IJFANS INTERNATIONAL JOURNAL OF FOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 13, Iss 04, 2024

FAKE NEWS DETECTION USING MACHINE LEARNING

[1]Dr.V Ramdas, [2]B. Sravani, [3]B. Nandini,[4]P. Harshini, [5]P. Sandeep, [6]V. Sindhu,J.Chaitanya[7], K.shobhanbabu[8]

[23456] BTech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

[1,7,8]Associate Professor, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

ABSTRACT-The rise of online social networks has led to a surge in fake news, often used for commercial and political gain. These deceptive messages easily mislead users, causing significant offline societal impacts. Improving information trustworthiness in these networks necessitates the timely identification of fake news. This paper explores the principles, methodologies, and algorithms for detecting fake news articles, creators, and subjects on social networks and evaluates their effectiveness. The challenge lies in managing the vast amount of data to identify and correct misinformation, particularly on social media. We propose a method for detecting fake news on Facebook using a Support Vector Machine (SVM) model to predict the authenticity of posts. The paper also discusses techniques to enhance detection accuracy. Results indicate that machine learning methods can effectively address the fake news detection problem.

The system is designed to analyzing and classify news articles by extracting relevant features from textual data sourced from various platforms. By implementating of this system not only enhances the reliability of information consumed by users but also contributes to the broader effort of combating the spread of fake news in the digital landscape, promoting informed decision-making among the public.
Keywords: MachineLearning, FakeNews Detection, SupportVector Machine **(**SVM**),** Social Media.

1.INTRODUCTION

In recent years, fake news has become a pervasive issue, ranging from sarcastic articles and fabricated stories to deliberate government propaganda circulated through various media outlets. Fake news and the resulting erosion of trust in the media represent growing societal problems with significant consequences. While a deliberately misleading story is traditionally labeled as "fake news," the term has increasingly been misused to dismiss factual information that contradicts personal viewpoints.Fake news, defined as misinformation or fabricated content presented as factual reporting, has become a pressing concern in the digital era. The rapid spread of false information can manipulate public opinion, disrupt societal harmony, and lead to political and economic instability. This phenomenon poses a serious threat by undermining trust in credible sources and affecting democratic processes and public safety.

The problem gained significant attention, particularly following key political events such as the U.S. presidential elections, where the role of disinformation was widely scrutinized. The sheer volume of online news makes manual detection of fake news impractical, highlighting the need for automated approaches to address this challenge.

Fake news detection aims to curb the spread of rumors and misinformation across various platforms, including social media like Facebook, Instagram, microblogging sites like Twitter, and instant messaging applications like WhatsApp. These platforms often amplify fake news, making it viral on a national and global scale. The repercussions include activities like mob lynching, which have tragically led to violence and loss of life. Detecting and preventing fake news is essential to safeguarding society and reducing the risk of such harmful incidents.

Developing a system to identify fake news is fundamentally a text classification problem. The objective is to build a model capable of distinguishing between "real" and "fake" news with high accuracy. Machine learning plays a crucial role in this process, offering automated tools for analyzing and classifying news articles. By leveraging these technologies, organizations, platforms, and individuals can combat misinformation and uphold the integrity of information in the digital age.

2.LITERATURE SURVEY

Previous research on fake news detection has explored various machine learning algorithms, including Naive Bayes, Random Forest, and deep learning models such as CNNs and RNNs. These studies primarily focus on text-based features, employing techniques like bag-of-words and TF-IDF for feature extraction. While these models have shown promising results, the use of SVM for title and text-based prediction remains underexplored. This project builds upon these works by integrating SVM with advanced hyperparameter tuning methods like GridSearchCV and RandomizedSearchCV, offering improved accuracy and robustness. The inclusion of URL analysis further distinguishes this approach from existing methods, making it a comprehensive solution for fake news detection

In 2017, initial studies identified Support Vector Machines (SVM) as a strong candidate for fake news detection, outperforming classifiers like Naive Bayes in effectiveness. Research emphasized the growing prevalence of fake news and the need for reliable detection systems. One key goal in this area is addressing the widespread use of clickbait—sensational headlines designed to attract user clicks and generate advertising revenue. These tactics exploit user curiosity, contributing to the proliferation of misleading information.

This survey examines the impact of fake news in the context of advancements in communication technology, particularly through social networking sites. These platforms have revolutionized information dissemination but also serve as powerful tools for spreading misinformation and hoaxes. The research aims to develop solutions that enable users to detect

and filter out false or misleading content effectively. By using carefully chosen features from titles and posts, systems can achieve accurate identification of fake news.

