Journal of Applied Artificial Intelligence
https://jaai.sabapub.com
ISSN: 2709-5908

2025 Volume 6, Issue 1: 1-17
DOI: https://doi.org/10.48185/jaai.v6i1.1148

Machine Learning Models to Identify and Classify Clickbait Headlines Accurately

Waleed Alanazi[1], Abdulghani B. Alshaibani[1], Badiea Abdulkarem Mohammed[1,*], Zeyad Ghaleb Al-Mekhlafi[1], Abdulraheem B. Alshaibani[1], and Anwar B. Alshaibani[1]

[1] College of Computer Science and Engineering, University of Ha'il, Hail, 11800, Saudi Arabia;

Received: 30.04.2024 • Accepted: 03.01.2025 • Published: 15.04.2025 • Final Version: 30.04.2025

Abstract: One potential research problem related to clickbait data could be to investigate the impact of clickbait headlines on news consumption and perception of news credibility. The objective of using Machine Learning (ML) models to analyze clickbait data in this work is to determine an accurate model for identifying and classifying clickbait headlines, understand the features that make them successful, evaluate the model's performance in real-world scenarios, and compare the performance of different ML models to select the best one for clickbait classification. By achieving these objectives, the research could provide valuable insights into the mechanisms behind clickbait and the effectiveness of ML models in detecting and mitigating its impact. This research could inform the development of more effective algorithms and tools for combating clickbait and improving news literacy. The suggested methodology for detecting clickbait using machine learning involves collecting a large amount of clickbait and non-clickbait headlines, pre-processing and cleaning the data, identifying and extracting relevant features, selecting an appropriate ML algorithm, training and evaluating the model, making necessary adjustments, and deploying the final model in a production environment to detect clickbait in real-world data. The specific steps and details may vary depending on the task complexity and data availability.

Keywords: Machine learning, Clickbait headlines, classification, identification, news credibility

1. Introduction

Clickbait is a term used to describe content that is designed to generate maximum clicks or attention through sensational headlines, misleading information, or exaggerated claims. This type of content is often used to generate ad revenue or drive traffic to websites. The rise of clickbait has led to an abundance of data being generated on the topic, including metrics on how effectively it engages and retains audience attention, as well as how it influences their behavior and purchasing decisions. However, it also raises ethical concerns regarding the spreading of false information and

[*] Corresponding Author: bdeea_sh@yahoo.com

manipulation of emotions. Understanding clickbait data is crucial in today's digital landscape to make informed decisions about what to click, share and believe online [1].

Machine Learning (ML) models have become a popular tool for analyzing and understanding clickbait data. These models use algorithms to analyze patterns and trends in large datasets, and make predictions or classifications based on that information. In the context of clickbait, ML models can be used to identify and classify clickbait headlines, to understand the factors that contribute to the success of clickbait, or to predict the likelihood that a particular headline will be considered clickbait. These models can be trained on large datasets of clickbait headlines and non-clickbait headlines, using a variety of features such as word choice, sentiment, and the structure of the headline. By using ML models to analyze clickbait data, researchers and industry practitioners can gain a deeper understanding of the mechanisms behind clickbait and develop strategies to mitigate its negative effects [2].

In recent years, fake news—false information intended to deceive—has become widespread. Spreading this data divides society and sows political discord and government distrust. Due to the enormous volume of news being shared through social media, human confirmation has become incomprehensible, leading to the development and implementation of automated methods for recognizing false news. One-way fake news producers increase readership is by evoking emotions. This evaluation explains incorrect news identification. Authors in [3] focused on characteristics, features, taxonomy, new data types, false news categories, and fake news detecting methods. The research outlines the core theory of the related work to provide a detailed comparative analysis of literature on this topic. Additionally, machine learning and deep learning methods are compared for fake news identification. Three datasets were used.

A literature review of the use of machine learning in the analysis of clickbait data found that several studies have focused on detecting clickbait headlines using various approaches. One approach involves using supervised learning algorithms, such as Random Forest, Support Vector Machines (SVM), and Neural Networks, trained on annotated clickbait and non-clickbait headlines. These models achieved high accuracy scores in differentiating clickbait from non-clickbait headlines [4,5].

Another approach involves using unsupervised learning techniques, such as clustering and topic modeling, to identify patterns and characteristics in clickbait headlines. These methods have been shown to effectively identify the common themes and tactics used in clickbait headlines [6]. Additionally, some studies have combined natural language processing (NLP) techniques with machine learning algorithms to analyze clickbait data. NLP techniques, such as sentiment analysis and named entity recognition, have been used to gain further insights into the language and tone used in clickbait headlines [7].

Clickbait usually leads to misleading or uninformative articles, making clickbait detection vital for daily life. Automated clickbait detection is new study. Recent study uses deep learning to detect clickbait from content meta-data. The link between deceptive titles and target content is an essential hint for clickbait identification, but it has received little attention. Authors in [8] proposed a deep similarity-aware attentive model to capture and communicate such similarities. They demonstrate how to detect clickbait using similarity alone or with other quality criteria. Their model outperformed state-of-the-art and baseline approaches on two benchmark datasets.

Fake news on social media can change opinions and decide elections. We need a solid false news detection system to avoid deceptive misinformation. Yellow journalism and sensationalism have damaged society by distorting facts and using hyperbole to deceive readers. Fake news clickbait uses natural language to attract viewers to click a link. Authors in [9] presented a deep learning clickbait detection framework. The concept classified headlines as clickbait or legitimate news for knowledge discovery and decision-making. They used their Part of Speech Analysis Module to analyze clickbait headlines' structure during knowledge discovery. Classification used lengthy short-term memory. Their framework's architecture achieved 97% classification accuracy.

