International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

FAKE NEWS DETECTION SYSTEM USING FEATURED-BASED OPTIMIZED MSVM CLASSIFICATION

#1 SK.HIMAMBASHA, #2 CH.HARIKRISHNA
#1 ASSISTANT PROFESSOR, #2 MCA SCHOLAR
DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS
QIS COLLEGE OF ENGINEERING & TECHNOLOGY, ONGOLE
VENGAMUKKAPALEM(V), ONGOLE, PRAKASAM DIST.,ANDHRA PRADESH

Abstract

The rapid spread of misinformation and deceptive content across online platforms has become a pressing concern in today's digitally connected world. Fake news can influence public opinion, disrupt democratic processes, and cause widespread panic or confusion. In response to this challenge, this project proposes a robust and intelligent fake news detection system utilizing a feature-based optimized Multi-class Support Vector Machine (MSVM) classification approach. The system extracts a rich set of features from news articles, including lexical patterns, syntactic structures, semantic representations, and contextual cues, ensuring a comprehensive understanding of the input text. These features are then optimized using advanced selection techniques to enhance model performance and reduce computational complexity. The MSVM classifier is trained to identify not only binary outcomes (fake or real) but also multiple categories such as satire, partially true, and misleading content, offering more nuanced and accurate classifications. This hybrid approach improves detection accuracy, scalability, and adaptability across different data sources and languages. By integrating machine learning with optimized feature engineering, the system presents a powerful solution for combating the spread of fake news and can be deployed in real-time environments such as social media platforms, fact-checking tools, and news monitoring applications. The proposed method addresses limitations of existing models and significantly contributes to the advancement of intelligent misinformation detection systems.

INTRODUCTION

In the digital era, the internet and social media platforms have revolutionized the way people consume information. While this advancement has enabled fast and widespread dissemination of news, it has also opened the door to the rampant spread of fake news. Fake news refers to false or misleading information presented as news, which is often intended to damage reputations, influence public opinion, or generate financial gain. The consequences of such misinformation are profound, ranging from social unrest and political manipulation to public health crises, especially when fake news is widely shared before it can be fact-checked.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Traditional approaches to fake news detection have relied heavily on manual fact-checking or simple keyword-based filtering methods. However, these methods are inefficient and inadequate given the volume and complexity of online data. Manual verification is time-consuming and unable to scale with the rapid growth of misinformation. Furthermore, fake news has become increasingly sophisticated, often mimicking legitimate sources in language and style, which makes it difficult for rule-based systems or basic machine learning techniques to accurately detect and classify.

To address these challenges, this project proposes an intelligent and efficient fake news detection system that leverages an optimized Multi-class Support Vector Machine (MSVM) classifier with advanced feature engineering. The system focuses on extracting and selecting meaningful textual features such as lexical cues, semantic relationships, and syntactic patterns, which are then fed into the MSVM for classification into multiple categories. This approach not only improves accuracy but also enhances the system's ability to handle diverse and large-scale datasets. By adopting this method, the system aims to provide a reliable tool for real-time fake news detection across various domains and platforms.

LITERATURE SURVEY

Title:*Fake News Detection on Social Media: A Data Mining Perspective*
Authors: Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu
Description:
This paper provides a comprehensive review of fake news detection using data mining techniques. It discusses the characteristics of fake news, challenges in detection, and outlines various feature extraction methods such as user behavior, text content, and network-based features. The authors suggest that a combination of content and social context can improve the accuracy of detection.

Title:*LIAR: A Benchmark Dataset for Fake News Detection*
Author: William Yang Wang
Description:
The paper introduces the LIAR dataset consisting of 12.8K labeled short statements from various political contexts. It emphasizes the need for fine-grained classification and evaluates several machine learning models, including logistic regression and SVMs, showing the effectiveness of feature-rich models in detecting fake news.

