International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fake News Detection Using NLP and Machine Learning: Evaluating Models for Identifying Misinformation on Social Media

Sujay Srinivas[1], Srikara S Shet[2]

[1] Asst. Prof. Surana college Autonomous
[2] MCA, Surana college Autonomous

ABSTRACT :

The spread of detrimental information and misinformation, especially on internet-based media, has presented a pressing threat to the quality of public debate and informed decision-making. This article provides a structured summary of the latest Natural linguistic Processing (NLP) and machine learning methods for malicious data detection, with consideration also for purely text-based as well as multimodal approaches using images and videos in addition to text. The study compares some of the model architectures used to identify forged data from conventional classifiers such as logistic regression and Support Vector Machines (SVM) and Random Forest to state-of-the-art deep learning models such as Long Short-Term Memory (LSTM) networks and transformer-based models such as BERT and CNN, which have been shown to be more accurate in misinformation tagging. Criteria such as precision, precision, recall, F1-score, and ROC-AUC are described to provide an all-rounded view of model effectiveness, whereas the work also vilifies real-world issues such as corpus imbalance, linguistic diversity, quick advancement of false data plans, and paucity in analyzing text and images or videos together. Through a critical review of the resources to be exploited in this assignment, the manuscript discusses leading datasets widely used to train and test identification models and outlines the relevance of high-quality data annotation during authenticity verification.

1 Introduction

Bogus data may be likened to a cunning imposter, pretending to be genuine news. This imposter has become a huge problem on the internet, and its effect is much more critical than merely tricking people. It can crush our trust on each other, upend political discussions, and even cause real-world problems such as social unrest or health crises. The issue becomes worse when it is against people who are already excluded or when it erupts at times of importance like during elections or epidemics. Unlike the conventional newspapers or television networks where they had editors to vet everything, platforms like Twitter and Facebook enable anybody to publish information immediately, making it easy for lies to spread at staggering speeds and volumes. In the past, all you could do to fight it was with slow, human fact-checking. But that was like trying to put out a forest fire using a watering can. Today, we have much smarter hardware. We've developed deep learning algorithms—digital sleuths like CNNs, LSTMs, and BERT. They can spot teeny-weeny clues, subtlety of meaning, and hidden relationships a human might miss. Even with these new powerful tools, we still have issues. The people responsible for fake news are always adapting their approach, and we must be careful that we don't become biased or censor the wrong thing. However, we've made tremendous strides in our war on misinformation. Long Short-Term Memory networks (LSTMs), and Transformer-based models like BERT, have also been promising. These models are capable of detecting subtle subtleties, context dependencies, and semantic relationships that are typically ignored by classical algorithms. With these advancements, challenges still exist, such as adversarial content, evolving information warfare tactics, and ethical concerns regarding censorship and bias.

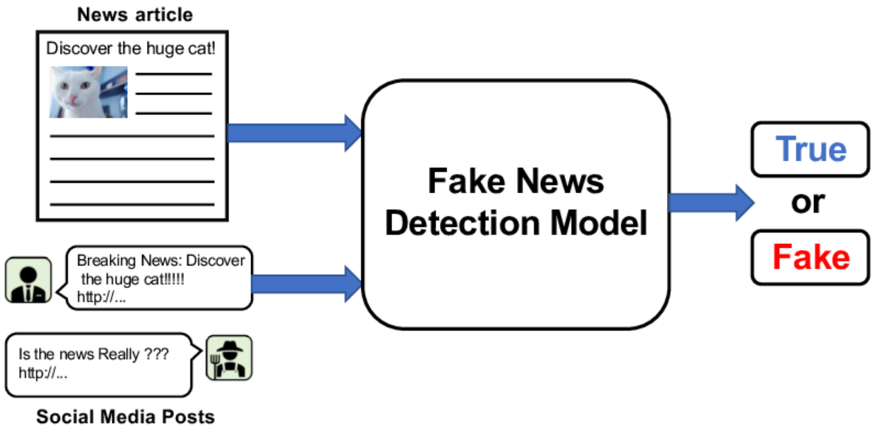
2 Motivation

These days, with the internet age, misinformation travels like wildfire on the net within no time at all, and it's a real threat to everything from public discourse to how we decide. Made-up stories, whether political, health-related, etc., travel faster than truth can in most situations, with a tendency to lead to confusion, hysteria, or even outright manipulation. Manually fact-checking is just too time-consuming to keep up with what's being printed.

