International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 14, No. 3, December 2025, pp. 941~949
ISSN: 2252-8776, DOI: 10.11591/ijict.v14i3.pp941-949

Empowering low-resource languages: a machine learning approach to Tamil sentiment classification

Saleem Raja Abdul Samad[1], Pradeepa Ganesan[1], Justin Rajasekaran[1], Madhubala Radhakrishnan[1], Peerbasha Shebbeer Basha[2], Varalakshmi Kuppusamy[3]
[1]Information Technology Department, College of Computing and Information Sciences, University of Technology and Applied Sciences, Shinas, Oman
[2]Department of Computer Science, Jamal Mohamed College, Tiruchirappalli, Affiliated to Bharathidasan University, Tamil Nadu, India
[3]Department of Computer Science and Engineering, SRM University, Chennai, India

Article Info

Article history:

Received Sep 18, 2024
Revised May 18, 2025
Accepted Jul 1, 2025

Keywords:

FastText
Natural language processing
Tamil sentiment classification
Text embedding
Word embedding
Word2Vec

ABSTRACT

Sentiment analysis is essential for deciphering public opinion, guiding decisions, and refining marketing strategies. It plays a crucial role in monitoring public sentiment, fostering customer engagement, and enhancing relationships with businesses' target audiences by analyzing emotional tones and attitudes in vast textual data. Sentiment analysis is extremely limited, particularly for languages like Tamil, due to limited application in diverse linguistic contexts with fewer resources. Given its global impact and linguistic diversity, addressing this gap is crucial for a more nuanced understanding of sentiments in India. In the context of Tamil, the need for sentiment analysis models is particularly crucial due to its status as one of the classical languages spoken by millions. The cultural, social, and historical nuances embedded in Tamil language usage require tailored sentiment analysis approaches that can capture the subtleties of sentiment expression. This paper introduces a novel method that assesses the performance of various text embedding methods in conjunction with a range of machine learning (ML) algorithms to enhance sentiment classification for Tamil text, with a specific focus on lyrics. Experiments notably emphasize FastText word embedding as the most effective method, showcasing superior results with a remarkable 78% accuracy when coupled with the support vector classification (SVC) model.

Corresponding Author:

Peerbasha Shebbeer Basha
Department of Computer Science, Jamal Mohamed College, Tiruchirappalli
Tamil Nadu. India
Email: bashapeer2003@gmail.com

1. INTRODUCTION

Tamil is one of the world's oldest ancient languages, with a great literary legacy spanning thousands of years. The Indian states of Tamil Nadu, Puducherry, and Sri Lanka all use Tamil as their official language. Its significance transcends geographical limitations, as there exists a worldwide Tamil-speaking diaspora [1]. Tamil literature serves as a crucial repository for the cultural heritage of the Tamil-speaking community, comprising prose, poetry, and ancient scriptures. Sentiment analysis is paramount in today's data-driven landscape, enabling organizations to decipher and respond to the sentiments expressed in vast volumes of textual data. Using natural language processing (NLP) techniques, sentiment analysis extracts valuable insights from customer reviews, social media interactions, and other textual sources, providing businesses with a nuanced understanding of public opinion. This information is instrumental in shaping strategic decisions, enhancing brand perception, and improving customer satisfaction. From market research and product development to brand reputation management and political analysis, sentiment analysis is a powerful tool across various domains, fostering informed decision-making and responsive communication in a rapidly evolving digital environment [2], [3].

Incorporating sentiment analysis specifically for Tamil is of profound significance, offering a nuanced comprehension of emotions and opinions within this vibrant linguistic community. A more refined understanding emerges by applying sentiment analysis to Tamil text, enabling effective communication strategies, content personalization, and targeted decision-making [4]. Businesses can leverage sentiment analysis on Tamil reviews and social media to tailor their products and services, resulting in elevated customer satisfaction. Moreover, sentiment analysis in Tamil holds considerable importance in political and social realms, providing policymakers and researchers with invaluable insights into public sentiment. This, in turn, contributes to more informed and insightful decision-making processes [5].