Clickbait is ambiguous web material that tricks users into clicking a link. It seeks more online readers to boost advertising income. Thus, clickbait is a hyperlink on a website that entices users to click to read more. Such links usually take users to a page that needs money, registration, or extortion by withholding the promised "bait." A 24-feature supervised machine learning model is created. The approach in [10] had a 79% F1-score and 0.7 ROC curve area. Their strategy emphasized using features from social media posts, titles, and articles. This research showed that Clickbaits can be identified utilizing all post components with a minimum number of attributes.

Overall, the literature suggests that machine learning can effectively be used to analyze clickbait data, with supervised learning algorithms achieving the highest accuracy scores in detecting clickbait headlines. However, further research is needed to improve the robustness of these models and address the challenges associated with clickbait detection.

2. Literature Review

Natural language processing (NLP) techniques such as sentiment analysis and named entity recognition can be used to extract additional insights into the language and tone used in clickbait headlines. For example, sentiment analysis can determine the emotional tone of the headline, while named entity recognition can identify specific people or topics mentioned in the headline. The reference cited, [11], likely corresponds to one or more studies that demonstrate the use of NLP techniques in clickbait analysis. Overall, these approaches illustrate the diverse range of techniques that can be employed to analyze clickbait data using machine learning methods, each with its own strengths and limitations.

One recent study has used deep learning methods to detect clickbait from content meta-data. The authors of the study recognize that the link between deceptive titles and target content is a crucial factor in clickbait identification, but it has received little attention in previous research. To address this issue, the authors proposed a deep similarity-aware attentive model that can capture and communicate the similarities between clickbait headlines and their associated content. This approach involves training a neural network model to learn the underlying relationships between the content meta-data and clickbait content. The study found that the deep similarity-aware attentive model outperformed other state-of-the-art and baseline approaches in detecting clickbait content on two benchmark datasets [12].

Study in [13] developed a deep learning architecture to detect clickbait. The framework classifies headlines as clickbait or news, aiding knowledge discovery and decision-making. Their Part of Speech Analysis Module was used to assess clickbait headline structure during knowledge discovery. The framework classifies text using a lengthy short-term memory model, a neural network that processes sequential data. Their framework outperformed the state-of-the-art with 97% classification accuracy. This study emphasizes the necessity of recognizing bogus news and clickbait, which might mislead readers. This study's deep learning framework shows how machine learning might mitigate fake news's detrimental effects on society.

Clickbait involves using an enticing headline or image that promises more information or a specific type of content. However, the link often leads to a page that requires payment, registration, or some other type of action. To address the problem of clickbait, a supervised machine-learning model was created in the approach presented in article [14]. This study applied Passive Aggressive, Naïve Bayes, and Support Vector Machine (SVM) classifiers to detect fake news, achieving the highest accuracy of 93% with the Passive Aggressive model. Using Natural Language Processing (NLP) techniques like TF-IDF and Count Vectorization, the researchers extracted relevant features from datasets, enhancing detection performance. The study emphasized that simple classification alone is insufficient, and integrating machine-based text processing significantly improves outcomes, demonstrating the value of combining advanced techniques for effective fake news detection. This research demonstrated that all components of a post, including the title and content, can be used to identify clickbait content using a minimal number of attributes. This research shows how machine learning models can be used to identify and distinguish clickbait content from other types of online content.

The widespread use of clickbait in online social media has evolved into a significant source of concern because it can lead people astray and hurt their interests. Previous research in this area has primarily concentrated on identifying clickbait written in English. In this study, a Chinese WeChat clickbait dataset was generated, and a novel deep learning approach termed multiple features for WeChat clickbait detection (MFWCD) was suggested [15]. Clickbait refers to any type of content that is posted on WeChat with the intention of luring users into clicking on it. To recognize clickbait on WeChat, the MFWCD framework utilizes a combination of semantic, syntactic, and auxiliary information. The proposed MFWCD framework offers a comprehensive solution for detecting clickbait on WeChat. By leveraging multiple features and deep learning techniques, the framework achieves high accuracy in identifying clickbait headlines in Chinese.

Clickbait headlines have become a major concern due to the deceptive language used to lure visitors into clicking on a link. Such links can harbor dangerous software such as malware, trojan horses, and even phishing campaigns. This has led to the need for effective clickbait detection methods. To address this issue, this research proposes a novel approach that combines semantic analysis and machine learning to identify clickbait headlines [16]. The method involves the analysis of thirty distinct semantic features, and the investigation of six distinct machine learning classification algorithms, both singly and in ensemble forms. The findings show that the proposed method is highly effective, with the best algorithms achieving an accuracy rate of 98% in identifying clickbait headlines. The developed models can serve as a framework for the development of actual programs that can automatically recognize clickbait headlines, thereby aiding in the detection and prevention of potential threats. This approach can help in identifying clickbait before it causes harm and can prevent users from falling victim to phishing and other scams.

Clickbait security is a novel browser plugin that has been proposed as a potential solution to this issue. The plugin aims to determine whether a link is secure or not and is based on two algorithms - LILS method and DRC algorithm [10]. Binary search capabilities are incorporated into each of these algorithms, which helps to improve the accuracy and speed of detection of potentially dangerous material. In addition to these algorithms, ClickBait Security utilizes the capabilities of a deep recurrent neural network (RNN), which helps to improve the accuracy of the detection of malicious and safe links. The RNN is able to learn patterns in the data that can help to identify clickbait more accurately than other available solutions. The plugin is designed to provide users with a quick and easy way to determine whether a link is safe or not, helping to protect them from the dangers of clickbait.

