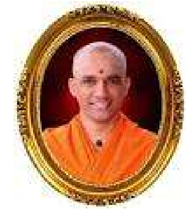
||Jai Sri Gurudev ||
BGSKH Education Trust (R.) – A unit of Sri AdichunchanagiriShikshanaTrust(R.)



BGS College of Engineering and Technology (BGSCET)
Mahalakshmipuram, West of Chord Road, Bengaluru-560086

PROJECT PHASE 1 (BCG685)
PRESENTATION
ON

Enhancing Media Credibility with Machine Learning-Based Fake News Detection

Team Members:
1. Diksha G(1MP22CG014)
2. Madhushree V(1MP22CG032)
3. Monisha S(1MP22CG035)
4. Divya Shankar(1MP23CG400)

Guide Name with Designation
Mrs.Sanjitha S Laad
(Assistant Professor)

Contents

Introduction

•Social media platforms and online news outlets enable the rapid spread of information; however, they also create opportunities for the dissemination of false or misleading content.

•To address this growing concern, Machine Learning (ML) techniques are utilized for automated fake news detection.

•Models such as Support Vector Machine (SVM) and Logistic Regression effectively analyze textual data to identify linguistic patterns, semantic inconsistencies, and deceptive cues that indicate fake news.

•These algorithms help in improving content verification processes, thereby minimizing the influence of misinformation.

•The implementation of ML-based detection systems enhances the credibility, accuracy, and trustworthiness of online information shared across digital platforms.

Objectives

• To develop an automated fake news detection system utilizing machine learning algorithms and deep learning models for accurate classification of news as real or fake.

• Model Training**:**
To train and optimize a language-independent or multilingual machine learning model capable of detecting fake news across multiple non-English languages using diverse datasets.

• User Interface Development**:**
To design a user-friendly web or mobile interface that integrates with the trained ML model, allowing users to input news articles or text and instantly receive predictions on their authenticity.

Literature Survey

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
1	MCWDST: A Minimum-Cost Weighted Directed Spanning Tree Algorithm for Real-Time Fake News Mitigation in Social Media.(2023)	- Fake News Detection-Convolutional Neural Network (CNN)-3BiLSTM Fake News. - Mitigation-Minimum Cost Weighted Directed Spanning Tree-(MCWDST).	- Developed two deep learning models:CNN-3BiLSTM Trained on 4 real-world datasets. - Detection Accuracy:98.37% on FNC dataset 96.38% on Kaggle dataset	- Transformer and graph-based embeddings improve fake news detection by enabling contextual understanding and real-time network-aware analysis.

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
2	Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)(2020).	- CNN-LSTM hybrid model combines (CNN) and (LSTM) for stance detection.	- Proposed a CNN-LSTM hybrid model for fake newsstance. Tested the model on the Fake News Challenge (FNC-1) dataset with 75,385 samples. - Achieved 97.8% accuracy using CNN-LSTM . - - Outperformed models like BERT (91.3%), XL Net (92.1%), RoBERT a (93.7%).	Expanding the dataset to include more diverse and multilingual news articles could further improve model generalization.
3	Big Data ML-Based Fake News Detection Using Distributed Learning(2023).	- CNN-LSTM - Big Data ML Paper: Random Forest (RF) - Logistic Regression (LR) - Decision Tree (DT) - Support Vector Machine (SVM)	- Developed CNN-LSTM and ML models for fake newsstance detection using FNC-1 dataset. - Used N-grams, Hashing TF, TF-IDF, and Count Vectorizer for feature extraction. - Ensemble (TF-IDF + LR): 93.78% accuracy, 92.45% F1-score.	Future research can focus on deep, unsupervised, and distributed learning approaches to enhance the accuracy and efficiency of fake news detection.

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
4	ArabFake: A Multitask Deep Learning Framework(2024)	- (MTL) framework for 3 tasks: fake news detection, categorization, and risk prediction. - Pretrained model: MARBERTv2 (based on Arabic tweets)	- Developed ArabFake Dataset with 2,495 manually labeled Arabic news articles. - Achieved F1 scores: 94.12% (detection), 84.92% (categorization), 88.91% (risk assessment).	Future work includes expanding Arabic dialects, adding multimodal analysis, and improving model explainability.
5	Fake News Detection Using Deep Learning (2024)	- RNNs: Sequential modeling. - LSTM, GRU, BiLSTM: Contextual understanding. - Hybrid Models: CNN+LSTM, CNN+BiLSTM.	- Datasets: LIAR, ISOT, PHEME, FakeNewsNet, BuzzFeed, PolitiFact. - Top-performing models: BiLSTM and BERT-based architectures achieved up to 99% accuracy.	Multilingual Detection: Extend beyond English language data. Explainability: Enhance model interpretability.- Real-time Detection: Enable live platform integration.

