Journal Homepage: www.ijarpr.com ISSN: 3049-0103 (Online)

International Journal of Advance Research Publication and Reviews

Vol 03, Issue 10, pp 457-459, October, 2026

TruthLens: An AI-Based Fake News Detection System

Sneha Navale[1],Arpita Gunjal[2], Mrs. Manisha Patil[3], Mr. Nitin Joshi[4]

Master of Science in Computer Science (M.Sc. Computer Science), Department of Computer Science, Indira University, Pune, Maharashtra, India

[3-4] Under the Guidance of:

ABSTRACT :

The rapid dissemination of online news motivates computational approaches for screening potentially misleading content. TruthLens is an academic project proposing an English-language binary news-text classifier and an accessible interface for presenting its predictions. The project materials identify the Kaggle Fake and Real News Dataset, distributed as Fake.csv and True.csv, as the intended data source. They document exploratory-analysis steps for inspecting file structure, missingness, duplicate records, class distribution and article length. The proposed modelling workflow cleans text, transforms it with term frequency–inverse document frequency (TF-IDF), and trains a supervised classifier, with Logistic Regression identified in the presentation as the preferred model. Accuracy, precision, recall, F1-score and a confusion matrix are specified for evaluation. A React-based interface, application backend, Python model API, PostgreSQL history and explainability are architectural intentions, not independently verified deployed features. Although the presentation names Logistic Regression as the best model, neither numerical test results nor model artefacts are supplied; accordingly, comparative superiority and predictive performance cannot yet be established. The contribution documented at this stage is a coherent design and preliminary data-analysis procedure, rather than a validated fake-news verification service. The paper specifies the missing experiments and implementation evidence required for a reproducible empirical submission.

Keywords—Fake News Detection; Natural Language Processing; TF-IDF; Logistic Regression; Text Classification; Explainability.

I. Introduction

Online news dissemination increases the practical need for scalable screening of potentially misleading articles. A learned text classifier estimates association with training labels, rather than independently verifying factual claims. TruthLens proposes a supervised, text-centred pipeline and a user-facing presentation layer. The source documents describe the project mainly in future-tense terms; the distinction between design and demonstrated execution is maintained throughout this manuscript.

II. Problem Statement and Research Objectives

The research problem is to classify English news text as Fake or Real using labelled examples while communicating the limits of such classification. The documented objectives are dataset inspection, preprocessing and TF-IDF representation, supervised model training and evaluation, and a web interface for article submission. Confidence display, explanations, prediction history and dashboards are secondary planned objectives; implementation evidence for these features was not provided.

III. Related Work and Literature Review

Shu et al. [1] introduced FakeNewsNet, combining news content with social and spatiotemporal context; it demonstrates information sources absent from a text-only design, rather than validating TruthLens. Athira et al. [2] reviewed explainable fake-news detection and highlighted the need to evaluate explanation quality, but their work does not establish that an explanation feature exists in this project. The ISOT research group describes a dataset assembled from truthful and unreliable news sources [3]; its source provenance raises potential domain and source-style confounding. Pedregosa et al. [4] document scikit-learn as a general-purpose implementation framework, not evidence of a particular project experiment. These sources motivate a transparent, carefully evaluated baseline; none supports a claim that TruthLens is novel or more accurate than prior systems.

IV. Research Gap and Motivation

The defensible motivation is a practical integration and evaluation question: can a lightweight text classifier be developed and its limitations communicated in a usable interface? The supplied literature does not establish that integrated classification, explanations and history are absent from previous research. Consequently this is an application-oriented project objective, not a proven first-of-its-kind research gap.

V. Proposed Methodology

The proposed workflow is: acquire the labelled CSV files; inspect missing and duplicate records; assign Fake and Real labels; combine and explore records; clean article text; partition the data; fit TF-IDF only on training text; transform held-out text using the fitted vectorizer; train Logistic Regression; evaluate predictions; persist the fitted preprocessing and model objects; and, subject to implementation, expose predictions through a web service. The report includes representative R code for inspection and plotting, but no printed output or execution log.

VI. Dataset Description

The synopsis identifies the ISOT-associated Kaggle Fake and Real News Dataset [3], with Fake.csv and True.csv. Title, text, subject and date are source attributes; Fake/Real is the assigned target label. WELFake is mentioned as an alternative, not as confirmed training data. The exact downloaded version, row counts, class balance, missingness, deduplication outcome, licence and input-field selection remain unverified.

