Comparative Study of Clickbait News Prediction Using Machine Learning

Fathul Qorib1,a), Eka Yuniar 2,b), Frida Kusumastuti3,c), Tutut Ismi Wahidar4,d), and Intan Mustafa5,e)

1Department of Communication, University of Tribhuwana Tunggadewi, Malang, Indonesia

2International Marketing Management, Jember State Polytechnic, Jember, Indonesia

3Department of Communication, University of Muhammadiyah Malang, Malang, Indonesia

4Department of Communication, Süieyman Demirel Üniversitesi, Türkiye

5Department of Communication, Nusa Nipa University, Sikka, Indonesia

a) Corresponding author: [fathul.qorib@unitri.ac.id](mailto:fathul.qorib@unitri.ac.id)

b) [sayayuniar@gmail.com](mailto:sayayuniar@gmail.com)

c) [frida@umm.ac.id](mailto:frida@umm.ac.id)

d) [d2340244002@ogr.sdu.edu.tr](mailto:d2340244002@ogr.sdu.edu.tr)

e) [intanmustafa15@gmail.com](mailto:intanmustafa15@gmail.com)

Abstract. News is a source of information that contains facts and current events happening around society. With valid and useful news, it can enhance the public's knowledge and awareness of the events around them, ensuring they are well-informed. However, among the circulating news, there are also news articles that can cause misunderstandings for readers due to a discrepancy between the headline and the content. These are known as clickbait news, which lure users to read but provide information that deviates from the headline. To address this misunderstanding, a system must check whether the news is valid or clickbait. This study involves developing a machine learning system using Decision Tree, Support Vector Machine, and Naïve Bayes algorithms. With this model, it is expected to assist users in identifying whether the news is clickbait or not. After testing the models, the results showed that the Decision Tree algorithm achieved an accuracy of 90.04%, the Support Vector Machine attained an accuracy of 96.14%, and the Naïve Bayes achieved an accuracy of 96.26%.

Keywords: Clickbait, Comparative studt, Machine Learning, News Predictions;

# **Introduction**

News is a report about recent events or incidents considered important or interesting to the public [1]. It can cover various topics such as politics, economy, sports, culture, technology, and more. The process of presenting news involves gathering information through interviews, observations, and research, which is then processed and presented objectively, accurately, and clearly. Mass media, both print and digital, play a crucial role in disseminating news to the public [2]. News must adhere to journalistic ethics, including principles of honesty, accuracy, and freedom from conflicts of interest. Good news not only informs but also educates, influences public opinion, and provides broader context on an issue [3]. However, with the advancement of information and communication technology, clickbait news has emerged, potentially causing misunderstandings in society.

Clickbait news is a type of news content designed to attract readers' attention and encourage them to click on links, often with sensational or misleading headlines. The primary goal of clickbait is to increase web traffic and generate ad revenue rather than to deliver meaningful or useful information [4]. Clickbait headlines often use emotional, dramatic, or provocative words to arouse readers' curiosity, but the article content often fails to live up to the headline's promise, providing shallow and superficial information [5]. This practice can undermine the credibility of the media that publish it and disappoint readers who feel deceived. Additionally, clickbait can disrupt user experience and divert attention from more important and meaningful news. Therefore, it is essential to be able to discern the news consumed to obtain useful information.

Artificial Intelligence (AI) and Machine Learning (ML) are integral components driving modern technological progress. AI, within computer science, focuses on creating systems capable of human-like tasks like speech recognition and decision-making [6]. ML, a subset of AI, develops algorithms enabling computers to learn from data for making predictions or decisions [7]. ML techniques encompass supervised, unsupervised, and reinforcement learning, allowing systems to identify data patterns and enhance performance without explicit programming [8]. A decision tree is a ML method for classification and regression, representing decisions and potential outcomes in a branching structure [9]. Internal nodes denote dataset features, branches signify decisions based on these features, and leaves indicate outcomes or class labels.