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
6	Advancing Fake News Detection: Hybrid Deep Learning With Fast Text and Explain(2024)	- ML models: SVM, DT, LR, RF, XGBoost. - DL models: LSTM, BiLSTM, GRU, CNN-LSTM. - Transformer models: BERT, RoBERTa, XLNet.	- Datasets used: WELFake, FakeNewsNet, FakeNewsPrediction. - CNN-LSTM model with FastText embeddings achieved best performance (F1-score up to 0.99).	Explore multilingual detection using models like mBERT, mT5, GPT. Expand to multi-label classification. Apply adversarial training to improve model robustness.
7	Deep vs. Shallow: A Comparative Study of Machine Learning and Deep Learning Approaches for Fake Health News Detection(2023)	- Machine Learning AdaBoost Random Forest,CBM & FBM categories. - Deep Learning Models:CNN-LSTM, Hybrid CNN-BiLSTM.	- Focuses on identifying fake news in the healthcare domain. - Deep Learning Accuracy:Hybrid CNN-LSTM F1 Score: 98.7%, Accuracy: 98.2%,Hybrid CNN-BiLSTM → F1 Score: 94%. - Machine Learning Accuracy:AdaBoost + Random Forest → F1 Score: 98.9%.	The inclusion of additional linguistic, semantic, and network-related features, along with multilingual datasets, could further enhance fake news detection across various domains.

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
8	Constructing a User-Centered Fake News Detection Model Using Classification Algorithms in Machine Learning Techniques(2023).	- Support Vector Machine (SVM) - (RF), Classification and Regression Tree (CART) - Neural Network (NNET).	- The proposed model leverages social capital to improve detection performance. - The (RF) algorithm achieved the highest accuracy: RF:94.1%,Logistic Regression (LR): 93.1%,CART: 98.8%,NNET: (accuracy not specified).	Integrating multilingual, cross-platform, and cultural features can improve fake news detection accuracy.
9	Utilization Strategy of User Engagements in Korean Fake News Detection(2022).	- K-FANG model (graph-based learning using Graph Neural Networks).	- Propose a fake news detection model using user engagement graphs and article content. - F1-Score: 88.5%K-FANG outperformed baseline content-based models by effectively incorporating user engagement.	Future research will use heterogeneous graphs with more entities, larger datasets, and improved evaluation methods for fake news detection.

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
10	A Review of Methodologies for Fake News Analysis(2023).	- Deeplearning : RNN LSTM- CNN ,MCNN-TFW ,Bi LSTM,TI-CNN, - Machine learning : SVM, random forest,Bayesian modelling.	- RNN LSTM- 90.10% accuracy, CNN ,MCNN-TFW :88.82 % accuracy ,Bi LSTM,TI-CNN,	Use Bayesian modeling for smarter, adaptable fake news detection. Enable fast fake news detection across multiple platforms.
11	OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection(2021)	- Deep Learning models: RNN, LSTM - Machine Learning models: Decision Tree,Logistic Regression,Random Forest , Naive Bayes	- A CNN-based model optimized for high accuracy in fake news detection - Dataset1 (Kaggle): Accuracy = 97.84%, F1 = 97.84% - Dataset2 (FakeNewsNet): Accuracy = 95.26%, F1 = 95.27% .	Apply OPCNN-FAKE to COVID-19 fake news detection. Incorporate multimodal data . Expand datasets to multi-language support.

SL NO	Name of the Research paper and Year	Methods/ Algorithms Used	Work carried Out and Results	Future Work
12	A Novel Temporal Footprints-Based Framework for Fake News Detection.(2024)	-Machine Learning 1.Random Forest 2.Decision Tree 3. Support Vector 4.Machine Naive - Bayes Deep Learning (DL) - (Long Short-Term Memory) - 98% accurancy ,FI score =0.97%.	- A novel framework uses both temporal and textual features, showing temporal data alone can help detect fake news. - Bi-LSTM model achieves near-perfect accuracy, emphasizing context- and time-aware detection. - (Long Short-Term Memory) - 98% accurancy ,FI score =0.97%.	Social network and user behavior as additional features Develop real time fake news detection system for deployment on news platforms. Improve temporal feature extraction to capture time references .

