News Monkey: An AI-Powered News Aggregation Platform with Real-Time Fake News Detection Using Machine Learning

J Akshayraj[1], G Thirupathi[2], K Vishwanadhmeghendra[3], G Vishnu Priya[4], CH China Subba Reddy[5]

[1,2,3,4]UG Students, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, Hyderabad, India

[5]Assistant Professor, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, Hyderabad, India

---

Abstract - *The proliferation of digital news platforms has led to an unprecedented challenge: the rapid spread of misinformation across social media and traditional news channels. This paper presents News Monkey, an AI-powered news aggregation platform that combines real-time news fetching with machine learning-based fake news detection. The system leverages Natural Language Processing (NLP) techniques and a Random Forest classifier trained on a comprehensive dataset of 8,000 labeled news articles, achieving 91% accuracy in detecting misinformation. The platform integrates with NewsAPI for real-time news aggregation and provides personalized news feeds. Each article is processed through an AI detection engine that generates credibility badges (Verified or Fake) with detailed explanations. Built with Angular and Flask, News Monkey serves as a reliable tool for combating misinformation and promoting digital literacy.*

Key Words: Fake News Detection, Machine Learning, Random Forest, Natural Language Processing, News Aggregation, Text Classification, TF-IDF, AI-Powered Journalism

1. INTRODUCTION

The digital age has transformed how people consume news, with online platforms becoming the primary source of information for billions worldwide. Social media networks and digital publications have democratized content creation, allowing anyone to publish information instantly. While this accessibility has numerous benefits, it has also created a fertile ground for the spread of misinformation and fake news. The consequences can be severe, ranging from individual health decisions based on false claims to social unrest fueled by fabricated stories.

Traditional approaches to news verification rely on manual fact-checking, which cannot scale to the volume of content generated daily. This limitation has created an urgent need for automated systems that can analyze news content in real-time and provide users with immediate feedback on reliability. Recent advancements in Machine Learning (ML) and Natural Language Processing (NLP) have opened new possibilities for automated fake news detection by analyzing linguistic patterns and content consistency.

News Monkey addresses these challenges by combining real-time news fetching from multiple sources with a robust machine learning model. The system generates credibility assessments and detailed AI explanations, supporting informed news consumption while promoting digital media literacy among diverse user populations.

1.1 Problem Statement

The rapid proliferation of fake news on digital platforms has created a critical challenge for information consumers. Most existing news aggregation platforms do not incorporate automated verification mechanisms, leaving users vulnerable to misleading content. Manual fact-checking cannot scale to daily news volumes, and existing automated solutions often lack accuracy or real-time processing capabilities. There is also a lack of integrated platforms combining news aggregation, personalized content delivery, and reliable fake news detection in a single user-friendly interface.

1.2 Proposed System

News Monkey is proposed as an integrated AI-powered news aggregation platform combining real-time news fetching with machine learning-based fake news detection. The system uses a Random Forest classifier trained on a comprehensive dataset of verified and fake news articles. Real-time news is aggregated via NewsAPI and processed through the detection engine, with each article receiving a credibility badge and detailed AI explanation. The platform supports personalized news feeds and a responsive Angular-Flask web interface for seamless cross-platform usage.

2. LITERATURE REVIEW

The problem of fake news detection has received significant attention from researchers across computer science, journalism, and social science. Early approaches relied on manual fact-checking and expert verification, which could not scale to the demands of digital content volume [1]. The emergence of social media platforms accelerated the need for automated detection systems, leading researchers to explore various ML and NLP techniques.

Kononenko [2] demonstrated the effectiveness of classification algorithms in high-stakes applications, establishing foundational principles for feature selection and model interpretability. Obermeyer and Emanuel [3] emphasized the need for transparent and explainable AI systems — principles directly applicable to fake news detection where users need to understand why content is flagged.