The researchers have presented a variety of methodologies; however, the challenges persist due to the shortage of labeled data and the intricate nature of the Tamil language. The introduction of code-mixing, integrating multiple languages in communication, adds complexity to sentiment analysis by incorporating diverse linguistic elements. Addressing these obstacles demands specialized approaches that account for the intricacies of both the Tamil language and code-mixed text, underscoring the necessity for dedicated research and methodologies in this distinctive domain. This paper uses word embedding methods and machine learning (ML) to classify Tamil lyrical tweet statements. Experiments notably emphasize FastText word embedding as the most effective method, showcasing superior results with a remarkable 78% accuracy when coupled with the support vector classification (SVC) model.

The subsequent sections of the paper are structured as follows: section 1 delineates the necessity of sentiment analysis for the Tamil language and identifies research gaps. Section 2 provides an overview of recent research endeavors. Section 3 outlines the proposed methodology. Section 4 showcases experimental results and engages in discussion. The paper concludes in section 5.

2. METHOD

This section offers an outline of current research efforts in the domain of Tamil sentiment analysis. Se *et al.* [6] utilized ML algorithms, including support vector machine (SVM), Maxent classifier, decision tree (DT), and Naive Bayes, to classify Tamil movie reviews into positive and negative categories. The dataset, collected from various web sources, incorporated features from TamilSentiwordnet, with SVM demonstrating superior performance, achieving an accuracy of 75.9%. Thavareesan and Mahesan [7] critically analyze recent literature on sentiment analysis employing Tamil text, concluding that SVM and recurrent neural network (RNN) classifiers using term frequency-inverse document frequency (TF-IDF) and Word2vec features outperform grammar rule-based classifications and other classifiers. Using different corpora and feature representation techniques, Thavareesan and Mahesan [8] experimented with various sentiment analysis approaches, including lexicon-based, supervised ML, hybrid, and clustering with bag of word methods. A maximum accuracy of 79% was attained for the UJ_Corpus_Opinions_Nouns corpus using FastText in the supervised ML approach, incorporating both basic and traditional features. Babu and Sri [9] introduced hybrid deep learning approaches, including convolutional neural network- bidirectional long short-term memory (CNN-BiLSTM), CNN-long short-term memory (LSTM), and CNN- bidirectional gated recurrent unit (BiGRU), leveraging tools supporting the Tamil language for data preparation. The models were evaluated based on metrics like accuracy, recall, and F1, revealing that CNN-BiLSTM achieved the highest accuracy (80.2%) and F1-score (0.64) compared to other models, effectively classifying sentiments in Tamil movie reviews. Kishore *et al.* [10] employed ML models for sentiment analysis in Tamil and Tulu languages, utilizing a code-mixed dataset from social media. Achieving 64% accuracy and a 43% macro F1 score for Tamil and 66% accuracy and 51% macro F1 score for Tulu with TF-IDF feature extraction, their study highlights the efficacy of the TF-IDF with logistic regression (LR) model, emphasizing its potential applications in addressing social issues and fostering inclusivity online. The research findings unequivocally indicate that combining ML models with word embeddings produces better results.

The proposed method is structured into two distinctive phases. In the initial phase, word embeddings are generated for each word in Tamil text using four diverse embedding methods: count, TF-IDF, Hashing, Word2Vec, and FastText. These techniques are employed to vectorize the Tamil text, providing a comprehensive representation of the semantic and contextual information present in the language. Moving to the second phase, eight distinct ML methods are employed. Each method is individually trained and tested to evaluate its classification performance, thereby enabling a thorough assessment of the effectiveness of the proposed approach, as depicted in Figure 1.

