

## FAKE PROFILE IDENTIFICATION USING MACHINE LEARNING : A REVIEW

**Dhruvi Patel**<sup>1,a</sup>, **Dr. Chintan Thacker**<sup>1,b</sup>, **Dr. Kamal Sutaria**<sup>1,c</sup>, **Hetal Bhaidasna**<sup>1,d</sup>

<sup>a</sup> Student of Computer Science Department, Parul Institute Of Engineering and Technology, Limda, Vadodara, Gujarat, India

<sup>b</sup> Assistant Professor Of Computer Science Department, Parul Institute Of Engineering and Technology, Limda, Vadodara, Gujarat, India

<sup>c</sup> Assistant Professor Of Computer Science Department, Parul Institute Of Engineering and Technology, Limda, Vadodara, Gujarat, India

<sup>d</sup> Head Of Department Computer Engineering-Diploma Studies, Parul Institute Of Engineering and Technology, Limda, Vadodara, Gujarat, India

<sup>a)</sup> dhruveepatel51@gmail.com

<sup>b)</sup> chintan.thacker19435@paruluniversity.ac.in

<sup>c)</sup> kamal.sutaria24554@paruluniversity.ac.in

<sup>d)</sup> hetal.bhaidasna@paruluniversity.ac.in

**ABSTRACT--** The popularity of the internet has led to its recognition as a crucial medium for social interaction, marketing, and advertising. There are millions of users, however some of them frequently mishandling the network by creating false identities. People are taking part social media site actively all over the world. But, it also needs to address the problem of bogus profiles. Fake account creation is the process of creating accounts using fake or stolen identity information and regularly made by people, software, or machines. They are employed in the spread of rumors and illegal actions like phishing and identity theft. This article discusses the different methods used to identify malicious accounts using different machine learning methods, including SVM, naive bayes, random forest, logistic regression, and K-nearest neighbors.

**Keywords:** — Social media, Fake accounts, Machine learning ,Supervised learning, Particle Swarm Optimization, Recursive Feature Elimination

## INTRODUCTION

The use of social media platforms is widespread for sending and receiving data. Everyone utilizes social media, whether it's to post pricey images, follow celebrities, or communicate with close and far friends. It is an excellent setting for socializing and exchanging information. [1]. They make use of them to engage with the material and data offered by platform users. People all over the world use social media towards communicate and share their ideas and creative works. Additionally, it gives them a stage on which to display their abilities and connect with others worldwide who share their interests [2]. Social media comes in many forms, including Instagram, LinkedIn, Twitter, Facebook, snapchat, and others. Social networking has become a daily requirement for the majority of people as technology advances. There are many benefits to using social media online, but there are also drawbacks. Due to problems like malware attacks and privacy threats, every social media user is at risk [3]. When attackers gain unauthorized access to private data and the user's consent is not obtained, a

privacy breach assault has occurred. Social media is increasingly being used to distribute malware thanks to user-to-user communication and advertising. Bots or fraudulent accounts are a significant issue as well [16]. A fake account is a representation of a person or business that impersonates another. Fake accounts come in two different varieties.

Duplicate accounts are backup accounts that a user creates for future usage without interfering with another user's actions. Unwanted profile or profile accounts Some fictitious accounts are made for organizational and business purposes. These accounts, however, are typically used for spamming and breaking social network terms of service. Numerous fictitious accounts have been made in order to promote businesses as well as spam users. Some accounts are made in order to encourage the growth of certain corporate entities or political organizations. The study recommends improved machine learning methods for identifying these phone Instagram profiles.

A fake account is basically an account on any social media platform where the details displayed are actually dishonest or even fraudulent. Misrepresentation on fake accounts, using false details, deceives the general public to spread inaccurate information or collect financial or personal information [8]. People set up accounts on various social networking sites to exchange social media content. In order to propagate false information without disclosing their identities, users frequently create accounts with false or anonymous information. Users frequently alter their accounts or create accounts in someone else's name (identity theft). There are certain specific financial advantages to creating false accounts. These phone identities are maintained by bots or automated programs, which aid in the deeper and faster dissemination of fake news online. On the network, fake accounts frequently connect and follow the posts of influencers. Through policies against impersonation, even social media like twitter, Facebook, and WhatsApp that are used for online social networking erase or freeze these phone accounts. Within hours of the explosion during the Boston Marathon, approximately 32,000 new accounts were created, of which twenty percent had their profile discontinued by twitter (Gupta et al. 2013). The majority of malicious profiles are created with the intention of spamming, phishing, and getting more followers. The fake accounts have all the tools necessary to commit online crimes. Identity theft and data breaches are substantial risks posed by fake accounts. All user information is transferred to faraway servers when users view the URLs sent by this fake profile, where it may be utilized against them. False profiles that claim to be from companies or individuals can harm their reputation and get them less likes and follows [9].