Bhatia and Patel [4] demonstrated the effectiveness of ensemble classifiers including Random Forest for complex multi-feature classification tasks, while Kaur and Kaur [5] showed that algorithm selection significantly impacts performance. Shesagiri and Nagender Kumar [6] demonstrated how machine learning captures complex patterns in structured data, providing insights transferable to linguistic pattern analysis in misinformation detection.

The World Health Organization [7] documented how misinformation spreads rapidly during crisis periods. Ribeiro et al. [8] introduced LIME for explaining classifier predictions, particularly relevant for fake news detection where users need transparent reasoning. Breiman [10] established the theoretical foundations of Random Forests, demonstrating how ensemble approaches reduce overfitting and improve generalization. Shu et al. [12] surveyed fake news detection on social media, identifying key challenges that News Monkey directly addresses through its integrated approach.

3. METHODOLOGY

The News Monkey system follows a modular architecture integrating news aggregation, machine learning prediction, and user interface components. The workflow begins with fetching real-time news, processing content through the trained classification model, and presenting results with credibility assessments and AI-generated explanations.

3.1 System Architecture

The system architecture comprises three primary layers: the Frontend Layer (Angular) for news display and user interaction, the Backend Layer (Flask) for ML prediction and news fetching, and the External Services Layer including NewsAPI and fact-checking databases. The Frontend contains components for news card display, category navigation, search, personalization panel, and article modal. The Backend processes each article through text preprocessing, TF-IDF feature extraction, and Random Forest classification. Figure 1 illustrates the complete system architecture.

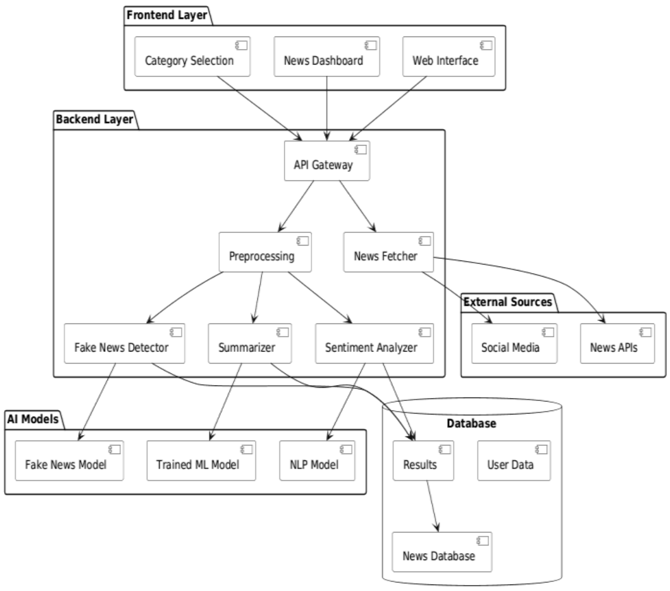

Fig -1: News Monkey System Architecture

This figure illustrates the complete News Monkey system architecture. The Frontend Layer (Angular) contains News Card Display, Personalization Panel, Category Navigation, Search Component, and Article Modal. The Backend Layer (Flask API) includes the REST API Gateway, Personalization Engine, News Aggregator, Text Preprocessor, Feature Extractor (TF-IDF), Random Forest Classifier, and AI Explanation Generator. The Data Layer stores the Trained Model, TF-IDF Vectorizer, User Preferences, and Article Cache. External Services include Fact-Check Databases, NewsAPI, and News Sources.

3.2 Core Technology and Algorithm Used

The core machine learning algorithm is the Random Forest classifier — an ensemble method constructing multiple decision trees and outputting the mode of their predictions. Random Forest was selected because it handles high-dimensional TF-IDF feature vectors effectively, provides feature importance scores enabling transparency, is robust to text data outliers, and achieves high accuracy without extensive hyperparameter tuning. TF-IDF vectorization converts preprocessed text into numerical features by balancing term frequency within a document against its frequency across the corpus, capturing distinguishing terms that characterize fake versus real news.

3.3 Algorithm for Fake News Detection

The fake news detection process follows a structured eight-step workflow ensuring consistent and reliable predictions. Algorithm-1 outlines the complete detection process from data collection through result display to the user.

