# DETECTING FAKE NEWS THROUGH LOGESTIC REGRESSION

Alan Joy [1], M G Alwin [2], Prabath A S [3], Sanjay P S [4], Vinay Krishna C S [5], Sana Shafi [6], Ajith P J [7], Bindu Anto [8]

[1] *Alan Joy, Student, Computer Engineering, KKMMPTC Mala, Kerala, India*
[2] *M G Alwin, Student, Computer Engineering, KKMMPTC Mala, Kerala, India*
[3] *Prabath A S, Student, Computer Engineering, KKMMPTC Mala, Kerala, India*
[4] *Sanjay P S, Student, Computer Engineering, KKMMPTC Mala, Kerala, India*
[5] *Vinay Krishna C S, Student, Computer Engineering, KKMMPTC Mala, Kerala, India*
[6] *Sana Shafi, Student, Computer Engineering, KKMMPTC Mala, Kerala, India*
[7] *Ajith P J, HOD Robotic Process Automation, KKMMPTC Mala, Kerala, India*
[8] *Bindu Anto, HOD Computer Engineering, KKMMPTC Mala, Kerala, India*

## ABSTRACT

*When people are exposed to fake news, they might believe false information. This can lead to misguided beliefs and actions based on inaccurate data, affecting their decision-making in areas such as health, politics, and finance. Fake news can shape public opinion and influence elections, policies, and social attitudes. It can sway people's views on important issues and contribute to polarization in society. Fake news often promotes divisive narratives, fostering hostility and division among different groups of people. This can lead to social unrest and conflicts. False information can influence stock markets, consumer behavior, and investment decisions, potentially causing financial losses for individuals and businesses. In the context of health, fake news can be particularly harmful. We use logestic regression algorithm. Logistic regression is a statistical method used for binary classification tasks, where the goal is to predict one of two possible outcomes, typically labeled as 0 and 1. It's a type of regression analysis that models the probability of the binary outcome. Logistic regression is widely used in various fields, including healthcare (predicting disease outcomes), marketing (customer churn prediction), and natural language processing (spam detection), among others. It's a fundamental algorithm in machine learning and serves as a building block for more complex models. This literature review also includes papers discussing algorithms for fake news detection, such as logstic regression, natural language processing and naïve baye Classifier.*

*Keyword : - Fake news, misinformation, logistic regression, natural language processing, naïve baye etc….*

## INTRODUCTION

In today's digital age, the rapid dissemination of information through online platforms has reshaped the way we perceive and interact with the world. However, this newfound connectivity has also given rise to a formidable adversary – fake news. The propagation of false or misleading information, often masquerading as credible news, has

become an alarming societal issue with far-reaching consequences. From influencing elections to undermining public trust in institutions, fake news poses a significant threat to the fabric of our information ecosystem.Detecting and mitigating the spread of fake news has thus become an imperative task in the realms of journalism, technology, and academia.In this paper, we delve into studies and articles about fake news detection and also cover some papers on logistic regression, which can be used as an effective algorithm for differentiating fake and real news.

## MILESTONES

Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., & Reyes Ortiz, J. L authored and published a paper titled "A public domain dataset for human Activity recognition using smartphones". [1] Human-centered computing is a burgeoning research area that seeks to comprehend human behavior and incorporate users and their social surroundings into computer systems. Within this framework, one of the most recent, demanding, and intriguing applications involves utilizing smartphones to sense human body movements, thus gathering contextual information about people's actions. In this study, we introduce an Activity Recognition database produced from recordings of 30 individuals performing Activities of Daily Living (ADL) while carrying a waist-mounted smartphone equipped with inertial sensors. This database has been made publicly available on a widely recognized online repository. Furthermore, we acknowledge the results obtained from applying a multiclass Support Vector Machine (SVM) to the dataset.

On the 13th of November,2019, "Human activity recognition with smartphones" was published by Kaggle. [2] group of 30 volunteers, aged between 19 and 48 years, participated in experiments. Each person engaged in six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) while carrying a Samsung Galaxy S II smartphone on their waist. The smartphone's embedded accelerometer and gyroscope captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. These experiments were video-recorded for manual labeling. The resulting dataset was randomly divided into two sets, with 70% of the volunteers contributing to the training data and the remaining 30% for the test data.