## RELATED WORK

M. Senthil Raja et. al [5]., recommends detection techniques as 3PS (Publicly Private Protected system) for detecting a fake account over an online social network by considering the behavior patterns of user account activity. It emphasizes the detection of fake or malicious users who users own various online social network accounts, send friend requests, and also share many posts as malicious. It considers the posts, follows, status updates, followers, and also posts. A fake or malicious user can be detected by distinguish between the threshold value of an attribute related to the user's personal profile along with network similarity exploration.

Kusum kumara Bharti et al. [6] recommends a system to identify a fake account on twitter by selecting the user profile characteristic feature and logistic regression with PSO, naive bayes and KNN algorithms used for the classification task. They used information gain, correlation-

based feature selection (CFS), minimum relevance maximum redundancy (MRMR) for feature selection. The dataset, which consisted of 6973 profile data and was split into training and testing sections, was gathered manually and in other ways utilizing the Twitter REST API.

Ginuseppe Sansonetti et al. [7], used a framework that performs a dual analysis: message reliability prediction and social network user profile reliability prediction. It features offline analysis utilizing deep learning techniques incorporated, as well as online analysis using real users to categories a reputable or unreliable user profile on Twitter. S. Uppada et al. [8] recommends a technique for the UN to use to spot fake news and photos. By rating users and articles, as well as, fake profile identification is utilized as a statistic to track user engagement trends. In order to create characteristics related to fake news images, forensic techniques and picture polarity analysis are merged. The SENAD approach and the CredNN model each had accuracy levels of 76.3% and 93.5% for fake news, respectively.

In [9] “researcher recommends a new bot identification methodology using deep neural networks and active learning. Modules for feature extraction, active learning, data gathering and labelling, and detection are included. In addition, this framework offered a most recent RGA deep neural network model for identification that makes use of ResNet, BiGRU, and attention mechanisms. The testing findings demonstrated that the suggested DABot framework is superior to existing detection methods in terms of effectiveness for identifying social” bots. Somya Ranjan Sahoo et al. [10] uses a framework which is based on chrome extension that identifies a fake twitter account. They also contrasted a number of tasks utilizing machine learning strategies to confirm user information gathered through manual and web crawler. The data set was obtained from the twitter profile using a web crawler and the twitter API. The profile data gathered includes the user's name, ID, number of status updates, friends list, number of favorites, and number of URLs cited in tweets. Then the data is grouped into test and training data in an 80:20 ratio. And passed the WEKA machine learning platform. The trust score from the features is calculated and then goes to the Chrome extension to identify the malicious profile. Through user distinctive analysis and the usage of a trust score, Chrome extensions produce a suspiciousness score for each user. They employ random forest and bagging to identify fraudulent profiles and generalize the performance of the Chrome plugin.

Rathore et al. [11] recommends using a deep walk method on rater graphs and a semi-supervised clustering technique that can take partial background knowledge into account, offer a top-down framework for identifying fake groups of prospective reviewers. Additionally, using temporal affinity, semantic traits, and sentiment analysis, this strategy can be expanded to find groups of opinion spammers on social media. Detection of potential spammer groups based solely on the underlying graph's topological structure. The reviewer ID, represented by feature vectors, is divided into various groups of spammer candidates using a modified version of the semi-supervised clustering process known as Pair wise after first obtaining a representation for each node in G using the Deep Walk method. confined K-Means. To identify groups of fraud reviewers from reviewer graphs, the framework was verified on a partial core data set consisting of 2207 fraudulent reviewer IDs belonging to 23 different reviewers.