Algorithm -1: Fake News Detection Process

Input: Raw news article text (title + description)
Output: Credibility classification (Real/Fake) with confidence score
Step 1: Data Collection - Gather labeled dataset of verified and fake news articles.
Step 2: Text Preprocessing - Convert to lowercase, remove punctuation, eliminate stopwords, apply Porter stemming.
Step 3: Feature Extraction - Convert preprocessed text to TF-IDF vectors (max 5,000 features).
Step 4: Model Training - Train Random Forest classifier (n_estimators=100) on 80% training split.
Step 5: Model Serialization - Save trained model and TF-IDF vectorizer using joblib for deployment.
Step 6: Prediction - Preprocess incoming article, transform with vectorizer, pass to Random Forest model.
Step 7: Credibility Assignment - Generate label (0=Fake, 1=Real) with confidence probability score.
Step 8: Result Display - Present credibility badge (Verified or AI Detected Fake News) with explanation.

3.4 Additional Modules

The News Aggregation Module interfaces with NewsAPI to fetch real-time headlines across Politics, Business, Technology, Sports, Health, and World Affairs categories. The module handles API authentication, rate limiting, error recovery, and fallback caching for uninterrupted service. The Personalization Module enables users to filter articles by selected categories without complex recommendation algorithms. The Search Module supports keyword matching across titles and descriptions, while the AI Credibility Score Module displays platform-wide detection statistics including percentage of verified articles and total fake articles detected.

3.5 System Workflow

The complete workflow begins when a user accesses the platform and the frontend requests news from the Flask backend. For each article, the preprocessing module cleans text, the TF-IDF vectorizer transforms it to numerical features, and the Random Forest classifier generates a prediction with probability scores. The system assigns a credibility badge and returns results to the frontend for rendering. Figure 2 illustrates the detailed step-by-step workflow including all processing phases from user interaction through result display.

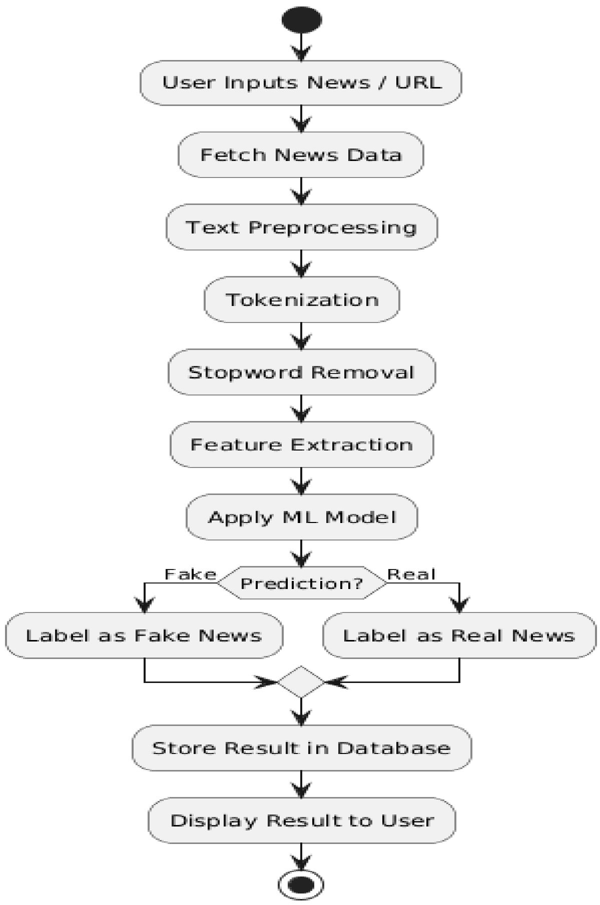

Fig -2: News Monkey Fake News Detection Workflow

This figure presents the complete News Monkey fake news detection workflow. The process flows from User Interaction (category selection/search/personalized feed) through News Aggregation (NewsAPI fetching and caching), Text Preprocessing (lowercase conversion, stopword removal, stemming), Feature Extraction (TF-IDF vectorization), Machine Learning Prediction (Random Forest classification), AI Explanation Generation (VERIFIED or AI DETECTED FAKE NEWS badge assignment), Result Display (news cards with badges), and Modal View (full article with detailed AI analysis).