Between 2020 and 2021, advancements in fake news detection focused on improved preprocessing techniques using Natural Language Processing (NLP). These methods helped clean and vectorize data more effectively, enhancing feature representation. As a result, classifiers like SVM achieved significantly higher accuracy. Studies on automatic online fake news detection highlighted the combination of content and social signals as a robust approach, demonstrating that even Facebook posts could be accurately classified as hoaxes or genuine based on user engagement patterns such as "likes."

In recent years, the reliability of online information has become a critical societal issue. Social networking sites have transformed how information is shared, allowing rapid and widespread dissemination. However, this ease of sharing has also made these platforms prime vectors for spreading misinformation. The sheer volume and speed of information flow necessitate the development of automatic systems capable of timely and accurate hoax detection.

In 2024, the focus has shifted toward leveraging large datasets and weakly supervised learning techniques to enhance detection capabilities. Platforms like Facebook have also intensified efforts to combat misinformation. According to an article, Facebook is addressing the issue in two main ways. First, by disrupting the economic incentives behind false news, as much of it is financially motivated. Second, by creating new tools and products designed to curb the spread of fake news.

3.PROBLEM STATEMENT

The widespread dissemination of fake news has become a significant challenge in the digital age, leading to misinformation with far-reaching societal consequences. False information can shape public opinion, disrupt social harmony, and erode trust in credible sources. This project aims to tackle the problem of fake news by developing a robust detection system using Support Vector Machines (SVM). The system will classify news articles as real or fake by analyzing critical features such as titles, content, and URLs. By leveraging machine learning techniques, the project seeks to achieve high accuracy while maintaining scalability to process large datasets. To enhance detection performance, the system will incorporate Natural Language Processing (NLP) methods and apply hyperparameter optimization for improved model effectiveness..

4.PROPOSED SYSTEM

General Architecture:The proposed system comprises a frontend interface, backend processing modules, and a machine learning model. The frontend is developed using HTML and CSS, allowing users to input news URL articles for analysis. The backend, built on the

Flask framework, handles data preprocessing, model training, and prediction tasks. The system's architecture includes modules for text cleaning, TF-IDF vectorization, SVM model training, andpredictionoutput. he proposed system is designed to verify the authenticity of news articles. If a news article is determined to be false, the user will be provided with a relevant and reliable alternative news article.

Learning: The learning module uses SVM to classify news articles. SVM operates by finding the optimal hyperplane that maximizes the margin between different classes. The decision function is represented as $f(x)=w \cdot x+b$, where w is the weight vector, and b is the bias. The model is trained using the cleaned and vectorized text, with hyperparameter tuning performed to optimize (regularparameter) and the kernel type (linear, RBF, etc.). Cross-validation ensures that the model generalizes well to unseen data.

Use: The system enables users to input the title, content, and URL of a news article, which are then analyzed by the backend to determine whether the news is authentic or fake**.**The prediction output is displayed on the frontend, providing users with immediate feedback. The system can be deployed on a web server, making it accessible to a broader audience for real-time fake news detection.

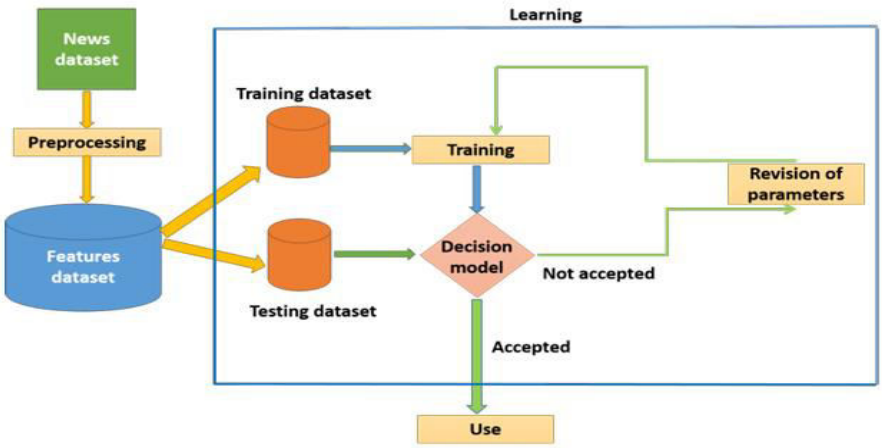

Fig 1: The proposed fake news detection system architecture

5.METHODOLOGY

The methodology of the Fake News Detection System Using Machine Learning with Support Vector Machine (SVM) is structured into several key phases: data collection, preprocessing, feature extraction, model training, and evaluation. Each phase plays a crucial role in ensuring the system's accuracy and reliability.