The authors Van der Maaten, L and Hinton, G published a paper titled. "Visualizing data using t-SNE. Journal of machine learning research". [3] We introduce a novel method known as "t-SNE," designed to visualize highdimensional data by assigning a position in a two or three-dimensional map to each data point. This approach is a variation of Stochastic Neighbor Embedding (Hinton and Roweis, 2002), which is more amenable to optimization and yields superior visualizations by mitigating the clustering of points in the map's center. T-SNE outperforms existing techniques in creating a unified map that reveals structure across various scales. This is especially valuable for high-dimensional data distributed across multiple, interconnected low-dimensional patterns, such as images of objects from various classes and view points. To visualize the structure of extensive datasets, t-SNE utilizes random walks on neighborhood graphs, enabling the implicit structure of all data to influence the display of a subset. We demonstrate t-SNE's performance across a wide range of datasets and compare it with numerous other nonparametric visualization methods, including Sammon mapping, Iso map, and Locally Linear Embedding. T-SNE consistently produces superior visualizations compared to these other techniques across nearly all datasets.

In September of 2011, the article "Activity classification using a single wrist-worn Accelerometer" [4] was published by Chernbumroong S, Atkins, A. S and Yu, H. In 2011 5th International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA) Proceedings. The automatic identification of human activity has opened up opportunities for offering personalized services in various fields like healthcare, security, and sports. With advancements in sensor technology, it's now possible to recognize activities discreetly and without intrusion. Sensor placement and wearability are crucial factors for successfully recognizing activities in open living spaces. In experiments, a single wrist-worn accelerometer was used to classify daily living activities, comparing the performance of two classification algorithms, Decision Tree C4.5 and Artificial Neural Network, using different sets of features. The results indicated that Decision Tree C4.5 consistently outperformed the

Neural Network across various feature sets. The highest accuracy of 94.13% was achieved using a feature set that included popular and accurate metrics such as mean, minimum, energy, and sample differences. This demonstrates the potential for automatic activity classification without movement constraints, discomfort, or sensorrelated stigma when using a wrist-worn accelerometer.

The article titled "Fake News Detection Using Naïve Bayes Classifier" [5] was published in 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering. This research paper outlines a straightforward method for identifying fake news by utilizing a naïve Bayes classifier. The approach was transformed into a software system and evaluated using a dataset of Facebook news posts. We attained a classification accuracy of around 74% on the test dataset, a respectable outcome given the model's simplicity. The article also discusses various ways to potentially enhance these results. These findings indicate that artificial intelligence techniques can be applied effectively to tackle the issue of fake news detection.

The authors Shu, Kai, et al published an article titled "Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter" [6] on 2017 Social media's role in news consumption is a double-edged sword. On one hand, it offers cost-effective, easily accessible, and rapidly disseminated news. On the other hand, it serves as a platform for the widespread circulation of "fake news," which consists of low-quality news intentionally containing false information. This proliferation of fake news can have profoundly negative consequences for individuals and society. Consequently, the detection of fake news on social media has recently gained significant attention as an emerging research field Detecting fake news on social media poses distinct challenges not addressed by traditional news media detection algorithms. Firstly, fake news is craftily designed to deceive readers with false information, making it a complex task to identify based solely on content. Hence, it becomes imperative to leverage auxiliary information, such as users' social interactions on social media, to aid in the detection process. Secondly, utilizing this auxiliary information is inherently challenging, given that users' engagements with fake news generate extensive, incomplete, unstructured, and noisy data recognizing the demanding and pertinent nature of fake news detection on social media, we have conducted this survey to facilitate further research on this issue. Our survey provides a comprehensive examination of fake news detection on social media, encompassing characterizations of fake news from psychological and social perspectives, existing algorithms viewed through a data mining lens, assessment metrics, and notable datasets. Furthermore, we delve into related research domains, highlight unresolved challenges, and outline prospective directions for future research in the realm of fake news detection on social media.

