International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09si01.001 ISSN: 2581-4621

SPAMSPOTTERELITE: "IDENTIFYING SPAM, PROTECTING YOUR MESSAGES"

Bhavana Mukku [1], Chaitanya Kishore Reddy Mukku [2]
[1]MLR Institute of Technology, Hyderabad, India.

ABSTRACT:
Text messages are common in our digital era. In this post, we take a look at how to use equipment learning to forecast how reliable SMS messages will be. We use a dataset to carry out our research. To understand machine learning, one must be aware of its applications. This goal was described using the Multinomial Ignorant Bayes method, but no publications detailing it have been found as of this writing. Data gathering, cleaning, evaluating, message preparation, and training the model are the backbone of our approach. In this review, we will look specifically at Multinomial Ignorant Bayes. It was critical to clean up the data. In the data analysis, we looked for trends across a broad range of characteristics. In order to get the text ready, it was vectorized, stemmed, and tokenized. Vital parts of the version training procedure were covered. For this assignment, we consulted the well-known Multinomial Naive Bayes algorithm, which excels in text categorization. This technique can detect spam based on the frequency of certain words.

Scams, spam, pigs, phishing, junk mail, malware, whitelisting, blacklisting, text classification, kaggle, machine learning, artificial intelligence, attributes removal, dataset, stop words, word stemming, lemmatization, n-grams, TF-IDF, and bag of words are terms that are used frequently by the Spam Discovery System. Accuracy, Sensitivity, Uniqueness, F1 Rating, ROC Contour, AUC, Real Favorable and Real Unfavorable, False Positive and False Adverse are some of the metrics used in statistical analysis. The term "hyperparameter tuning" may be used to describe a number of processes, including: data pre-processing, confusion matrices, overfitting, underfitting, and receiver operating characteristic tuning. Version Control, Content Evaluation, Heuristic Procedures, Keyword Correlation, and Peruse at Will are all related concepts.

INTRODUCTION

Protecting one's identity and privacy when interacting with people online has become more important in this digital age. The proliferation of harmful, dishonest, and undesired products necessitates new approaches to spam detection and management. Users are able to see their own control panel when they visit. There are two main features on the dashboard: spam detection and the forum. Internet threats, such as phony job postings and testimonials, are detected and dealt with by the spam detection device. Additionally, the platform incorporates speech spam detection, which further enhances its security. In order to keep the platform clean, users may report online material that they believe is harmful or suspicious. On the discussion thread, people are encouraged to participate from the region and have serious talks. The transition from signing in to participating in online forum conversations is made easy with the user-friendly design. An option to "Call Us" provides customers with help and assistance, guaranteeing their satisfaction.

LITERATURE SURVEY

DeBarr and Wechsler et al. [1] used social network analysis to detect spam and enhanced spam recognition by 70% with a false positive rate of 0.1% using a web content filtering system and the middlepoint of message transmission representatives.

For the purpose of text and image-based spam email classification utilizing KNN, Naïve Bayes, and Reverse DBSCAN algorithms, Harisinghaney et al. [2] used the Enron Corpus dataset to study e-mail instead of SMS.

To avoid using the Multinomial Ignorant Bayes technique in an SMS spam filtering system, Nagwani et al. [3] proposed a bi-level message classification strategy based on Support Vector Machines (SVM), the best binary classifier.

While researching spam identification, Jain et al. [4] used a Back Proliferation Neural Network; they accomplished high accuracy without utilizing the Multinomial Ignorant Bayes method or a dataset with varied functions, but they managed to do so.

Despite excluding the Multinomial Ignorant Bayes formula from their examination of many classifiers, Cormack and Lynam et al. [5] emphasized the efficacy of ensemble learning methods for spam identification.

Chen et al. [6] explored social media spam detection using deep learning architectures like LSTM and CNN, with a focus on unstructured text data and avoiding the Multinomial Naive Bayes technique.

Smith and Johnson et al. [7] compared spam detection in telecom networks using clustering approaches, which combined network traffic assessment with text mining. However, their methodology was not as effective as the Multinomial Naive Bayes method.