Y. Fang et al. [12] suggest a technique using dynamic knowledge graphs to find fake feedback. Using a newly developed neural network model termed conditional two-way long-term short-term memory, the first four types of entities are extracted by embedding phrase vectors/double words based on the characteristics of online product reviews. Next, time series-related

properties are incorporated into the knowledge graph building process to generate dynamic graph networks. They first extracted many different sorts of entities using the ST-BLSTM algorithm, after which they provided the most recent MI-based criteria for evaluating the connections between entities. They subsequently produced a dynamic graph network by mixing the time sequence factors. The second showed the value of evaluating the truthfulness and commodity quality of the rather by calculating trustworthiness, honesty, great commodity degree, and reliability ratings utilizing novel methodologies. The technique demonstrated how review portrait data reflects the reasons of bogus reviews and provides crucial information. The technique demonstrated how review portrait data reflects the root reasons of false reviews and offers crucial hints regarding various reviewer types.

Kayode “SakariyanAdewole et al.[15] presents With 23 features in the dataset, three classification algorithms—such as random forest, support vector machine (SVM), and multilayer perceptron (MLP)—are utilized to detect spam accounts on Twitter social media to improve the first detection of spammers” clusters.

Ruan, N. et al. [20], propose long-short memory neural network and an AdaBoost model are combined in the geolocation-based profile recognition prototype to analyse user account and geolocation data. Its two sub-models are the geolocation detection model and the profile detection framework. Geolocation DM receives geolocation features and applies LSTM to analyse geolocation sequences in order to create user identity prediction scores. Account-DM receives account feature input and analyses account feature using AdaBoost. The linear classifier SVM generates the final evaluation of user IDs using the profile-DM and geolocation-DM prediction scores as input. This technique may reliably and effectively identify fraudulent reviews, according to the study, which used a huge dataset from Yelp.

#### **Summary of related works:**

The most recent updates, posts, comments, images, etc. are investigated using OSN identities made for experiments. SVM-NN classification algorithm used for detection of fake account [5].

Classification of account as genuine or fake on twitter. Naïve bayes, decision tree , logistic regression with PSO are used [6].

The dissemination of fake news is identified by the SENAD model, which also identifies bogus accounts that were set up to do so. CredNN is the framework suggested for use with fake pictures [8].

A revolutionary DABot architecture is suggested using deep neural networks and active learning to identify social bots [9].

Identifying dangerous content and accounts in the Chrome environment, a false account detector protects the twitter user account from attackers [10].

Use core view graph's topological structure to identify potential spammer groups. The identification of potential fake reviewer groups using a semi-supervised clustering technique and the deep walk methodology [11].

Used “neural network model termed sentence vector embedding conditioned bidirectional long short-term memory, four different sorts of entities are extracted [12].”

Classification methods including SVM, KNN, Naive Bayes, and XGBoost. Results of categorization are recorded in a data dictionary and utilised to identify fake IDs based on profile IDs.[13]

The dataset was created using 17 metadata elements from both actual and phone Instagram profiles. Random Forest, Logistic Regression, Multi-Layer perceptron, AdaBoost, SVM, and neural network are the classification algorithms employed [14].

## METHODOLOGY

### Datasets

Yelp dataset[19]: “It is a dataset with partially annotations that includes user profiles, review information, and store data. Yelp's screening algorithm has marked reviews for the highlighted part as false or unimportant. Contains 107,264 tagged reviews out of 760,212 total reviews and 3,142 tagged users out of 16,941 total users. Among the 107,624 evaluations with labels, 20,267 are fake.”

Enron dataset[15]: Enron's “email communication network encompasses all email exchanges inside a dataset of about 500,000 emails. Email addresses make up the nodes of the network, and if email address  $i$  sends at least one email to email address  $j$ , there is an undirected edge from  $i$  to  $j$  in the graph. The dataset is available at <https://www.cs.cmu.edu/~./enron/>.”

Instagram fake spammer genuine account[13] :This dataset contains 10 columns such as followers, following, post count, username length, profile picture, description length, private, public, external url. The dataset is available at <https://www.kaggle.com/free4ever1/instagram-fake-spammer-original-accounts>.

The Fake Project.[6,16]: The IIT-CNR in Pisa, Italy, researchers' research project produced the Fake Project dataset. It is made up of 469 volunteer accounts that CAPTCHA has verified are owned by humans.