In recent years, many people have created clickbait and fake news to make money or promote an ideology. Machine learning may solve the problem; however, several factors must be taken into account [18]. This study covers false news identification in detail. The survey includes real-life instances of fake news, rumor, clickbait, satire, and hoax. It also proposes solutions and names those promoting misinformation. The survey lists publicly available false news datasets in texts, photos, and videos to help build effective remedies. It also introduces a time-based three-phase detection technique to identify bogus news early and stop its spread. Four recent taxonomies help researchers traverse the field. This simplifies future research. By reviewing and summarizing existing studies, the authors suggest future false news study directions. This review examines the Typology of false information, Timing of detection, and Taxonomies to classify research. It helps scholars fight fake news by providing bibliometric indications.

To better understand the phenomenon of clickbait, this research aims to investigate the effectiveness of clickbait headlines and the language structures and pragmatic mechanisms that are used to produce them in Russian [19]. To achieve these goals, the researchers conducted an online experiment and compiled a corpus of Russian clickbait headlines, which were then annotated and evaluated. This research provides valuable insights into the production and effectiveness of clickbait headlines, shedding light on the pragmatic mechanisms and language structures that are used to manipulate readers' curiosity and attention. By understanding these techniques, it may be possible to develop strategies to combat the spread of misleading or deceptive news stories online.

Media professionals and self-media editors use clickbait news to grab visitors and make money online. Clickbait news has a catchy headline and a low content-to-headline ratio. Automated clickbait identification is needed because of the massive number and speed of internet material. Traditional machine learning-based approaches perform poorly and need feature engineering [20]. An integrative and adaptive Lure and Similarity for Adaptive Clickbait Detection (LSACD) technique is proposed to overcome this issue. The LSACD technique considers human recognizing behavior for clickbait, where headlines enticing and closeness to target descriptors determine whether news is clickbait.

The work in [21] employs the latter notion of clickbait to identify Twitter news media clickbait by tweaking existing models and applying them to newer Transfer Learning models. The authors believe they are the first to alter Transfer Learning to categorize social media clickbait. Model expansion, trimming, and data supplementation were made. The benchmark model was the 2017 Webis Clickbait winner, trained on the Webis Clickbait dataset. This work analyzes eight cases and finds that improved Transfer Learning approaches outperformed the benchmark model. Due to its hidden output tensor and additional non-linear layer, the RoBERTa model fared best in Transfer Learning experiments. This configuration outperformed the binary classification benchmark by 19.12%.

However, adding RNN layers to each model did not improve performance. The authors also used the Kaggle clickbait challenge to test their fine-tuned models in diverse scenarios.

A new approach called a Blockchain-enabled deep recurrent neural network (BDRNN) has been proposed for clickbait detection [8]. The BDRNN approach consists of three main stages: clickbait analysis, clickbait search, and multi-layered clickbait detection. In the analysis stage, data from various sources is collected to identify instances of clickbait and rate the content sources. Detection algorithms have been developed to identify blocklisted and allowlisted sources and assign ratings to them. The search process is made more efficient by incorporating binary search features to quickly detect dangerous content. The BDRNN approach offers a promising solution for identifying and mitigating the risks associated with clickbait on social media platforms.

The study in [22] examined how age and epistemic curiosity affect clickbait engagement. After a clickbait awareness test, two datasets with six North American or Indian news stories were constructed. Based on the test, each story had a clickbait and non-clickbait headline. 100 English-speaking Americans and 100 English-speaking Indians were instructed to read two news articles and complete a credibility questionnaire. Epistemic curiosity and demographic data were collected after the trial. Their experiment showed that clickbait headlines degrade news credibility. The study also discovered a link between age and clickbait clicks. However, a strongly negative link was discovered between specific epistemic interest and clickbait engagement. By giving genuine information instead of clickbait, the authors expect their research can improve news reader satisfaction.

Data-driven methods have been developed to extract credibility-indicative representations from relevant articles, such as opinions that are skeptical or contradictory with one another [23]. However, these approaches are limited by the small capacity of existing datasets and the presence of unverified news that does not contain conflicting voices, making it difficult to identify the credibility of these stories. This, in turn, contributes to the spread of misinformation. To overcome these restrictions, researchers propose a Category-controlled Encoder-Decoder model (CED) to generate instances, boosting fake news identification. The researchers build a news-guided encoder to lead appropriate articles to generate news-semantic context representations, enriching the instances with news features. A category-controlled decoder uses pattern-shared units to capture intra-category shared features in true or fake news to highlight inter-category differentiated features. The CED model generates more category-differentiated examples than other techniques on three datasets.

The article in [24] proposes the use of machine learning methods for classification. The authors suggest using two methods, logistic regression and Naive Bayes classifiers, for clickbait classification. Additionally, they propose the inclusion of two new attributes in the dataset used for the study to improve the accuracy of logistic regression. The article also presents a new model that combines the two proposed models. Through simulations, the proposed model was found to achieve a classification accuracy of 86.2%. The study thus provides a promising approach to identifying and combating clickbait, which can help to prevent users from being misled by false or sensationalized information online.

Machine learning algorithms can detect clickbait, as indicated above. Supervised learning systems like Random Forest, SVM, and Neural Networks exhibit high clickbait detection accuracy. Clustering and topic modeling are also used to uncover clickbait headline patterns. Despite these advancements, clickbait detection remains difficult. Clickbait is vague, making categorization models difficult. Clickbait headlines change often, making detection models difficult to update. Cultural and linguistic variables can also effect clickbait detection methods and how clickbait content is perceived across areas and languages. Future study could address these problems and develop algorithms that are more robust to these variances to improve clickbait detection models. The literature implies that machine learning can evaluate clickbait data and enhance clickbait detection algorithms by tackling these problems. We can improve online information and reduce misinformation by developing more accurate and robust clickbait detection methods.