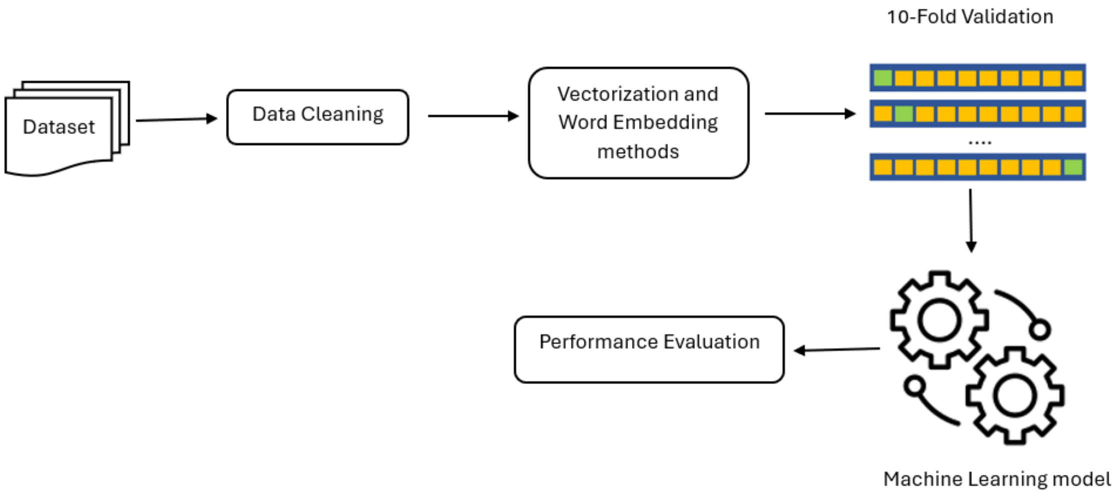

Figure 1. Proposed method

2.1. Dataset

The dataset comprises 1,000 tweets sourced from Twitter, each labeled "Happy" or "Sad." This dataset primarily aims to facilitate binary classification, sentiment analysis, and other NLP tasks. Notably, the distribution of labels is balanced, with 50% of the tweets categorized as "Sad" and the remaining 50% as "Happy." This balanced representation ensures the dataset is well-suited for training and evaluating models in sentiment analysis and related NLP applications [11].

2.2. Data cleaning

Data cleaning in the context of Tamil text (lyrics) involves the removal of special symbols, alphabets, and numbers to ensure a refined and standardized dataset. This process enhances the quality of the text data by eliminating unnecessary elements that may not contribute to the sentiment analysis task. The cleaning steps typically include stripping away punctuation marks, special characters, and numeric values, leaving behind a preprocessed text more conducive to accurate sentiment analysis. This meticulous cleaning aids in creating a streamlined and uniform dataset, optimizing the subsequent analysis and model training stages.

2.3. Vectorizing or embedding

Vectorizing or embedding Tamil text involves transforming textual data into numerical representations that capture semantic and contextual information. Various techniques, such as count vectorization, TF-IDF, Hashing, Word2Vec, and FastText, are employed to convert each word or document into a corresponding vector in a high-dimensional space [12], [13]. These vectorization methods are crucial in preparing the Tamil text for ML models, facilitating tasks like sentiment analysis. By representing words or documents as numerical vectors, these techniques enable algorithms to process and understand the inherent patterns and relationships within the language, contributing to more effective and accurate computational analysis of Tamil textual data.

2.3.1. Count vectorizer

Count vectorization is a simple technique for transforming text data into numerical format [14]. It represents each document as a vector of term frequencies. For Tamil sentiment classification, each unique word in the corpus is assigned an index, and the count of occurrences of each word is used to construct a vector representation.

$$Count(t, d) = Number\ of\ occurrences\ of\ term\ t\ in\ document\ d$$

2.3.2. TF-IDF vectorizer

The TF-IDF approach assesses a word's significance in a document about a group of documents [15]. It applies weights to words using TF and IDF.

$$TFIDF(w, d, D) = TF(w, d) x\ IDF(w, D)$$

Where:

$$\text{TF}(\text{w}, \text{d}) = \frac{\text{Number of occurences of word } w \text{ in document } d}{\text{Total number of words in } d}$$

$$\text{IDF } (\text{w}, \text{d}) = \log \left(\frac{\text{Total number of documents in } D}{\text{Number of documents containing word } w}\right)$$

2.3.3. Hashing vectorizer

Hashing vectorizer is a text vectorization technique that employs a hash function to map words directly to feature indices, converting them into fixed-size vectors [15].

2.3.4. Word2Vec (continuous bag of words)

Word2vec is a popular word embedding technique used in NLP tasks, such as sentiment analysis [16]. In the context of Tamil sentiment classification, Word2Vec employs the continuous bag of words (CBOW) model. CBOW predicts a target word based on its context, aiming to understand the distributional semantics of words.

2.3.5. FastText

FastText is an extension of Word2Vec that considers sub-word information. It breaks words into smaller n-gram sub-words and represents them as the sum of these sub-word embeddings [17]. This particularly benefits languages like Tamil, where words can have complex morphological structures.