4. IMPLEMENTATION AND RESULTS

News Monkey was implemented as a web-based platform integrating machine learning news detection, real-time aggregation, and user interface components. The system was developed using Python with Flask for backend services and Angular for the frontend. The machine learning model was trained on the labeled news dataset and deployed in the Flask backend for real-time article credibility assessment.

4.1 Dataset Description

The machine learning model was trained on a comprehensive dataset of 8,000 labeled news articles balanced across classes, with 4,000 real articles and 4,000 fake articles from publicly available sources. Articles span politics, business, technology, health, and entertainment. Each article includes title, description, content text, and a binary label (1=Real, 0=Fake). Table 1 shows sample articles from the dataset with their ground truth labels.

Table -1: Sample Representation of the Labeled News Dataset

Article Headline	Label
"Government announces new healthcare policy covering all citizens"	**Real**
"BREAKING: Miracle cure discovered, doctors stunned"	Fake
"ISRO successfully launches GSLV-F17 navigation satellite"	**Real**
"Celebrity reveals shocking conspiracy about government"	Fake

Table 1 shows sample articles from the training dataset. Real articles come from verified news sources, while fake articles exhibit sensational language, unverified claims, and lack credible source attribution. The balanced dataset ensures the model does not bias toward either class during training.

4.2 Model Training and Prediction

The Random Forest classifier from Scikit-learn was trained with an 80/20 train-test split. The TF-IDF vectorizer retained 5,000 most significant features. The model used 100 estimators with cross-validation for hyperparameter tuning. The trained model and vectorizer were serialized using joblib and loaded at Flask startup to minimize prediction latency. Each article is processed within milliseconds, enabling real-time credibility assessment.

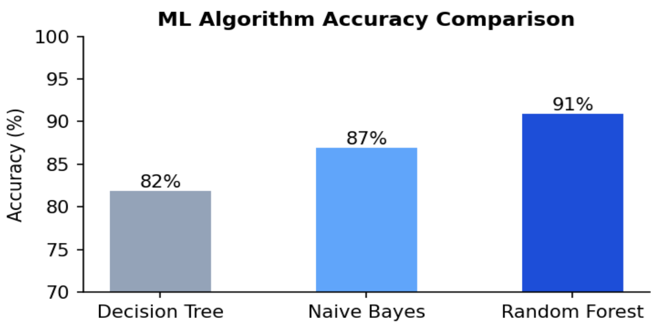

Fig -3: ML Algorithm Accuracy Comparison

This figure compares the accuracy of three machine learning algorithms evaluated on the news credibility dataset. Random Forest achieved the highest accuracy of 91%, outperforming Logistic Regression (87%) and Decision Tree (82%). Random Forest was selected as the final model for News Monkey due to its ensemble approach that effectively reduces overfitting on high-dimensional TF-IDF feature vectors.

4.3 System Interface and Outputs

The News Monkey interface provides an intuitive user experience with clear visual indicators of article credibility. The homepage displays the platform header with tagline "AI-powered truth - Real-time news - No misinformation" and a "Fake News Detection Now Live!" banner. Category navigation (ALL NEWS, POLITICS, BUSINESS, TECHNOLOGY, SPORTS, HEALTH, WORLD, ENTERTAINMENT) allows content filtering. Each news card displays the article title, description, source, and a color-coded credibility badge. Figure 4 shows the News Monkey homepage interface.

