NAYEL@DravidianLangTech-2025: Character N-gram and Machine Learning Coordination for Fake News Detection in Dravidian Languages

Hamada Nayel[1,2], Mohammed Aldawsari[1], Hosahalli Lakshmaiah Shashirekha[3]

[1]Department of Computer Engineering and Information, Faculty of Engineering,
Prince Sattam Bin Abdulaziz University, Wadi Addawasir, Saudi Arabia
[2]Department of Computer Science, Faculty of Artificial Intelligence,
Benha University, Egypt
[3]Department of Computer Science, Mangalore University, India

Correspondence: hamada.ali@fci.bu.edu.eg

Abstract

This paper introduces the detailed description of the submitted model by the team NAYEL to Fake News Detection in Dravidian Languages shared task. The proposed model uses a simple character n-gram TF-IDF as a feature extraction approach integrated with an ensemble of various classical machine learning classification algorithms. While the simplicity of the proposed model structure, although it outperforms other complex structure models as the shared task results observed. The proposed model achieved a f1-score of 87.5% and secured the 5th rank.

1 Introduction

The growth of social media platforms have significantly contributed to the widespread issue of fake news across the globe. Fake news, which refers to intentionally deceptive or inaccurate information circulated through internet, presents serious challenges to public trust, governance, and societal health (Ashraf et al., 2022). While most research in fake news detection has concentrated on widely spoken languages like English, the distinct linguistic and cultural characteristics of regional languages have been largely overlooked (Nayel and Amer, 2021). Dravidian languages, spoken mainly in southern India, have received research attention in the context of automatically fake news detection (Subramanian et al., 2024; Devika et al., 2024; Subramanian et al., 2023).

Dravidian languages, such as Tamil, Telugu, Kannada, and Malayalam, are highly diverse in terms of their linguistic structures, including syntax, semantics, and morphology. These languages create unique difficulties for natural language processing (NLP) models, especially when applied to tasks like fake news detection. The variety in regional dialects, differences in scripts, and the influence of local culture further complicate the process of distinguishing between true and false information (Hegde et al., 2024, 2023).

This paper explores the submitted model to the Fake News Detection in Dravidian Languages shared task. This shared task is divided into two subtasks, the first subtask aimed at classifying a a given social media text into original or fake. While, the second subtask aimed at detecting the fake news from Malayalam News into five fake categories as well as original. Our team participated in the first subtask, and have submitted three runs.

The rest of the paper demonstrates the structure of the submitted model and the experimental results that have been produced in the devolvement phase.

2 Literature Review

Research works in fake news detection have gained interesting in last few years according to the massive usage of social media platforms.

Nayel and Amer (2021) used a simple Term Frequency-Inverse Document Frequency (TF-IDF) framework to extract the features of Urdu tweets and integrate with a linear classifier that achieved f1-score of 67.9% and outperformed all the submitted runs. The basic frame work that combines TF-IDF and ML algorithms has been used efficiently in various tasks such as text classification (Ashraf et al., 2024) and word level language identification (Ismail et al., 2022; Fetouh and Nayel, 2023). In the era of large language models (LLMs), they have been adapted for fake news detection (Hu et al., 2024; Su et al., 2024).

The research work has been done in fake news detection in Dravidian languages varies from the classical machine learning approaches (K et al., 2024), deep learning approaches (M et al., 2024) and transformer-based approach (Tabassum et al., 2024). K et al. (2024) explored character n-gram model with classical ML algorithms such as Linear Regression (LR), Support Vector Machines (SVMs), Naive Bayes (NB) and an ensemble model.

Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, pages 657–660

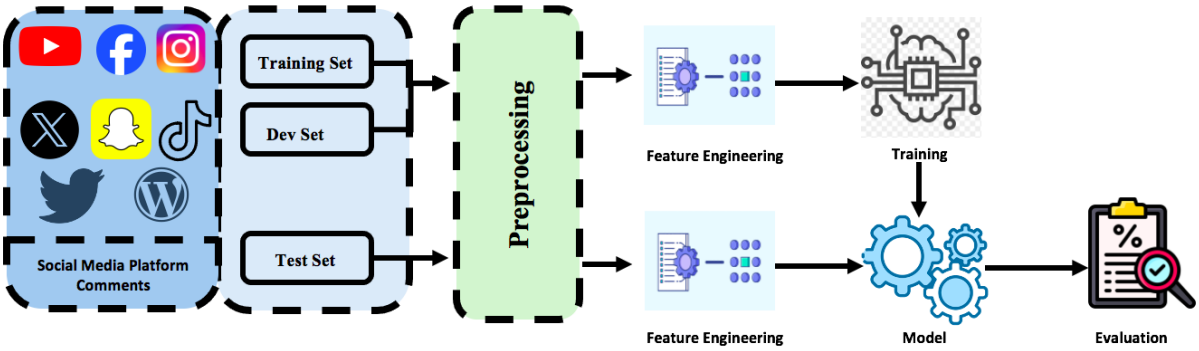

Figure 1: The structure of the proposed ML-based model

3 Dataset

The dataset has been used in this task was collected from various social media platforms such as X (formerly Twitter), Facebook etc. Detailed description of dataset, the methodology has been used to collect comments statistics and detailed analysis are given in (Subramanian et al., 2025). Table 1 shows the statistics of the dataset for both subtasks.