2.4. Cross validation and performance evaluation

Following the vectorization or embedding process, a comprehensive 10-fold cross-validation [18] is conducted using eight diverse ML schemes, such as SVM, LR, Naive Bayes, random forest (RF), DT, gradient boosting (GB), extreme gradient boosting (XGB), and AdaBoost [19]-[23]. This meticulous evaluation assesses the performance of each model in the context of Tamil sentiment classification. The performance is measured across various evaluation metrics, including accuracy, precision, recall, and F1 score.

Each ML model undergoes rigorous testing and training across the 10 folds, ensuring a robust assessment of its efficacy in handling the complexities of Tamil sentiment analysis. The chosen evaluation metrics provide a holistic view of the model's performance, accounting for aspects such as overall correctness (accuracy), capability to appropriately identify positive instances (precision), capability to seizure all positive instances (recall), and the harmonic mean of precision and recall (F1 score) [24]-[30]. This multifaceted evaluation strategy helps identify the appropriate ML model for achieving optimal sentiment classification results in the Tamil context.

3. RESULTS AND DISCUSSION

The experimental results of the proposed method were conducted using Python and Jupyter Notebook, leveraging the popular SKlearn package for ML functionalities. Eight distinct ML models, namely LR, SVM, Naive Bayes, DT, RF, XGB, GB, and AdaBoost, were employed for evaluation.

Table 1 presents the experimental results for the count vectorizer. Subsequent tables, Tables 2 to 5, showcase the outcomes of TF-IDF, Hashing vectorizer, Word2Vec, and FastText embeddings, respectively. These tables collectively provide a comprehensive overview of the performance of different vectorization techniques and word embeddings in the sentiment analysis task for the Tamil language. Each table encapsulates the evaluation metrics, including accuracy, precision, recall, and F1 score, offering insights into the effectiveness of the corresponding methodologies.

Table 1. Performance of count vectorizer

Model	Accuracy	Precision	Recall	F1-score
LR	68.38	68.56	71.11	69.20
SVC	68.59	68.54	73.64	70.12
GNB	57.36	62.36	32.62	42.25
DT	63.75	64.48	64.82	63.92
RF	70.45	73.78	69.13	69.86
GB	69.17	69.38	75.21	70.92
AdaBoost	64.65	65.57	66.02	64.21
XGB	67.30	68.24	68.75	67.52

FastText, a powerful word embedding technique, leverages pre-trained models for Tamil text, accessible on the FastText website. The performance results indicate that FastText word embeddings outshine other methods, demonstrating superior effectiveness. Specifically, with the SVC model, FastText achieves an impressive 78% accuracy and F1-score, underscoring its prowess in capturing nuanced semantics and contributing significantly to the success of sentiment analysis in the Tamil language.

Table 2. Performance of TF-IDF vectorizer

Model	Accuracy	Precision	Recall	F1-score
LogR	68.59	68.60	71.89	69.57
SVC	67.89	67.45	72.08	69.14
GNB	57.26	62.01	36.75	45.51
DT	62.16	63.34	64.02	62.56
RF	68.89	69.08	71.69	69.53
GB	69.57	69.09	75.21	71.05
AdaBoost	66.91	68.15	67.38	66.80
XGB	65.43	66.30	66.58	65.37

Table 3. Performance of Hashing vectorizer

Model	Accuracy	Precision	Recall	F1-score
LogR	66.23	66.98	68.35	66.98
SVC	65.93	66.83	67.76	66.61
GNB	58.83	59.14	54.63	56.14
DT	58.92	60.49	56.76	57.90
RF	67.51	69.08	67.38	67.38
GB	68.77	69.33	74.23	70.53
AdaBoost	64.05	65.71	63.47	63.22
XGB	65.04	65.81	66.78	65.51

Table 4. Performance of Word2Vec embedding

Model	Accuracy	Precision	Recall	F1-score
LogR	63.68	64.98	56.76	60.49
SVC	66.39	72.49	50.46	59.25
GNB	59.12	71.24	27.10	38.91
DT	56.91	55.75	58.53	56.99
RF	64.91	65.20	61.26	63.04
GB	66.52	66.41	63.78	64.98
AdaBoost	59.49	59.07	57.01	57.86
XGB	62.46	61.53	63.29	62.26

