**Title - Knowledge-based Data Processing for Multilingual Natural Language Analysis**

**T5**

this paper proposes multilingual data processing using feature extraction with classification . input text data has been collected based on various languages and processed . extracted data has been extracted using histogram equalization based global local entropy (HEGLE) and classified using Kernel-based Radial basis Function (Ker\_Rad\_BF) this paper presents solutions to the multilingual sentiment analysis issue by implementing algorithms . the paper concludes that the proposed neural network could be used in a variety of ways .

**Bart -Knowledge-based Data Processing for Multilingual Natural Language Analysis. Natural Language Processing (NLP) aids the empowerment of intelligent machines by enhancing human language understanding for linguistic-based human-computer communication. This paper proposes multilingual data processing using feature extraction with classification using deep learning architectures. For the HASOC dataset, the proposed HEGLE\_ Ker\_Rad\_BF achieved an accuracy of 98%, a precision of 97%, a recall of 90.5%, an f-1 score of 85%, RMSE of 55.6%, and a loss curve analysis attained 44%. For the TRAC dataset, a proposed neural network achieved an accurate 98%, the precision attained is 97%, the Recall is 91%, the F-1 scored is 87%, and the RMSE is 55%. For a variety of other languages, we present solutions to the multilingual sentiment analysis issue in this research article by implementing algorithms. We compare precision factors to discover the optimum option for mult bilingual sentiment analysis. The proposed algorithms are based on a deep learning architecture called Ker\_ Rad\_BF. These architectures could be utilized to process natural language. For a range of different languages, this article presents a set of algorithms that can be used to process the input text data based on various languages and remove missing values and null values. For more information on the proposed algorithms.**

**Gemini -**-Multilingual Natural Language Processing (NLP) enables machines to understand human language for effective communication. This paper presents a multilingual data processing approach using Histogram Equalization based Global Local Entropy (HEGLE) for feature extraction. These features are then classified using Kernel-based Radial basis Function (Ker\_Rad\_BF). The proposed HEGLE\_Ker\_Rad\_BF model achieved high accuracy (98%) and precision (97%) in sentiment analysis tasks on HASOC and TRAC datasets. This approach provides an effective solution for multilingual NLP problems.

**Gpt 4-**This paper delves into the enhancement of Natural Language Processing (NLP) capabilities through knowledge-based data processing, specifically targeting multilingual natural language analysis. The primary focus is on empowering intelligent machines to better understand human languages, thereby improving human-computer interactions. The research leverages recent advancements in computational power and the increased availability of extensive linguistic datasets to drive the development of data-driven methods for automatic semantic analysis across multiple languages.

The methodology introduced in this study involves a sophisticated process of data handling and analysis. Initially, multilingual text data is collected and subsequently cleansed of any missing or null values to ensure the integrity of the data for processing. The cleaned data is then subjected to feature extraction using Histogram Equalization based Global Local Entropy (HEGLE), a novel approach aimed at enhancing the quality and reliability of the data analysis. Following this, the extracted features are classified using a Kernel-based Radial Basis Function (Ker\_Rad\_BF), a technique chosen for its effectiveness in handling the complexities of multilingual data.

The paper presents a detailed comparative analysis of the performance of these techniques using two datasets: HASOC and TRAC. For the HASOC dataset, the combined HEGLE and Ker\_Rad\_BF approach achieved impressive results, including an accuracy of 98%, precision of 97%, recall of 90.5%, and an F-1 score of 85%. Additionally, the Root Mean Square Error (RMSE) was recorded at 55.6%, and the loss curve analysis showed a 44% rate. Similar high-performance metrics were observed with the TRAC dataset, where the system also achieved an accuracy of 98%, precision of 97%, recall of 91%, F-1 score of 87%, and an RMSE of 55%.

These results underscore the effectiveness of the proposed multilingual data processing framework, highlighting its potential in addressing the challenges of sentiment analysis across different languages. The study not only contributes to the field of NLP by providing a robust model for semantic analysis but also sets a benchmark for future research in multilingual sentiment analysis.

**Gpt-3** This paper delves into the realm of Knowledge-based Data Processing for Multilingual Natural Language Analysis, emphasizing the pivotal role of Natural Language Processing (NLP) in advancing human-computer communication through intelligent machines. Leveraging recent advancements in processing power and the abundance of linguistic data, the study underscores the growing need for data-driven approaches to automatic semantic analysis. The research introduces a novel methodology for multilingual data processing, employing feature extraction and classification through deep learning architectures. By collecting and refining input text data from diverse languages, the study utilizes Histogram Equalization based Global Local Entropy (HEGLE) for data extraction and Kernel-based Radial basis Function (Ker\_Rad\_BF) for classification. The proposed architectures exhibit promise in natural language processing, particularly in addressing multilingual sentiment analysis challenges. Through algorithm implementation and precision comparisons, the study achieves remarkable results, with the HEGLE\_Ker\_Rad\_BF model attaining an accuracy of 98%, precision of 97%, recall of 90.5%, and an f-1 score of 85% on the HASOC dataset. Similarly, on the TRAC dataset, the model achieves an accuracy of 98%, precision of 97%, recall of 91%, and an f-1 score of 87%. These findings underscore the efficacy of the proposed approach in multilingual sentiment analysis, showcasing its potential for enhancing natural language processing tasks.