Task	Class	Train	Test	Dev
Task A	Fake	1599	507	406
	Original	1658	512	409
Task B	Half True	145	24	–
	False	1251	149	–
	Partly False	44	14	–
	Mostly False	242	63	–

Table 1: Statistics of the dataset

4 Methods

The proposed model composite of an ensemble of three base classifiers namely; SVMs, NB and linear classifier with Stochastic Gradient Descent (SGD) as an optimization algorithm. Majority voting mechanism has been used to combine the outputs of the base classifiers. As shown in figure 1, the general structure of the base classifier consists of dataset pre-processing, feature engineering, model training and evaluation.

- *Data pre-processing*
  Data pre-processing or text cleaning aims at omitting the unwanted texts such as stopwords, repeated letters, emojis and any uninformative tokens. In our model we employed a simple pre-processing procedure that removes the repeated characters, emojis and stopwords.

- *Feature Engineering*
  This phase involves feature extraction and selection. We employed a simple character n-gram TF-IDF for feature extraction. A wide range of n-grams including 3-gram, 4-gram, 5-gram and 6-gram have been extracted for the given text. Each token is consumed as a feature This approach reported improved results in .

- *Model Training*
  A set of classical machine learning classification algorithms have been implemented. The set of classifiers are: SVM, Naive Bayes and SGD. A voting-based ensemble model has been implemented using aforementioned classifiers as base classifiers. Ensemble learning is a machine learning technique that aggregates several individual models to produce more accurate predictions than a single model alone (Nayel and Shashirekha, 2017).

- *Evaluation*
  Performance evaluation of the model has been measured by f1-score, which is widely used in such cases of text classification. F1-score is a harmonic mean of precision (P) and recall (R) and is calculated as follows:

$$f1 - score = \frac{2PR}{R + P}$$

The general structure of voting-based ensemble model is shown in figure 2. The output of the base classifiers are input to a majority voting function.

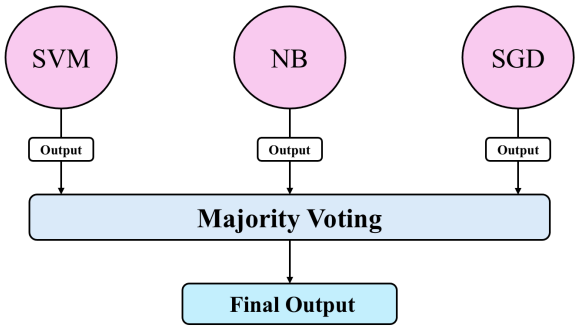

Figure 2: The structure of the majority voting ensemble model

5 Experimental Setting and Results

In this section experimental setting hyper parameters and running environment have been discussed. The code is freely available on GitHub repository[1].

A free package for `python` implementation of classical machine learning algorithms text features extraction `sklearn`[2] has been used to implement proposed model. In development phase and to fit the hyper parameters, the development set provided by shared task organizers has been used. To get the same results at each run, we utilized the package `random` and set the parameter `random_state` at 42.

The results of the applying the proposed models on development set are given in Table 2. The

Classifier	Precision	Recall	F1-score
SGD	0.89	0.86	0.87
NB	0.84	**0.88**	0.85
SVM	**0.90**	0.85	**0.88**
voting	0.89	0.86	**0.88**

Table 2: Results of the proposed model on development set

results on development set show that SGD-based model reported the minimum results in terms of all metrics. While, SVM outperforms other models in terms of precision. NB-based model reported the highest recall. SVM and voting classifiers reported the highest f1-score. Voting-based model, as it is clear, utilize the strength of all classifiers.

For test set, the majority voting based model outperforms all baseline model and reported f1-score of 0.875.

[1] https://github.com/hamadanayel/NAYEL_DRAVIDIAN

[2] https://scikit-learn.org

6 Conclusion

Fake news detection is a vital task in the era of social media expansion especially for low resources languages such as Dravidian languages. This work proposed a basic model that uses character n-gram TF-IDF and ML algorithms. The results obtained by the proposed model is a promising according to its simplicity and the low computational resources have been spent.

The model can be improved by applying more pre-processing steps as well as surveying more classical ML algorithms. In addition, LLMs can be tested in this case.

7 Limitations

The proposed model as clear uses character n-gram TF-IDF as a feature extraction approach, which is not efficient in the sense of semantic features. This is a lexicographical feature, while the meaning of the token is not used. In addition, The quality of dataset is a very important factor to develop an amenable systems.

LLMs can be applied to the task, but the limited computational resources access leads to applying the classical computational models such as ML and deep learning.