Title:*A Survey of Fake News Detection: Fundamental Theories, Detection Methods, and Opportunities*
Authors: Xinyi Zhou, Reza Zafarani
Description:
This survey outlines the different types of fake news, existing detection models, and the theoretical foundations such as psychological theories and social theories. It classifies detection methods into content-based and context-based approaches and stresses the importance of machine learning and deep learning in modern detection systems.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Title:*CSI: A Hybrid Deep Model for Fake News Detection*
Authors: NataliRuchansky, SungyongSeo, Yan Liu
Description:
CSI (Capture, Score, Integrate) is a hybrid model that considers both content and user behavior. It combines recurrent neural networks and user activity modeling to identify fake news with higher accuracy. This model emphasizes temporal and user interaction features for better detection performance.

Title:*Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking*
Authors: Hannah Rashkin, Eunsol Choi, Jin Yea Jang, SvitlanaVolkova, Yejin Choi
Description:
This study investigates linguistic patterns and stylistic features in deceptive versus truthful statements. It highlights the role of sentiment, subjectivity, and exaggeration in identifying misinformation, helping to improve feature selection in fake news classification.

Title:*Automatic* ***Deception Detection: Methods for Finding Fake News***
Authors: Nicole J. Conroy, Victoria L. Rubin, Yimin Chen
Description:
The authors present various deception detection methods focusing on linguistic cues and stylistic writing patterns. The study evaluates traditional classifiers like SVM and Naïve Bayes using n-gram and syntactic features, proving their effectiveness in early-stage fake news detection.

System Analysis

Existing System

In the existing landscape of fake news detection, most systems rely on conventional binary classification techniques using basic machine learning algorithms such as Logistic Regression, Naive Bayes, Decision Trees, and traditional Support Vector Machines (SVM). These models primarily utilize content-based features like word frequency, TF-IDF scores, and sentiment polarity extracted from the text of news articles. While these methods are relatively simple and fast to implement, their performance is often limited due to shallow feature representations and an inability to handle complex linguistic patterns or context-aware semantics present in fake news articles.

Furthermore, many existing systems fail to incorporate optimization techniques for feature selection, leading to high-dimensional input data that may contain noise or irrelevant information. This often results in overfitting or underfitting, reducing the generalization capability of the models on unseen data. In addition, these systems typically apply standard SVMs in binary classification without considering multi-class categorization or the intricacies involved in news spread behavior, source credibility, or writing style, which are crucial in distinguishing fake news from genuine news more accurately.

Another major drawback of current systems is the lack of integration with real-time or

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

large-scale data sources like social media platforms. Most systems operate offline on static datasets, and their models are not adaptive to evolving misinformation tactics. Additionally, few systems provide robust interpretability or explanations for their predictions, making them less trustworthy for public deployment. Thus, while existing fake news detection systems have laid the foundation, there is still a strong need for enhanced models that incorporate optimized feature selection and advanced classification mechanisms like Multi-class SVMs to improve both accuracy and scalability in real-world applications.

Disadvantages of Existing Systems

1. **Limited Feature Representation:**
   Most existing systems rely heavily on basic textual features such as word count, TF-IDF, or sentiment polarity. These features may not capture the deeper semantic and contextual meaning of the text, reducing the system's ability to accurately detect sophisticated or well-crafted fake news.
2. **No Feature Optimization:**
   Many models do not include feature selection or dimensionality reduction techniques, resulting in high-dimensional and noisy input data. This can degrade performance, slow down processing time, and cause overfitting or underfitting of the model.
3. **Binary Classification Only:**
   Several systems use binary classifiers that only distinguish between "fake" and "real" news. They lack the ability to handle more complex or multi-class scenarios (e.g., satire, partially true news), which limits their flexibility in real-world applications.
4. **Poor Generalization Across Domains:**
   Models trained on one type of dataset often perform poorly when applied to news from other domains or platforms. This lack of generalization results from over-reliance on dataset-specific vocabulary or structure.
5. **Lack of Real-time Detection Capabilities:**
   Most existing systems are trained and tested on offline datasets. They cannot process live data streams from social media or news feeds, making them impractical for real-time fake news mitigation.
6. **No User Behavior or Source Analysis:**
   These systems often ignore external factors such as user credibility, source reliability, or how the news is spreading across networks—factors which can significantly enhance the accuracy of fake news detection.
7. **Low Interpretability:**
   Many models function as black boxes, offering little to no explanation for their predictions. This limits user trust and makes it hard to justify or verify classification results in sensitive scenarios.