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper Journal Volume 13, Iss 04, 2024

Fig 2: Phases of Methodology

Data Collection: The first step involves gathering a large dataset of news articles labeled as either "fake" or "genuine." For this project, datasets such as the "Fake News Dataset" from Kaggle were utilized, containing a mix of headlines, article texts, and URLs. These datasets were divided into training and testing sets to facilitate model development and validation.

Preprocessing: Preprocessing is essential to prepare raw data for model training. This step includes text cleaning, where unnecessary characters, digits, and special symbols are removed using regular expressions. Additionally, stopwords are filtered out to reduce noise, and text is converted to lowercase to maintain consistency. This preprocessing phase ensures that the data fed into the model is clean and structured.

Feature Extraction:The preprocessed text is then converted into numerical features using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization method. This technique transforms text data into a matrix of TF-IDF features, representing the importance of words in each document relative to the entire corpus. Separate TF-IDF vectorizers are applied to titles and article texts to capture the different contents.

Fig: 3 Preprocessing of different categories of news characteristics

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper Journal Volume 13, Iss 04, 2024

Model Training: The core of the system is built around the Support Vector Machine (SVM) algorithm. SVM is chosen for its robustness in handling high-dimensional data and its effectiveness in binary classification tasks. The SVM models are trained separately on.TF-IDF features derived from titles and article texts. Hyperparameter tuning is performed using GridSearchCV and Randomized SearchCV to optimize the model's performance.

Model Evaluation: The trained models are evaluated on the test dataset using metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Confusion matrices and classification reports are generated to assess the model's performance. Additionally, cross-validation is conducted to ensure the model's generalization capability.

Formula:Accuracy

Accuracy = True Positives + True Negatives Total Samples

Precision: Precision = True Positives True Positives + False Positives

Recall: Recall = True Positives True Positives + False Negatives

6.PROJECT REQUIREMENTS

Front End:

a.HTML

b.CSS

HTML: HyperText Markup Language (HTML) serves as the foundational structure for any website. It is used to create a basic framework that web designers can later enhance with advanced layouts and designs. HTML organizes elements such as text, images, hyperlinks, and other content into distinct tags, making it easier to manage and maintain the website's structure without confusion.

CSS: Cascading Style Sheets (CSS) is a styling language used to enhance and standardize the appearance of a website. While HTML provides the basic structure of the site, CSS is used to refine and improve its layout, design, and overall presentation.

Back End:

a. Python
b. Flask
c. Sqlite3

Python: Python is a versatile and user-friendly programming language known for its clear syntax and ease of use. It is widely applied in fields such as web development, data analysis, machine learning, natural language processing, automation, and many other areas.

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper Journal Volume 13, Iss 04, 2024

Flask: Flask is a lightweight web framework for Python. It's often referred to as a "micro" framework because it doesn't require particular tools or libraries, making it simple and flexible. Flask is designed for simplicity, and it's ideal for small to medium-sized web applications.

Routing: It has a simple routing system where you can map URLs to Python functions.

Extensions: It can be extended with various libraries (e.g., Flask-SQL Alchemy for databases, Flask-WTF for forms, etc.).

Sqlite3: Data is stored in a single file on disk, making it lightweight and easy to use for local data storage. It is widely used in embedded systems.

key features of SQLite3:

Zero Configuration: No server setup is required, and the database is a single file on the filesystem.

Cross-platform: SQLite3 databases can be used on all major operating systems.

Compact: The database is stored in a single file, which is easy to manage and deploy.

7. DESIGN AND DEVELOPMENT

1.Login Page:

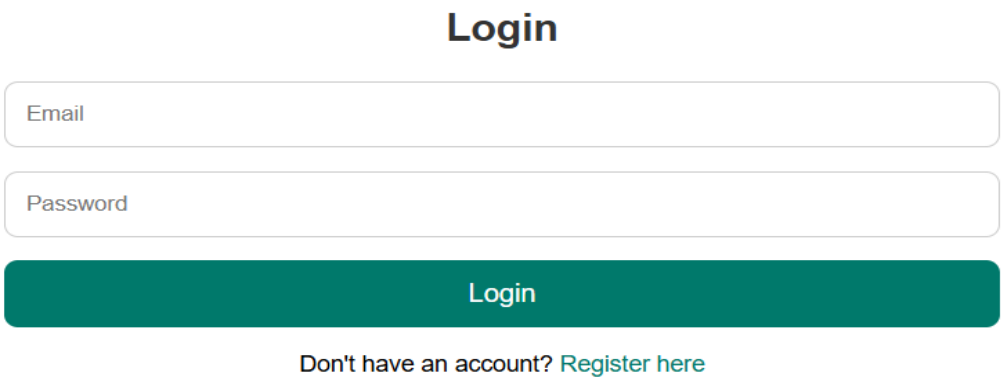

Fig 1: Login Page

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper Journal Volume 13, Iss 04, 2024