That is why we need smart, automated systems that can detect and flag suspicious spoofed information in real-time. By using technology such as Natural Language Processing (NLP) and Machine Learning (ML), we can create intelligent systems that monitor usage of language and information sharing. We can design these systems so that they can differentiate between what is real and what is not. By trying and testing various models on this job, we are able to make the technology more precise and help make the web a better place for us all. This is so significant because it allows us to build good, fact-based tools that can help empower people, businesses, and even governments to battle the spread of misinformation responsibly and ethically.

3 Literature Review

Platforms such as social media and online forums are superhighways of disinformation, particularly in the health sphere. One of our chief methods of combating it is with a type of technology known as Natural Language Processing (NLP), which lends itself to studying various languages, such as English, Italian, and Dutch. [1] NLP employs highly sophisticated methods to find the most important topics, emotions, and sentiments expressed in what human beings write. But there is a significant gap in this research: we have been studying so much of written language and Western languages that we know less about how to identify misleading information within oral language or non-Western languages. The issue is made worse by how simple it is to produce and disseminate misinformation. It causes a great deal of stress for news sources and businesses because this fake information is so well-written that it's difficult to tell it apart from reality.[2] In this research, we engage various models looking for this false information, ranging from classic machine learning to current deep models such as LSTMs and CNNs. While these more advanced models are extremely promising, they come with a significant failing: they require an enormous quantity of data to train on and significant computational power to execute.[3]



4 False Data Identification

4.1 Definition

False data identification is the process of detection and elimination of knowingly false or deceptive data, which is presented as real news False data is actually false news that's created with purpose so that it appears real. It's usually employed to make people believe in something, confuse them, or get attention. It's such a big deal now because it goes so rapidly viral on the internet, so that people can't help but lose track of what is authentic and what isn't.

Think of "false data identification" as a task for a super-geniuses detective. It's the detective's duty to locate and purge out false information that's masquerading as genuine news. The false information is not only done for the mere purpose of misinforming, but it's usually made so that people will believe or feel something, create confusion, or simply just make drama.[4]

4.2 Types

Fake news can be categorized into three major types. Here's a easy way to look at them:[5]

1. Fabrication

This is a blatant lie created by an individual or a group. The person who created it knows that it is untrue, and they normally put a catchy and appealing headline (known as clickbait) to trick people into clicking on the article.

2. Hoax

A hoax is a more complex lie that spreads among numerous individuals, some of whom take it to be the truth. It's more advanced in the sense that it aims to deceive a greater number of people and spreads on several websites and mass media. As an example, during an election process, there may be a hoax spreading on Twitter, Facebook, and blog sites such that individuals are made to think that a candidate had a fake history.

3. Satire

Satire is bogus news that's actually humorous. The source themselves know it's not true, and they are offering it as humorous or exaggerated commentary. The issue is, when this joke gets passed on to people who don't have the sense of humour of the source, they may not catch the joke and instead believe that the satirical piece of news is true.