Decision tree construction begins by selecting attributes to split data into homogeneous subsets, typically using metrics like entropy or information gain for classification, and mean squared error for regression. Support Vector Machine (SVM), another ML approach for classification and regression, excels in finding optimal hyperplanes that separate data classes with maximal margin [10]. SVM seeks a hyperplane in multidimensional space that maximizes the distance between closest data points from different classes, termed support vectors. Naive Bayes, a probabilistic ML technique for classification, is noted for its efficiency in managing large datasets due to its assumption—though often oversimplified—that each dataset feature independently influences class probabilities based on Bayes' Theorem [11], despite real-world interdependencies among features.

This studies makes a speciality of the use of Artificial Intelligence, especially Machine Learning, to categorise and expect clickbait information articles primarily based totally on their headlines. The aim is to increase a version that allows customers distinguish clickbait from proper information, thereby mitigating the effect of clickbait [12] [13]. The system gaining knowledge of version on this observe will appoint 3 algorithms: Decision Tree, Support Vector Machine (SVM), and Naïve Bayes. The Decision Tree set of rules is selected for its interpretability, permitting builders to apprehend the intent in the back of the version`s decisions. It handles each specific and numerical information properly and may seize complex relationships among functions withinside the dataset. SVM is chosen for its cappotential to locate the ideal hyperplane in complicated characteristic spaces, making sure clean separation among clickbait and non-clickbait information with a maximal margin, even in instances wherein information isn't linearly separable. Naïve Bayes is applied because of its effectiveness in leveraging conditional possibilities of found functions. Despite its assumption of characteristic independence (which simplifies computations), Naïve Bayes gives green information class and meets the requirement for truthful and quick version interpretation.

Aufar et al. (2020) carried out a examine specializing in sentiment evaluation of feedback on YouTube the usage of Decision Tree and Random Forest algorithms [14]. Their studies aimed to streamline the identity of high quality and poor feedback in the YouTube context. The examine applied a 70% schooling and 30% trying out facts split, revealing that the Decision Tree set of rules carried out barely better accuracy in comparison to Random Forest. Specifically, the Decision Tree set of rules carried out an accuracy of 89.4%, while Random Forest carried out 88.2%. In a separate examine via way of means of Shah et al. (2020) [15], Logistic Regression, Random Forest, and K-Nearest Neighbors algorithms have been in comparison for textual content category purposes. The number one goal turned into to become aware of the handiest set of rules for this task. According to their findings, Logistic Regression emerged because the pinnacle performer with an outstanding category accuracy of 97% after trying out.

# **Methods**

## **DATASET**

In this study, a version of the system study can be developed to expect and classify clickbait information primarily based on headlines. The versioning procedure requires a usable dataset of 32,000 datasets. This study will use a public dataset from Kaggle.com, with a total of 1000 datasets containing clickbait information headlines from various sources including WikiNews, New York Times, The Guardian, The Hindu, BuzzFeed, Upworthy, ViralNova, Thatscoop, Scoopwhoop, and ViralStories. The pattern of this dataset is evident in FIGURE 1.

