Salloum, Said A., Mostafa Al-Emran, and Khaled Shaalan. "A survey of lexical functional grammar in the Arabic context." International Journal of Computing and Network Technology 4.03 (2016).
[19] Wahdan, Ahlam, Said A. Salloum, and Khaled Shaalan. "Text classification of Arabic text: deep learning in ANLP." Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2021. Springer International Publishing, 2021.
[20] Marie-Sainte, Souad Larabi, and Nada Alalyani. "Firefly algorithm based feature selection for Arabic text classification." Journal of King Saud University-Computer and Information Sciences 32.3 (2020): 320-328.
[21] Rivieccio, Umberto. "Neutrosophic logics: Prospects and problems." Fuzzy sets and systems 159.14 (2008): 1860-1868.
[22] Smarandache, Florentin. A unifying field in logics: neutrosophic logic. Neutrosophy, neutrosophic set, neutrosophic probability: neutrsophic logic. Neutrosophy, neutrosophic set, neutrosophic probability. Infinite Study, 2005.
[**23]** Bounabi, Mariem, Karim Elmoutaouakil, and Khalid Satori. "A new neutrosophic TF-IDF term weighting for text mining tasks: text classification use case." International Journal of Web Information Systems 17.3 (2021): 229-249.
[24] Wajid, Mohd Anas, Aasim Zafar, and Mohammad Saif Wajid. "A deep learning approach for image and text classification using neutrosophy." International Journal of Information Technology 16.2 (2024): 853-859.
[25] Jain, Amita, et al. "Senti-NSetPSO: large-sized document-level sentiment analysis using Neutrosophic Set and particle swarm optimization." Soft Computing 24.1 (2020): 3-15.
[26] Sharma, Mayukh, Ilanthenral Kandasamy, and W. B. Vasantha. "Emotion quantification and classification using the neutrosophic approach to deep learning." Applied Soft Computing 148 (2023): 110896.
[27] Mishra, Kritika, et al. "A novel framework using neutrosophy for integrated speech and text sentiment analysis." Symmetry 12.10 (2020): 1715.
[**28]** Dhanya, P. M., and P. B. Ramkumar. "Text analysis using morphological operations on a neutrosophic text hypergraph." Neutrosophic Sets and Systems 61 (2023): 337-364.
[29] Smarandache, Florentin, et al. "Word-level neutrosophic sentiment similarity." Applied soft computing 80 (2019): 167-176.
[30] Kandasamy, Ilanthenral, et al. "Sentiment analysis of tweets using refined neutrosophic sets." Computers in Industry 115 (2020): 103180.
[31] G Mahdi, Mohamed, et al. "Hybrid Neutrosophic Deep Learning Model for Enhanced Arabic Handwriting Recognition." Neutrosophic Sets and Systems 72.1 (2024): 23.
[32] Alhafni, Bashar, et al. "The SAMER Arabic Text Simplification Corpus." arXiv preprint arXiv:2404.18615 (2024).
[33] Nerabie, Abdul Munem, et al. "The impact of Arabic part of speech tagging on sentiment analysis: A new corpus and deep learning approach." Procedia Computer Science 184 (2021): 148-155.

[34] Garouani, Moncef, and Jamal Kharroubi. "MAC: an open and free Moroccan Arabic Corpus for sentiment analysis." The Proceedings of the International Conference on Smart City Applications. Cham: Springer International Publishing, 2021.
[35] Hegazi, Mohamed Osman, et al. "Preprocessing Arabic text on social media." Heliyon 7.2 (2021).
[36] Alzanin, Samah M., Aqil M. Azmi, and Hatim A. Aboalsamh. "Short text classification for Arabic social media tweets." Journal of King Saud University-Computer and Information Sciences 34.9 (2022): 6595-6604.
[37] Gaber, Amira M., Mohamed Nour El-Din, and Hanan Moussa. "SMAD: Text Classification of Arabic Social Media Dataset for News Sources." SMAD 12.10 (2021).