

**Detection of Spam Tweets in Twitter**

Submitted by: P.S.S.N.Chandra

PGDM

2111234

**INTRODUCTION**

An OSN (Online Social Network) i.e., a Web-based service that allows individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with people they have a connection with, and access and browse their list of connections as well as those formed by others in the system. Some of the Online Social Networks (OSNs) that are widely popular right now are Facebook, WhatsApp, Twitter, Instagram etc.

Along with the growth of the social networks, increased the number of spammers. Spammers are the users who manipulate the platforms to broad cast unwanted or malicious messages. Twitter is a microblogging service where users can post 280-character messages called tweets.

As of 2022, every second, on the average, around 6,000 tweets or 350,000 tweets sent per minute or 500 million tweets sent every day or, 200 billion tweets per annum are present facts (Beveridge, 2022). Thanks to this huge growing trend, this Online Social Network has attracted many users along -side spammers. Web Attacks that have appeared on Twitter are Scam, Spam, Phishing etc., Spam may be a sort of Platform Manipulation. Platform Manipulation is taken into account as an activity that's intended to negatively impact the people’s experience on Twitter. This encompasses issues that are undesired or reiterated. Spam can include malicious automation and other sorts of platform manipulation like fake accounts. Spamming is the method of sending several unsolicited messages to a large range of recipients for the purpose of commercial advertising, non-commercial preaching, any unlawful purpose, or just sending the same message to the same user repeatedly.

Shortened URL is included in most of the Spam Tweets to trick users into clicking on it. Also, they tend to tweet similar trending topics to attract a larger audience since resources, such as tweets can be shared with each other. This type of Web Attacks not only disturbs the user experience but also causes a whole internet damage which may possibly cause temporary of Internet Services all over the Globe.

To handle the consequences, User can report a spam by clicking their home page. Then accordingly the spam accounts are suspended. However, as the Total number of Tweets sent per Day are 500 million in 2022, around 10% of those tweets are of Spam Tweets (Kumar, 2022). This has become a major problem on finding an appropriate Solution.