4.3 Fake Data Matrix [6]

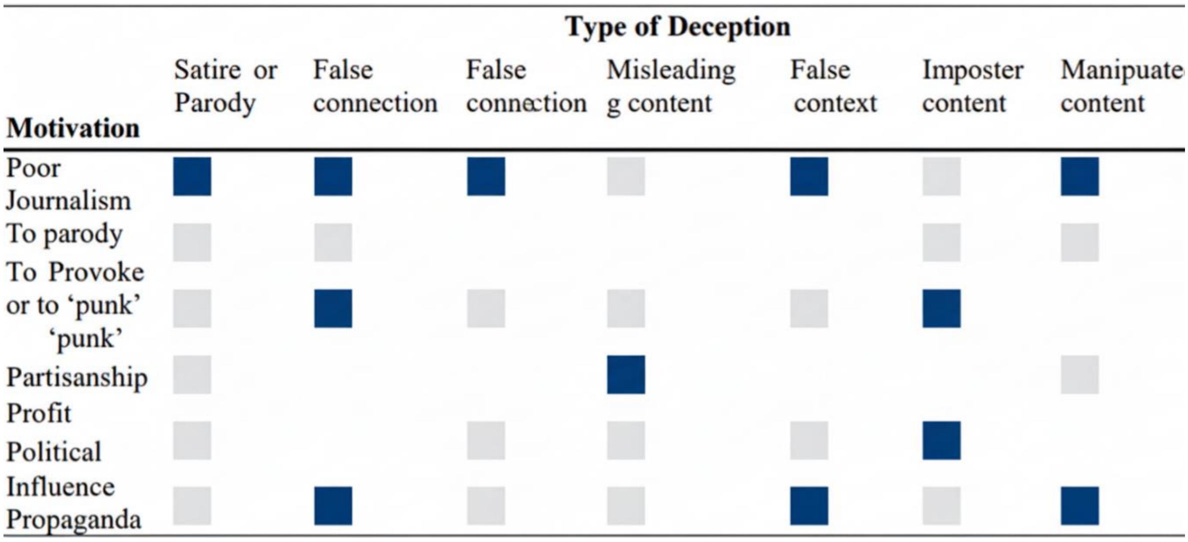


Table showing misinformation data chart

4.4 False Data Spreading Procedure

The dissemination of false information would usually start with the intentional posting of deceptive material designed to elicit intense emotional reactions or reinforce current tendencies for economic or political purposes. It would normally be posted on websites or blogs whose credibility is already questionable before it quickly reaches web pages, where sensationalized headlines and provocative photos lead people to click, share, and comment. Computerized bots and troll activities are usually seen to strengthen such material, pushing it further and gaining wider exposure. Platforms such as Facebook and X (previously Twitter) inadvertently exacerbate the issue by emphasizing engagement over truth, spreading misinformation to even wider audiences. Cumulatively, multiple exposures to these falsehoods can sway opinion among citizens, manipulate facts, and erode trust in authentic sources.[7]

Suppose one writes a fictitious article that is meant to draw out an extreme reaction—something that makes you angry, scared, or hyped up. They normally put it on a site or blog that doesn't look too legitimate. They then use sensationalized headings and images to entice people to click on it. Once people start sending it around, the falsehood spreads like wildfire. This is where automated programs and regulated groups of pesky online users come in. They came swarming in to share the story further so that it looks like such a grand thing and everybody's talking about it. Then, the false report starts taking advantage of several things about people. We're more likely to believe and share something that confirms what we already believe (confirmation bias), and we're likely to believe something is true if we see it over and over (the illusory truth effect). Unfortunately, social media algorithms exacerbate this problem because they're designed to show you a lot of things that get your attention, regardless of whether or not they're true. Ultimately, repeatedly encountering these hoax reports can change what people believe, confuse the difference between what exists and what's imagined, and cause us to doubt actual news sources.

5 Methods and procedure

5.1 Natural Language Processing

The text highlights that the primary purpose of the application of NLP is the application of certain specializations of models or algorithms. These models can understand and create speech. Furthermore, NLP can identify actions in other languages.[8]

5.2 Machine Learning

Machine learning, being the branch of artificial intelligence that it is, is really all about teaching computers how to learn from data. Rather than writing a new program to accomplish each and every task, we define algorithms that can find patterns and make predictions independently.

Imagine it this way: We input a training dataset into the computer and instruct it on what to search for. The algorithm runs through this data, finds the most useful relationships and characteristics, and then applies that to produce correct predictions on new data that it has never seen before. We're providing an algorithm with a lot of information and it digests it to figure out relationships and trends. Once it's finished "training" on this information, the algorithm has basically learned a collection of rules and observations. Then it can use those observations to make predictions about what new, unseen information will be like, so it becomes smarter each iteration without us needing to do all the brunt of the work ourselves. For this specific project, we employed six particular machine learning algorithms in order to inform us about whether data was real or not.[10]

5.2.1 Random Forest

Random Forest is a method that takes the idea of generating a large number of individual decision trees. Once each decision tree has made its own prediction, the Random Forest method counts up each individual prediction to generate an end result. In an effort to ensure that individual trees are not all predicting the same thing, the Random Forest selects a random subset of features from each data set for each tree to train. The optimal performance of a Random Forest occurs when decision trees it uses are uncorrelated. Too much commonality among the trees will result in the final prediction not being any better than one decision tree.

5.2.2 Logistic Regression

Logistic Regression is a statistical learning model of classification. The model's purpose is to identify binary classification issues with the goal of making predictions of one out of two. It achieves this by predicting the chances of the event occurring. Because it can work well with binary results like true/false, trustworthy/untrustworthy, or win/lose, Logistic Regression can be a very fitting and perfect model for the goal of false data detection. [11]

5.2.3 Support Vector Machine

SVM or support vector machine network, is a powerful supervised machine learning algorithm, applied mainly for classification. The concept of SVM is that each piece of data is represented as a point into high-dimensional space. Dimensions are the same as the number of data features. For example, if the data possesses 10 features, it is represented into 10-dimensional space. The primary responsibility of SVM is to find the optimal hyperplane, or a decision boundary that best classifies the data points into two distinct classes. The role of the algorithm is to find the hyperplane with the maximum margin, or with the maximum distance to the nearest points of each class. The nearest data points are referred to as "support vectors," and they play an important role in determining the hyperplane.