**Elitsumm-** Recent developments in Natural Language Processing (NLP) have leveraged enhanced processing power and vast linguistic datasets to improve semantic analysis through data-driven methods. This paper proposes a multilingual data processing approach using deep learning architectures for feature extraction and classification, focusing on preprocessing multilingual text data to eliminate missing and null values.Amaar et al. proposed a method to detect fraudulent job postings using machine learning and natural language processing, employing TF-IDF, BoW, and ADASYN, achieving 99.9% accuracy [1]. Esmaeilzadeh et al. demonstrate that Apache Spark\'s Spark-SQL API significantly outperforms its RDD API in runtime performance for feature extraction from large text datasets like English Wikipedia, highlighting its efficiency in large-scale NLP feature engineering [2]. Trappey et al. developed an AI-driven patent summarization system using machine learning and natural language processing, automating key information extraction and overcoming manual abstraction challenges. The system demonstrated high precision and recall, enhancing knowledge accessibility and supporting innovation [3]. Lavanya and Sasikala discussed the effectiveness of deep learning techniques, particularly CNN and RNN, for text classification in social healthcare networks. They highlighted how these models process unstructured data, utilize NLP to analyze and extract features, and improve accuracy and performance in extracting meaningful information from large health datasets [4]. Wang et al. discuss a novel anomaly detection method in system logs using NLP (Word2vec, TF-IDF) and LSTM, demonstrating superior performance over traditional machine learning algorithms by effectively capturing contextual semantics [5]. Bawazeer et al. proposed a novel radial basis function (RBF) approach using Hermite expansion, enhancing stability and accuracy without needing parameter optimization. This method improves interpolation, derivative approximation, and PDE solutions, showing significant benchmarks in accuracy and stability [6]. Madeira et al. propose a methodology using machine learning and NLP to classify human factors in aviation incidents, achieving high predictive accuracy with Micro F1 scores of 0.900, 0.779, and 0.875. Techniques include text pre-processing, label spreading, and SVMs, with recommendations for larger datasets in future studies [7]. Jain et al. explore knowledge-based data processing in multilingual natural language analysis, emphasizing advancements in NLP and the importance of knowledge-driven approaches for tasks like text classification, machine translation, and information extraction [8]. Chotirata & Meesad proposed a methodology using part-of-speech tagging, feature selection, and word embedding to enhance Thai wh-question classification. Their approach, integrating POS tagging with SVM and CNN models, achieved high accuracy on TREC-6 and Thai datasets [9]. Marie-Sainte et al. discuss Arabic Natural Language Processing (ANLP), which uses machine learning (ML) techniques to develop tools for analyzing Arabic in both written and spoken forms. They emphasize the importance of understanding ML methodologies and the complexity of the Arabic language [10]. The text pertains to the title and editorial details of the "Proceedings of International Conference on Recent Trends in Computing ICRTC 2021," without an abstract or detailed content [11]. Kumar et al. proposed a hybrid approach combining SVM, NLP, and probabilistic neural networks for detecting email phishing, enhancing accuracy by integrating feature extraction and classification techniques to address the significant cybersecurity threat posed by socially engineered attacks [12]. Alwehaibi & Roy investigated the impact of various pre-trained word vectors on Arabic text classification using LSTM-RNN, demonstrating significant accuracy improvements with data from Twitter [13]. García-Méndez et al. proposed a novel system that combines Natural Language Processing and Machine Learning to detect temporality in financial news, distinguishing between past events and predictions. This enhances financial decision-making by identifying predictive knowledge with high precision [14]. Shi et al. propose a deep-learning traffic classification method using NLP-inspired features to identify multiple video streaming sources simultaneously. The method shows strong performance in binary and multilabel classification, including zero-shot learning, validated on a large dataset [15]. Pal and Patel investigated the classification of 154 Hindi poetries using NLP and five machine learning techniques: SVM, Naïve Bayes, Decision Tree, Random Forest, and K-NN. Their study, employing k-fold cross-validation and holdout methods, found SVM, NB, and Random Forest to be most effective [16]. Muntean & Donea discussed an advanced Multi-Agent System that uses NLP to improve human-machine interfaces by integrating a Collective User Knowledge-Base and GPS Database, enhancing user experience and information precision [17].