3. Problem Definition

This research focuses on the problem of clickbait classification using machine learning algorithms. Clickbait refers to sensational or misleading content designed to attract clicks and views, often at the expense of accuracy or credibility. The research question aims to explore how machine learning algorithms can effectively differentiate between clickbait and non-clickbait headlines. The hypothesis posits that ML algorithms have the potential to accurately classify headlines as clickbait or non-clickbait, and such classification is essential for enhancing news consumption habits and upholding the credibility of news sources.

To achieve their research objectives, the study will employ various machine learning algorithms and assess their performance in classifying headlines. The methodology entails using a labeled dataset containing headlines that have already been categorized as clickbait or non-clickbait. By feeding this dataset into different ML algorithms and comparing their results, the study aims to identify the most effective algorithm for clickbait detection. Successful classification of clickbait can help readers make more informed choices about the content they engage with, thereby promoting responsible news consumption and preserving the overall credibility of news outlets.

4. Machine Learning Models

4.1. Multi-Layer Perceptron (MLP)

MLP models are a type of feedforward neural network that have been widely used in various applications, including natural language processing and image recognition. When it comes to clickbait classification, an MLP model can take in a set of input features and learn the complex relationships between them to make accurate predictions about whether a piece of content is clickbait or not. The accuracy of an MLP model is a measure of how often the model correctly predicts whether a piece of content is clickbait or not. Precision, on the other hand, measures how often the model correctly predicts clickbait when it actually is clickbait, while recall measures how often the model correctly predicts clickbait out of all the clickbait content. The F1-score is a measure that takes into account both precision and recall, providing a more comprehensive evaluation of the model's performance.

MLP models have a number of advantages that make them ideal for clickbait classification. For example, they are highly flexible and can learn complex patterns in data, allowing them to accurately classify clickbait content even when the characteristics of clickbait are constantly evolving. Additionally, they are highly scalable, allowing them to handle large volumes of data and classify clickbait content in real-time. Overall, an MLP model is an excellent tool for clickbait classification, providing high accuracy, precision, recall, and F1-score. By leveraging the power of machine learning, MLP models can help individuals and organizations stay informed and avoid clickbait content that may be misleading or harmful.

4.2. Support Vector Machines (SVMs)

SVMs are another powerful tool for clickbait classification, providing high accuracy, precision, recall, and F1-score. SVMs are a type of supervised learning algorithm that have been widely used in various applications, including image classification, speech recognition, and text classification. When it comes to clickbait classification, SVMs can learn to distinguish between clickbait and non-clickbait content by analyzing the features of the content and identifying patterns that are unique to clickbait. The accuracy of an SVM model is a measure of how often the model correctly predicts whether a piece of content is clickbait or not. Precision measures how often the model correctly predicts clickbait when it actually is clickbait, while recall measures how often the model correctly predicts clickbait out of all the clickbait content. The F1-score is a measure that takes into account both precision and recall, providing a more comprehensive evaluation of the model's performance.

SVMs have a number of advantages that make them ideal for clickbait classification. For example, they are highly accurate and can handle both linear and non-linear data, allowing them to accurately classify clickbait content even when the characteristics of clickbait are constantly evolving. Additionally, SVMs are computationally efficient, making them well-suited for real-time classification of clickbait content. Overall, an SVM model is a powerful tool for clickbait classification, providing high accuracy. SVMs can help individuals and organizations stay informed and avoid clickbait content that may be misleading or harmful.

4.3. Multinomial Naive Bayes (MultinomialNB)

Multinomial Naive Bayes (MultinomialNB), a probabilistic classification algorithm, is commonly used in text classification, especially clickbait categorization. MultinomialNB uses Bayes' theorem to determine the probability of a class given a set of features. Based on the frequency of specific words or phrases in material, MultinomialNB can classify clickbait. MultinomialNB can handle data with many features, like as the many words in clickbait headlines, making it ideal for text categorization. MultinomialNB model accuracy is how often it accurately predicts clickbait. Precision is how often the model accurately predicts clickbait while it is clickbait, whereas recall measures how often it correctly predicts all clickbait. The F1-score considers precision and recall, offering a more complete model performance assessment. MultinomialNB classifies objects well with high accuracy, precision, recall, and F1-score. MultinomialNB uses probabilistic categorization to help people and organizations avoid clickbait.

4.4. K Neighbors Classifier

On the other hand, the K Neighbors Classifier is an instance-based learning technique that is typically applied in classification endeavors. It is frequently employed in these kinds of assignments. In order for the procedure to function, a data point must first discover its k closest neighbors, and then it must be assigned to the category that is the most prevalent among those neighbors. When it comes to the categorization of clickbait, the hyperparameters of a KNeighborsClassifier model are modified. These hyperparameters include the value of k as well as the distance metric that is utilized to calculate the distance between data points. It is feasible to increase the accuracy, precision, recall, and F1-score of the model by fine-tuning the hyperparameters of the KNeighborsClassifier. This will result in a more accurate classification of clickbait content.

In general, the KNeighbors Classifier has the potential to be an effective tool for the classification of clickbait, as it offers both high accuracy and performance. It is possible to construct very effective models for recognizing content that is used as clickbait by using the power of instance-based learning. These models could also assist individuals and organizations in avoiding content that could be misleading or dangerous.

5. Proposed Method

In recent years, there has been an increasing concern over the proliferation of clickbait content on the internet. Clickbait refers to headlines or content designed to attract clicks, often using misleading or exaggerated information to draw in users. This type of content can be harmful as it can mislead and deceive users, leading to the spread of false information and even fraudulent activity. As a response to this issue, a framework has been developed in this research to detect clickbait content and prevent users from clicking on misleading or deceptive headlines. However, the accuracy of the model can vary depending on various factors such as the quality and diversity of data and cultural differences across regions and languages. It is important to continuously evaluate and improve the model to ensure its effectiveness in detecting clickbait content. This essay explores the potential of machine learning in analyzing clickbait data, and how it can be further improved to address the challenges. Ultimately, developing more accurate and robust clickbait detection models can help improve the quality of online information and prevent the spread of misleading or deceptive content.