**LITERATURE REVIEW**

* (Resul & Abdullah, 2017), in their paper titled A Survey of Spam Detection Methods on Twitter, tried to identify new features to identify the tweets as Spams. They opined that is in order to provide a spam-free environment, tweets of spammers are needed to be detected and filtered as well as the owners. It is vital to limit false positive detections in order to avoid legitimate users from being labelled as spammers. In their paper, the features of Twitter spam detection and proposed approaches in the literature are discussed with considering their advantages and disadvantages. In addition, the old Twitter features that are widely employed by Twitter spam detection algorithms are emphasized. Some new Twitter features are also introduced that, to the best of our knowledge, have not been addressed in any previous literature.
* (Nasira , Malik.M, & Qaisar, 2016), in their paper titled Sentiment Based Twitter Spam Detection, tried to propose a spam detection approach for twitter based on sentimental features. They opined that by using twitter API they collected their dataset of 29 most trending topic in 2012. They proposed sentimental and some content based features which will help in identifying spam tweets and return spam filtered result set when user visit twitter with good accuracy rate. They evaluate the usefulness of their suggested features in spam detection by using five traditional classifiers like BayesNet, Naive Bayes, Random Forest, Support Vector Machine (SVM) and J48 schemes. Their experiments results shows that Naive Bayes, J48 and Random Forest classifier gives over all best performance than the others.
* (Monika , Divya, & Sanjeev, 2014), in their paper titled Techniques to Detect Spammers in Twitter- A Survey, tried to detect Spam profiles in OSNs. In this paper they opined many methods that have been developed and used by various researchers to find out spammers in different social networks. According to the literature analyzed, the majority of the work has been done utilizing classification algorithms such as SVM, Decision Tree, Naive Bayesian, and Random Forest. Detection has been done using either user-based or content-based characteristics, or a mix of the two. A few authors also offered unique detecting features. All of the methods were verified on a very limited dataset and were never tested with other combinations of spammers and non-spammers. When compared to employing simply user-based or content-based characteristics, combining features for spam detection performed better in terms of accuracy, precision, recall, and so on.
* (Rohini, Sunil, & Jaiswal, 2018), in their paper titled Improving Spam Detection on Online Social Media with hybrid classification techniques on Twitter platform, tried to use the Naïve Bayes theorem classifier and develop a speaker organisations to exclude and not exclude spam. In this paper they opined that Using ML algorithm SVM (Support vector machine) and NB are used to Improving Spam Detection on Online Social Media with hybrid classification techniques on the Twitter platform. In this dissertation, the System provides a fundamental evaluation of ML algorithms on the detection of streaming spam tweets. In this evaluation, the system works on offline tweets and real-time tweets which are timely updated. Feature discretization was discovered as a key pre-process in ML-based spam detection by the system.
* (N.Noor, S.Vishnu, Suman, & G.Tejaswi, 2020), in their paper titled Spam Detection Framework for Twitter using ML, tried to Detect Spam Accounts using ML Algorithms. In this paper they opined that spam has become one of the main issue and it should be solved in every social networking site. Spam detection framework in twitter using machine learning help people to solve their spam issues in an easy and accurate way. This project detects the spam to 95% and spammers can be easily blocked.
* (Claudia, Federica, Paolo, & Rodolfo, 2014), in their paper titled Machine Learning Techniques applied to Twitter Spammers Detection, tried to Detect spam tweets using machine learning techniques. In this paper they opined that SVM and Random Forest algorithm are used to find the spam in the data set and they came to the conclusion that random forest performance is more when compared with other techniques.
* (Vanyashree, Anvaya, V.M.Sukanya, & S.Rachana, 2020), in their paper titled Text-Based Spam Tweets Detection Using Neural Networks, tried to Detect Spam Tweets Using Neural Network. In this Paper is methods such as Naive-Bayes Classifier and Artificial Neutral Network are used. They opined that Performance study of these two algorithms, shows that Artificial Neural Network performs better than Naive Bayes Classification algorithm.
* (Nan, Guanjun, Junyang, & Paul, 2020), in their paper titled near real-time twitter spam detection with machine learning techniques, tried to Detect Real-Time Spam Tweets using ML Algorithms utilizing Parallel Computing Techniques. In this paper, Machine Learning along with utilizing parallel computing techniques are used. Machine Learning Techniques such as Random Forest, SVM are used. They opined that among them, Random Forest has more stable performance when compared to other technique.
* (Jagtap, Dr.Bamu, & B.A.Khansole, 2017), in their paper titled Machine Learning Approach for Spam Tweets Detection, tried to Detect Of Real-Time Spam Tweets Using Naïve Bayes or SVM. In this Journal they have opined that NB and SVM cannot accurately give us the output as NB and SVM have different options including the choice of kernel function for each. They are also both susceptible to parameter tuning (i.e. different parameter selection can significantly change their output). The work of Naive Bayes classifier is better than SVM classifier for the taken dataset.
* (K.Subba & E.Srinvasa, 2019), in their paper titled Detecting Spam Messages in Twitter Data by Machine learning Algorithms using Cross-Validation, tried of Classifying messages as Spam or Ham. In their paper to avoid spam messages, they proposed a methodology by using machine learning algorithms and to develop an approach using a set of content-based features and prepared a spam detection model using the Support vector machine algorithm (SVM) and Naive Bayes classification algorithm and made use of precision, recall, and F measure metrics to measure the performance of the model.
* (Miss.Shukla & Prof.D.B.Kshirsagar, 2016), in their paper titled Design of Machine Learning Approach For Spam Tweet Detection, tried to categorize the Spam and Non-spam tweets by machine learning approach. In their paper classifier system-based approach is used to solve the detection of spam messages, the classification model is mainly based on a machine learning algorithm which gives the output in the form of the binary value. The most important phase of the project is feature extraction to add more benefits to the system. They believe that the performance evaluation is carried out on a huge dataset of roughly 600 tweets to identify the spammer, and that the system also assists in categorizing spam and non-spam messages.
* (Vinodhini.M, Prithvi.D, & Balaji.S, 2020), in their paper titled Spam Detection Framework using ML Algorithm, tried to Determine whether or not a specific message in the dataset is spam using a set of machine learning algorithms. They have used Bayes Network, Naive Bayes, K-nearest neighbour identified the spams and spammers present in a twitter dataset with the help of machine learning algorithms and NLP concepts. They believe that by reading the spam, all of the spammer's details are accessed and presented, which aids in detecting additional spams, spammers, and their message writing style.
* (Wu, Liu, Zhang, Xiang, & Yang, 2017), in their paper titled Twitter spam detection based on deep learning, proposed a novel technique based on deep learning techniques to address the challenges. They learned syntax of each tweet through WordVector Training Mode. They then constructed a binary classifier based on the preceding representation dataset. Then studied the performance of different classifiers, and then compared their method to other existing text-based methods. They found that their method largely outperformed existing methods. They further compared their method to non-text-based detection techniques.
* (Abkenar, Sepideh, Mostafa , Mohammad, & Ebrahim, 2020), in their paper titled Twitter Spam Detection: A Systematic Review, focuses on comparing the existing research techniques on Twitter spam detection systematically. They proposed a taxonomy based on different feature selection methods and analyses, namely content analysis, user analysis, tweet analysis, network analysis, and hybrid analysis. Then, they presented numerical analyses and comparative studies on current approaches, coming up with open challenges that help researchers develop solutions in this topic. A hybrid classification method for Twitter spam detection based on differential evolution and random forest is used in this paper.
* (Borse, Dipalee, & Swati, 2022), in their paper titled State of Art on Twitter Spam Detection, their survey has been divided into three parts such as spam detection, real-time spam detection, and spammer detection. They also discussed different Twitter features used for spam detection, effectiveness, and challenges in current research work. They have used ensemble learning framework for spam detection.
* (Devi & Kumar, 2022), in their paper titled Stochastic Gradient Boosting Model for Twitter Spam Detection, they aim to improve the detection of spam in the social networks. To analyse Twitter data sets in the English language, this study recommends employing statistically based characteristics that are modelled using the supervised boosting technique known as stochastic gradient boosting. The performance of their proposed model is evaluated using simulation results.