Table 5. Performance of FastText embedding

Model	Accuracy	Precision	Recall	F1-Score
LogR	75.00	76.00	73.00	75.00
SVC	78.00	78.00	77.00	78.00
GNB	71.00	74.00	66.00	69.00
DT	63.00	64.00	63.00	63.00
RF	74.00	76.00	72.00	74.00
GB	74.00	75.00	73.00	74.00
AdaBoost	71.00	71.00	70.00	71.00
XGB	75.00	76.00	74.00	75.00

Figure 2 illustrates the receiver operating characteristic (ROC) area under the curve (AUC) [24], [25] for FastText word embeddings coupled with the SVC model. On the other hand, Figure 3 presents the confusion matrix associated with this configuration. These visualizations offer insightful information into the model's discriminatory power and its performance in distinguishing between different sentiment classes in Tamil text.

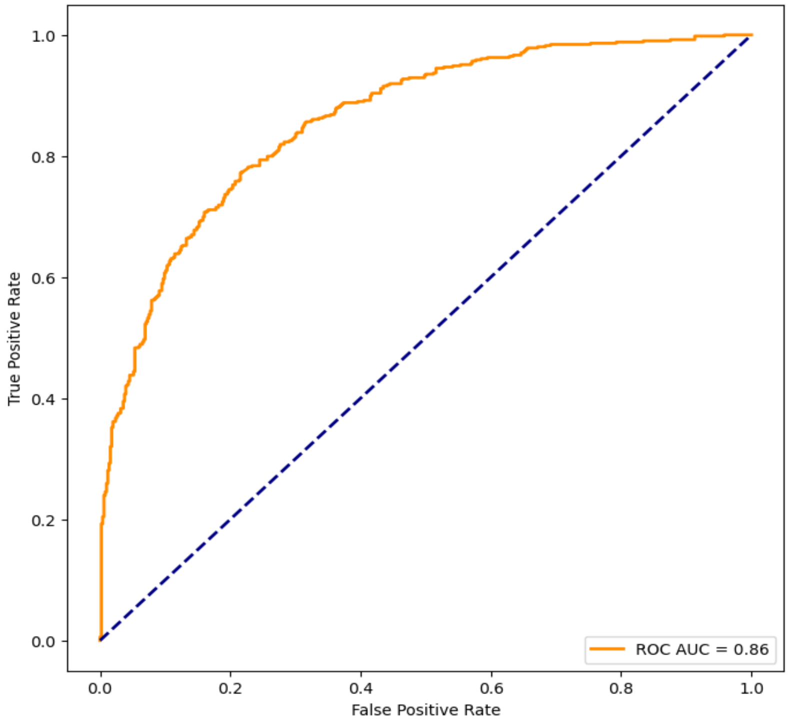

Figure 2. ROC-AUC curve for FastText word embedding with SVC model

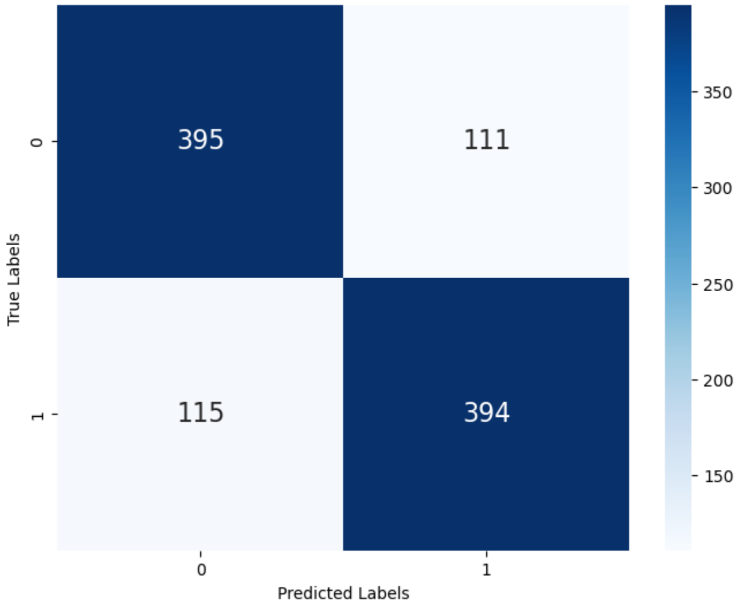

Figure 3. Confusion matrix for FastText word embedding with SVC model

4. CONCLUSION

The proposed methodology systematically evaluates diverse text embedding methods alongside various ML algorithms to enhance sentiment classification for Tamil text, focusing particularly on lyrics. Our findings underscore the effectiveness of FastText word embeddings through extensive experimentation, showcasing superior performance with a notable 78% accuracy when employed with the SVC model. The results affirm the significance of leveraging FastText embeddings for sentiment analysis in Tamil lyrics, emphasizing their potential to capture nuanced linguistic nuances and achieve commendable classification accuracy.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Saleem Raja Abdul Samad	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Pradeepa Ganesan		✓	✓	✓		✓		✓	✓	✓	✓	✓		
Justin Rajasekaran	✓						✓			✓	✓		✓	✓
Madhubala Radhakrishnan					✓					✓			✓	
Peerbasha Shebbeer Basha		✓	✓	✓	✓		✓			✓		✓		✓
Varalakshmi Kuppusamy	✓							✓		✓			✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors declare no conflict of interests

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

REFERENCES

[1] B. Krishnamurti, “Tamil language,” *Britannica*. https://www.britannica.com/topic/Tamil-language (accessed Feb. 26, 2024).

[2] Q. A. Xu, V. Chang, and C. Jayne, “A systematic review of social media-based sentiment analysis: emerging trends and challenges,” *Decision Analytics Journal*, vol. 3, p. 100073, Jun. 2022, doi: 10.1016/j.dajour.2022.100073.