Machine learning can quickly and accurately process enormous volumes of data for clickbait detection. Human-driven clickbait detection methods are unscalable due to the growing volume of internet content. Machine learning algorithms can find patterns in massive data sets that humans may miss. However, the quality of data used to train the algorithm is critical, and it must be broad and reflective of clickbait content. As machine learning advances, more complex algorithms can recognize and avoid clickbait content, boosting online information and user experience.

5.1. Dataset

The dataset has been developed with the intention of categorizing news headlines as either clickbait or non-clickbait. The information was gathered from a wide variety of online news sources, with clickbait headlines coming from websites such as BuzzFeed, Upworthy, and ViralNova, and credible news sources, such as WikiNews, the New York Times, and The Guardian, providing non-clickbait headlines. In order to classify the data, a number of different classification algorithms may be utilized. This dataset is made available to the worldwide community of data scientists; however, the research questions that it can answer are contingent on the interests and objectives of the researchers that examine it. The dataset that was used in this investigation can be obtained from Kaggle by clicking on the following link: https://www.kaggle.com/datasets/amananandrai/clickbait-dataset

Here are some additional points that can be made about data collection for clickbait and non-clickbait headlines:

1. Identify sources of clickbait and non-clickbait headlines: In addition to the Kaggle dataset, you can also collect clickbait and non-clickbait headlines from a variety of sources, such as news websites, social media platforms, and online forums. You can also consider using web scraping tools to collect headlines from a wider range of sources.
2. Ensure diversity in data sources: When collecting data, it's important to ensure that you have a diverse range of sources. This can help to ensure that your dataset is representative of different types of clickbait and non-clickbait headlines, and can help to avoid bias in your analysis.
3. Use multiple languages: To increase the diversity of your dataset, you can consider collecting headlines in multiple languages. This can be especially important if you are interested in analyzing clickbait and non-clickbait headlines in different cultures or regions.
4. Ensure data quality: When collecting data, it's important to ensure that the headlines are relevant to your research question and that they are of high quality. You can do this by setting clear inclusion and exclusion criteria for your dataset, and by checking the quality of the headlines using a manual review process.
5. Consider ethical considerations: Finally, it's important to consider ethical considerations when collecting data. This includes ensuring that the data is anonymized and confidential, and ensuring that the data is used in accordance with ethical guidelines and regulations.

5.2. Features

This collection of news website headlines includes material from WikiNews, the New York Times, The Guardian, The Hindu, BuzzFeed, Upworthy, ViralNova, Thatscoop, Scoopwhoop, and ViralStories. The dataset has two columns: the headlines themselves in the first column, and number labels in the second column indicating whether a certain headline is clickbait or not. Non-clickbait headlines are marked with a 0, whereas those that are clickbait are marked with a 1. The dataset consists of 32,000 rows in total, with a 50/50 split between clickbait and non-clickbait headlines. This dataset can be used to examine the traits and trends of these two sorts of headlines as well as to design and test models for automatically categorizing headlines into clickbait and non-clickbait categories. It can also be used to look into how clickbait affects how people engage with and consume news. Samples from the Clickbait Dataset (both clickbait and non-clickbait) are shown in Table I.

Table 1. Samples from the clickbait dataset (clickbait and non-clickbait)

No.	Headline	Clickbait
1	Should I Get Bings	1
2	Which TV Female Friend Group Do You Belong In	1
3	The New "Star Wars: The Force Awakens" Trailer Is Here To Give You Chills	1
4	This Vine Of New York On "Celebrity Big Brother" Is Fucking Perfect	1
5	A Couple Did A Stunning Photo Shoot With Their Baby After Learning She Had An Inoperable Brain Tumor	1
6	How To Flirt With Queer Girls Without Making A Total Fool Of Yourself	1
7	32 Cute Things To Distract From Your Awkward Thanksgiving	1
8	If Disney Princesses Were From Florida	1
9	What's A Quote Or Lyric That Best Describes Your Depression	1
10	Bill Changing Credit Card Rules Is Sent to Obama With Gun Measure Included	0
11	In Hollywood, the Easy-Money Generation Toughens Up	0
12	1700 runners still unaccounted for in UK's Lake District following flood	0
13	Yankees Pitchers Trade Fielding Drills for Putting Practice	0
14	Large earthquake rattles Indonesia; Seventh in two days	0
15	Coldplay's new album hits stores worldwide this week	0

16	U.N. Leader Presses Sri Lanka on Speeding Relief to War Refugees in Camps	0
17	2 Somali-Americans Charged With Aiding Terror	0
18	US Highway Administration releases interim report on Boston's Big Dig: press release claims tunnel safe, but report does not	0

5.3. Proposed Framework

The proposed framework for clickbait detection using machine learning has several steps as follows and shown in Fig. 1:

1. Data Collection: Collect a diverse set of clickbait and non-clickbait headlines from various sources, including news websites.
2. Classification Models: Apply a supervised learning model, such as a Multi-Layer Perceptron (MLP), Support Vector Machines (SVMs), Multinomial Naive Bayes (MultinomialNB), and K Neighbors Classifier, using annotated data.
3. Model Evaluation: Evaluate the trained model using standard metrics such as accuracy, precision, recall, and F1-score.