Item	Documented status
Source	Kaggle Fake and Real News Dataset (proposed)
Files	Fake.csv; True.csv (presentation)
Target	Fake / Real
Attributes	Title; Text; Subject; Date; assigned Label
Rows and split	Not supplied

VII. Text Preprocessing and TF-IDF Feature Extraction

The proposed cleaning steps are missing-value handling, punctuation and unnecessary-character removal, lowercasing, optional stop-word removal, normalization and duplicate checking. No executed preprocessing configuration or tokenization settings were supplied. TF-IDF represents a document using term occurrence weighted against document frequency. A conventional formulation is $tfidf(t,d)=tf(t,d) \times \log(N/df(t))$; actual scikit-learn smoothing, normalization, n-gram and vocabulary settings must be reported from code. The vectorizer must be fitted exclusively on training records to prevent leakage.

VIII. Logistic Regression Classification

Logistic Regression is the initially proposed classifier and the presentation labels it the preferred model without numerical support. For feature vector x, binary logistic regression models $P(y=1|x)=1/(1+\exp(-(w \cdot x+b)))$. Its linear decision function is compatible with sparse TF-IDF features, and coefficients can be inspected as associations with class labels. The positive-class convention, regularization, solver, random seed and decision threshold have not been documented. Coefficients alone do not prove factual truth or causal explanation.

IX. System Architecture and Implementation

Figure 1 is a proposed architecture, not a verified deployment: news input (React + Vite/Tailwind CSS) → Node.js/Express application service → Python/FastAPI model service → saved text preprocessing, TF-IDF and classifier → prediction response. PostgreSQL is intended for users, submissions, predictions, scores and timestamps. Authentication, dashboard, history and explainability remain proposed until code, screenshots and end-to-end tests are supplied. No external authoritative claim-verification integration is documented.

X. Experimental Setup and Evaluation Metrics

The presentation names Logistic Regression, Multinomial Naive Bayes, SVM, Random Forest and Decision Tree as evaluated models, but the report provides only an empty results template and no run outputs. Therefore the actual scope of executed comparisons is unconfirmed. A reproducible experiment requires file hashes/version, class counts, cleaning rules, leakage-safe split, seed, model parameters, software versions, and held-out predictions. Define TP, TN, FP and FN for an explicitly chosen positive class. $Accuracy=(TP+TN)/(TP+TN+FP+FN)$; $Precision=TP/(TP+FP)$; $Recall=TP/(TP+FN)$; $F1=2PR/(P+R)$. Report per-class scores and the confusion matrix.

XI. Results and Discussion

The presentation refers to article-length, class and subject histograms but the supplied text contains neither plots nor counts. The R assignment in the report provides commands for generating descriptive results, not the resulting values. “Best Model: Logistic Regression” is an unverified project assertion because accuracy and F1 fields literally read “Actual result”. No measured predictive or deployment performance can responsibly be reported.

Required result	Current evidence
Dataset size / class distribution	Not provided
Train/test proportion and seed	Not provided
Logistic Regression accuracy / precision / recall / F1	Not provided

Confusion matrix	Not provided
Other classifier comparison	Named in slides; numerical output absent
SHAP/explanation examples	No implementation evidence
Web application tests/screenshots	Not provided

XII. Limitations and Threats to Validity

A text-only classifier may learn source, topic, formatting and temporal cues rather than factual veracity. Duplicate or near-duplicate articles across partitions can inflate scores; fitting the vectorizer before splitting also leaks information. The intended dataset is topic- and source-dependent, so transfer to new publishers, events, languages and manipulated text requires separate testing. Predicted scores are not calibrated probabilities without calibration assessment. No evidence currently supports deployed authentication, secure retention or SHAP-based explanations. The absent numerical results preclude an empirical effectiveness conclusion.

XIII. Conclusion

The documents establish a proposed TruthLens pipeline centred on labelled English news, TF-IDF and Logistic Regression, with preliminary EDA instructions and a planned web architecture. They do not establish measured classification performance or a completed explainable application. TruthLens should be presented as a dataset-dependent screening design pending reproducible experiments, not a factual verification system.

XIV. Future Scope

Execute and archive a leakage-safe experiment; compare only genuinely trained baselines on an identical held-out partition; inspect errors by source and topic; assess temporal and external-dataset generalization; calibrate scores if shown as probabilities; implement and validate explanation methods before naming them; and test the integrated API, interface, authentication, history and retention controls. Multilingual and multimodal analysis are longer-term extensions, not current capabilities.

REFERENCES

[1] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, “FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media,” Big Data, vol. 8, no. 3, pp. 171–188, 2020. doi: 10.1089/big.2020.0062.

[2] Athira A. B., S. D. Madhu Kumar, and A. M. Chacko, “Towards Smart Fake News Detection Through Explainable AI,” arXiv:2207.11490, 2022. https://arxiv.org/abs/2207.11490

[3] ISOT Research Group, University of Victoria, “Fake News Detection Datasets.” https://onlineacademiccommunity.uvic.ca/isot/datasets/ ; project-proposed Kaggle distribution: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

[4] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. https://www.jmlr.org/papers/v12/pedregosa11a.html