SYSTEM ARCHITECTURE

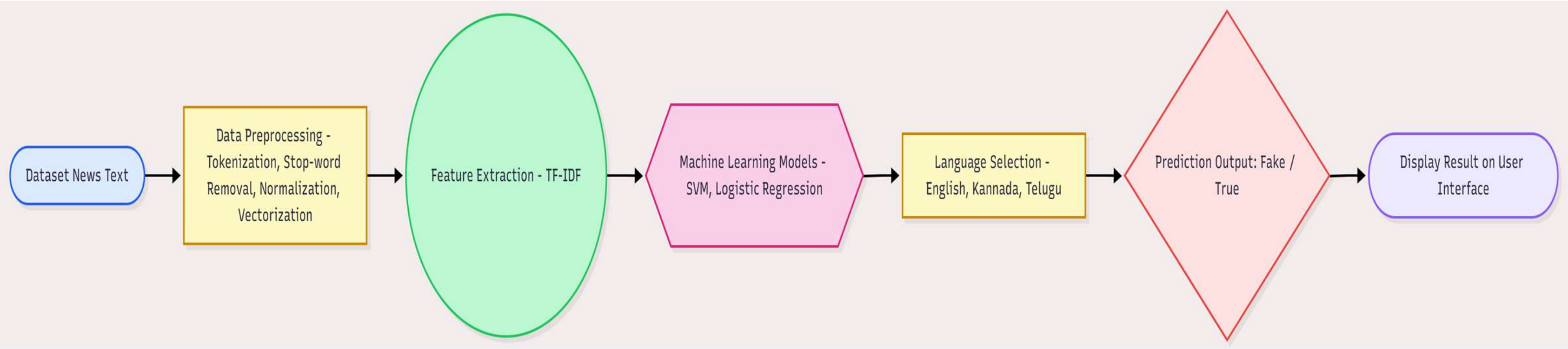


Fig 1 : SYSTEM ARCHITECTURE

•**Dataset (News Text):**
The system begins with collecting raw news articles or text data from various sources.

•**Data Preprocessing:**
Involves cleaning and preparing the text through steps like tokenization, stop-word removal, normalization, and vectorizafake or truetion to make it suitable for analysis.

•**Feature Extraction (TF-IDF):**
Converts textual data into numerical form using TF-IDF (Term Frequency–Inverse Document Frequency), which helps identify the importance of words in the dataset.

•**Machine Learning Models (SVM, Logistic Regression):**
These algorithms are trained to classify news as fake or true based on extracted features.

•**Language Selection (English, Kannada, Telugu):**
Supports multiple languages for a wider application of fake news detection.

•**Prediction Output (Fake / True):**
The model outputs whether the given news text is.

•**Display Result on User Interface:**
The final result is displayed to the user through an interface for easy understanding and interaction.

PROPOSED METHODOLOGY

1. Algorithms Used in Fake News Detection System

Purpose:
Classifies news articles as Fake or True by finding the best hyperplane that separates the two classes in high-dimensional space.

Working Steps (Algorithmic Flow):

1. Input: Preprocessed feature vectors from the text data.
2. Map input features into a high-dimensional space (if necessary using a kernel function).
3. Identify the optimal hyperplane that maximizes the margin between the Fake and True classes.

4.For a new article, compute which side of the hyperplane it lies on.
5.Assign class label:
    Fake if it falls on one side of the hyperplane
    True if it falls on the other side
Output the predicted class.

Advantages:
Effective for high-dimensional text data.
Can handle non-linear classification using kernels.
Robust against overfitting if the margin is maximized.

2. Logistic Regression Algorithm

Purpose:
Predicts the probability that a news article belongs to Fake or True class.

Working Steps (Algorithmic Flow):
Input: Preprocessed and vectorized feature vectors.
Compute weighted sum of input features.
Pass the sum through the sigmoid function:

$$P(Y = 1) = \frac{1}{1 + e^{-(wX+b)}}$$

Here, $Y = 1$represents True, $Y = 0$represents Fake.
Set a threshold (commonly 0.5) to assign class:
  True if $P \geq 0.5$
  Fake if $P < 0.5$
Output the predicted class.

Advantages:
Simple, fast, and interpretable.
Works well for binary classification tasks.
Outputs probability scores which can be useful for confidence measurement.

3.Algorithm Comparison

Feature	SVM	Logistic Regression
Type	Margin-based classifier	Probabilistic classifier
Handles high-dimensional data	Yes	Moderate
Output	Class label only	Probability + class label
Complexity	Higher	Lower
Use Case	High-dimensional text	Simple, interpretable models

Observation:

- Both algorithms are effective for fake news detection.
- SVM provides slightly better accuracy on high-dimensional text data.
- Logistic Regression offers interpretability and probabilistic predictions, making it useful for understanding model confidence.

IMPLEMENTATION

1. Dataset Description:

- **Source:** Kaggle – *ISOT Fake News Detection Dataset*.
- **Files:** Separate CSV files for True and False news articles.

- **Languages Supported:**

    - English (original statements)
    - Kannada (statements_kn column) – translated from English
    - Telugu (statements_te column) – translated from English

- **Label Column:**

- Indicates the veracity of the news:
    - `1` → True news
    - `0` → Fake news

- **Data Expansion:**

    - English statements translated to Kannada and Telugu using Google Translate.
    - Dataset now supports multilingual fake news detection.