PROPOSED SYSTEM

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

The proposed system aims to overcome the limitations of existing fake news detection approaches by introducing a **feature-rich, optimized Multi-class Support Vector Machine (MSVM) classification model**. This system leverages advanced Natural Language Processing (NLP) techniques to extract both shallow and deep features from the news text. These include lexical features (TF-IDF, n-grams), syntactic patterns, readability scores, and sentiment attributes, which together provide a comprehensive representation of the article's content. This multifaceted feature set ensures that the model captures subtle cues present in fake news, such as sensational tone, writing inconsistencies, and manipulation tactics.

To improve model performance and reduce noise, **feature optimization techniques** such as Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), or Genetic Algorithms (GA) are applied. These techniques help in selecting the most relevant and high-impact features, reducing computational overhead and improving classifier accuracy. The optimized feature set is then used to train a **Multi-class Support Vector Machine** instead of a basic binary SVM. The MSVM is capable of classifying news into multiple categories (e.g., Real, Fake, Satire, Misleading), making it more adaptable for diverse real-world applications. Hyperparameter tuning using GridSearchCV or similar optimization frameworks ensures that the model is well-regularized and avoids overfitting.

The final system is designed to be scalable and ready for real-time integration with social media monitoring tools and news aggregation platforms. It is evaluated using standard metrics such as Accuracy, Precision, Recall, F1-Score, and AUC-ROC to ensure robustness and reliability. Additionally, the system can be enhanced with explainability tools like LIME or SHAP to provide users with transparent reasons behind classification decisions. By combining optimized features with a powerful classification engine, the proposed system offers a significant advancement over traditional fake news detection models, ensuring higher accuracy, broader scope, and real-world usability.

Advantages of the Proposed System

1. **Improved Accuracy through Feature Optimization:**
   The system employs advanced feature selection and optimization techniques (like PCA or Genetic Algorithms), which remove irrelevant or redundant data, significantly boosting the accuracy and efficiency of the classification model.
2. **Multi-class Classification Support:**
   Unlike traditional binary classifiers, the use of a **Multi-class Support Vector Machine (MSVM)** allows the system to classify news into multiple categories such as *fake*, *real*, *satire*, and *misleading*, making it more practical and adaptable in real-world scenarios.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

3. **Comprehensive Feature Extraction:**
   The model extracts a rich set of features including TF-IDF, n-grams, sentiment scores, and readability metrics, enabling a deeper understanding of the textual content and improving detection of deceptive writing patterns.
4. **Scalability and Real-time Readiness:**
   The system is designed with scalability in mind and can be integrated with real-time data sources like social media platforms or news feeds, allowing it to detect fake news as it spreads.
5. **Reduced Computational Complexity:**
   By applying feature optimization, the system minimizes dimensionality, which in turn lowers training time and memory usage, making it efficient and suitable for deployment in resource-constrained environments.
6. **High Generalization Capability:**
   Through optimized training and robust feature engineering, the model is capable of generalizing well across different datasets and news domains, unlike many existing systems that are domain-specific.
7. **Enhanced Trust and Interpretability:**
   The system can be integrated with explanation tools (such as LIME or SHAP) to provide users with understandable insights on why a particular article is classified as fake, thereby improving transparency and trust.

Implementation

The implementation of the Fake News Detection System focuses on identifying whether a news article is genuine or fake using Feature-Based Optimized Multi-Class Support Vector Machine (MSVM) classification techniques. The system analyzes textual content, linguistic patterns, and news source credibility to classify news accurately.