References

Nsrin Ashraf, Hamada Nayel, Mohammed Aldawsari, Hosahalli Shashirekha, and Tarek Elshishtawy. 2024. BFCI at AraFinNLP2024: Support vector machines for Arabic financial text classification. In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 446–449, Bangkok, Thailand. Association for Computational Linguistics.

Nsrin Ashraf, Hamada Nayel, and Mohamed Taha. 2022. Misinformation detection in arabic tweets: A case study about covid-19 vaccination. *Benha Journal of Applied Sciences*, 7(5):265–268.

K Devika, B Haripriya, E Vigneshwar, B Premjith, Bharathi Raja Chakravarthi, et al. 2024. From dataset to detection: A comprehensive approach to combating malayalam fake news. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 16–23.

Ahmed M. Fetouh and Hamada Nayel. 2023. BFCAI at coli-tunglish@fire 2023: Machine learning based model for word-level language identification in code-mixed tulu texts. In *Working Notes of FIRE 2023 -*

Forum for Information Retrieval Evaluation (FIRE-WN 2023), Goa, India, December 15-18, 2023, volume 3681 of *CEUR Workshop Proceedings*, pages 205–212. CEUR-WS.org.

Asha Hegde, F Balouchzahi, Sharal Coelho, Shashirekha H L, Hamada A Nayel, and Sabur Butt. 2024. Coli@fire2023: Findings of word-level language identification in code-mixed tulu text. In *Proceedings of the 15th Annual Meeting of the Forum for Information Retrieval Evaluation*, FIRE '23, page 25–26, New York, NY, USA. Association for Computing Machinery.

Asha Hegde, Fazlourrahman Balouchzahi, Sharal Coelho, H. L. Shashirekha, Hamada A. Nayel, and Sabur Butt. 2023. Overview of coli-tunglish: Word-level language identification in code-mixed tulu text at FIRE 2023. In *Working Notes of FIRE 2023 - Forum for Information Retrieval Evaluation (FIRE-WN 2023), Goa, India, December 15-18, 2023*, volume 3681 of *CEUR Workshop Proceedings*, pages 179–190. CEUR-WS.org.

Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. 2024. Bad actor, good advisor: Exploring the role of large language models in fake news detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(20):22105–22113.

Shimaa Ismail, Mai K. Gallab, and Hamada Nayel. 2022. BoNC: Bag of n-characters model for word level language identification. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, pages 34–37, IIIT Delhi, New Delhi, India. Association for Computational Linguistics.

Manavi K, Sonali K, Gauthamraj K, Kavya G, Asha Hegde, and Hosahalli Shashirekha. 2024. MUCS@DravidianLangTech-2024: Role of learning approaches in strengthening hate-alert systems for code-mixed text. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 252–256, St. Julian's, Malta. Association for Computational Linguistics.

Madhumitha M, Kunguma M, Tejashri J, and Jerin Mahibha C. 2024. TechWhiz@DravidianLangTech 2024: Fake news detection using deep learning models. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 200–204, St. Julian's, Malta. Association for Computational Linguistics.

Hamada Nayel and Ghada Amer. 2021. A simple n-gram model for urdu fake news detection. In *Working Notes of FIRE 2021 - Forum for Information Retrieval Evaluation, Gandhinagar, India, December 13-17, 2021*, volume 3159 of *CEUR Workshop Proceedings*, pages 1150–1155. CEUR-WS.org.

Hamada Nayel and H. L. Shashirekha. 2017. Improving NER for clinical texts by ensemble approach using segment representations. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, pages 197–204, Kolkata, India. NLP Association of India.

Jinyan Su, Claire Cardie, and Preslav Nakov. 2024. Adapting fake news detection to the era of large language models. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1473–1490, Mexico City, Mexico. Association for Computational Linguistics.

Malliga Subramanian, , B Premjith, Kogilavani Shanmugavadivel, Santhia Pandiyan, Balasubramanian Palani, and Bharathi Raja Chakravarthi. 2025. Overview of the Shared Task on Fake News Detection in Dravidian Languages: DravidianLangTech@NAACL 2025. In *Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, B Premjith, K Vanaja, S Mithunja, K Devika, et al. 2024. Overview of the second shared task on fake news detection in dravidian languages: Dravidianlangtech@ eacl 2024. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 71–78.

Malliga Subramanian, Bharathi Raja Chakravarthi, Kogilavani Shanmugavadivel, Santhiya Pandiyan, Prasanna Kumar Kumaresan, Balasubramanian Palani, Muskaan Singh, Sandhiya Raja, Vanaja, and Mithunajha S. 2023. Overview of the shared task on fake news detection from social media text. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Nafisa Tabassum, Sumaiya Aodhora, Rowshon Akter, Jawad Hossain, Shawly Ahsan, and Mohammed Moshiul Hoque. 2024. Punny_Punctuators@DravidianLangTech-EACL2024: Transformer-based approach for detection and classification of fake news in Malayalam social media text. In *Proceedings of the Fourth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*, pages 180–186, St. Julian's, Malta. Association for Computational Linguistics.