Preprocessing Applied:

1.Lowercasing :
Converting all text to lowercase to maintain consistency and avoid redundant representations.

2. Removal of unwanted characters :
Eliminating punctuation marks, numeric values, URLs, and other non-alphabetic symbols to reduce noise.

3. Stopword removal :
Removing common words that carry minimal semantic value for classification in English, Kannada, and Telugu.

4. Tokenization :
Segmenting each sentence into individual words (tokens) to facilitate feature extraction and vectorization.

2. Algorithm Formulas:

a) Support Vector Machine (SVM)
Decision function:

$$f(x) = w \cdot x + b$$

Classification Rule:

$$\text{Class} = \begin{cases} 1 & \mathit{if}\ f(x) \geq 0\ (\mathit{True}) \\ 0 & \mathit{if}\ f(x) < 0\ (\mathit{Fake}) \end{cases}$$

Objective: Maximize margin $\frac{2}{||w||}$subject to:

$$y_i(w \cdot x_i + b) \geq 1$$

b) Logistic Regression (LR)
Sigmoid function:

$$P(Y = 1 \mid X) = \frac{1}{1 + e^{-(wX+b)}}$$

Classification Rule:

$$\text{Class} = \begin{cases} 1 & P \geq 0.5 \ (\mathit{True}) \\ 0 & P < 0.5 \ (\mathit{Fake}) \end{cases}$$

Loss function (Binary Cross-Entropy):

L=−N1i=1∑N[yilog(Pi)+(1−yi)log(1−Pi)]

3. Model Training & Evaluation

- **Training:** Models trained separately for English, Kannada, and Telugu datasets.

- **Algorithms Used:**

- **SVM** – Linear kernel
- **Logistic Regression** – solver: lbfgs, max_iter=100

Accuracy Results:

Language	SVM Accuracy	Logistic Regression Accuracy
English	99.24%	98.52%
Kannada	97.51%	96.63%
Telugu	96.79%	96.22%

Observation:

- SVM consistently performs slightly better than Logistic Regression.
- Multilingual support ensures high accuracy across all languages.