Elezioni2013 dataset[16]: “In order to assist a project for a psychological research done in partnership with the Universities of Perugia and Sapienza in Rome, the #elezioni2013 dataset was created. With the exception of the following categories, this study added a collection of 1481 Twitter accounts with various professional backgrounds: Two sociologists from the University of Perugia in Italy manually classified the selected accounts of political parties, journalists, and bloggers as belonging to human” beings.

PolitiFact dataset[4,7]: It is a popular fact check website that contain 21,152 statements in the dataset have been verified as accurate by professionals. All of the claims are divided into one of six groups: true, mostly true, half true, mainly false, false, and trousers on fire.

MIB dataset[17]: “The dataset was created using management information, and it included 3474 real accounts and 3351 fake accounts. The dataset used E13 and TFP for genuine accounts and TWT, INT, and FSF for fraudulent ones. For automatic" extraction, the data is saved in CSV format.

### Attribute Selection methods

Particle Swarm Optimization [6]: Dr. Eberhart and Kennedy created Particle Swarm Optimization in 1995. It encourages birds to flock together. The algorithm, which is stochastic and population-based, begins with randomly initialized solutions. Each individual population solution is viewed as a prospective contender for the problem's solution. The objective function evaluates each particle's fitness.”

At each repetition, the particle updates its position once the swarm is first randomly initialised with a collection of particles. Each particle has two optimum values related to it. The terms

"pbest" and "gbest" refer to the best position and value each individual particle in the population has achieved, respectively. Moreover, they have a velocity that, starting from the present point, describes the passage of the particles through the search space. The algorithm starts with randomly generated solutions. Each randomly generated solution is considered as a potential candidate solution. The solutions are updated over the number of iterations and the best solution is considered as the final solution of the algorithms."

Recursive Feature Elimination : It is well-liked method because simple to set up and use, and because it is efficient in choosing the features (columns) in a training dataset that are more or more useful in predicting the target variable. In order to effectively use RFE, two configuration decisions must be made: how many features to pick and the methodology to be applied to the feature selection process. It is possible to look at both of these parameters, even though their proper configuration does not significantly affect how well the strategy performs.

**Machine Learning Models**

Naïve Bayes[2,7,13]:“ A supervised learning strategy based on the bayes theorem for classification problems. It primarily functions as a text categorization system with a big training set. A classification method that is among the simplest and most effective ones currently in use. Fast machine learning models that can anticipate outcomes accurately can be developed with its assistance. As a probabilistic classifier, it based its forecasts on the likelihood that an object will occur. Theorem in mathematical terms can be” written as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

Where, P(A|B) represents posterior, P(B|A) represents the likelihood probability, P(B) is the marginal probability, while P(A) denotes the prior.

Logistic Regression[2,10]: It is a statistical model that is used to predict the likelihood that a statement is true or incorrect, will pass or fail, will win or lose, etc. can be extended to multiple class models to determine whether an image contains what type of object, where each Detected object has only two probabilities, either 0 or 1.

The function of logistic regression is given as:

$$P(x) = \frac{1}{1 + e^{-(x-\mu)/s}} \tag{2}$$

Where  $\mu$ =location parameter and  $s$ =state parameter

The modified expression is expressed below:

$$P(x) = \frac{1}{1 + e^{-(\beta_0-\beta_1x)}} \tag{3}$$

Where,  $\beta_0=-\mu/s$  (the intercept), and  $\beta_1 = 1/s$ . Further, one may define the "fit" to  $y_k$  at a given  $x_k$  as[20]:

$$p_k = (x_k) \tag{4}$$

Where,  $p_k$  are the probabilities that the corresponding  $y_k$  [20].

Support Vector Machine[7,9,10]:This linear model is applied to both regression and classification models. Capable of solving both linear and nonlinear problems, it also works well for many practical problems and is shown in the figure for illustration. His idea is simple and he says that an algorithm generate a line or hyper-plane that classifies the data. Its major

objective is to identify the boundary between several dataset classifications. The support vector represents points closer to the hyper plane that affect its orientation and position[22].