* (Chowdhury, Ratul, Kumar, Banani, & Samir, 2020), in their paper titled A Method Based on NLP for Twitter Spam Detection, this paper addresses the aforementioned problem of user profiles by emphasising both profile data and content-based spam detection. To that purpose, this study makes three important contributions. First and foremost, extensive use of natural language processing (NLP) approaches. Second, this dataset was put into a unique cutting-edge hybrid machine learning model that was constructed entirely utilising a combination of machine learning and deep learning approaches. Finally, a novel methodology based on logistic regression is proposed and supported by analytical formulations. This opens the path for the custom-built dataset to be processed and associated probabilities that distinguish legitimate users from spammers to be determined.
* (Vijayalakshmi, K Tanvi, L Rao, & Y Sindhu, 2022), in their paper titled Spam Detection in Twitter using Multinomial Naïve-Bayes Classifier, their approach leverages tweet text to detect spammers. This social network spam has piqued the interest of many experts, who have proposed several ideas for spam categorization and identification. The primary goal of this effort is to create a precise Twitter spam detection system for accurate spammer and non-spammer identification, as well as to improve the security of every Twitter user.
* (Rodrigues, et al., 2022), in their paper titled Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques, their proposed work is to develop a system that can determine whether a tweet is “spam” or “ham” and evaluate the emotion of the tweet. After preprocessing the tweets, the extracted features are identified using several classifiers, including decision tree, logistic regression, multinomial naive Bayes, support vector machine, random forest, and Bernoulli nave Bayes for spam identification. Stochastic gradient descent, support vector machine, logistic regression, random forest, naive Bayes, and deep learning methods, such as the basic recurrent neural network (RNN) model and the long short-term memory (LSTM) model, are examples of deep learning methods, bidirectional long short-term memory (BiLSTM) model, and 1D convolutional neural network (CNN) model are used for sentiment analysis. The performance of each classifier is analyzed. The classification findings demonstrated that the characteristics collected from the tweets may be utilised to determine if a given tweet is spam or not and to develop a learning model that would correlate tweets with a specific emotion.
* (R. Priyanka & Dr. Bhuvana, 2022), in their paper titled A Semi-Supervised Learning Approach for Tackling Twitter Spam Drift, they proposed a system with semi-supervised learning approach (SSLA) has been proposed to tackle the challenge. The novel method learns the domain structure from unlabeled data. To deal with the drift, the study uses a live Twitter feed of data. Pre-processing of live downloaded data is tagged, and machine learning is used to differentiate between spam and non-spam users. The data is saved in the cloud and may be viewed by the user from any location. Experiment results on many machine learning algorithms were obtained, and the best in terms of accuracy for the specified task were discovered.
* (Rosita & Jenifer, 2022), in their paper titled Multi-Objective Genetic Algorithm and CNN-Based Deep Learning Architectural Scheme for effective spam detection, they proposed a Multi-Objective Genetic Algorithm and a CNN-based Deep Learning Architectural Scheme (MOGA–CNN–DLAS) for the predominant Twitter spam detection process. The proposed MOGA-CNN-DLAS is tested in terms of accuracy, precision, recall, F-Score, RMSE, and MAE by altering the ratio of training data using three actual datasets, including the Twitter 100k dataset and the ASU dataset.
* (Zhang Zhijie, Rui, & Jin, 2020), in their paper titled Detection of Social Network Spam Based on Improved Extreme Learning Machine proposed Twitter spam characteristics such as user attribute, content, activity, and relationship, and a novel spam detection algorithm called the Improved Incremental Fuzzy-kernel-regularized Extreme Learning Machine (I2FELM) is designed based on regularised extreme learning machine, which is used to detect Twitter spam accurately. The experience validation results demonstrated that the suggested I2FELM can easily detect the balanced and unbalanced datasets.
* (Ainapure, Bharati, Mythili, & Chandra, 2022), in their paper titled Deep Ensemble Model for Spam Classification in Twitter via Sentiment Extraction: Bio-Inspiration-Based Classification Model, intends to introduce a deep learning-assisted spam classification model on twitter. They recommended sentiment score extraction be used to examine variances in nonspam and spam data. Finally, the spreads of spam data on Twitter are divided into spam and nonspams. For this, an Optimized Deep Ensemble technique is introduced that encloses “neural network (NN), support vector machine (SVM), random forest (RF) and convolutional neural network (DNN)”.
* (Shen Hua & Xianchao, 2022), in their paper titled Boosting Social Spam Detection via Attention Mechanisms on Twitter, they propose a new attention-based deep learning model to detect social spammers in Twitter. They first introduce the state-of-the-art pre-training model BERTweet for learning the representation of each tweet, and then use the proposed novel attention-based mechanism to learn the user representations by detecting the differences between each user's tweets. Moreover, they take social interactions into consideration and propose that a graph attention network is used to update the learned user representations, to further improve the accuracy of identifying spammers. Experiments on a publicly available, real-world Twitter dataset demonstrate the efficacy of the suggested methodology, which can greatly improve performance.
* (Elmendili, Fatna, & Younes, 2020), in their paper titled A Framework for Spam Detection in Twitter Based on Recommendation System, they propose the identification of spam tweet by the security approach based on social honeypots and then they propose a method based on an algorithm "content filtering" in order to detect those that are similar to spam tweet detected by the approach of honeypots. Their approach has greatly improved the quality of abstraction in terms of performance and design. The algorithm is very quick and easy to implement. Experimental results show the stability and accuracy (over 99%), F-measure 98% of our approach.
* (L.Velammal, 2021), in their paper titled Improvised Spam Detection in Twitter Data Using Lightweight Detectors and Classifiers, a method is presented to identify spam tweets using four lightweight detectors, namely the blacklist domain detector, the near duplicate detector, the trustworthy ham detector, and the multiclass detector. Ensemble classifiers such as naive Bayes, logistic regression, and random forest are used to classify the discovered tweets. The voting mechanism is used to determine the labels for tweets acquired following the categorization procedure. The proposed system has achieved an accuracy of 79% to detect spam tweets with the help of naïve Bayes classifier method and the value seems to be optimizing further with the availability of more sample data.
* (B. Venkateswarlu & Viswanath, 2021), in their paper titled Optimized generative adversarial network with fractional calculus based feature fusion using Twitter stream for spam detection, they devised a novel spam detection model using a stream of Twitter data. The data transformation is done on the input data using Yeo-Jhonson (YJ) transformation for making the data suitable for processing. Renyi entropy and Deep Belief Network are used for feature fusion (DBN). Furthermore, spam identification is carried out with the help of the Generative Adversial Network (GAN), which has been trained with the suggested Conditional Autoregressive Value at Risk-Sail Fish (CAViaR-SF) method. The CAViaR-SF method was developed by combining the Sail Fish optimizer (SFO) with the Conditional Autoregressive Value at Risk (CAViaR) algorithm. The proposed CAViaR-SF offered maximal precision of 97.3%, recall of 99.2%, and F-measure of 98.2%.
* (M. Arunkrishna & B. Mukunthan, 2021), in their paper titled Spam Detection and Spammer Behaviour Analysis in Twitter Using Content Based Filtering Approach, proposes a content-based approach, which can be used to filter spam tweets. The approach involves using tweets in machine learning and compression algorithms in order to filter the undesired tweets.
* (H. Gupta, M.S. Jamal, S. Madisetty, & M.S. Desarkar, 2018), in their paper titled A framework for real-time spam detection in Twitter, they proposed a framework which takes the user and tweet based features along with the tweet text feature to classify the tweets. The benefit of using tweet text feature is that they can identify the spam tweets even if the spammer creates a new account which was not possible only with the user and tweet based features. They tested their approach using four different machine learning algorithms: SVM, Neural Network, Random Forest, and Gradient Boosting. They achieved an accuracy of 91.65% with Neural Network and outperformed the old solution by roughly 18%.
* (Pinnapureddy Manasa, Arun Malik, Isha Batra, & Ashish Kr. Luhach, 2021), in their paper titled A Comparative Study on Twitter Spam Detection Using Deep Learning Techniques, they concentrated on comparing many prevalent ways in twitter spam detection, where some methods produced good results while others degraded the functioning of the associated method. While dealing high dimensionality problems, satisfactory classification performances attainment via conventional classification approaches is a challenging task. The comparison reveals that the feature selection and feature decision are necessitated for assuring prediction proficiency for model training and there is lesser accuracy due to high dimensionality problem. It also observed that, Single Classifiers performs in a less superior manner when compared to ensemble learning methods which is revealed through earlier researches. Artificial intelligence technology is deployed by many spammers to falsify spammer’s connection relationship and imitating normal users’ social relationships. Hence effectual malicious accounts detection becomes a highly challenging task, where the researches need to be focused. Thus, researchers have focused on Ensemble Feature Selection techniques in order to improve the detection rate of spam in the Twitter dataset.