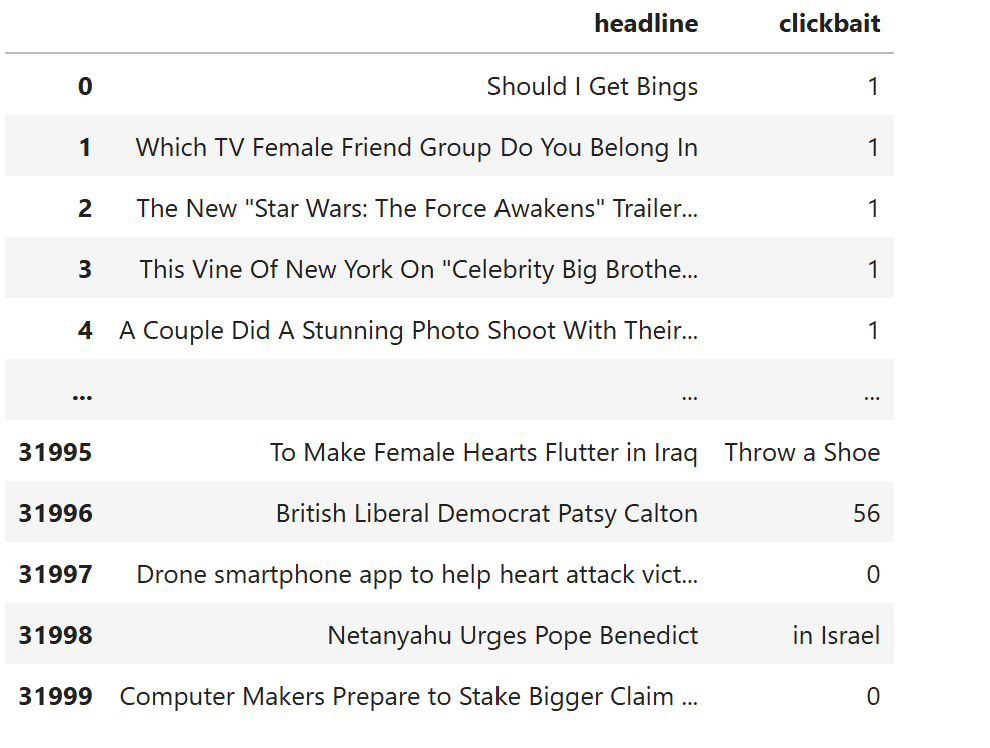


Figure 1. Dataset visualization

**FIGURE 1** indicates the dataset used on this studies to construct the system gaining knowledge of version for clickbait information prediction. As visible in **FIGURE 1**, the information used withinside the version improvement manner includes columns: the "text" column, which includes the information headlines, and the "label" column, which shows whether or not the headline is clickbait or not. The dataset used consists of a complete of 32,000 entries saved in a .csv file, comprising 16,001 valid (non-clickbait) information headlines and 15,999 clickbait information headlines.

## **TF-IDF VECTORIZER**

In growing a device mastering version that leverages Natural Language Processing (NLP), changing textual content facts into numerical or vector representations is vital for allowing device mastering algorithms to manner and examine the facts. This transformation helps the identity and extraction of styles from textual content-primarily based totally facts. TF-IDF (Term Frequency-Inverse Document Frequency) is a method used to assign weights to phrases in a file through giving better values to key phrases applicable to every category [16].

The preprocessing steps concerned on this manner consist of sentence splitting, lowercasing, tokenization, filtering, and stemming. TF-IDF calculates phrase weights primarily based totally on their frequency and significance in a file, with the TF-IDF Vectorizer computing those weights and changing them into TF-IDF values. The goal of changing facts right into a vector area version is to permit the usage of device mastering algorithms on textual content facts [17]. Term Frequency (TF) measures how frequently a phrase seems in a file, at the same time as Inverse Document Frequency (IDF) assesses the importance of the phrase throughout all documents [18]. The mathematical formulation for calculating TF-IDF may be discovered in Eq 1, 2, and 3. Result for TF-IDF Vectorizer is 32.000 dataset with 24.679 data token **FIGURE 2**.