RESULTS

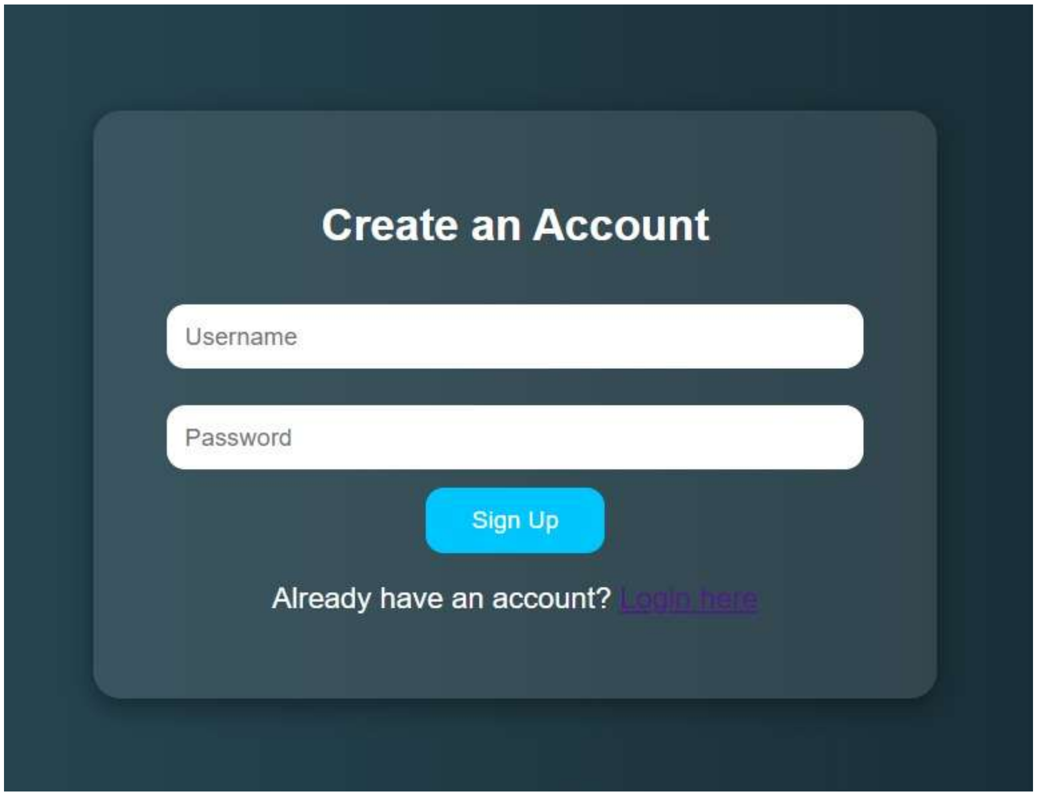


Fig1:Sign Up

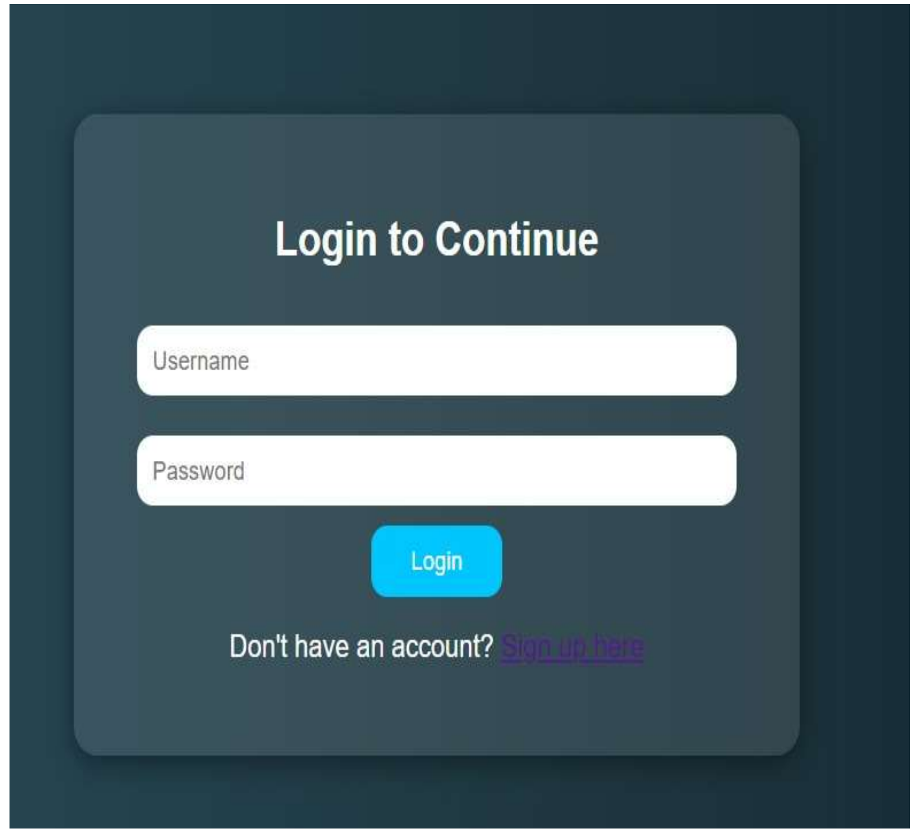


Fig 2:Login Page

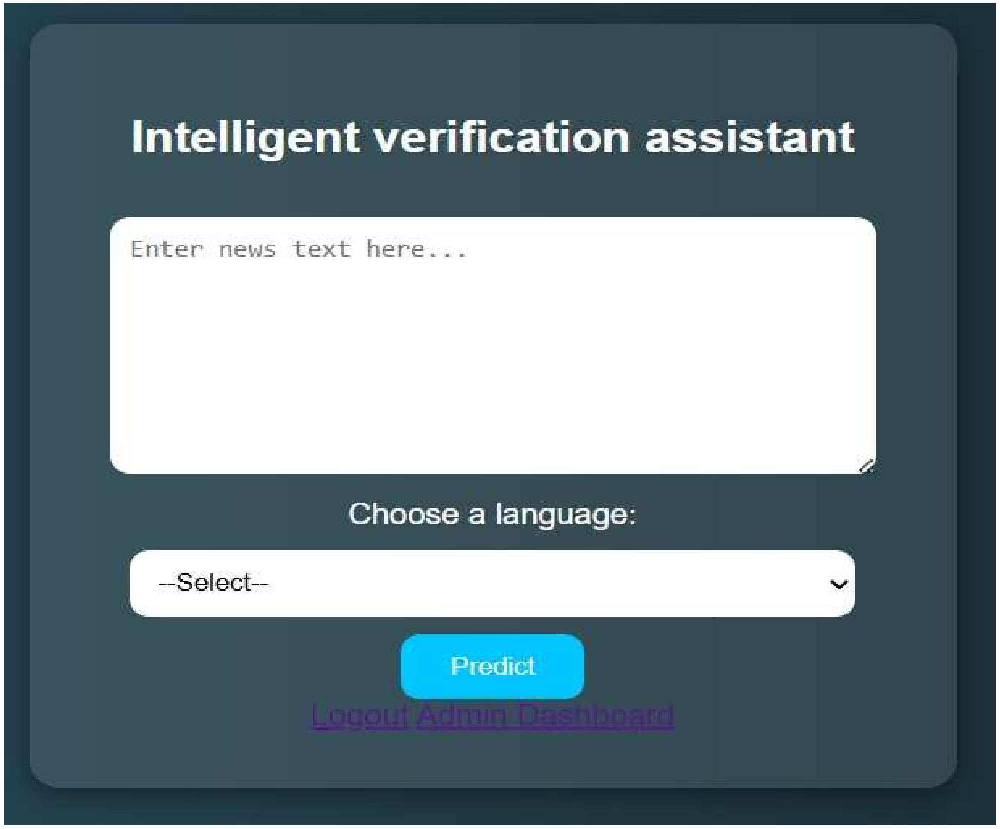