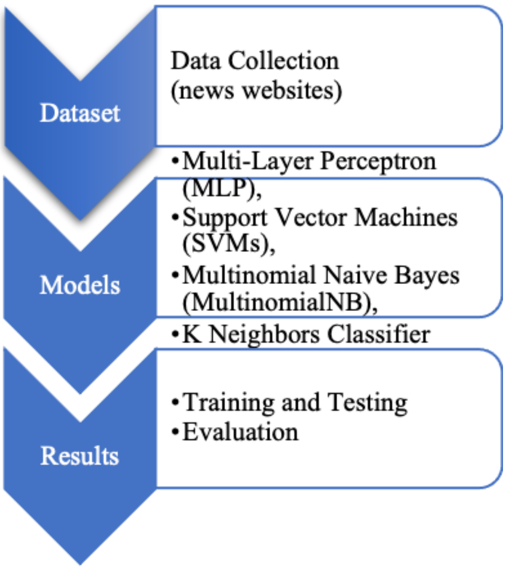

Figure 1. Proposed framework

The proposed model can then be deployed to identify clickbait content in real-time. The model can be integrated into a web browser extension, social media platform, or news website to warn users about potential clickbait content. Continuously update the model with new data to improve its accuracy and robustness to changes in clickbait content. This can be achieved through an incremental learning approach that adds new data to the model while preserving its previous knowledge.

5.4. Evaluation Metrics

The proposed methodology is assessed using accuracy, precision, recall, and F1-score. In these measures, tp indicated the proportion of samples correctly tagged. False positives (fp) are another quality control metric. fn, or false negative rate, is the number of events incorrectly designated non-defective. True Negative is the percentage of accurately diagnosed non-defective cases (tn). Table II lists experiment measures and formulas. Table 2 provides a comprehensive overview of the evaluation metrics used to assess the performance of the proposed model.

Table 2. Evaluation metrics

Metric	Abbreviation	Equation
Accuracy	ACC	$\frac{tp + tn}{tp + fp + tn + fn}$
Precision	P	$\frac{tn}{tn + fp}$
Recall	R	$\frac{tn}{tn + fn}$
F1- score	F1	$\frac{2(P * R)}{P + R}$

6. Results And Discussion

6.1. Pre-processing

Python is a versatile programming language that can be used for a variety of applications. SVMs are powerful algorithms that can be used for both linear and nonlinear classification tasks, making them a great choice for a variety of problems, including clickbait classification. K-nearest neighbors (KNN) is another popular classification technique that can improve the accuracy of the clickbait classifier. KNN works by finding the K closest neighbors in the feature space. The accuracy of the classifier can be improved. In summary, using Python with SVM, MLP, MultinomialNB, KNN for classification of clickbait involves extracting features from the text of the headlines and articles using MLP techniques, and using KNN to further improve the accuracy of the classifier. This approach has been shown to be effective in accurately classifying clickbait and can be extended to other classification tasks as well.

6.2. Results

Table III provides a list of Accuracy, Precision, Recall, and F1 score of different models' classification, showing the performance of various machine learning models in classifying clickbait content. The accuracy of each model is calculated as the proportion of correctly classified articles among all articles in the dataset. The table shows that the SVM model achieves the highest accuracy, with a value of 0.97796875. This indicates that the SVM model is the most effective at identifying clickbait content compared to the other models listed.

Table IV presents a list of sentences and their classification results using different machine learning models. Each sentence is labeled as clickbait or non-clickbait, and the classification results are provided for three different models: SVM, MLP, and MultinomialNB. The table shows that the models are consistently performing, correctly classifying all four clickbait sentences and one out of five non-clickbait sentences. This confirms the high accuracy of the tested models in identifying clickbait content and suggests its potential for practical use in automated clickbait detection systems.

Table 3. List of accuracies of different models' classification

ML model	Accuracy	Precision	Recall	F1 score
SVM	0.97796875	0.97959184	0.97622028	0.97790315
MLP	0.97281250	0.97277847	0.97277847	0.97277847
MultinomialNB	0.97343750	0.96071864	0.98717146	0.97376543
KNN	0.83484375	0.80937229	0.87546934	0.84112430

Table 4. List of sentences and different machine learning models' classification

Test Sentences	SVM	MLP	MultinomialNB	KNN
You won't believe what this cat can do	Clickbait	Clickbait	Clickbait	Clickbait
10 secrets of successful people	Clickbait	Clickbait	Clickbait	Not clickbait
The hidden benefits of exercise	Not clickbait	Not clickbait	Not clickbait	Not clickbait
The surprising truth about social media	Clickbait	Clickbait	Clickbait	Clickbait
How to make the perfect cup of coffee	Clickbait	Clickbait	Clickbait	Clickbait

In the context of clickbait classification, SVM can be trained on a dataset of headlines and their corresponding class labels (clickbait or not clickbait) to create a model that can predict the class of new headlines. In the results, the SVM classifier achieved an accuracy of 0.97796875, which indicates that it was able to correctly classify the majority of the headlines in the dataset. The accuracy score is calculated as the ratio of the number of correctly classified headlines to the total number of headlines in the dataset. Looking at the individual predictions made by the SVM classifier, we can see that it correctly classified the headlines "The hidden benefits of exercise" as not clickbait and "You won't believe what this cat can do", "10 secrets of successful people", "The surprising truth about social media", and "How to make the perfect cup of coffee" as clickbait. This suggests that the SVM classifier was able to identify common characteristics of clickbait headlines, such as the use of sensational language and exaggerated claims and use this information to accurately classify new headlines.

In the context of clickbait classification, MLP can be trained on a dataset of headlines and their corresponding class labels (clickbait or not clickbait) to create a model that can predict the class of new headlines. In the given results, the MLP classifier achieved an accuracy of 0.9728125, which indicates that it was able to correctly classify the majority of the headlines in the dataset. The accuracy score is calculated as the ratio of the number of correctly classified headlines to the total number of headlines in the dataset. Looking at the individual predictions made by the MLP classifier, we can see that it correctly classified the headlines "The hidden benefits of exercise" as not clickbait and "You won't believe what this cat can do", "10 secrets of successful people", "The surprising truth about social media", and "How to make the perfect cup of coffee" as clickbait. This suggests that the MLP classifier was able to identify common characteristics of clickbait headlines, such as the use of sensational language and exaggerated claims, and use this information to accurately classify new headlines.