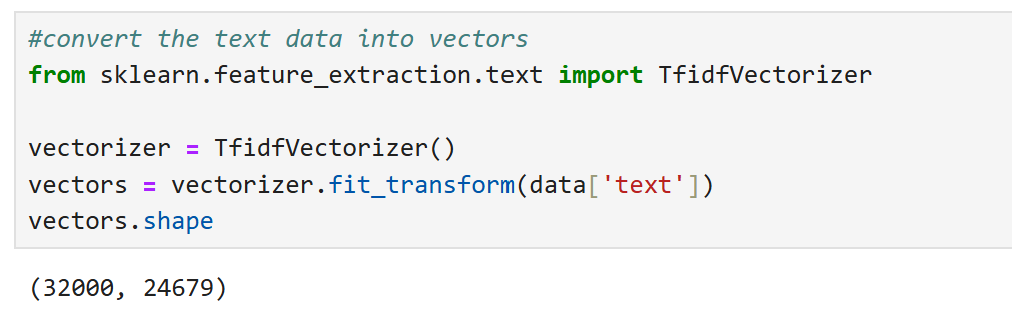


Figure 2. Dataset tf-idf vectorizer

|  |  |
| --- | --- |
|  | (1) |

|  |  |
| --- | --- |
|  | (2) |

|  |  |
| --- | --- |
|  | (3) |

## **DECISION TREE**

The choice tree is a gadget getting to know set of rules used for each class and regression tasks. This set of rules conceptually fashions choices and their outcomes throughout numerous situations in a branching tree structure. Each inner node withinside the tree represents an characteristic or function from the dataset, the branches constitute choices primarily based totally on that characteristic, and the leaf nodes constitute the very last magnificence labels or outcomes [19]. Constructing a choice tree begins offevolved through choosing the simplest characteristic to cut up the records into greater homogeneous subsets, usually the use of metrics together with facts advantage or the Gini index to decide the most advantageous cut up [20].

This records department is achieved recursively till all records inside subsets have the equal label or meet preventing situations like most tree intensity or minimal samples in line with leaf. For visualization, talk to **FIGURE 3** for a illustration of a choice tree.

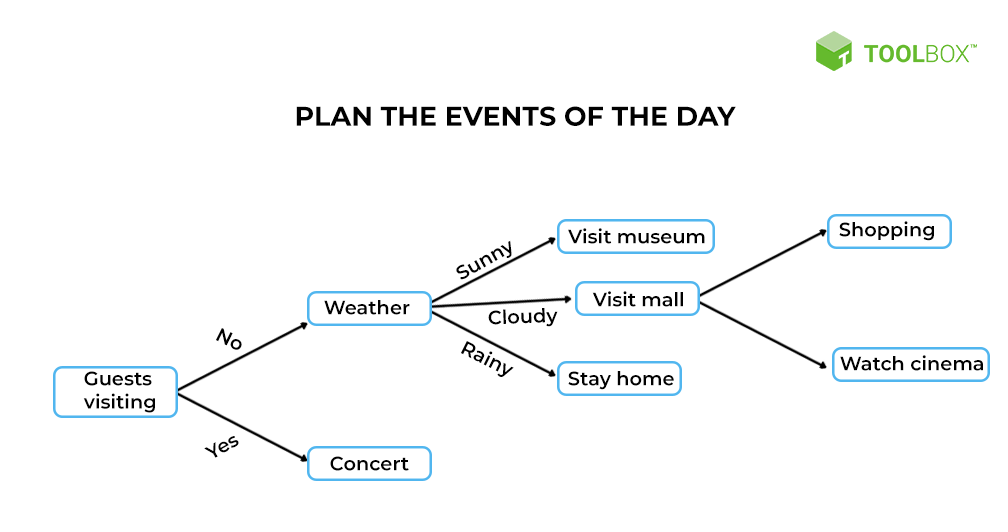


Figure 3. Decision tree process

The foremost gain of selection bushes is their excessive interpretability; the ensuing version is simple to apprehend and visualize [21]. This makes them appropriate for contexts wherein version selections want to be really defined to stakeholders. Additionally, selection bushes can cope with each express and numerical records with out requiring complicated records normalization or transformation. However, their weak point lies of their tendency to overfit, in particular if the tree grows too deep and captures noise withinside the schooling records. To deal with this, strategies together with pruning or the use of ensemble techniques like Random Forest may be applied. To higher apprehend the type manner of a selection tree, the pseudocode is supplied below.

(1)

(2)

(3)

(4) Select node X which has the highest information gain

(5) Split node into sub node

## **SUPPORT VECTOR MACHINE**

Support Vector Machine (SVM) is a regularly used device gaining knowledge of method for each type and regression tasks, identified for its cappotential to discover an surest hyperplane that separates statistics factors into wonderful lessons with the most important margin [11] [22]. SVM does this through finding a hyperplane in a multidimensional area that maximizes the space among the closest statistics factors from distinctive lessons, referred to as guide vectors [16] [23]. When statistics can't be linearly separated, SVM makes use of kernel strategies to map the statistics into higher-dimensional areas wherein linear separation will become possible. Common kernels consist of linear, polynomial, radial foundation function (RBF), and sigmoid [24]. For a visible illustration of the SVM process, please confer with **FIGURE 4**.