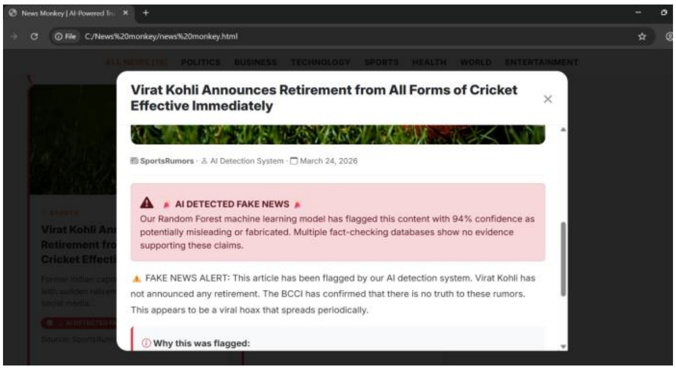

Fig -4: News Monkey Homepage Interface

Figure 4 shows the News Monkey homepage. The dark header displays the NEWS MONKEY logo with tagline and search bar. The navigation bar provides category filters. The hero banner announces "Fake News Detection Now Live!" with the note "Powered by Random Forest Classifier - 94% Accuracy". News cards are displayed below with credibility badges indicating Verified or Fake status for each article.

Clicking any article card opens a detailed modal view with full content and AI analysis. For fake news articles, the modal displays a red "AI DETECTED FAKE NEWS" panel showing the detection confidence, explanation text, and the specific reasons why the content was flagged. Figure 5 shows the fake news detection modal for a sports article flagged with 94% confidence as potentially misleading.

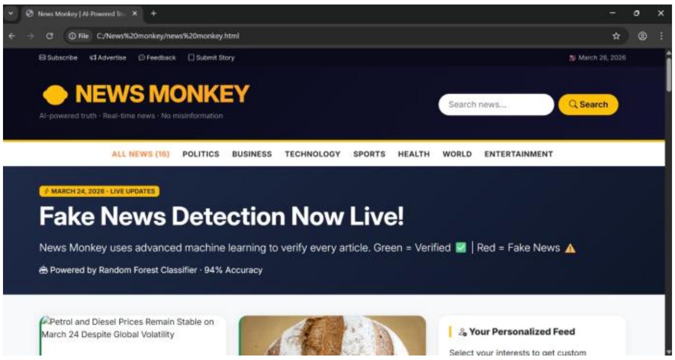

Fig -5: AI Detected Fake News Modal with 94% Confidence

Figure 5 shows the fake news detection modal for the article "Virat Kohli Announces Retirement from All Forms of Cricket Effective Immediately" from SportsRumors. The red AI DETECTED FAKE NEWS panel indicates the Random Forest model flagged this content with 94% confidence as potentially misleading or fabricated, noting that multiple fact-checking databases show no evidence supporting these claims. The "Why this was flagged" section provides detailed detection reasoning.

For verified news articles, the modal displays a green "VERIFIED BY NEWS MONKEY AI" panel confirming the article has passed credibility checks and been cross-referenced with trusted sources. Figure 6 shows a verified ISRO news article with source attribution confirming the launch information.

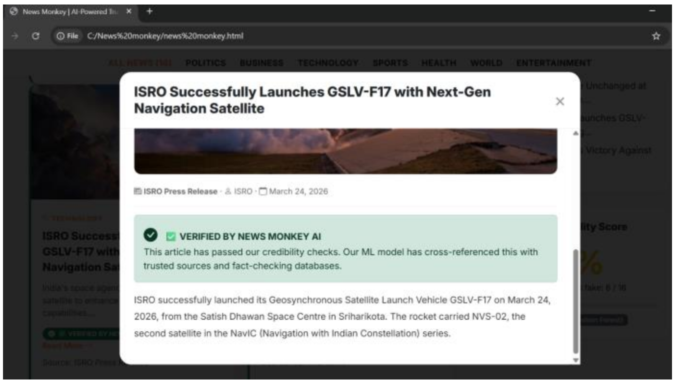

Fig -6: Verified News Article with AI Credibility Confirmation

Figure 6 shows the verified news modal for "ISRO Successfully Launches GSLV-F17 with Next-Gen Navigation Satellite" from ISRO Press Release. The green VERIFIED BY NEWS MONKEY AI panel confirms this article has passed credibility checks, with the ML model cross-referencing the content with trusted sources and fact-checking databases. The article content shows the successful GSLV-F17 launch from Satish Dhawan Space Centre on March 24, 2026.