[3] “Sentiment analysis explained: from theory to real-world applications,” *Archive*, 2023. https://archive.com/blog/sentiment-analysis (accessed Feb. 26, 2024).

[4] R. Padmamala and V. Prema, “Sentiment analysis of online Tamil contents using recursive neural network models approach for Tamil language,” in *2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)*, Aug. 2017, pp. 28–31, doi: 10.1109/ICSTM.2017.8089122.

[5] S. Anbukkarasi and S. Varadhaganapathy, “Analyzing sentiment in Tamil tweets using deep neural network,” in *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2020, pp. 449–453, doi: 10.1109/ICCMC48092.2020.ICCMC-00084.

[6] S. Se, R. Vinayakumar, M. A. Kumar, and K. P. Soman, “Predicting the sentimental reviews in Tamil movie using machine learning algorithms,” *Indian Journal of Science and Technology*, vol. 9, no. 45, Dec. 2016, doi: 10.17485/ijst/2016/v9i45/106482.

[7] S. Thavareesan and S. Mahesan, “Review on sentiment analysis in Tamil texts,” *Journal of Science*, vol. 9, no. 2, pp. 1–9, Dec. 2018, doi: 10.4038/jsc.v9i2.14.

[8] S. Thavareesan and S. Mahesan, “Sentiment analysis in Tamil texts: a study on machine learning techniques and feature representation,” in *2019 14th Conference on Industrial and Information Systems (ICIIS)*, Dec. 2019, pp. 320–325, doi: 10.1109/ICIIS47346.2019.9063341.

[9] R. Babu and S. Sri, “Sentiment analysis in Tamil language using hybrid deep learning approach,” National College of Ireland, 2022.

[10] K. K. Ponnusamy, C. Rajkumar, P. K. Kumaresan, E. Sherly, and R. Priyadharshini, “VEL@DravidianLangTech: sentiment analysis of Tamil and Tulu,” in *DravidianLangTech 2023 - 3rd Workshop on Speech and Language Technologies for Dravidian Languages, associated with 14th International Conference on Recent Advances in Natural Language Processing, RANLP 2023 - Proceedings*, 2023, pp. 211–216, doi: 10.26615/978-954-452-085-4_030.

[11] R. C. Moore, D. P. W. Ellis, E. Fonseca, S. Hershey, A. Jansen, and M. Plakal, “Dataset balancing can hurt model performance,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10095255.

[12] S. R. A. Samad *et al.*, “Analysis of the performance impact of fine-tuned machine learning model for phishing URL detection,” *Electronics*, vol. 12, no. 7, p. 1642, Mar. 2023, doi: 10.3390/electronics12071642.