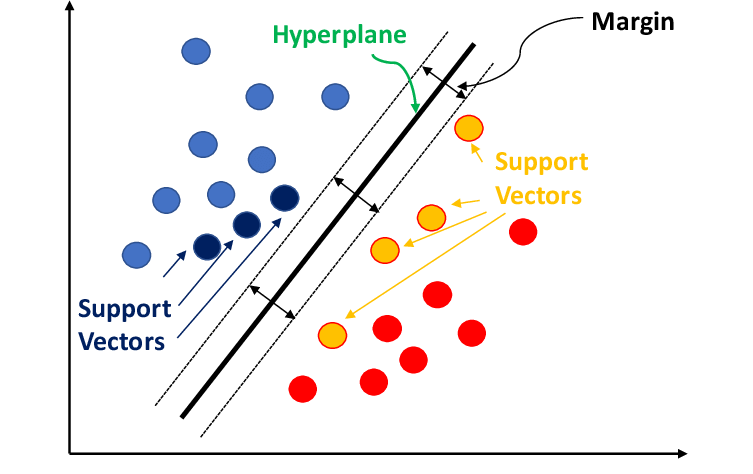


Figure 4. SVM process visualization

One of the number one blessings of SVM is its cappotential to efficaciously function in excessive-dimensional areas whilst last green even if the range of functions exceeds the range of samples [25]. Additionally, SVM well-knownshows excellent generalization abilties and may deal with imbalanced datasets via way of means of adjusting the hyperplane thru regularization parameters. However, SVM might also additionally emerge as much less powerful with very big datasets because of excessive computational complexity and may require tremendous parameter tuning to acquire ultimate performance.

## **NAÏVE BAYES**

Naive Bayes is a probabilistic gadget getting to know technique used for type, recognised for its simplicity and performance in dealing with massive datasets. This set of rules is primarily based totally on Bayes` Theorem with the "naive" assumption that every characteristic withinside the dataset contributes independently to the possibility of a specific magnificence, even though in reality, those capabilities can be interdependent [26]. The type technique with Naive Bayes entails calculating the posterior possibility of a category primarily based totally at the earlier possibility of that magnificence and the chance of the present capabilities [27]. The major strengths of Naive Bayes encompass its pace and capacity to address datasets with a massive wide variety of capabilities, in addition to its first rate overall performance even if the independence assumption amongst capabilities is regularly now no longer met. Moreover, Naive Bayes is notably powerful for textual content type troubles together with unsolicited mail filtering, sentiment analysis, and record categorization. However, this technique can be much less correct if capabilities are notably correlated, as its underlying independence assumption will become invalid. The mathematical components for Naive Bayes is furnished in Eq 4.

|  |  |
| --- | --- |
| = \* \* … \* \* P(z) | (4) |

Point 4 shows the mathematical formula of the Naive Bayes algorithm. The formula depicts the calculation of Naive Bayes with several independent features, where Y = (Y\_1, Y\_2, …, Y\_n) represents the features used, including within class z.

## **CONFUSION MATRIX**

In growing a system studying version for predicting clickbait news, comparing its overall performance is crucial. The confusion matrix is a essential device for this assessment, which includes True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values [28]. These values permit for the computation of key overall performance metrics along with precision, bear in mind, and f1-rating. Precision measures the accuracy of the version in figuring out fantastic cases, bear in mind suggests the version`s capacity to successfully perceive all fantastic cases, and the f1-rating is the harmonic imply of bear in mind and precision [29]. Detailed formulation for calculating precision, bear in mind, and f1-rating are supplied in Eq 5, 6, and 7.

|  |  |
| --- | --- |
|  | (5) |

|  |  |
| --- | --- |
|  | (6) |

|  |  |
| --- | --- |
|  | (7) |

## **PROPOSED METHOD**

In this study, a machine learning model will be developed to predict clickbait news [30]. The flow of the proposed method for developing the model in this research is illustrated in **FIGURE 5**.