FIG 4: Predicting Page

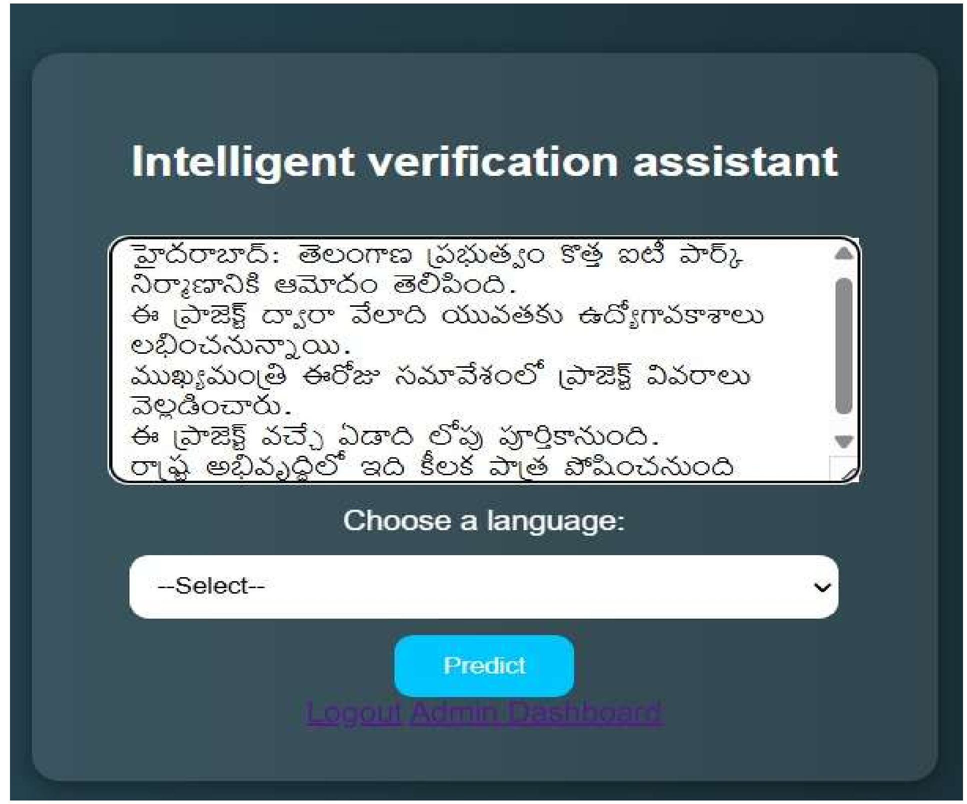


Fig 5:texting Page

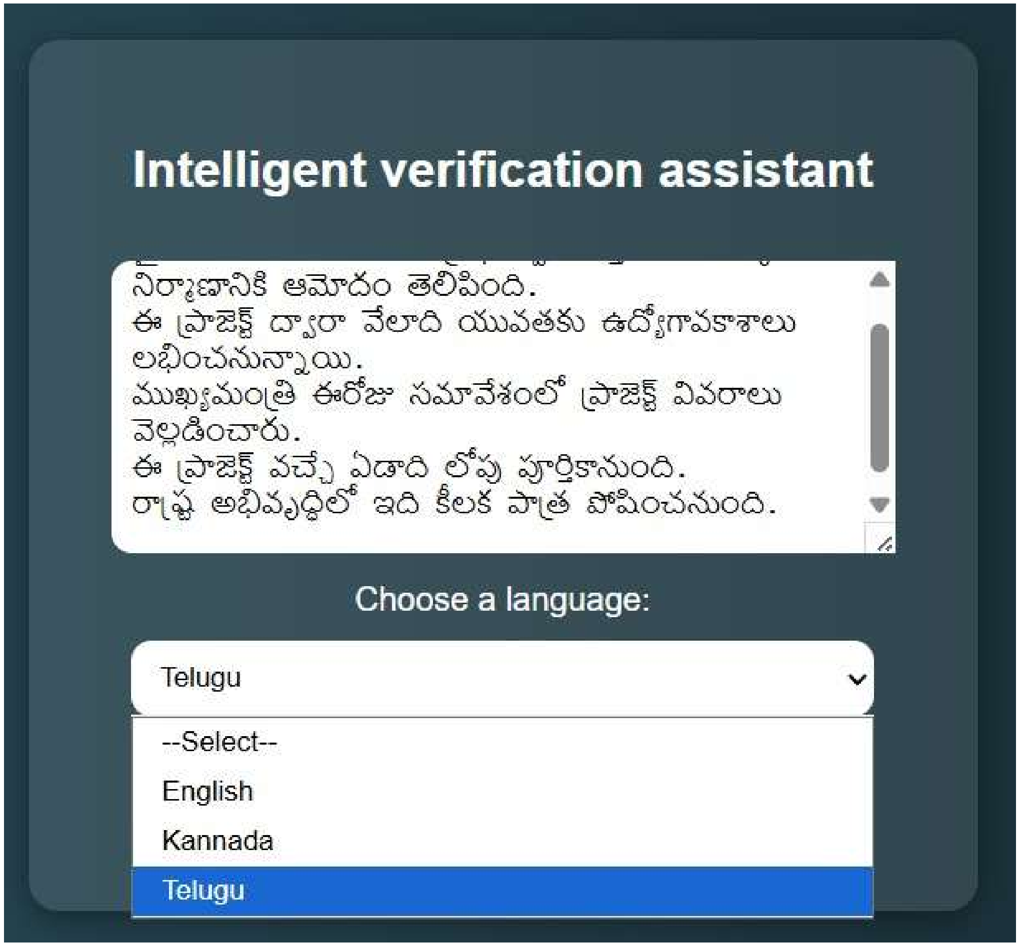


FIG 5: Language Selection page

Fig 6: Prediction Result Page