[13] S. R. A. Samad, P. Ganesan, J. Rajasekaran, M. Radhakrishnan, H. Ammaippan, and V. Ramamurthy, “SmishGuard: leveraging machine learning and natural language processing for smishing detection,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 11, pp. 586–593, 2023, doi: 10.14569/IJACSA.2023.0141160.

[14] A. S. S. Raja, G. Pradeepa, S. Mahalakshmi, and M. S. Jayakumar, “Natural language based malicious domain detection using machine learning and deep learning,” *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol. 23, no. 2, pp. 304–312, Apr. 2023, doi: 10.17586/2226-1494-2023-23-2-304-312.

[15] A. S. Raja, S. Peerbashab, Y. M. Iqbal, B. Sundarvadivazhagan, and M. M. Surputheen, “Structural analysis of URL for malicious URL detection using machine learning,” *Journal Of Advanced Applied Scientific Research*, vol. 5, no. 4, pp. 28–41, Jul. 2023, doi: 10.46947/joaasr542023679.

[16] S. Sivakumar, L. S. Videla, T. Rajesh Kumar, J. Nagaraj, S. Itnal, and D. Haritha, “Review on Word2Vec word embedding neural net,” in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Sep. 2020, pp. 282–290, doi: 10.1109/ICOSEC49089.2020.9215319.

[17] N. Badri, F. Kboubi, and A. H. Chaibi, “Combining FastText and Glove Word embedding for offensive and hate speech text detection,” *Procedia Computer Science*, vol. 207, pp. 769–778, 2022, doi: 10.1016/j.procs.2022.09.132.

[18] J. Kaliappan, A. R. Bagepalli, S. Almal, R. Mishra, Y.-C. Hu, and K. Srinivasan, “Impact of cross-validation on machine learning models for early detection of intrauterine fetal demise,” *Diagnostics*, vol. 13, no. 10, p. 1692, May 2023, doi: 10.3390/diagnostics13101692.

[19] P. Dhanalakshmi, G. A. Kumar, B. S. Satwik, K. Sreeranga, A. T. Sai, and G. Jashwanth, “Sentiment analysis using VADER and logistic regression techniques,” in *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, Feb. 2023, pp. 139–144, doi: 10.1109/ICISCoIS56541.2023.10100565.

[20] M. Ahmad, S. Aftab, M. Salman, and N. Hameed, “Sentiment analysis using SVM: a systematic literature review,” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, 2018, doi: 10.14569/IJACSA.2018.090226.

[21] P. P. Surya and B. Subbulakshmi, “Sentimental analysis using Naive Bayes classifier,” in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, Mar. 2019, pp. 1–5, doi: 10.1109/ViTECoN.2019.8899618.

[22] M. Aufar, R. Andreswari, and D. Pramesti, “Sentiment analysis on Youtube social media using decision tree and random forest algorithm: a case study,” in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, Aug. 2020, pp. 1–7, doi: 10.1109/ICoDSA50139.2020.9213078.

[23] A. Sawarn, Ankit, and M. Gupta, “Comparative analysis of bagging and boosting algorithms for sentiment analysis,” *Procedia Computer Science*, vol. 173, pp. 210–215, 2020, doi: 10.1016/j.procs.2020.06.025.

[24] F. S. Nahm, “Receiver operating characteristic curve: overview and practical use for clinicians,” *Korean Journal of Anesthesiology*, vol. 75, no. 1, pp. 25–36, Feb. 2022, doi: 10.4097/kja.21209.

[25] S. V Balshetwar, A. RS, and D. J. R, “Fake news detection in social media based on sentiment analysis using classifier techniques,” *Multimedia Tools and Applications*, vol. 82, no. 23, pp. 35781–35811, Sep. 2023, doi: 10.1007/s11042-023-14883-3.

[26] S. Amiriparian *et al.*, “The MuSe 2024 multimodal sentiment analysis challenge: social perception and humor recognition,” *arxiv preprint: 2406.07753*, Jun. 2024, doi: 10.48550/arXiv.2406.07753.

[27] J. Cabezas, D. Moctezuma, A. Fernández-Isabel, and I. Martín de Diego, “Detecting emotional evolution on Twitter during the COVID-19 pandemic using text analysis,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 13, p. 6981, Jun. 2021, doi: 10.3390/ijerph18136981.

[28] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, “Multimodal sentiment analysis: a systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions,” *Information Fusion*, vol. 91, pp. 424–444, 2023, doi: 10.1016/j.inffus.2022.09.025.