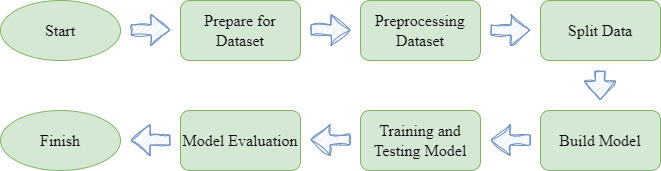


Figure 5. Workflow Process

**FIGURE 4** illustrates the workflow of growing a system mastering version able to predicting clickbait news. The first step on this technique entails making ready the dataset, which on this look at is received from the internet site kaggle.com as a public dataset. Following dataset preparation, the following step is information preprocessing. In this research, information preprocessing entails casting off prevent phrases or much less beneficial phrases in building a sentence and changing textual content information into numerical vectors for processing the use of system mastering algorithms, because the algorithmic version constructed can simplest understand styles primarily based totally on numeric information [31].

After preprocessing the information, the following step is dividing the information into education information to educate the version to understand styles and checking out information to assess the overall performance of the built version. In this look at, the information could be cut up into 70% education information and 30% checking out information. Subsequently, after information processing, the following step is to construct the system mastering version. The parameters used for the Decision Tree, SVM, and Naïve Bayes fashions are furnished in **TABLE 1**. Once the version creation technique is completed, education and checking out of the constructed version may be accomplished the use of the preprocessed education and checking out information. Consequently, after checking out the version, the overall performance of the version may be evaluated the use of the confusion matrix.

Table . Machine Learning Model Parameter

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Decision Tree** | **Support Vector Machine** | **Naïve Bayes** |
| Parameter | Criterion is 'entropy'  Splitter is 'best'  depth(max) is None  samples split(min) is 2  samples leaf(min) is 1  weight fraction leaf(min) is 0.0  features(max) is None  random state is 42  leaf nodes(max) is None  impurity decrease(min) is 0.0  class weight is None  ccp alpha is 0.0 | C is 1.0  Kernel is 'sigmoid'  Degree is 3  Gamma is 1.0  coef0 is 0.0  shrink is True  prob is False  tolerance is 0.001  weight per class is 'balanced'  verbose is False  max iteration is -1  shape for decision function is 'ovr'  random state is None | Alpha is 1.0  Binarizing is 0.0  fit prior is True  prior for class is None |

# **Results and Discussion**

In this study, a gadget gaining knowledge of version is evolved to are expecting and classify clickbait information headlines. The version improvement might be performed the use of the Python programming language and the Jupyter Notebook IDE for scripting. After dataset preparation, the following step entails getting rid of forestall phrases or beside the point phrases to streamline the education and trying out processes, decreasing the wide variety of processed phrases from the e-mail textual content. Following the elimination of forestall phrases, the textual content records is converted into vector shape. This step is critical as gadget gaining knowledge of fashions can best technique numerical records for education and trying out, necessitating the conversion to vectors to permit sample analysis. Subsequently, the dataset is cut up into 70% for education and 30% for trying out. This department lets in the version to study styles from the education records, and version overall performance is evaluated the use of the trying out records to evaluate class capabilities. After trying out, the effects are represented withinside the shape of a confusion matrix, which may be discovered in **FIGURE 5**.

|  |  |
| --- | --- |
| (a) | (b) |
| (c) | |

Figure 6. Confusion matrix decision tree (a), svm (b) and naive bayes (c)

**FIGURE 6** shows the results of the confusion matrix testing using the decision tree, support vector machine (SVM), and Naïve Bayes algorithms. It can be seen from **FIGURE 5** that each model achieved predominantly correct predictions compared to incorrect ones, indicating that the built models are predictive and proficient in recognizing patterns well, thus effectively distinguishing between clickbait and non-clickbait headlines. Based on the testing results with the decision tree, there were 4333 correct predictions and 467 incorrect predictions for the non-clickbait class, while for the clickbait class, there were 4311 correct predictions and 489 incorrect predictions.

For SVM, there were 4667 correct predictions and 133 incorrect predictions for the non-clickbait class, and for the clickbait class, there were 4562 correct predictions and 238 incorrect predictions. In the case of Naïve Bayes, there were 4572 correct predictions and 228 incorrect predictions for the non-clickbait class, and for the clickbait class, there were 4469 correct predictions and 131 incorrect predictions. Based on these results, it can be observed that the Naïve Bayes model had the least incorrect predictions compared to the Decision Tree and SVM algorithms. After the model testing process, precision, recall, and F1-score metrics can be calculated based on the confusion matrix values obtained earlier. The results for precision, recall, and F1-score after testing are provided in **TABLE 2**.