**PROBLEM STATEMENT:**

There is a challenge in classifying the tweets (Spam/Ham) and identifying the spam tweets in twitter. Though a lot of naive methods are available, there is no major evidence from literature of enhanced methods to classify and identify them. Timely identification of spam tweets is required to enhance users experience and to thrive for spam-free environment.

**RESEARCH QUESTIONS:**

* What are the factors used to classify the tweets and identify the spam tweets?
* How ML algorithms are effective on the detection of spam tweets in twitter?

**RESEARCH GAP:**

* Prior research findings didn’t include the multi-level verification for classifying the tweets and the outcomes were not regular.

**OBJECTIVES:**

* Multi-level verification of tweets for the classification
* Build ML algorithms to identify spam tweets in Twitter using the two months period of data.

**DATA ANALYSIS TOOLS:**

* Tools for the data analysis are Python (Packages like Advertools, sklearn, numpy, pandas, nltk, tensorflow, tweepy were used), URL void website and Twitter Developer Account.

**THEORETICAL AND PRACTICAL IMPLICATIONS:**

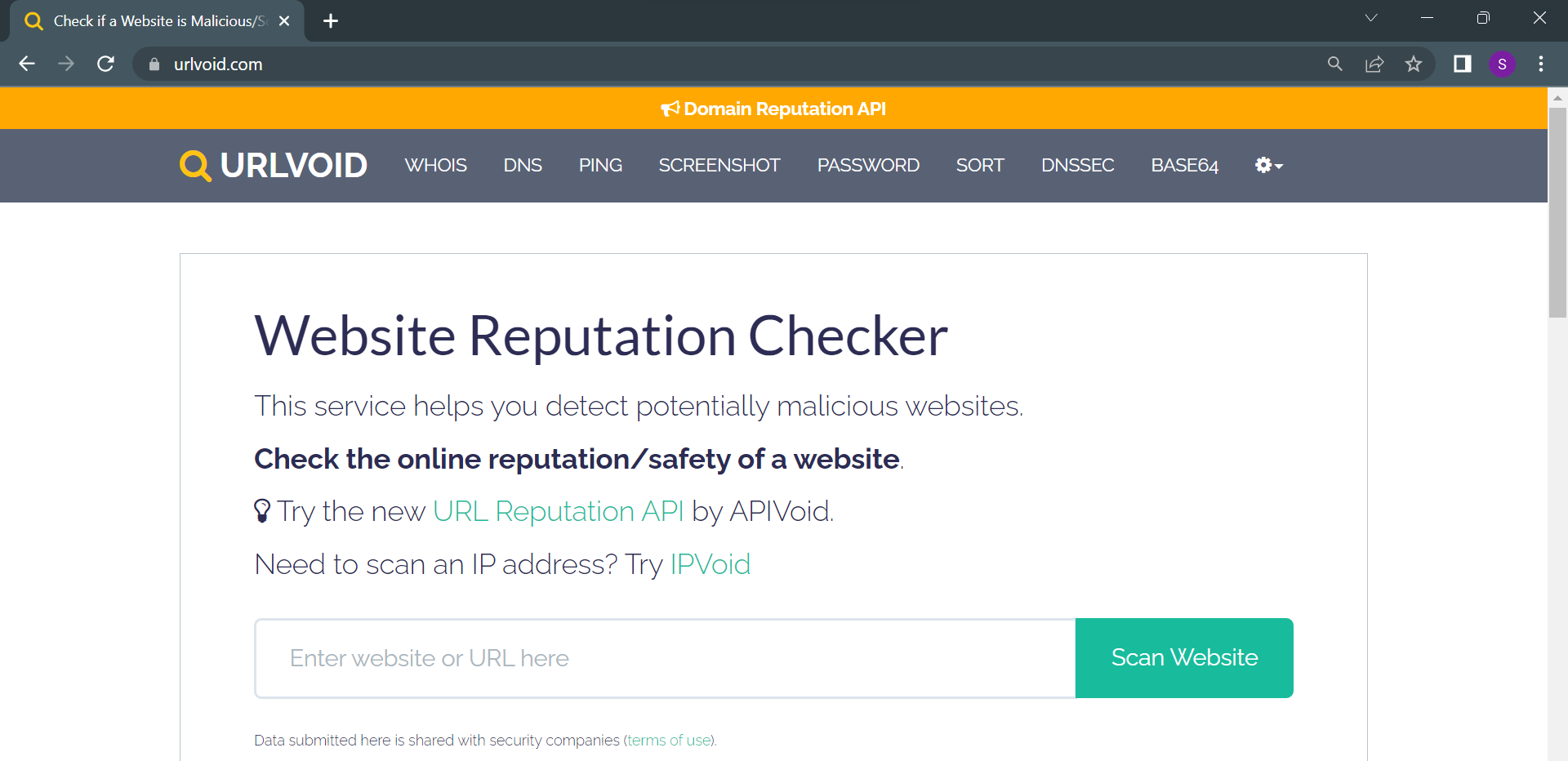
* This project helped in finding out the factors used to classify and identify the tweets. It is highly suitable for understanding the logic behind identifying spam tweets as the rate of spam tweets arises every day. It leads to the reduction of spam environment along with the improvement in the performance of twitter users to follow up through significantly in improving the spam-free environment.