The proposed system helps reduce the spread of misinformation on digital platforms and improves information reliability.

1. Data Collection

The first stage involves collecting news datasets from various online sources such as:

- News Websites
- Social Media Platforms
- Online Blogs
- Fact-Checking Portals
- Public News Datasets

The collected dataset may contain:

- News Title
- News Content
- Author Information
- Publishing Date
- Source URL
- Category of News

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

- User Comments
- Labels (Fake/Real)

These attributes help train the fake news detection model.

2. Data Preprocessing

The collected news data is cleaned and prepared before classification.

Preprocessing steps include:

- Removing duplicate articles
- Handling missing values
- Removing stop words
- Tokenization
- Stemming and Lemmatization
- Converting text to lowercase
- Noise and punctuation removal

This improves text quality and model accuracy.

3. Feature Extraction

Important textual and linguistic features are extracted from news articles.

Features Used

Linguistic Features

- Word frequency
- Sentence structure
- Grammar patterns
- Use of sensational words

Statistical Features

- TF-IDF values
- N-gram analysis
- Keyword occurrence frequency

Semantic Features

- Contextual meaning
- Sentiment analysis
- Topic modeling

Source-Based Features

- Website credibility
- Author reliability
- Publishing behavior

Feature extraction improves fake news classification performance.

4. Feature Optimization

Feature optimization techniques are applied to select the most relevant features and reduce unnecessary data.

Common optimization techniques include:

- Principal Component Analysis (PCA)
- Chi-Square Feature Selection
- Information Gain
- Genetic Algorithms
- Recursive Feature Elimination (RFE)

Optimized features improve classification efficiency and reduce computational complexity.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

5. MSVM Classification Model Development

The optimized features are used to train a Multi-Class Support Vector Machine (MSVM) classifier.

MSVM Functions

The MSVM model:

- Separates fake and real news categories
- Handles multi-category news classification
- Improves classification boundaries
- Maximizes prediction accuracy

The model learns patterns from historical fake and genuine news datasets.

6. Model Training and Testing

The dataset is divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

Training Phase

The MSVM classifier learns patterns and relationships between extracted features and news labels.

Testing Phase

The trained model is tested using unseen news articles to evaluate performance.

Performance metrics include:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix
- ROC-AUC Score

Methodology

The methodology of the proposed Fake News Detection System follows a Feature-Based Machine Learning classification approach using optimized MSVM algorithms.

Step 1: Problem Identification

The rapid spread of fake news through social media and digital platforms creates misinformation and public confusion. Traditional manual verification methods are time-consuming and inefficient. The proposed system aims to automate fake news detection using optimized Machine Learning techniques.

Step 2: Requirement Analysis

The following requirements are analyzed:

- News dataset requirements
- Text processing requirements
- Feature extraction techniques
- Feature optimization methods
- Classification algorithm requirements

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Step 3: Dataset Preparation

News datasets containing fake and genuine articles are collected and divided into:

- Training Dataset
- Validation Dataset
- Testing Dataset

The dataset is preprocessed for better classification performance.

Step 4: Feature Engineering and Optimization

The methodology includes:

1. Extract linguistic and semantic features
2. Apply TF-IDF and N-gram analysis
3. Perform feature optimization
4. Select most relevant features

This improves detection efficiency and reduces noise.

Step 5: MSVM Classification Implementation

The Machine Learning workflow includes:

1. Input news article
2. Preprocess text data
3. Extract optimized features
4. Train MSVM classifier
5. Predict fake or real news
6. Generate classification results

Step 6: Performance Evaluation

The system is evaluated based on:

- Classification accuracy
- Detection speed
- False positive rate
- Precision and recall
- Feature optimization efficiency

Technologies Used

- Python
- Machine Learning Algorithms
- MSVM Classification
- Scikit-learn
- TensorFlow
- Natural Language Processing (NLP)
- Pandas &NumPy
- Flask / Django

Results

To run project double click on ‘run.bat’ file to start python server and then will get below page