Table . Testing result

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1-Score** | **Accuracy** |
| Decision Tree | 90% | 90% | 90% | 90.04% |
| Support Vector Machine | 96% | 96% | 96% | 96.14% |
| Naïve Bayes | 96% | 96% | 96% | 96.26% |

**TABLE 2** shows the assessment metrics acquired after the checking out manner at the fashions. From the checking out metric consequences, it is able to be located that the Support Vector Machine (SVM) and Naïve Bayes fashions constructed completed the identical Precision, Recall, and F1-Score values of 96%, which might be better as compared to the Precision, Recall, and F1-Score values of the Decision Tree version constructed, which have been 90%. This suggests that the SVM and Naïve Bayes fashions built on this take a look at have properly accuracy in predicting data, carry out nicely in successfully predicting all instructions from each issue or goal withinside the dataset, and still have properly harmonic or balanced values among precision and recall. However, the SVM and Naïve Bayes fashions constructed have mild variations in checking out accuracy. Specifically, the SVM version completed a checking out accuracy of 96.14% in predicting clickbait news, even as the Naïve Bayes version completed a checking out accuracy of 96.26%. These consequences imply that the Naïve Bayes version built plays barely higher as compared to the SVM set of rules constructed, yielding extra correct prediction version precision.

# **Conclusion**

In this study, a system learning model was developed to categorize and predict clickbait titles using Decision Tree, Support Vector Classification, and Naïve Bayes algorithms. Model testing on clickbait prediction found an accuracy of 90.04% for Decision Tree, 96.14% for Support Vector Classification, and 96.26% for Naïve Bayes. Based on these findings, the Naïve Bayes version, which is included with TF-IDF Vectorizer, validated the best overall performance and was the most accurate compared to Decision Tree and Support Vector Machine modes. This shows that the Naïve Bayes version developed in this test excels in predicting clickbait titles with precision. So a good algorithm model using Support Vector Classification For future study directions, it is recommended to explore deep learning models and opportunities for textual information processing strategies such as word2vec to compare and examine unique models for their effectiveness in clickbait prediction.

# **Acknowledgments**

The authors would like to extend their gratitude to the Department of Communication at the University of Tribhuwana Tunggadewi and the Department of International Marketing Management at Jember State Polytechnic for their support in this research. We also appreciate the Department of Communication at the University of Muhammadiyah Malang and the Department of Communication at Süieyman Demirel Üniversitesi for their invaluable contributions and collaboration. This work would not have been possible without the dedication and efforts of all parties involved.

# **References**

1. Boukes, M. and R. Vliegenthart, *A general pattern in the construction of economic newsworthiness? Analyzing news factors in popular, quality, regional, and financial newspapers.* Journalism, 2020. **21**(2): p. 279-300.

2. Anwar, A., et al., *Role of mass media and public health communications in the COVID-19 pandemic.* Cureus, 2020. **12**(9).

3. Cavaliere, P., *From journalistic ethics to fact-checking practices: defining the standards of content governance in the fight against disinformation.* Journal of Media Law, 2020. **12**(2): p. 133-165.

4. Munger, K., *All the News That’s Fit to Click: The Economics of Clickbait Media.* Political Communication, 2020. **37**(3): p. 376-397.

5. Munger, K., et al., *The (Null) Effects of Clickbait Headlines on Polarization, Trust, and Learning.* Public Opinion Quarterly, 2020. **84**(1): p. 49-73.

6. Sarker, I.H., *AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems.* SN Computer Science, 2022. **3**(2): p. 158.

7. Taye, M.M. *Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions*. Computers, 2023. **12**, DOI: 10.3390/computers12050091.

8. Sarker, I.H., *Machine Learning: Algorithms, Real-World Applications and Research Directions.* SN Computer Science, 2021. **2**(3): p. 160.