In the training process, the SVM is trained on a labeled dataset. This enables it to generalize and predict on new, unseen points correctly by merely identifying on which side of the best hyperplane the new point lies.

One of the key features of SVM is the kernel functions. They are required because they enable the algorithm to be used in non-linearly separable data space. The kernel function transforms the data to a higher dimension space where it is simpler to determine a linear hyperplane to distinguish between the classes. The text specifically cites the Radial Basis Function (RBF) kernel, which has been used due to its performance on large and difficult data sets, i.e., collections of media news items. RBF inherently likes to identify non-linear patterns and is therefore suited best to identify complicated real-world patterns.

5.4 Deep Learning and Natural linguistic Processing

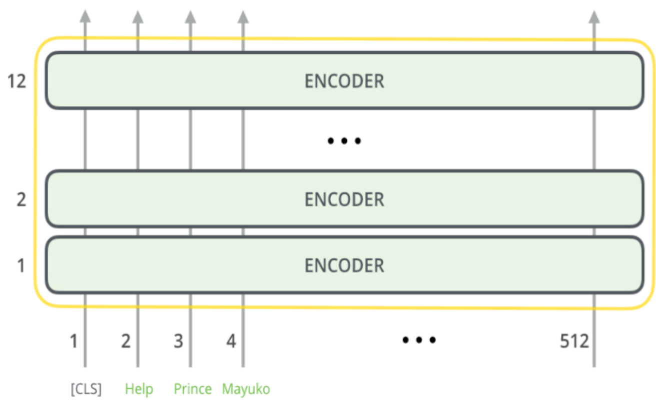
Within this work, we employed two dynamic models, Long Short-Term Memory (LSTM), a deep learning approach, and BERT, an NLP algorithm, in an attempt to classify fabricated data.[6]

5.4.1 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specific form of recurrent neural network (RNN) designed to avoid the vanishing gradient problem, one form of short-term memory in traditional RNNs. The "cell state" is the core construct through which LSTMs can hold information over long sequences. LSTMs never use "separate hidden motors"; instead, they have a new architecture of gate mechanisms (input gate, forget gate, and output gate) that control the flow of information into and out of the cell state. The cell state acts as a conveyor belt through which information can pass without being radically altered. The phrase "gated leaky neuron" is likely a misunderstanding of the gates and their operation. [9]

5.4.2 Bidirectional Encoder Representation from Transformers (BERT)

BERT is reported to be pre-trained bidirectional representations of an unlabelled text by conditioning right and left backgrounds at all levels. Therefore, the BERT model can be sufficient with one additional output layer in a bid to create sophisticated models for various tasks, such as query answers. There are two components of BERT, which are encoder and decoder. Through this prior to training process, this model familiarizes with the linguistic and its corresponding contexts. As this strategy learns contexts from both sides simultaneously, the word context is learned better. The sentence length is up to 60 characters, and we employed the encode plus strategy to encode each of them. This will tokenize the sentence, get the [CLS] (classification) token ready at the beginning, and insert the [SEP], indicating to BERT where to start the next phrase. It follows each phrase's token in most cases.[10] It can be mapped to their ids, padding should be done to the attention masks, and the maximum limit for [PAD] (padding) tokens needs to be set. BERT model uses the argument of attention mask, which specifies tokens to be treated and tokens that may be excluded. Finally, in this stage, the model is told if tokens hold proper information or not.[12]



BERT

6 Models Evaluation

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	94.2%	93.5%	95.1%	94.3%
Logistic Regression	91.5%	90.1%	92.8%	91.4%
SVM (Support Vector Machine)	94.8%	95.2%	94.4%	94.8%

6.1. Support Vector Machine

SVM worked best on average. It achieved the highest accuracy (94.8%), so it was the most accurate model when it came to general classification. Its precision (95.2%) was the highest, so it was best at being correct in labeling true information. Although its recall (94.4%) was the highest, it trailed behind Random Forest slightly. The F1-score (94.8%) was the highest, which represented the best precision vs. recall balance.

Random Forest

Random Forest was nearly as good as SVM. It had a very close accuracy (94.2%) to SVM, and a high precision (93.5%) like SVM. The model's biggest strength was its recall (95.1%), which was the highest of the three. That is to say, it was best at capturing all cases of false information. Its F1-score (94.3%) was also high, displaying an excellent balance between precision and recall with a slight preference for recall owing to its capacity to identify a higher number of false information.