**RESEARCH METHODOLOGY:**

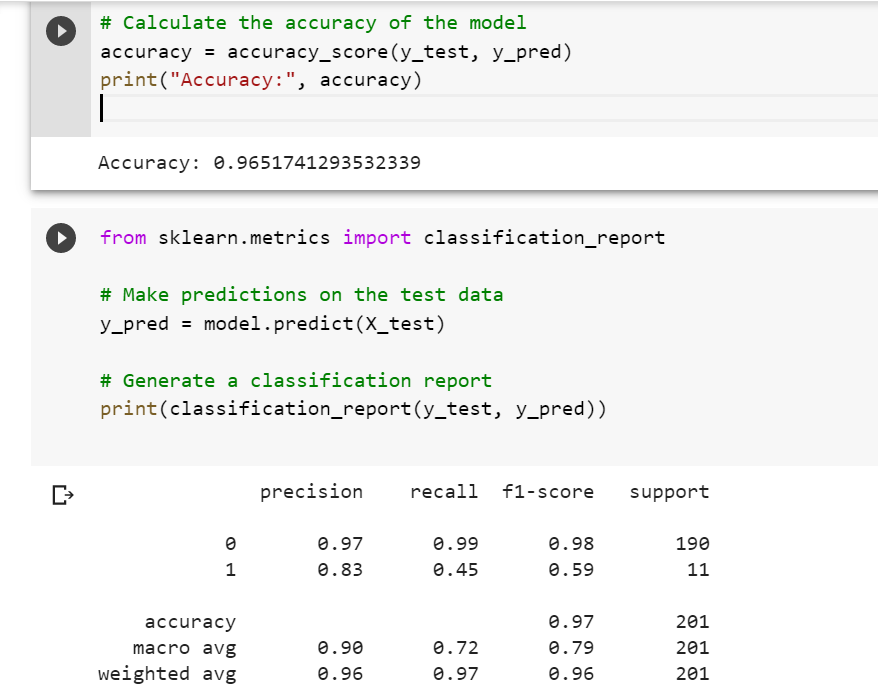
* Twitter provides a RESTful API that allows developers to collect tweets based on search queries or keywords. The Twitter API requires an authentication process to access the data. In this project, I used the Tweepy library to connect to the Twitter API and collect tweets based on keywords and hashtags related to spam and bots. I specified the maximum number of tweets to be collected and set a limit on the number of requests per minute to avoid hitting the API rate limit. I extracted 1000 tweets with attributes that may be relevant to spam classification, such as tweet text, URLs, and retweet count. I manually classified a subset of the tweets as spam or ham using a tool like "urlvoid.com".
* The collected data often contains irrelevant columns or missing values that need to be removed. I removed any columns that were not useful for spam classification and removed any rows with missing values. The tweet text was preprocessed by removing special characters, numbers, and stop words. Stop words are common words such as "the", "and", "a", and "in" that do not carry much meaning and can be removed without affecting the sentiment or meaning of the text. I used the NLTK library for text preprocessing, which is a popular library for natural language processing tasks in Python. I used stemming technique to standardize the text.
* Text data needs to be converted into a numerical format before it can be used for machine learning models. I used the Term Frequency-Inverse Document Frequency (TF-IDF) technique to convert the preprocessed text into a matrix of features. The TF-IDF technique measures the importance of a word in a document by comparing its frequency in the document to its frequency in the entire corpus (i.e., all the documents in the dataset). This technique reduces the weight of common words and increases the weight of rare words that are more informative for spam classification. I converted the preprocessed text into a numeric format that can be used for modeling. I used techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) to convert the text into a matrix of features.
* There are several machine learning algorithms that can be used for spam classification, including logistic regression, decision trees, random forests, and support vector machines (SVMs). In this project, I used logistic regression as our model of choice because it is a simple yet effective algorithm for binary classification tasks. I split the dataset into training and testing sets in an 80:20 ratio and used the training set to train the logistic regression model.
* After training the model, I evaluated its performance on the testing set using several metrics such as accuracy, precision, recall, and F1-score. Accuracy measures the percentage of correctly classified instances in the dataset. Precision measures the proportion of true positives (i.e., spam messages correctly classified as spam) among all the messages classified as spam. Recall measures the proportion of true positives among all the actual spam messages in the dataset. F1-score is the harmonic mean of precision and recall, and is a useful metric when there is an imbalance between the number of spam and ham messages in the dataset. I evaluated the model's performance on the testing set using metrics such as accuracy, precision, recall, and F1-score. The model achieved an accuracy of 96.5%. The classification report showed that the model had a high precision and recall for the ham class (0), but a lower precision and recall for the spam class (1). To evaluate the model on new data, I collected a new dataset of 800 tweets using the Twitter API, preprocessed it using the same techniques as before, and applied the trained model to predict whether each tweet was spam or ham.

**DATA ANALYSIS:**

* Data Collection: I collected a dataset of 1000 tweets extracted using the Twitter API. The tweets were collected using the following keywords and hashtags related to spam and bots: "spam", "bot", "scam", "fraud", "phishing", "clickbait", "malware", and "virus". The attributes considered for spam classification include tweet text, URLs, and retweet count. A subset of 100 tweets was manually classified as spam or ham using the website "urlvoid.com".