**Litllm -In the burgeoning field of Natural Language Processing (NLP), researchers are continually striving to enhance the interfaces between humans and machines, leveraging innovative techniques to improve user interaction and information accuracy. The proposed paper aims to contribute to this domain by introducing an intelligent Multi-Agent System that integrates NLP modules with a Collective User Knowledge-Base and GPS Database to furnish precise responses to user inquiries. In crafting the related work section for this document, it is crucial to explore a spectrum of scholarly contributions that contextualize and underscore the necessity and innovation of the proposed methodology.1. Feature Engineering and Computational Efficiency One pivotal aspect of NLP involves feature engineering, a computationally intensive process, especially when dealing with large datasets. The research [1] evaluates the runtime performance of different APIs within the Apache Spark framework, demonstrating that SparkSQL offers superior efficiency over RDD for feature extraction tasks. This finding is particularly relevant as it underscores the importance of selecting optimal computing frameworks to handle extensive text data, a consideration that must also be factored into the proposed system's design to ensure its responsiveness and scalability.2. Fraudulent Job Ad Detection Using NLP and Machine Learning,The detection of fraudulent job advertisements using NLP and machine learning techniques [2] highlights the practical applications of NLP in safeguarding users against online scams. By employing techniques like TF-IDF and BoW for feature extraction and experimenting with various machine learning models, this study not only illustrates the efficacy of NLP in real-world applications but also discusses the challenge of dataset imbalance. The use of ADASYN to address this issue informs the proposed system's approach to handling uneven data distributions, ensuring robust model training and performance.\*\*3. Patent Document Analysis\*\*In the domain of intellectual property, NLP and machine learning are leveraged to automate the summarization of patent documents [3]. This process, traditionally manual, is both time-consuming and subjective. By automating it, the research illustrates NLP's capability to distill essential information from complex texts, a feature that can be integrated into the proposed system to enhance its ability to analyze and utilize structured data effectively from patents or similar documents.\*\*4. Deep Learning in Health Data Analysis\*\*The application of deep learning models like CNNs and RNNs in analyzing complex health data [4] provides insights into the advanced capabilities of NLP techniques in interpreting intricate patterns and structures within large datasets. For the proposed system, adopting similar deep learning frameworks could potentially improve the analysis and processing of user queries and information extraction from the knowledge base.5. Anomaly Detection in System Logs,Exploring anomaly detection in system logs using LSTM models [5] presents an application of NLP in monitoring and maintaining system integrity. The effective use of LSTM to parse and understand system logs for anomalies could be analogous in the proposed system to detect errors or inconsistencies in user queries or the knowledge database, enhancing reliability.\*\*6. Enhancing Human Factor Analysis in Aviation\*\*Research on classifying human factors in aviation incident reports using NLP and machine learning [6] demonstrates the utility of NLP in safety-critical applications. This methodology could inspire similar use cases in the proposed system, where understanding and classifying user input accurately is crucial for system performance and safety.7. Arabic Text Sentiment Analysis\*\*The study of Arabic text sentiment analysis using LSTM-RNN [7] highlights the challenges and solutions in processing non-English languages. This is particularly relevant for the proposed system if multilingual support is envisioned, ensuring inclusivity and broad applicability.8. Detecting Phishing Emails\*\*The approach to detecting phishing emails using SVM [8] illustrates how NLP can be employed to protect users from cyber threats. Integrating a similar feature in the proposed system could safeguard users from potential scams or malicious entities through intelligent analysis of communication content.9. Market Screening via Text Analysis\*\*The novel system for detecting temporality in financial news [9] to aid in market screening shows how NLP can be used to extract actionable insights from complex, dynamic data sources. This capability could be adapted for the proposed system to provide real-time, context-aware information to users based on current events or data trends. Traffic Classification with NLP Techniques.Lastly, the use of NLP techniques in traffic classification [10] demonstrates the versatility of NLP in different domains, including encrypted traffic analysis. This could inform security features within the proposed system, ensuring secure and private user interactions.In summary, the related works discussed provide a broad spectrum of applications and challenges that enrich the understanding and implementation possibilities for the proposed intelligent Multi-Agent System. By drawing lessons from these diverse applications, the proposed methodology can be optimized for performance, accuracy, scalability, and user safety, making it a significant contribution to the field of human-machine interfaces.**

Model T5 and Bart , Gemini -1.5 , GPT-4o and Gpt-3.5 and LitLLm and ELitsumm

Readability Relevancy Grammaticality Focus Consistency Coherence Integrity

Fluency