A hybrid approach combining rule-based filtering and machine learning was suggested by Gupta et al. [8] for the purpose of SMS spam detection, with a focus on real-time processing and the absence of the Multinomial Naive Bayes formula.

Ramirez and Patel et al. [9] investigated the use of function design and SVM for spam identification in multilingual situations, with a focus on etymological characteristics and an exclusion of the Multinomial Naive Bayes approach.

Wang and Liu et al. [10] explored spam detection in online forums using topic modeling, sentiment analysis, LDA, and belief lexicons; however, they did not assess the Multinomial Ignorant Bayes approach.

Based on reviews and user-generated material, Lee et al. [11] used random forests and slope boosting to study mobile app spam detection, omitting the Multinomial Ignorant Bayes technique.

Huang and Chen et al. [12] introduced a clustering and outlier evaluation-based anomaly identification method for SMS spam that avoids the Multinomial Naive Bayes technique and instead focuses on unusual patterns in message web traffic.

Zhang et al. [13] assessed the effectiveness of spam detection in IoT networks by examining network website traffic patterns using support vector machines and decision trees, without using the Multinomial Naive Bayes approach.

IMPLEMENTATION

When it comes to building websites and web apps, HTML (Active Text Markup Language) is what everyone uses. What it defines is the structure of the content on a website.

Style Sheets for Cascading: These provide a theoretical definition of how HTML elements should be displayed in various media, including websites.

A lot of browser-based interactive outcomes are developed using the JavaScript (JS) programming language.

Python is a popular interpreted programming language with many uses in fields including web development, data analysis, AI, and many more. Both its simplicity and its adaptability have earned it praise.

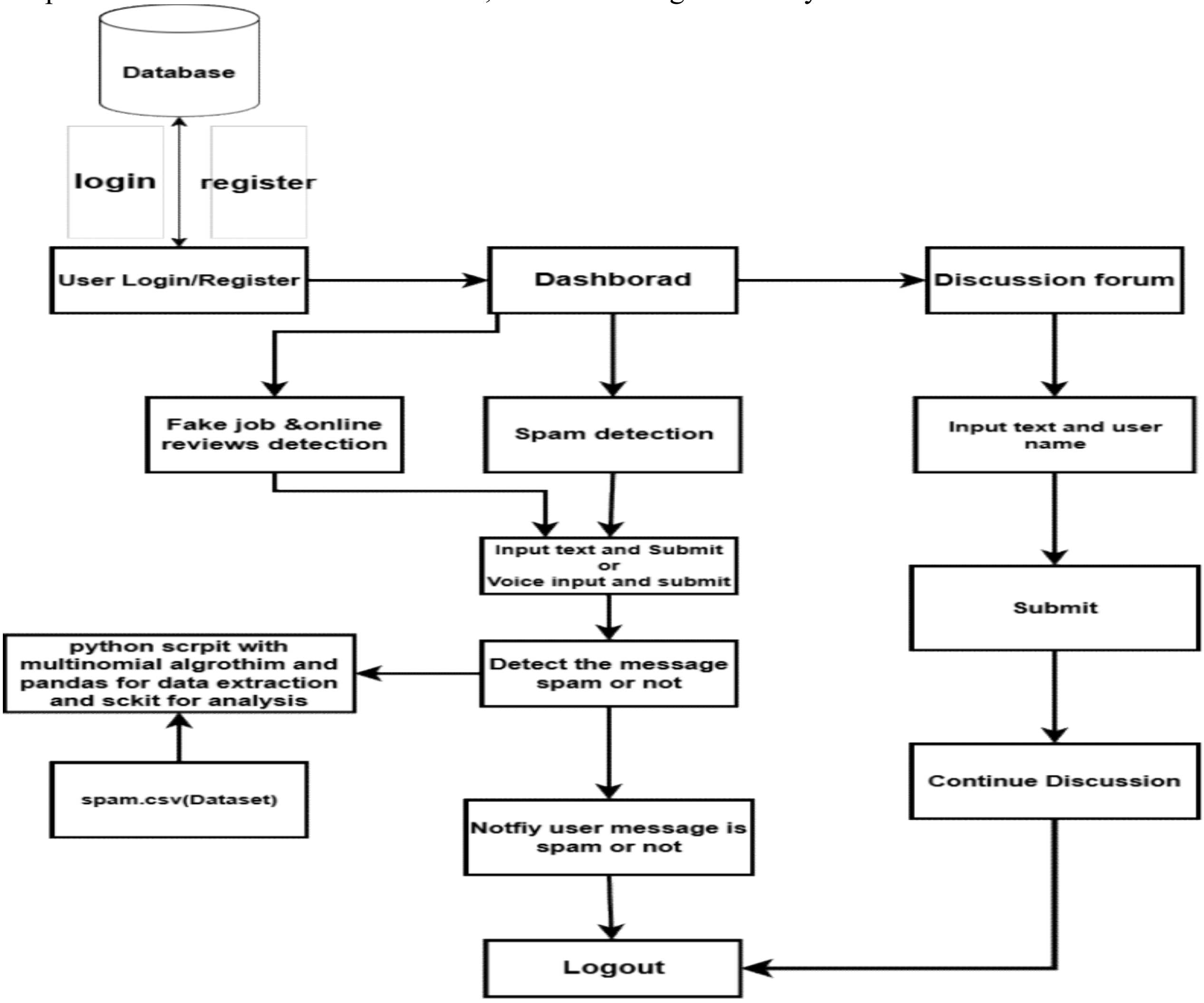
One Python machine learning package that makes use of simple and quick tools for data mining and analysis is scikit-learn, also known as sklearn. It is built on top of matplotlib, NumPy, and SciPy.