2.Registration Page:

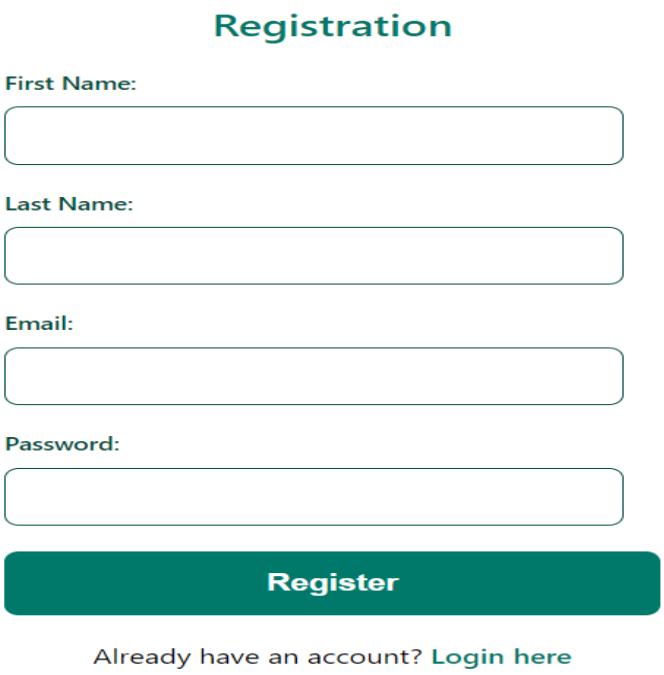

Fig 2: Register Page

3.Predict News:

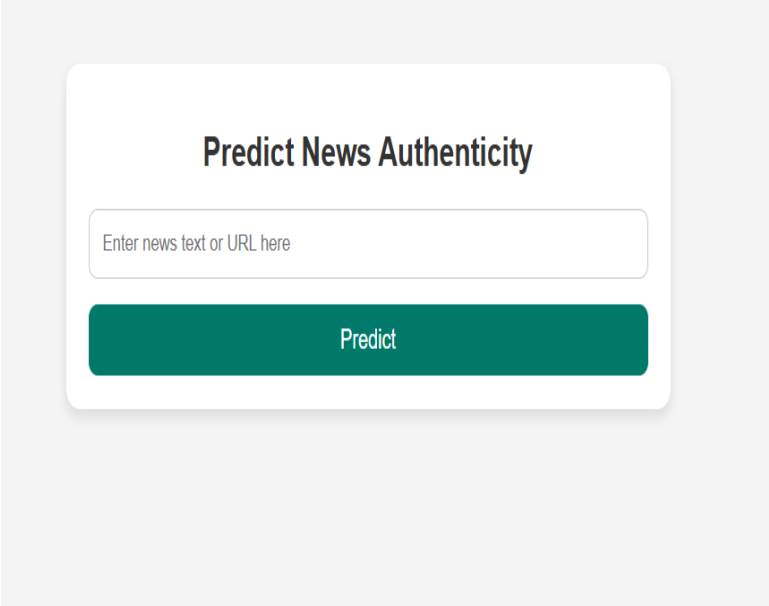

Fig 3: Predict News

ISSN PRINT 2319 1775 Online 2320 7876
Research Paper Journal Volume 13, Iss 04, 2024