In 2017, Ruchansky, Natali, Sungyong Seo, and Yan Liu. CSI: published a paper "A Hybrid Deep Model for Fake News Detection." **[7]** The issue of fake news has garnered attention from both the general public and the academic community. This misinformation has the potential to influence public opinion and can be exploited by malicious actors to manipulate events like elections. Detecting fake news automatically is a crucial but challenging problem that is not yet fully understood. However, there are three commonly accepted aspects of fake news: the content of the article, the reactions it receives from users, and the sources that promote it. Previous efforts have mainly focused on addressing one of these aspects, limiting their effectiveness and applicability. In this study, we introduce a model that combines all three aspects to improve the accuracy of automated fake news detection. Specifically, we consider the behaviors of both users and articles, as well as the collective behavior of users who spread fake news. Building on these three aspects, we present a model named CSI, consisting of three modules: Capture, Score, and Integrate. The first module, based on user responses and article content, employs a Recurrent Neural Network to capture the temporal patterns of user engagement with a given article. The second module learns source characteristics based on user behavior, and these components are integrated within the third module to classify an article as genuine or fake. Experiments conducted on real-world data show that CSI achieves higher accuracy compared to existing models and uncovers meaningful hidden features of both users and articles.

The article named “Distributed representations of sentences and documents International Conference on Machine Learning”[8 ]was published by the authores Le, Quoc, and Tomas Mikolov on 2014. Numerous machine learning algorithms necessitate input to be expressed as a fixed-length feature vector. In the context of textual data, one of the most prevalent approaches is the bag-of-words representation. Despite their widespread use, bag-of-words models exhibit two significant limitations: they disregard word order and semantics. To illustrate, words like “powerful,” “strong,” and “Paris” are treated equally. In this research, we introduce an unsupervised algorithm for acquiring vector representations of sentences and text documents. This algorithm represents each document using a dense vector, which is trained to predict words within the document. This construction grants our algorithm the capability to address the shortcomings of bag-of-words models. Empirical findings demonstrate that our method surpasses both bag-of-words models and other text representation techniques. Ultimately, we achieve new state-of-the-art results in various text classification and sentiment analysis tasks.

Shu, Kai, H. Russell Bernard, and Huan Liu authored and published a paper named “Studying Fake News via Network Analysis: Detection and Mitigation.” [9] arXiv preprint arXiv:1804.10233 on 2018. Social media has gained popularity as a source for news consumption due to its convenience, rapid sharing, and affordability. However, it also facilitates the widespread spread of “fake news,” which contains intentionally false information. The presence of fake news on social media has detrimental societal consequences and presents distinct challenges. To address these challenges, numerous existing studies utilize various network-related features to identify and combat fake news. Essentially, the ecosystem of news distribution on social media encompasses three aspects: the content, the social interactions, and the temporal aspect. In this chapter, we will examine network characteristics relevant to the study of fake news, introduce commonly used network types, and discuss how these networks can aid in the detection and mitigation of fake news on social media.

“Natural-language processing” [10] is an article published by T. Patten and P. Jacobs, in IEEE Expert, vol. 9, no. 1, pp. 35-, Feb. 1994, doi: 10.1109/64.295134. The study of processing natural language like English has long been a central concern within the field of artificial intelligence. This importance stems from language’s fundamental role in human intelligence and its vast range of potential applications. Many successful knowledge representation and inference techniques used in knowledge-based systems originated from research in natural language processing. However, realizing practical language-processing applications has remained a challenge. The special series on natural-language processing aims to spotlight language processing and its practical uses. It does so by showcasing recent techniques applied to real-world problems, identifying research ripe for practical implementation, and highlighting promising combinations of natural-language processing with emerging technologies. Each of the four articles in the series offers insights into the current state of the field and emphasizes the practical significance of recent research.

K. Jiang and X. Lu, are the authors of article titled “Natural Language Processing and Its Applications in Machine Translation: A Diachronic Review” [11] was published in 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI), Chongqing City, China, 2020, pp. 210-214, doi: 10.1109/IICSPI51290.2020.9332458. Natural language processing, an integral component of artificial intelligence technology, draws from various disciplines, including linguistics, computer science, and mathematics. The rapid advancements in this field strongly support the research on machine translation. This paper begins by introducing the fundamental concepts and core content of natural language processing and provides a brief overview of the historical and global progress in NLP research. Subsequently, it outlines the three stages of machine translation and their current research status. Throughout history, the development trajectory of natural language processing closely parallels that of machine translation, with both fields complementing each other. Building on this, the paper examines the applications of natural language processing in machine translation, identifies challenges, and forecasts trends in the domain of natural language processing. Finally, the author explores the relationship between machine translation and human translation in the era of artificial intelligence and envisions the future prospects of machine translation.