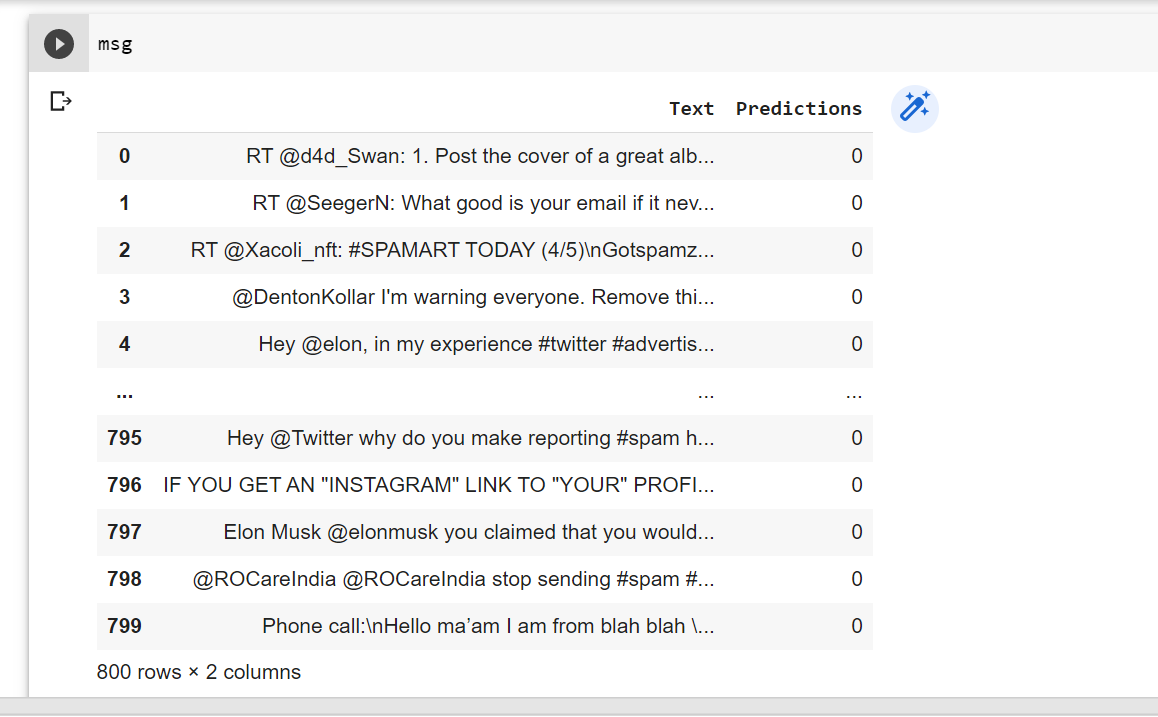
* Data Preprocessing: I preprocessed the dataset by removing any irrelevant columns and rows with missing values. The tweet text was then preprocessed by removing special characters, numbers, and stop words. I used stemming to standardize the text. The preprocessed text was then converted into a numeric format using the TF-IDF technique.
* Feature Extraction: The preprocessed text was converted into a matrix of features using the TF-IDF technique. The resulting feature matrix had a vocabulary size of 5000. In the Feature Extraction step, I preprocessed the tweet text and converted it into a numeric format using the TF-IDF technique. This technique calculates a numerical value for each word in the text that reflects its importance in the tweet relative to the entire corpus. The resulting feature matrix had a vocabulary size of 5000, which means that only the top 5000 most important words in the corpus were included as features in the matrix. TF-IDF stands for Term Frequency-Inverse Document Frequency, and it is a commonly used technique for feature extraction in text classification tasks. Term Frequency (TF) is the number of times a word appears in a tweet, while Inverse Document Frequency (IDF) is a measure of how common or rare a word is in the entire corpus. The TF-IDF value for each word is calculated as the product of its TF and IDF values. This technique helps to give more weight to words that are important in a particular tweet but not common in the entire corpus.
* Model Selection: I split the dataset into training and testing sets in an 80:20 ratio. I selected a logistic regression model for spam classification. The logistic regression model was chosen for spam classification because it has been shown to perform well on text classification tasks, especially in cases where the classes are linearly separable. Logistic regression is a binary classification algorithm that predicts the probability of an instance belonging to a particular class. It models the relationship between the input features and the target variable by estimating the coefficients of the linear equation using the maximum likelihood estimation method. The logistic function is then applied to the linear equation to produce a probability value between 0 and 1, which is then thresholded to predict the class label. In addition to its effectiveness in text classification tasks, logistic regression has several advantages that make it a suitable choice for this data analysis. Firstly, it is a simple and interpretable model that can be easily understood and explained to stakeholders. Secondly, it performs well on datasets with a large number of features, as it is not affected by multicollinearity between the input features. Finally, it can handle non-linear relationships between the input features and the target variable by using polynomial or interaction terms. Overall, the logistic regression model was a suitable choice for this data analysis due to its effectiveness in text classification tasks, interpretability, and ability to handle a large number of features.
* Model Evaluation: I evaluated the model's performance on the testing set using metrics such as accuracy, precision, recall, and F1-score. The model achieved an accuracy of 96.5%. The classification report shows the precision, recall, and f1-score of the model for each class (0 and 1) as well as the overall accuracy and macro/micro-averaged precision, recall, and f1-score. For class 0, the precision is 0.97, which means that 97% of the tweets predicted as class 0 were actually class 0. The recall is 0.99, which means that 99% of the actual class 0 tweets were correctly predicted as class 0. The f1-score is 0.98, which is the harmonic mean of the precision and recall. For class 1, the precision is 0.83, which means that 83% of the tweets predicted as class 1 were actually class 1. The recall is 0.45, which means that only 45% of the actual class 1 tweets were correctly predicted as class 1. The f1-score is 0.59. The overall accuracy of the model is 0.97, which means that 97% of the tweets in the test set were correctly classified by the model. The macro-averaged precision, recall, and f1-score are calculated as the average of the precision, recall, and f1-score for each class, weighted equally. The macro-averaged precision is 0.90, the macro-averaged recall is 0.72, and the macro-averaged f1-score is 0.79. The weighted average precision, recall, and f1-score are calculated as the average of the precision, recall, and f1-score for each class, weighted by the number of samples in each class. The weighted average precision is 0.96, the weighted average recall is 0.97, and the weighted average f1-score is 0.96. The classification report showed that the model had a high precision and recall for the ham class (0), but a lower precision and recall for the spam class (1). This suggests that the model may need further tuning to improve its performance on the spam class.