In above screen python server started and now open browser and enter URL as http://127.0.0.1:8000/index.html and then press enter key to get below page

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

In above screen click on 'New User Sign up' link to get below page

In above screen user is entering sign up details and then press button to get below page

In above screen user sign up completed and now click on 'User Login' link to get below page

In above screen user is login and after login will get below page

In above screen click on 'Load Fake News' link to load dataset and then will get below page

In above screen dataset loaded where first column contains news data and second column contains Class label and FALSE or TRUE. Now click on 'Run MPCA and Firefly' link to extract and select features from dataset and then will get below page

In above screen in first two lines can see number of records and features available in dataset and then MPCA extracted 300

features and then Firefly selected 160 important features and now click on 'Run MSVM Algorithm' link to train algorithms and then will get below page

Conclusion

In conclusion, the Fake News Detection System using a feature-based optimized Multi-class Support Vector Machine (MSVM) classification provides a comprehensive and efficient approach to tackling the pervasive issue of fake news on digital platforms. By leveraging a rich set of textual features—including lexical, syntactic, and semantic attributes—and applying advanced feature optimization techniques, the system enhances the overall classification accuracy while reducing computational complexity. The choice of MSVM as the core classifier enables the model to effectively differentiate between multiple classes such as real news, fake news, satire, and misleading content, moving beyond simple binary classification and offering more nuanced insights. This makes the system highly adaptable and useful in a variety of real-world scenarios where different types of misinformation may be encountered. The rigorous training and validation processes ensure that the model generalizes well across diverse datasets and maintains robustness even as news styles and sources evolve over time.

Moreover, the system's scalable architecture and potential for real-time integration make it suitable for deployment in social media monitoring, news aggregators, and fact-checking platforms, where timely detection of fake news is crucial. The implementation of explainability tools further adds value by providing transparency in decision-making, thereby enhancing user trust and facilitating broader adoption by journalists, researchers, and everyday users. By addressing key challenges faced by existing fake news detection systems—such as limited feature scope, high computational cost, and binary classification constraints—this proposed system marks a significant advancement in the field. Ultimately, it contributes meaningfully to curbing the spread of misinformation, promoting accurate information dissemination, and supporting informed public discourse in an increasingly digital and interconnected world.

References

1. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36. https://doi.org/10.1145/3137597.3137600
2. Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 422-426.

https://doi.org/10.18653/v1/P17-2067

3. Zhou, X., &Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys (CSUR)*, 53(5), 1-40. https://doi.org/10.1145/3395046
4. Cortes, C., &Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. https://doi.org/10.1007/BF00994018
5. Kaur, H., & Singh, A. (2020). Fake News Detection using Machine Learning Approaches: A Review. *International Journal of Computer Applications*, 176(23), 1-7. https://doi.org/10.5120/ijca2020920249
6. Ribeiro, M. T., Singh, S., &Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778
7. Jain, S., Kumar, N., & Gupta, D. (2021). Fake News Detection Using Recursive Feature Elimination and Support Vector Machine. *International Journal of Advanced Computer Science and Applications*, 12(5), 401-408. https://doi.org/10.14569/IJACSA.2021.0120548
8. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1), 1-4. https://doi.org/10.1002/pra2.2015.145052010082
9. Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797-806. https://doi.org/10.1145/3132847.3132877
10. Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2931-2937. https://doi.org/10.18653/v1/D17-1317

He earned his Master of Computer Applications (MCA) from Anna University, Chennai. With a strong research background, He has authored and co-authored research papers published in reputed peer-reviewed journals. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits

\

AUTHORS PROFILE

Mr.HimambashaShaik is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh.

CH.HARIKRISHNA is a postgraduate student pursuing a MCA in the Department of Computer Applications at QIS College of Engineering & Technology,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Ongole an Antonomous college in Prakasam dist. He completed his undergraduate degree in BCA (COMPUTERS) from ANU. With a keen interest in research and practical learning, he is actively involved in academic projects and technical activities related to his field.