Random Forest[13]: A common ML approach that uses the SML technique. It is also applied to machine learning challenges involving classification and regression. In order to handle complicated issues and enhance model performance, it is created using the study of ensemble learning[23]. Instead of relying just on one decision tree, this sort of classifier combines several with different subsets of the dataset and improves the predicted accuracy of the dataset on average. It bases its output on the results of the majority of votes cast using prediction. The accuracy and prevention of overfitting are directly inversely correlated with the number of trees.

K-Nearest Neighbor [7,13]: One of the simplest machine learning algorithms is based on the supervised learning approach. Based on the supposition that the new instance and present instances are comparable, this methodology assigns the new instance to the category that best matches the existing categories[25].When all of the previously saved data has been stored, a new data point is categorized using this method based on similarity. This suggests that employing this method, new data can be accurately and efficiently categorized. Although the technique can be applied to regression problems as well, classification issues are where it is most frequently used.

**Differential Analysis**

**Comparative ML Approches**

ML Approches	Pros.	Cons.
Naïve Bayes[2,7,13]	Performs better than other algorithms in multi-class predictions. It can be used for both binary and multi-class classifications.	Because Naive Bayes makes the assumption that each element is independent or unrelated, it is unable to identify the connections between elements.
Logistic Regression[2,10]	Logistic regression is incredibly simple to use, interpret, and train. Multinomial regression, which provides a natural probabilistic perspective of class predictions, is simply expanded to many classes.	Since overfitting might happen, it shouldn't be utilized if there are less observations than elements. It has a linear decision surface, hence it cannot address non-linear issues. Real-world situations rarely involve linearly separable data.
Support Vector Machine[7,9,10]	When there are more dimensions than samples, the method is successful. It can be applied to multi-class and	It performs poorly when the target classes overlap and the data set contains more noise.

	binary classifications.	In case of large data set, it does not perform as well because the training time is longer.
Random Forest [13]	With both categorical and continuous data, random forest performs well.  Missing values can be handled automatically using random forest.	The random forest algorithm can change significantly with a small change in the data.  Calculations, as compared to other algorithms, can be considerably more complex.
K-Nearest Neighbor [7,13]	Both classification and regression issues can be solved using it.  Since there are no underlying assumptions, it is perfect for non-linear data.	Given that it stores all of the training data, associated computation costs are substantial.  Large memory storage is necessary.

**CONCLUSION**

In this research review different methodologies are discussed for spot fake accounts on online social media. We have discussed different dataset used for identifying fake account. The article comprehensively reviewed the machine learning techniques developed for fake account identification. Out of the several ML algorithms, LR, NB, SVM, RF, KNN are found to be majorly used for fake profile analysis. It was observed that to enhance accuracy of an algorithm, one can tune the algorithm by setting its parameters according to the data or by reducing the size of the dataset with the help of attribute selection techniques. The observation suggests that among the discussed algorithms logistic regression performed well a smaller dataset or less complex dataset as compared to naïve bayes, whereas for multidimensional dataset, random forest is most suitable. So, in future the system can be implemented with attributes selection method and try with different datasets.

**REFERENCES**

Majed Alrubaian, Muhammad Al-Qurishi, Mohammad Mehedi Hassan, and Atif Alamri, A Credibility Analysis System for Assessing Information on Twitter, IEEE (Aug 2018) <https://ieeexplore.ieee.org/document/7551232>

Van Der Walt, Jan Eloff and Estée. Using machine learning to detect fake identities: bots vs humans. IEEE (2018) <https://ieeexplore.ieee.org/document/8265147>

Tri HadiyahMuliawati, Perdana, Rizal Setya, and Reddy Alexandro. Bot spammer detection in Twitter using tweet similarity and time interval entropy. (Research Gate)

Karatas, Arzum, and SerapŞahin. A Review on Social Bot Detection Techniques and Research Directions. (2015)