Fig 7:Accuracy Result

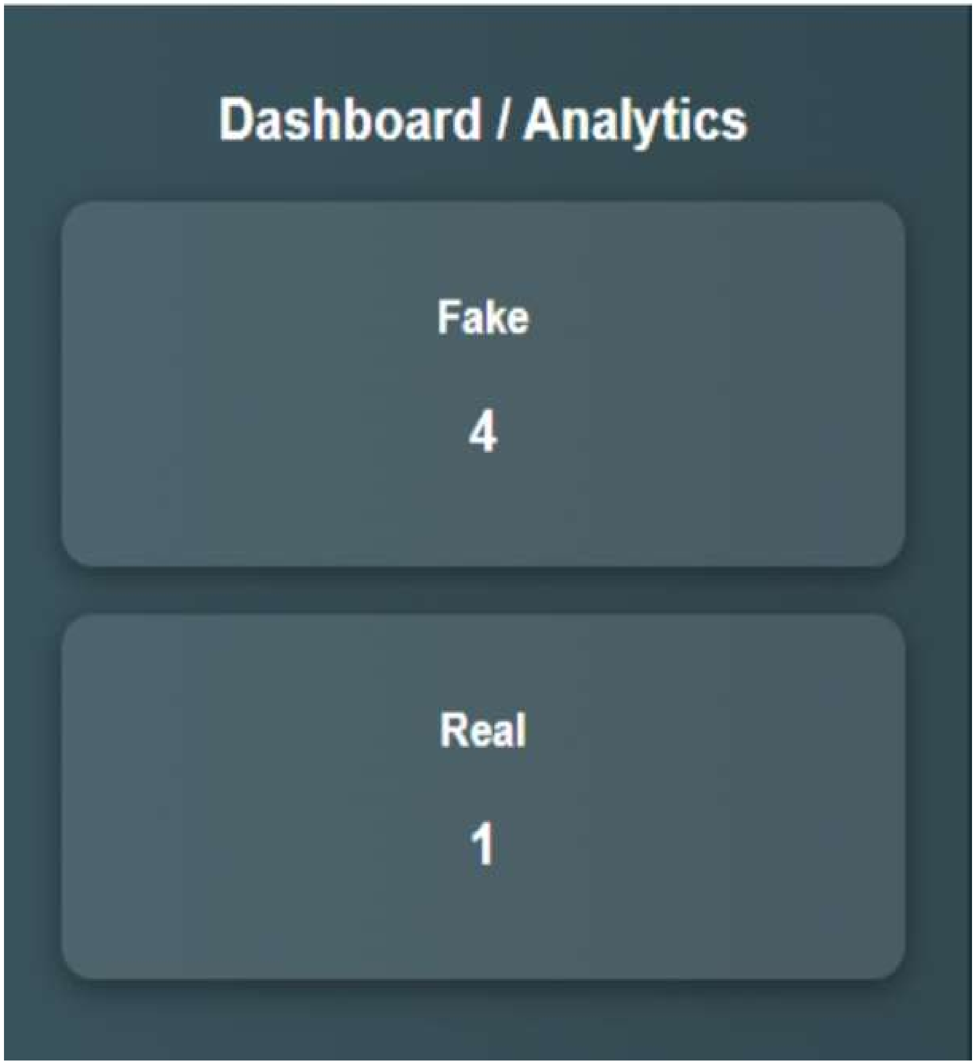


Fig 8: Dashboard

Fig 9: Language-wise Prediction count

CONCLUSION AND FUTURE SCOPE

1.Conclusion :