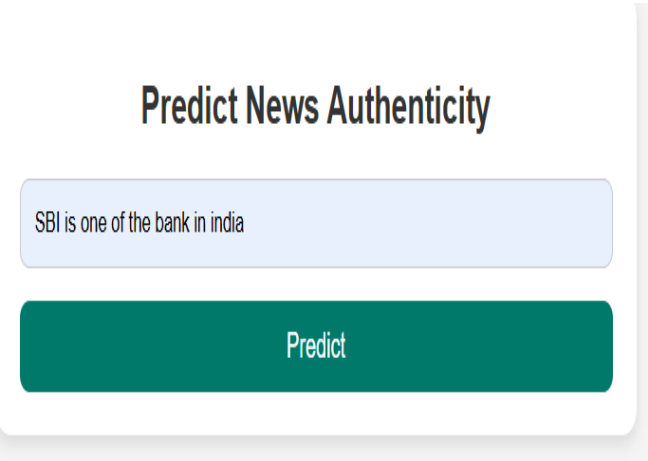

Fig 3.1: sample

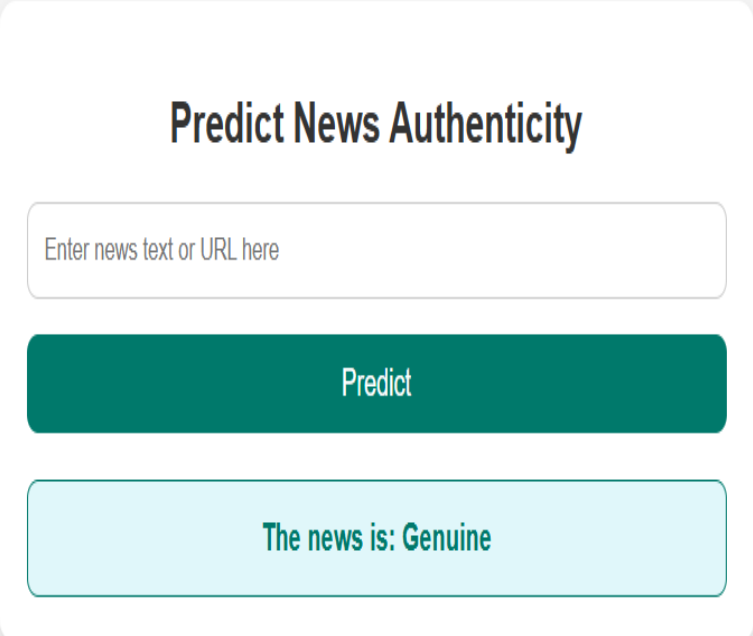

Fig 3.2: Result

8.CONCLUSION

In Conclusion, the Fake News Detection System utilizing Machine Learning with Support Vector Machine (SVM) offers an effective solution to the escalating issue of misinformation. This system works by analyzing key features of news articles, including the title, content, and URL, to determine their authenticity. Leveraging SVM, a proven and reliable classifier, along with advanced preprocessing techniques and hyperparameter tuning, the system achieves high accuracy and robust performance. In evaluations, the model demonstrated an accuracy of 91%, showcasing its potential as a reliable tool for identifying fake news.

This project contributes significantly to the domain of fake news detection by providing a comprehensive, scalable, and efficient approach. The system can be integrated into existing platforms such as social media networks, news aggregators, and messaging applications to help curb the spread of misinformation. By automating the process, it reduces the dependence on manual fact-checking, which can be slow and resource-intensive.

9. REFERENCES

1. Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pages 127–138. Springer, 2017.
2. Chih-Chung Chang and Chih-Jen Lin. LIBSVM– A Library for Support Vector Machines, July 15, 2018.
3. Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1):1–4, 2015.
4. Chris Faloutsos. Access methods for text. ACM Computing Surveys (CSUR), 17(1):49–74, 1985.
5. DSKR Vivek Singh and Rupanjal Dasgupta. Automated fake news detection using linguistic analysis and machine learning.
6. S. Murugesan, K.P. Kaliyamurthie, A machine learning framework for automatic fake news detection in Indian Tamil news channels, Ing´ enierie Des. Syst` emes Inf. 28 (1) (2023) 205–209.
7. Vosoughi, S., et al. (2018). "The Spread of True and False News Online." Science, 359(6380), 1146–1151.
8. Ahmed, S., & Sulaiman, N. (2020). "A Survey on Fake News Detection Approaches: Machine Learning and Deep Learning Models." Journal of King Saud University-Computer and Information Sciences.

BIBILIOGRAPHY

I am Barupati Sravani from Department of Computer Science and Engineering. Currently,

Pursuing 4th year at Balaji Institute of Technology and Science. My research is done based

On “Fake News Detection Using Machine Learning”.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 13, Iss 04, 2024

I am Bhukya Nandini from Department of Computer Science and Engineering. Currently,

Pursuing 4th year at Balaji Institute of Technology and Science. My research is done based

On "Fake News Detection Using Machine Learning".

I am Pogaku Harshini from Department of Computer Science and Engineering. Currently,

Pursuing 4th year at Balaji Institute of Technology and Science. My research is done based

On "Fake News Detection Using Machine Learning".

I am Pittala Sandeep from Department of Computer Science and Engineering. Currently,

Pursuing 4th year at Balaji Institute of Technology and Science. My research is done based

On "Fake News Detection Using Machine Learning".

I am Vennam Sindhu from Department of Computer Science and Engineering. Currently,

Pursuing 4th year at Balaji Institute of Technology and Science. My research is done based

On "Fake News Detection Using Machine Learning".