- Senthil Raja, M., & Arun Raj, L. (2021). Detection of malicious profiles and protecting users in online social networks. *Wireless Personal Communications*, 1-18.
- Bharti, K. K., & Pandey, S. (2021). Fake account detection in twitter using logistic regression with particle swarm optimization. *Soft Computing*, 25(16), 11333-11345.
- G. Sansonetti, F. Gasparetti, G. D'aniello and A. Micarelli, "Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection," in *IEEE Access*, vol. 8, pp. 213154-213167, 2020, doi: 10.1109/ACCESS.2020.3040604
- Uppada, S. K., Manasa, K., Vidhathri, B., Harini, R., & Sivaselvan, B. (2022). Novel approaches to fake news and fake account detection in OSNs: user social engagement and visual content centric model. *Social Network Analysis and Mining*, 12(1), 1-19.
- Wu, Y., Fang, Y., Shang, S., Jin, J., Wei, L., & Wang, H. (2021). A novel framework for detecting social bots with deep neural networks and active learning. *Knowledge-Based Systems*, 211, 106525.
- Sahoo, S. R., & Gupta, B. B. (2021). Real-time detection of fake account in twitter using machine-learning approach. In *Advances in computational intelligence and communication technology* (pp. 149-159). Springer, Singapore.
- Rathore, P., Soni, J., Prabakar, N., Palaniswami, M., & Santi, P. (2021). Identifying groups of fake reviewers using a semisupervised approach. *IEEE Transactions on Computational Social Systems*, 8(6), 1369-1378.
- Fang, Y., Wang, H., Zhao, L., Yu, F., & Wang, C. (2020). Dynamic knowledge graph based fake-review detection. *Applied Intelligence*, 50(12), 4281-429.
- Preethi Harris; J Gojal; R Chitra; S Anithra," Fake Instagram Profile Identification and Classification using Machine Learning", 2021 2nd Global Conference for Advancement in Technology (GCAT) | 978-1-6654-1836-2/21/\$31.00 ©2021 IEEE | DOI: 10.1109/GCAT52182.2021.9587858
- Muñoz, S. D., & Pinto, E. P. G. (2020, December). A dataset for the detection of fake profiles on social networking services. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 230-237). IEEE.
- Chaudhary, A., Mittal, H., & Arora, A. (2019, February). Anomaly detection using graph neural networks. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 346-350). IEEE.
- Khaled, S., El-Tazi, N., & Mokhtar, H. M. (2018, December). Detecting fake accounts on social media. In *2018 IEEE international conference on big data (big data)* (pp. 3672-3681). IEEE.
- Chakraborty, P., Shazan, M. M., Nahid, M., Ahmed, M. K., & Talukder, P. C. (2022). Fake Profile Detection Using Machine Learning Techniques. *Journal of Computer and Communications*, 10(10), 74-87.
- Laleh, N., Carminati, B., & Ferrari, E. (2016). Risk assessment in social networks based on user anomalous behaviors. *IEEE Transactions on Dependable and Secure Computing*, 15(2), 295-308.
- Latha, P., Sumitra, V., Sasikala, V., Arunarasi, J., Rajini, A. R., & Nithiya, N. (2022, March). Fake Profile Identification in Social Network using Machine Learning and NLP. In *2022 International Conference on Communication, Computing, and Internet of Things (IC3IoT)* (pp. 1-4). IEEE.

Ruan, N., Deng, R., &Su, C. (2020). GADM: Manual fake review detection for O2O commercial platforms. *Computers & Security*, 88, 101657.

“Logistic regression,” Wikipedia, 15-Jan-2023. [Online]. Available: [https://en.wikipedia.org/wiki/Logistic\\_regression](https://en.wikipedia.org/wiki/Logistic_regression). [Accessed: 17-Jan-2023].

“Logistic regression,” Wikipedia, 15-Jan-2023. [Online]. Available: [https://en.wikipedia.org/wiki/Logistic\\_regression#/media/File:Exam\\_pass\\_logistic\\_curve.jpg](https://en.wikipedia.org/wiki/Logistic_regression#/media/File:Exam_pass_logistic_curve.jpg). [Accessed: 17-Jan-2023].

“Support Vector Machine (SVM) algorithm - javatpoint,” [www.javatpoint.com](http://www.javatpoint.com). [Online]. Available: <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>. [Accessed: 17-Jan-2023].

“Machine learning random forest algorithm - javatpoint,” [www.javatpoint.com](http://www.javatpoint.com). [Online]. Available: <https://www.javatpoint.com/machine-learning-random-forest-algorithm>. [Accessed: 17-Jan-2023].

“K-Nearest Neighbor (KNN) algorithm for Machine Learning - Javatpoint,” [www.javatpoint.com](http://www.javatpoint.com). [Online]. Available: <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>. [Accessed: 17-Jan-2023].