[29] R. Geethanjali and A. Valarmathi, “A novel hybrid deep learning IChOA-CNN-LSTM model for modality-enriched and multilingual emotion recognition in social media,” *Scientific Reports*, vol. 14, no. 1, p. 22270, Sep. 2024, doi: 10.1038/s41598-024-73452-2.

[30] K. Fountoulakis *et al.*, “Self-reported changes in anxiety, depression and suicidality during the COVID-19 lockdown in Greece,” *Journal of affective disorders*, vol. 279, pp. 624–629, 2021, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33190113/.

BIOGRAPHIES OF AUTHORS

Saleem Raja Abdul Samad is a faculty member in the IT Department, University of Technology and Applied Sciences, Shinas, Sultanate of Oman. He hold his Ph.D. degree in 2017 in the broad area of data mining and M.Tech – information technology degree from Bharathidasan University, Tamil Nadu, India. He has over 18 years of teaching and research experience. Delivered technical talk on natural language processing, data science, mobile programming, and IoT. He has been supervising applied research projects and graduate student projects in cyber security. He has completed CCNA, HCAI, and Microsoft Associate certifications. He has published more than 55 research papers in international and national journals, conferences, and book chapters. His research interests include cyber security, machine learning, deep learning, natural language processing, image processing, and the IoT. He can be contacted at email: saleem.abdulsamad@utas.edu.om.

Pradeepa Ganesan is a faculty member in the IT Department at the University of Technology and Applied Sciences, Shinas, Sultanate of Oman. She holds her MCA degree from Madurai Kamaraj University, Tamilnadu, India, her M.Phil degree from Bharathidasan University, Tamilnadu, India, and her Ph.D. from Vels Institute of Science, Technology and Advanced Studies, Tamilnadu, India. She has 20 years of teaching experience. Her research interests include cybersecurity, natural language processing, machine learning, and deep learning. She can be contacted at email: pradeepa.ganesan@utas.edu.om.

Justin Rajasekaran is working as a faculty member in University of Technology and Applied Sciences, Shinas, Oman. He has authored several research papers and chapters in reputed national and international journals and presented research papers in national and international conferences under various domains viz. Cloud computing in smart farming and its application, students' performance using data mining techniques, mobile application testing, malicious link detection using natural language processing. He has participated in various seminars, workshops, and Faculty Development Programs. He has also participated in designing curriculum for AI and data science. He can be contacted at email: justin.rajasekaran@utas.edu.om.

Madhubala Radhakrishnan currently serves as a Lecturer in the IT Department at the University of Technology and Applied Science–Shinas in Oman. She earned her Doctor of Philosophy degree from Vels Institute of Science, Technology and Applied Science, India 2021. Research interests include cloud computing, artificial intelligence, networks, IoT, service-oriented architecture, and data structures. She's been a teacher for twenty-four years. Her research has been presented at conferences, in book chapters, and in thirteen international journals. She has been serving as reviewer for several reputed journals. She can be contacted at email: balamadh@gmail.com.

Peerbasha Shebbeer Basha is an Assistant Professor of Computer Science Department at Jamal Mohamed College, Affiliated with Bharathidasan University, Tamil Nadu. India. He has 14 years of teaching experience including one year of international experience as a senior lecturer handling IT course of Southern Cross University, Australia. He holds his Ph.D. degree and qualified with NET and SET in computer science and applications. He holds MCA, M.Phil, MBA, and M.Tech. He has received Indo Asian Marvin Minsky Distinguished Innovative Researcher Award in computer science. He has published more tha twelve research articles in reputed international computer science journals. His area of research expertise includes machine learning and deep learning. He is a reviewer and editorial board member for many international journals related with computer science, engineering and information technology. He can be contacted at email: bashapeer2003@gmail.com.

Varalakshmi Kuppusamy holds a B.Tech. and an M.Tech. in information technology. She has served as a lecturer at Anna University-Affiliated Colleges and SRM University. In her academic career, she has guided numerous B.Tech. and M.Tech. student projects, focusing on areas such as artificial intelligence, image processing, networking, and data structures. Her extensive experience in teaching and mentoring has made significant contributions to the academic and professional growth of her students. She can be contacted at email: varamkelranga@gmail.com.