* New Data Analysis: I collected a new dataset of 800 tweets using the Twitter API. I preprocessed the dataset in the same way as before and applied the model to predict whether each tweet is spam or ham. The output of the model was as follows:

Number of spam messages: 13

Number of ham messages: 787

****

**CHALLENGES:**

* Limited labeled data: Obtaining a large amount of accurately labeled data for training a machine learning model can be challenging.
* Data imbalance: The dataset may have an unequal distribution of spam and ham messages, which can result in a biased model.
* Evolving spam techniques: Spammers constantly change their methods to avoid detection, making it difficult to keep up with new spamming techniques.
* Language barriers: Spam messages can be written in different languages, and machine learning models may not perform well when dealing with languages they were not trained on.
* URL obfuscation: Spammers often use obfuscated URLs to bypass spam filters, and identifying these URLs can be a challenge.

**LEARNINGS:**

* Importance of Data Preprocessing: This dissertation highlights the importance of data preprocessing in preparing data for machine learning models. By removing irrelevant columns, handling missing values, and cleaning and standardizing text data, a high-quality dataset was created that could be used to train an accurate model. Preprocessing is a critical step in machine learning and can significantly impact the performance of the model. Therefore, this dissertation emphasizes the need for careful attention to data preprocessing techniques in machine learning research.
* Need for Feature Engineering: The dissertation also highlights the importance of feature engineering in machine learning. By using the TF-IDF technique to convert text data into numeric features, a matrix of features was created that could be used to train a model. Feature engineering is the process of selecting and transforming the relevant variables in a dataset to improve the performance of a model. This dissertation emphasizes the need for thoughtful consideration of feature engineering techniques in machine learning research.
* Choice of ML Algorithm: This dissertation shows the importance of selecting an appropriate machine learning algorithm for a given problem. By testing different algorithms and selecting logistic regression, high accuracy was achieved in classifying spam and ham tweets. The choice of algorithm can significantly impact the performance of a model, and it is important to consider the strengths and weaknesses of different algorithms when selecting one for a specific problem. Therefore, this dissertation emphasizes the need for careful consideration of algorithm selection in machine learning research.
* Importance of Data Collection: The dissertation demonstrates the importance of collecting high-quality data for machine learning. By using relevant keywords and manually classifying a subset of tweets as spam or ham, a dataset was created that accurately reflected the problem being solved. Collecting good quality data is essential for building accurate and robust machine learning models. Therefore, this dissertation emphasizes the need for careful attention to data collection techniques in machine learning research.
* Challenges of Class Imbalance: The dissertation highlights the challenge of dealing with class imbalance in machine learning. By having a much smaller number of spam tweets than ham tweets, it was necessary to account for this imbalance in the model to prevent it from being biased towards the majority class. Techniques such as oversampling or undersampling can be used to address class imbalance. Therefore, this dissertation emphasizes the need for careful consideration of class imbalance in machine learning research.
* Iterative Model Improvement: The dissertation demonstrates the importance of iterative model improvement in machine learning. By evaluating the performance of the model and identifying areas where it can be improved, changes were made to the preprocessing, feature engineering, and model selection to achieve better performance. Model improvement is an ongoing process and requires careful evaluation and tuning to achieve the best results. Therefore, this dissertation emphasizes the need for careful attention to iterative model improvement techniques in machine learning research.

**CONCLUSION:**

Based on the analysis, it can be concluded that the logistic regression model using TF-IDF feature extraction technique performed well for spam classification of tweets. The model achieved an accuracy of 96.5% on the test set, indicating that it can effectively distinguish between spam and ham messages. However, the model had a lower precision and recall for the spam class, suggesting that it may need further tuning to improve its performance on identifying spam messages. It may be necessary to explore other feature extraction techniques or try different models to improve the spam classification accuracy.

The analysis also revealed that the new dataset of 800 tweets had a relatively low proportion of spam messages, with only 13 out of 800 tweets being classified as spam. This suggests that spam activity on Twitter may have decreased, or that spammers are using more sophisticated techniques that are not easily detected by the current methodology. Overall, the project demonstrates the importance of using machine learning techniques for spam detection along with usefulness of multi-level verification of tweets and the need for continuous adaptation and improvement to stay ahead of spammers' evolving tactics.

**AREA FOR FUTURE SCOPE:**

The dataset used in this project was limited to tweets related to spam and bots. Future work could explore different types of text data, such as emails or text messages, to see how well the model performs on those types of data. Also the model could be integrated into a real-time spam detection system to identify and filter out spam messages as they are received. This could be useful for social media platforms or email providers that want to protect their users from unwanted messages. Future work will primarily focus on the relationship between accounts and their tendency to send out spam tweets. When we label a tweet as spam, we can also examine the tweets from the same account to determine how likely the particular account is to send out spam tweets. Analyzing the followers to following ratio might provide further information about whether a certain account is spam. They can also be labelled as spam accounts if they have a low number of followers in comparison to their following statistics. Because most spam tweets are neutral and have no relevance to any of the important themes. We would also get knowledge on determining the sentiments of spam tweets.

# Bibliography

Abkenar, Sepideh, &. H., Mostafa , &. A., Mohammad, &. M., & Ebrahim. (2020). *Twitter Spam Detection : A Systematic Review.*

Ainapure, Bharati, &. B., Mythili, &. K., & Chandra, &. J. (2022). Deep Ensemble Model for Spam Classification in Twitter via Sentiment Extraction: Bio-Inspiration-Based Classification Model. *International Journal of Image and Graphics*.