The Survey Paper: "Study of Natural Language Processing and its Recent Applications" [12] published by B. D. Shivahare, A. K. Singh, N. Uppal, A. Rizwan, V. S. Vaathsav and S. Suman on 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-5, doi: 10.1109/CISCT55310.2022.10046440 Natural Language Processing (NLP) constitutes a research field focused on the development of computer systems capable of intelligent native language comprehension. The examination of natural languages, like English, has consistently represented a fundamental challenge in research feasibility. This is due to the pivotal role language plays in human intelligence and the multitude of potential applications it offers. NLP encompasses critical functions such as machine translation, speech analysis, automatic summarization, discourse resolution, and speech recognition. Its purpose is to enhance human-computer interaction by providing a more intuitive virtual computer interface. This paper presents a comprehensive introduction to NLP and delves into its significant applications.

The article named "Natural language processing: state of the art, current trends and challenges" [13] published by Diksha Khurana, Aditya Koli, Kiran Khatter & Sukhdev Singh in 14 july 2022 In recent times, there has been a significant surge in interest surrounding natural language processing (NLP), which involves the computational representation and analysis of human language. Its applicability has expanded across a wide range of fields, including but not limited to machine translation, email spam detection, information extraction, summarization, medical applications, and question answering. This paper begins by categorizing NLP into four distinct phases, exploring different NLP levels and components, along with tracing the historical development of NLP. It then delves into an extensive examination of the current state of the art, showcasing the diverse applications of NLP, discussing ongoing trends, and addressing associated challenges. Finally, the paper concludes with a discourse on available datasets, models, and evaluation metrics within the realm of NLP.

## CONCLUSION

Integrating machine learning into fake news detection systems enhances the efficiency, accuracy, and scalability of the process, making it a valuable tool in the fight against misinformation and disinformation. Using logic regression in the system makes it simpler, transparent, and faster, making it suitable for real-time or high throughput fake news detection systems. Fake news detection also benefits from the incorporation of advanced machine learning techniques like natural language processing (NLP) and deep learning, which can handle the intricate patterns and nuances of textual data more effectively.

## REFERENCES

[1]. Anguita, D., Ghio, A., Oneto, L., Parra Perez, X., & Reyes Ortiz, J. L. (2013). "A public domain dataset for human Activity recognition using smartphones".
[2]. Learning, U. C. I. M. (2019, November 13). "Human activity recognition with smartphones".
[3]. Van der Maaten, L., \& Hinton, G. (2008). "Visualizing data using t-SNE. Journal of machine learning research", 9(11).
[4]. Chernbumroong, S., Atkins, A. S., \& Yu, H. (2011, September). "Activity classification using a single wristworn Accelerometer". In 2011 5th International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA) Proceedings (pp. 1-6).
[5]. "Fake News Detection Using Naïve Bayes Classifier"- 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering.
[6]. Shu, Kai, et al. "Fake news detection on social media: A data mining perspective."ACM SIGKDD Explorations Newsletter"19.1 (2017): 22-36. 2.

[7]. Ruchansky, Natali, Sungyong Seo, and Yan Liu."CSI: A Hybrid Deep Model for Fake News Detection." Proceedings Of the 2017.

[8]. Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents" International Conference on Machine Learning. 2014.

[9]. Shu, Kai, H. Russell Bernard, and Huan Liu."Studying Fake News via Network Analysis: Detection and Mitigation." arXiv preprint arXiv:1804.10233 (2018).

[10]. T. Patten and P. Jacobs, "Natural-language processing", in IEEE Expert, vol. 9, no. 1, pp. 35-, Feb. 1994, doi: 10.1109/64.295134.

[11]. K. Jiang and X. Lu, "Natural Language Processing and Its Applications in Machine Translation: A Diachronic Review", 2020 IEEE 3rd International Conference of Safe Production and Informatization (IICSPI), Chongqing City, China, 2020, pp. 210-214, doi: 10.1109/IICSPI51290.2020.9332458.

[12]. B. D. Shivahare, A. K. Singh, N. Uppal, A. Rizwan, V. S. Vaathsav and S. Suman, "Survey Paper: Study of Natural Language Processing and its Recent Applications", 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-5, doi: 10.1109/CISCT55310.2022.10046440.

[13]. "Natural language processing: state of the art, current trends and challenges", Diksha Khurana, Aditya Koli, Kiran Khatter & Sukhdev Singh in 14 july 2022.