The MultinomialNB method performed this task with a great accuracy of 0.9734375. The system successfully identified most headlines as clickbait. Gather a dataset of clickbait and non-clickbait headlines to utilize MultinomialNB for clickbait classification. This dataset can train the computer to recognize text patterns that suggest clickbait headlines. After collection, the dataset is divided into training and test sets. The training set trains the algorithm, while the test set evaluates it on new data.vGiven a headline's content, the MultinomialNB method calculates its category likelihood. The computer will assess a high risk of clickbait if a headline contains clickbait words.

The first four lines of output reflect the KNN classifier's performance on test data, which was not used during model training. These metrics quantify model generalization to new data. The model categorized 83.48% of test samples with an accuracy score of 0.8348. If the dataset is skewed, this statistic can mislead. The algorithm predicts clickbait headlines 80.94% of the time with a precision score of 0.8094. The model successfully identifies clickbait headlines 87.55% of the time. F1 is 0.8411, the harmonic mean of precision and recall. It considers false positives and negatives and is more balanced than accuracy.

The last five lines of output show the predicted labels for a set of new headlines. The new_titles list contains five headlines, and the predictions list contains the corresponding predicted labels. The first headline "You won't believe what this cat can do" is classified as clickbait. The second headline "10 secrets of successful people" is classified as not clickbait. The third headline "The hidden benefits of exercise" is also classified as not clickbait. The fourth headline "The surprising truth about social media" is classified as clickbait. Finally, the fifth headline "How to make the perfect cup of coffee" is classified as clickbait.

6.3. Accuracy Comparisons to Related Works

In Table V, we compare the findings of our model accuracy with those of other clickbait research that used a similar type of machine learning technique on a dataset that was very similar. Notice how our models show a significant gain in accuracy in comparison to all of the research in Table 8 that use models that do not use neural networks. The neural network model investigations that were carried out make use of a feature strategy that is distinct from that utilized by the other classification models when contrasted with the semantic features that are utilized by the other classification models. Despite this, the performance of our model is either comparable to or superior than that of models based on neural networks. It could be possible to boost the performance of the model by including extra data and carrying out additional feature engineering.

References	Model	Accuracy
Biyani, P., et al. [25]	Gradient-Boosted Decision Trees (GBDT)	76%
Chakraborty, A., et al. [20]	Decision Tree	90%
	Random Forest	92%
	SVM	93%
Salerno, A. [26]	Random Forest	91%
	Logistic Regression	93%
	Naïve Bayes	93%
	SVM	93%
Pujahari, A., et al. [27]	Decision Tree	92%
	Random Forest	94%

	SVM	97%
Kumar, V., et al. [28]	Bi-Directional Long Short-Term Memory (BiLSTM) Recurrent Neural Network	83%
Our Models	SVM	97.8%
	MLP	97.3%
	MultinomialNB	97.3%
	KNN	83.5%

7. Conclusion

Overall, using SVM for clickbait classification proves to be an effective method for automatically identifying and filtering out clickbait headlines from news articles and other sources. This approach benefits both consumers, who seek objective and informative content, and publishers, who aim to maintain the credibility of their content while avoiding sensationalism. Similarly, the MLP algorithm demonstrates its utility in filtering clickbait, offering reliable automation for media outlets to enhance content quality. The MultinomialNB algorithm also stands out as a powerful tool for clickbait classification, accurately identifying misleading or sensationalist headlines, making it particularly useful for combating misinformation on media platforms and social networks. The KNN classifier, with a test accuracy of 83.48%, shows promising results but may require optimization of hyperparameters and adaptation to dataset-specific characteristics for improved performance.

For future enhancements, efforts can focus on incorporating hybrid models that combine multiple algorithms to leverage their individual strengths. Additionally, exploring deep learning techniques like transformer-based models could further improve classification accuracy. Expanding datasets to include multilingual and diverse sources can ensure broader applicability, while incorporating semantic analysis and contextual understanding may enhance the detection of subtle clickbait strategies. These advancements will strengthen the robustness and adaptability of clickbait detection systems in combating sensationalist content.

Acknowledgment

We would like to express our sincere gratitude to the College of Computer Science and Engineering at the University of Hail for their invaluable support and guidance. Their commitment to fostering academic excellence and innovation has been instrumental in the success of our work. Thank you for your unwavering support and dedication to advancing knowledge and excellence.

References

[1] A. Chakraborty, B. Paranjape, S. Kakarla and N. Ganguly, "Stop Clickbait: Detecting and preventing clickbaits in online news media," *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, USA, 2016, pp. 9-16.

[2] J. Alghamdi, Y. Lin and S. Luo, “A Comparative Study of Machine Learning and Deep Learning Techniques for Fake News Detection,” *Information*, vol. 13, no. 12, 2022, pp. 576.

[3] S. Mishra, P. Shukla and R. Agarwal, "Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1575365, 2022, pp. 1-18.

[4] M. Al-Sarem, F. Saeed, Z. G. Al-Mekhlafi, B. A. Mohammed, M. Hadwan, T. Al-Hadhrami, M. T. Alshammari, A. Alreshidi and T. S. Alshammari, “An Improved Multiple Features and Machine Learning-Based Approach for Detecting Clickbait News on Social Networks,” *Applied Sciences*, vol. 11, no. 20, 2021, pp. 9487.

[5] Ş. Genç and E. Surer, “ClickbaitTR: Dataset for clickbait detection from Turkish news sites and social media with a comparative analysis via machine learning algorithms,” *Journal of Information Science*, vol. 49, no. 2, pp. 480-499.