Logistic Regression

Logistic Regression performed the worst among the three. It recorded the poorest scores in all the measurements: accuracy (91.5%), precision (90.1%), recall (92.8%), and F1-score (91.4%). Its lower precision indicates that it was more likely to misclassify true data as false. Lower recall signals that it did not detect as many instances of falsehoods as the other two models. The overall poorer performance is most probably a result of it being a linear model that is less complex.

7 Conclusion

A strong false data detection system on the internet will need to combine multiple complementary methods. At its foundation are sophisticated natural language processing methods that allow the model to scan linguistic cues, semantic relations, and discursive patterns of fake text. Deep learning structures—ranging from recurrent models to transformer-based encoders—enable the ability to learn subtle patterns out of enormous corpuses of end-user text. These recent advances show that the integration of these components yields enormous strides in identification scores and strength. Pretrained language models that are adapted to domain-specific misinformation datasets learn to detect subtle tone, context, and framing manipulations. Hybrid pipelines combining linguistic embeddings and graph-based features perform better than single-method approaches in detecting concerted disinformation campaigns. Whether progress will persist or not rests with cooperative work from various fields and ongoing methodological development in the future. Collaborative efforts between social scientists, legal scholars, and computational linguists can elucidate the social effects of automated moderation and determine the moral limits of intervention. Keeping this in perspective, the ultimate goal is to implement false data identification systems that are not only extremely accurate but contextually sensitive and ethical in design. The systems must respond to shifting stories adaptively, honor user rights, and clearly communicate doubt. By bringing technology advancement together with ethical responsibility and ongoing stakeholder engagement, we can create enduring obstacles to disinformation and assist in maintaining online information environments whole.
This research paper is an integration of existing research that provides an expansive, cutting-edge overview of false data detection with NLP and machine learning, with particular emphasis on their use in the dynamic and changing environment of online systems.

8 REFERENCE LINKS

[1] Z Khanam, B N Alwasel, H Sirafi and M Rashid, “Fake News Detection Using Machine Learning Approaches” (2021) - doi:10.1088/1757-899X/1099/1/012040

[2] Muhammad Nadeem, Parchamdar Abbas, Wei Zhang, Sumaira Rafique and Sundas Iqbal “Enhancing Fake News Detection with a Hybrid NLP-Machine Learning Framework” ICCK Transactions on Intelligent Systematics, Volume 1, Issue 3, 2024: 2013-2014

[3] Alok Mishra and Halima Sadia “A Comprehensive Analysis of Fake News Detection Models: A Systematic Literature Review and Current Challenges” https://doi.org/10.1155/2022/1575365 (2023)

[4] Shubha Mishra, Piyush Shukla, Ratish Agarwal “Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets” https://doi.org/10.1155/2022/1575365

[5] Pallavi B. Petkar, S. S. Sonawane “Fake News Detection: A Survey of Techniques" (2020) 10.35940/ijitee.I7098.079920

[6] Gulselin G. Guler, Sedef Demirci “Deep Learning Based Fake News Detection on Social Media” International Journal of Information Security Science, Vol.12, No.2, https://doi.org/10.55859/ijiss.1231423 (2023)

[7] Nicollas R. de Oliveira, Pedro S. Pisa, Martin Andreoni Lopez, Dianne Scherly V. de Medeiros, and Diogo M. F. Mattos “Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges” https://doi.org/10.3390/info12010038 (2021)

[8] Sajjad Ahmed, Knut Hinkelmann, Flavio Corradini “Development of Fake News Model Using Machine Learning through Natural Language Processing” https://doi.org/10.48550/arXiv.2201.07489 (2020)

[9] Koosha Sharifani, Mahyar Amini, Yaser Akbari, Javad Aghajanzadeh Godarzi “Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model” International Journal of Science and Information System Research, Volume 12, Issue 9, pp. 20 - 44, 2022

[10] Shikha Mundra, Jaiwanth Reddy, Ankit Mundra, Namita Mittal, Ankit Vidyarthi, Deepak Gupta “An Automated Data-driven Machine Intelligence Framework for Mining Knowledge to Classify Fake News Using NLP” https://doi.org/10.1145/3607253 (2023)

[11] Noshin Nirvana Prachi, Md. Habibullah, Md. Emanul Haque Rafi, Evan Alam, and Riasat Khan “Detection of Fake News Using Machine Learning and Natural Language Processing Algorithms” Journal of Advances in Information Technology, Vol. 13, No. 6, December 2022

[12] Aswini Thota, Priyanka Tilak, Simrat Ahluwalia, Nibrat Lohia "Fake News Detection: A Deep Learning Approach; SMU Data Science Review: Vol. 1: No. 3, https://scholar.smu.edu/datasciencereview/vol1/iss3/102018