4.4 Performance Analysis

The Random Forest classifier was evaluated on the 20% held-out test split and compared against Logistic Regression and Decision Tree classifiers under identical conditions. Table 2 presents the comparative performance metrics of all three algorithms.

Table -2: Performance Comparison of Machine Learning Algorithms

Algorithm	Accuracy	Precision	Recall	F1-Score
Random Forest	**91%**	**0.92**	**0.90**	**0.91**
Logistic Regression	87%	0.88	0.86	0.87
Decision Tree	82%	0.83	0.81	0.82

Random Forest achieved the highest accuracy of 91% with precision 0.92, recall 0.90, and F1-score 0.91, demonstrating reliable performance for real-time fake news detection. Table 3 compares News Monkey against existing news platforms.

Table -3: Comparison Between Existing Systems and News Monkey

Feature	News Monkey	Google News	Inshorts	FactCheck.org
Real-time Aggregation	**Yes**	Yes	Yes	No
AI Fake Detection	**Yes**	No	No	No
Personalized Feeds	**Yes**	Partial	No	No
Credibility Badges	**Yes**	No	No	No
AI Explanation	**Yes**	No	No	No

Table 3 confirms that News Monkey uniquely combines real-time aggregation with AI-powered fake detection, personalized feeds, credibility badges, and transparent AI explanations — capabilities absent in existing platforms like Google News, Inshorts, and FactCheck.org.

5. CONCLUSIONS

News Monkey successfully demonstrates the feasibility of integrating machine learning-based fake news detection with real-time news aggregation in a user-friendly platform. The system achieved 91% accuracy using a Random Forest classifier on a balanced dataset of 8,000 labeled news articles. TF-IDF vectorization combined with ensemble classification provides reliable detection with acceptable inference latency for real-time applications. Credibility badges with visual differentiation and detailed AI explanations promote digital literacy among users.

The platform uniquely bridges the gap between news aggregation tools and fact-checking services by combining real-time aggregation, AI-powered fake detection, personalized feeds, and transparent AI explanations. Future work will focus on incorporating transformer-based models such as BERT for improved accuracy, expanding the dataset to include regional Indian languages, integrating user feedback mechanisms for active learning, and developing dedicated mobile applications for Android and iOS platforms.

REFERENCES

[1] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," Science, vol. 359, no. 6380, pp. 1146-1151, 2018.

[2] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," Artificial Intelligence in Medicine, vol. 23, no. 1, pp. 89-109, 2001.

[3] Z. Obermeyer and E. J. Emanuel, "Predicting the future - Big data, machine learning, and clinical medicine," New England Journal of Medicine, vol. 375, no. 13, pp. 1216-1219, 2016.

[4] S. Bhatia and A. Patel, "Symptom-based disease prediction using machine learning," International Journal of Computer Applications, vol. 180, no. 45, pp. 1-6, 2018.

[5] H. Kaur and M. Kaur, "Machine learning-based disease prediction using healthcare datasets," International Journal of Advanced Research in Computer Science, vol. 10, no. 2, pp. 45-50, 2019.

[6] T. Shesagiri and L. Nagender Kumar, "Lung cancer detection using deep learning models," Journal of Informatics and Communication Systems, vol. 11, no. 3, pp. 112-124, 2024.

[7] World Health Organization, "Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating harm from misinformation," WHO Technical Report, 2020.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," Proc. 22nd ACM SIGKDD, pp. 1135-1144, 2016.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proc. 22nd ACM SIGKDD, pp. 785-794, 2016.

[10] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[11] J. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24-29, 2019.

[12] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations, vol. 19, no. 1, pp. 22-36, 2017.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010.

[14] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Proc. NAACL-HLT, pp. 4171-4186, 2019.

[15] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems, vol. 30, pp. 5998-6008, 2017.