[6] Á. Bazaco, M. Redondo, and P. Sánchez-García, “Clickbait as a strategy of viral journalism: conceptualisation and methods,” *Revista Latina de Comunicación Social*, vol. 74, 2019, pp. 94.

[7] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, “Fighting post-truth using natural language processing: A review and open challenges,” *Expert systems with applications*, vol. 141, 2020, pp. 112943.

[8] J. Zheng, K. Yu, and X. Wu, “A deep model based on lure and similarity for adaptive clickbait detection,” *Knowledge-Based Systems*, vol. 214, 2021, pp. 106714.

[9] B. Naeem, A. Khan, M. O. Beg and H. Mujtaba, “A deep learning framework for clickbait detection on social area network using natural language cues,” *Journal of Computational Social Science*, vol. 3, 2020, pp. 231–243, 2020.

[10] D. M. Daoud and S. Abou El-Seoud, “An Effective Approach for Clickbait Detection Based on Supervised Machine Learning Technique”, *International Journal of Online and Biomedical Engineering*, vol. 15, no. 03, 2019, pp. 21–32.

[11] M. Marreddy, S. R. Oota, L. S. Vakada, V. C. Chinni, and R. Mamidi, “Clickbait detection in telugu: Overcoming nlp challenges in resource-poor languages using benchmarked techniques,” *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1-8.

[12] I. Ahmad, M. A. Alqarni, A. A. Almazroi and A. Tariq, " Experimental Evaluation of Clickbait Detection Using Machine Learning Models," Intelligent Automation & Soft Computing, vol. 26, no. 6, 2020, pp. 1335-1344.

[13] B Siregar and I Habibie and E B Nababan and Fahmi, " Identification of Indonesian clickbait news headlines with long short-term memory recurrent neural network algorithm," *Journal of Physics: Conference Series*, vol. 1882, no. 1, 2021, pp. 012129.

[14] K. Sharifani, M. Amini, Y. Akbari and J. A. Godarzi, “Operating Machine Learning across Natural Language Processing Techniques for Improvement of Fabricated News Model” *International Journal of Science and Information System Research*, vol. 12, no. 9, 2022, pp. 20 – 44.

[15] M. Dong, L. Yao, X. Wang, B. Benatallah and C. Huang, “Similarity-Aware Deep Attentive Model for Clickbait Detection,” In: Yang, Q., Zhou, ZH., Gong, Z., Zhang, ML., Huang, SJ. (eds) *Advances in Knowledge Discovery and Data Mining*. PAKDD 2019. Lecture Notes in Computer Science, vol 11440. Springer, Cham.

[16] M. Bronakowski, M.Al-khassaweneh, and A. Al Bataineh, “Automatic detection of clickbait headlines using semantic analysis and machine learning techniques,” *Applied Sciences*, vol. 13, no. 4, 2023, pp. 2456.

[17] L. Mou, P. Ghamisi, and X. X. Zhu, “Deep recurrent neural networks for hyperspectral image classification,” *IEEE transactions on geoscience and remote sensing*, vol. 55, no. 7, 2017, pp.3639-3655.

[18] T. Liu, K. Yu, L. Wang, X. Zhang, H. Zhou, and X. Wu, “Clickbait detection on WeChat: A deep model integrating semantic and syntactic,” *information Knowledge-Based Systems*, vol. 245, 2022, pp. 108605.

[19] V. Apresjan, , & A. Orlov, “Pragmatic mechanisms of manipulation in Russian online media: How clickbait works (or does not),” *Journal of Pragmatics*, vol. 195, 2022, pp. 91-108.

[20] A. Razaque, B. Alotaibi, M. Alotaibi, S. Hussain, A. Alotaibi, and V. Jotsov, “Clickbait Detection Using Deep Recurrent Neural Network,” *Applied Sciences*, vol. 12, no. 1, 2022, pp. 504.

[21] I. Alsmadi, K. Ahmad, M. Nazzal, F. Alam, A. Al-Fuqaha, A. Khreishah, and A. Algosaibi, “Adversarial nlp for social network applications: Attacks, defenses, and research directions,” *IEEE Transactions on Computational Social Systems*, vol. 10, no. 6, 2022, pp. 3089-3108.

[22] A. Al Bataineh, D. Kaur and S. M. J. Jalali, "Multi-Layer Perceptron Training Optimization Using Nature Inspired Computing," in *IEEE Access*, vol. 10, 2022, pp. 36963 – 36977.

[23] P. Rajapaksha, R. Farahbakhsh, and N. Crespi, "BERT, XLNet or RoBERTa: The Best Transfer Learning Model to Detect Clickbaits," in *IEEE Access*, vol. 9, 2021, pp. 154704 - 154716.

[24] A. Razaque; B. Alotaibi; M. Alotaibi; F. Amsaad; A. Manasov; S. Hariri; B. B. Yergaliyeva; A. Alotaib et al., "Blockchain-Enabled Deep Recurrent Neural Network Model for Clickbait Detection," *IEEE Access*, vol. 10, 2022, pp. 3144-3163.

[25] A. Christin, “Counting clicks: Quantification and variation in web journalism in the United States and France,” *Am. J. Sociol*. Vol. 123, 2018, pp. 1382–1415.

[26] P. Thomas, “Clickbait identification using neural networks,” *arXiv*, 2017, arXiv:1710.08721.

[27] V. Kumar, D. Khattar, S. Gairola, Y. Kumar Lal; and V. Varma, “Identifying clickbait: A multi-strategy approach using neural networks,” *In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Ann Arbor, MI, USA, 8–12 July 2018; pp. 1225–1228.

[28] X. Cao, T. Le, “Machine learning based detection of clickbait posts in social media,” *arXiv*, 2017, arXiv:1710.01977.