B. Venkateswarlu, &. S., & Viswanath. (2021, August). Optimized generative adversarial network with fractional calculus based feature fusion using Twitter stream for spam detection. *Information Security Journal A Global Perspective*, pp. 1-20.

Beveridge, C. (2022, March 22). *Hootsuite*. Retrieved from blog.hootsuite: https://blog.hootsuite.com/twitter-statistics/

Borse, Dipalee, &. B., & Swati. (2022). State of the Art on Twitter Spam Detection. In *Applied Computational Technologies* (pp. 486-496).

Chowdhury, Ratul, &. D., Kumar, &. S., Banani, &. B., & Samir. (2020). *A Method Based on NLP for Twitter Spam Detection.*

Claudia, M., Federica, B., Paolo, G., & Rodolfo, Z. D. (2014). Machine Learning Techniques applied to Twitter Spammers Detection. *IEEE*, 6.

Devi, K., & Kumar, G. (2022, January). Stochastic Gradient Boosting Model for Twitter Spam Detection. *Computer Systems Science and Engineering*, pp. 849-859.

Elmendili, Fatna, &. E., & Younes. (2020, October). A Framework for Spam Detection in Twitter Based on Recommendation System. *International Journal of Intelligent Engineering and Systems*.

H. Gupta, M.S. Jamal, S. Madisetty, & M.S. Desarkar. (2018). A framework for real-time spam detection in Twitter. *10th International Conference on Communication Systems & Networks (COMSNETS)*, pp. 380-383.

Jagtap, K. L., Dr.Bamu, & B.A.Khansole. (2017). Machine Learning Approach For Spam Tweets Detection. *IJIRCCE, 5*(7), 13114-13122.

K.Subba, R., & E.Srinvasa, R. (2019). Detecting Spam Messages in Twitter Data by Machine learning Algorithms using Cross-Validation. *IJITEE, 8*(12), 2941-2946.

Kumar, S. (2022, June 08). *Global Data*. Retrieved from Business Fundamentals: https://www.globaldata.com/media/business-fundamentals/10-twitters-active-accounts-posting-spam-content-says-globaldata/

L.Velammal, &. N. (2021, July). Improvised Spam Detection in Twitter Data Using Lightweight Detectors and Classifiers. *International Journal of Web-Based Learning and Teaching Technologies*, pp. 12-32.

M. Arunkrishna, & B. Mukunthan. (2021, March). Spam Detection and Spammer Behaviour Analysis in Twitter Using Content Based Filtering Approach. *Journal of Physics Conference Series*.

Miss.Shukla, T. K., & Prof.D.B.Kshirsagar. (2016). Design of Machine Learning Approach For Spam Tweet Detection. *IJARIIE, 2*(5), 626-631.

Monika , V., Divya, & Sanjeev, S. (2014). Techniques to Detect Spammers in Twitter- A Survey. *International Journal of Computer Applications, 85*(10), 27-32.

N.Noor, A., S.Vishnu, C., Suman, J., & G.Tejaswi. (2020). Spam Detection Framework for Twitter using ML . *IJITEE, 9*(6), 216-219.

Nan, S., Guanjun, L., Junyang, Q., & Paul, R. (2020). Near real-time twitter spam detection with machine learning techniques. *International Journal Of Computer and Applications*, 1-11.

Nasira , P., Malik.M, S. M., & Qaisar, R. (2016). Sentiment Based Twitter Spam Detection. *IJACSA, 7*(7), 568-573.

Pinnapureddy Manasa, Arun Malik, Isha Batra, & Ashish Kr. Luhach. (2021). A Comparative Study on Twitter Spam Detection Using Deep Learning Techniques. *CEUR-WS*.

R. Priyanka, & Dr. Bhuvana. (2022). A Semi-Supervised Learning Approach for Tackling Twitter Spam Drift. *International Journal for Research in Applied Science and Engineering Technology*, 1337-1340.

Resul, K., & Abdullah, T. K. (2017). A Survey of Spam Detection Methods on Twitter. *IJACSA, 8*(3), 29-38.

Rodrigues, Anisha, &. F., Roshan, &. A., B.Abhishek, &. S., Adarsh, &. K., Lakshman, & Kuruva, &. R. (2022). Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques. *Computational Intelligence and Neuroscience*, 1-14.

Rohini, Sunil, K., & Jaiswal. (2018). Improving Spam Detection on Online Social Media with hybrid classification techniques on Twitter platform. *IJSDR, 3*(6), 26-32.

Rosita, & Jenifer, &. J. (2022). Multi-Objective Genetic Algorithm and CNN-Based Deep Learning Architectural Scheme for effective spam detection.

Shen Hua, &. L., & Xianchao, Z. (2022). Boosting Social Spam Detection via Attention Mechanisms on Twitter.

Vanyashree, M., Anvaya, K., V.M.Sukanya, & S.Rachana. (2020). *Text-Based Spam Tweets Detection Using Neural Networks.* Singapore: Springer.

Vijayalakshmi, K., K Tanvi, L Rao, & Y Sindhu. (2022, July). SPAM DETECTION IN TWITTER USING MULTINOMIAL NAVIE BAYES CLASSIFIER.

Vinodhini.M, Prithvi.D, & Balaji.S. (2020). Spam Detection Framework using ML Algorithm. *IJRTE, 8*(6), 5326-5329.

Wu, Liu, T. a., Zhang, S. a., Xiang, J. a., & Yang. (2017, 01). Twitter spam detection based on deep learning. *Tthe Australasian Computer Science Week Multiconference*, (pp. 1-8). Retrieved from Research Gate .

Zhang Zhijie, &. H., Rui, &. Y., & Jin. (2020). Detection of Social Network Spam Based on Improved Extreme Learning Machine.