- Developed a robust fake news detection system using SVM and Logistic Regression.
- Supports multilingual news detection: English, Kannada, and Telugu.
- Achieved high accuracy across all languages (SVM performing best).
- Flask interface allows easy real-time predictions.
- Modular architecture ensures scalability, maintainability, and flexibility for future enhancements.

2.Future Scope :

- **Advanced Models:** Incorporate Deep Learning (LSTM, BERT) for better contextual analysis.
- **Automatic News Updates:** Real-time scraping from news portals and social media.
- **Explainability:** Provide reasoning behind predictions for better user trust.
- **Mobile-Friendly Platform:** Extend to mobile apps for wider access.
- **Language Expansion:** Support additional regional and international languages.

REFERENCES

1. C. -O. Truică, E. -S. Apostol, R. -C. Nicolescu and P. Karras, "MCWDST: A Minimum-Cost Weighted Directed Spanning Tree Algorithm for Real-Time Fake News Mitigation in Social Media," in IEEE Access, vol. 11, pp. 125861-125873, 2023, doi:10.1109/ACCESS.2023.3331220

2. M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi and B. -W. On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)," in IEEE Access, vol. 8, pp. 156695- 156706, 2020, doi: 10.1109/ACCESS.2020.3019735.

3. A. Altheneyan and A. Alhadlaq, "Big Data ML-Based Fake News Detection Using Distributed Learning," in IEEE Access, vol. 11, pp. 29447-29463, 2023, doi: 10.1109/ACCESS.2023.3260763.

4. A. M. K. Shehata, M. Nasser Al-Suqri, N. Eldin Mohamed Elshaiekh Osman, F. Hamad, Y. Nasser Alhusaini and A. Mahfouz, "ArabFake: A Multitask Deep Learning Framework for Arabic Fake News Detection, Categorization, and Risk Prediction," in IEEE Access, vol. 12, pp. 191345-191360, 2024, doi:10.1109/ACCESS.2024.3518204

5. M.Q. Alnabhan and P. Branco, "Fake News Detection Using Deep Learning: A Systematic Literature Review,“ in IEEE Access, vol. 12, pp. 114435-114459, 2024, doi: 10.1109/ACCESS.2024.3435497.
6. E Hashmi, S. Y. Yayilgan, M. M. Yamin, S. Ali and M. Abomhara, "Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI," in IEEE Access, vol. 12, pp. 44462-44480,2024,doi: 10.1109/ACCESS.2024.3381038
7. T. Mahara, V. L. H. Josephine, R. Srinivasan, P. Prakash, A. D. Algarni and O. P. Verma, "Deep vs. Shallow: A Comparative Study of Machine Learning and Deep Learning Approaches for Fake Health News Detection," in IEEE Access, vol. 11, pp. 79330-79340, 2023, doi: 10.1109/ACCESS.2023.3298441.
8. M. Park and S. Chai, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques," in IEEE Access, vol. 11, pp. 71517-71527, 2023, doi: 10.1109/ACCESS.2023.3294613.
9. M. Kang, J. Seo, C. Park and H. Lim, "Utilization Strategy of User Engagements in Korean Fake News Detection," in IEEE Access, vol. 10, pp. 79516-79525, 2022, doi: 10.1109/ACCESS.2022.3194269

10. M. Tajrian, A. Rahman, M. A. Kabir and M. R. Islam, "A Review of Methodologies for Fake News Analysis," in IEEE Access, vol. 11, pp. 73879-73893, 2023, doi: 10.1109/ACCESS.2023.3294989.

11. H. Saleh, A. Alharbi and S. H. Alsamhi, "OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection," in IEEE Access, vol. 9, pp. 129471-129489, 2021, doi: 10.1109/ACCESS.2021.3112806.

12. A. Raza, S. Ur Rehman Khan, R. Sher Afgun Usmani, A. Kumar Das and S. Ashraf Chaudhry, "A Novel Temporal Footprints-Based Framework for Fake News Detection," in IEEE Access, vol. 12, pp. 172419-172428, 2024, doi:10.1109/ACCESS.2024.3490558.