9. Charbuty, B. and A. Abdulazeez, *Classification based on decision tree algorithm for machine learning.* Journal of Applied Science and Technology Trends, 2021. **2**(01): p. 20-28.

10. Bansal, M., A. Goyal, and A. Choudhary, *A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning.* Decision Analytics Journal, 2022. **3**: p. 100071.

11. Sheth, V., U. Tripathi, and A. Sharma, *A Comparative Analysis of Machine Learning Algorithms for Classification Purpose.* Procedia Computer Science, 2022. **215**: p. 422-431.

12. Bazaco, Á., M. Redondo, and P. Sánchez-García, *Clickbait as a strategy of viral journalism: conceptualisation and methods.* Revista Latina de Comunicación Social, 2019(74): p. 94.

13. Sun, Y. and C. Liang, *How Headline Strategies and News Sources Affect Number of User Clicks on WeChat.* Journalism Practice, 2023. **17**(8): p. 1696-1721.

14. Aufar, M., R. Andreswari, and D. Pramesti. *Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study*. in *2020 International Conference on Data Science and Its Applications (ICoDSA)*. 2020.

15. Shah, K., et al., *A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification.* Augmented Human Research, 2020. **5**(1): p. 12.

16. Rizwan, A., et al. *WR-SVM Model Based on the Margin Radius Approach for Solving the Minimum Enclosing Ball Problem in Support Vector Machine Classification*. Applied Sciences, 2021. **11**, DOI: 10.3390/app11104657.

17. Liu, H., X. Chen, and X. Liu, *A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis.* IEEE Access, 2022. **10**: p. 32280-32289.

18. Jiang, Z., et al., *Text Classification Using Novel Term Weighting Scheme‐Based Improved TF‐IDF for Internet Media Reports.* Mathematical Problems in Engineering, 2021. **2021**(1): p. 6619088.

19. Priyanka and D. Kumar, *Decision tree classifier: a detailed survey.* International Journal of Information and Decision Sciences, 2020. **12**(3): p. 246-269.

20. Mishra, S., et al. *Performance Evaluation of a Proposed Machine Learning Model for Chronic Disease Datasets Using an Integrated Attribute Evaluator and an Improved Decision Tree Classifier*. Applied Sciences, 2020. **10**, DOI: 10.3390/app10228137.

21. Streeb, D., et al., *Task-Based Visual Interactive Modeling: Decision Trees and Rule-Based Classifiers.* IEEE Transactions on Visualization and Computer Graphics, 2022. **28**(9): p. 3307-3323.

22. Sharma, A., et al., *Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis.* Neuroscience Informatics, 2021. **1**(3): p. 100012.

23. Wu, C., et al. *Clickbait Detection with Style-Aware Title Modeling and Co-attention*. in *Chinese Computational Linguistics*. 2020. Cham: Springer International Publishing.

24. Nurwaha, D., *Comparison of kernel functions of support vector machines: A case study for the solar cell output power prediction.* International Journal of Energy Applications and Technologies, 2020. **7**(1): p. 1-6.

25. Mustafa Abdullah, D. and A. Mohsin Abdulazeez, *Machine Learning Applications based on SVM Classification A Review.* Qubahan Academic Journal, 2021. **1**(2): p. 81-90.

26. Chen, H., et al., *Improved naive Bayes classification algorithm for traffic risk management.* EURASIP Journal on Advances in Signal Processing, 2021. **2021**(1): p. 30.

27. Blanquero, R., et al., *Variable selection for Naïve Bayes classification.* Computers & Operations Research, 2021. **135**: p. 105456.

28. Rahmad, F., Y. Suryanto, and K. Ramli, *Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification.* IOP Conference Series: Materials Science and Engineering, 2020. **879**(1): p. 012076.

29. Heydarian, M., T.E. Doyle, and R. Samavi, *MLCM: Multi-Label Confusion Matrix.* IEEE Access, 2022. **10**: p. 19083-19095.

30. Genç, Ş. 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, 2023. **49**(2): p. 480-499.

31. Ghosh, S., *Identifying click baits using various machine learning and deep learning techniques.* International Journal of Information Technology, 2021. **13**(3): p. 1235-1242.