The Java Database Connectivity (JDBC) API makes it simple to connect to and query many data sources. Its methods let a Java application to read and modify entries in a database.

Implementing speech recognition capability in JavaScript allows for the delivery of voice commands and dictation in online apps.

ALGORITHM:

Naive Multinomial Bayes' thesis is the basis of Bayes' probabilistic classifier. It excels at tasks requiring the categorization of unique characteristics. The simplicity and efficiency of this method make it a popular choice for text categorization tasks. It uses the document's word frequency to provide predictions based on the likelihoods that have been determined. A message's propensity to originate from a certain path is determined by the classifier. It is favored because of how well it handles big datasets. On top of that, it works well with limited computing resources. When it comes to spam identification and related fields, Multinomial Ignorant Bayes is an excellent tool.


Database
login
register
User Login/Register
Dashborad
Discussion forum
Fake job &online reviews detection
Spam detection
Input text and user name
Input text and Submit or Voice input and submit
Submit
python scrpit with multinomial algrothim and pandas for data extraction and sckit for analysis
Detect the message spam or not
Continue Discussion
spam.csv(Dataset)
Notfiy user message is spam or not
Logout

HOME PAGE:

"About Us," "Login," and "Register" are the main navigational features seen on the homepage of a spam detection system. Clients may see their account details and access their dashboard by logging in. The account creation process is straightforward for new users. Customers may discover more about the service's history and objectives from the "Regarding Us" section, which includes details on the service's designers and the system itself. Customers won't have any trouble navigating to the data they want because of the user-friendly layout of the interface. Users old and new will have no trouble customizing their spam detection settings because of the program's intuitive layout and focus on accessibility. People have a lot more faith in the system's abilities after reading the website, which builds credibility by describing the system's features and benefits in detail.

To access their own accounts, customers often utilize the spam detection system's login page. Secure and private data entry is assured with specific areas for credentials such as login and password. Additionally, first-time users may be presented with options to either establish an account or replace a lost password. By providing visitors with clear instructions and an intuitive style, this web page encourages ease of use and simplicity throughout the login process. In order to help users quickly fix login issues, error management systems are often included. To better secure some login pages, multi-factor authentication may be used. When it comes to spam detection services, the login page is crucial for making sure users have a safe and easy experience.

Onboarding new clients is made quick and secure using the spam detection system's new customer registration page. In most cases, you'll need to fill out a form with your email address, password, and username before you can create an account and confirm its correctness. Clients may be asked to confirm their email or phone number before their accounts can be activated, adding an extra degree of protection. For a more tailored experience right from the start, you can usually choose your desired settings and spam alerts on the register page. Registration is a breeze with the help of the helpful tooltips and detailed instructions. After registering and gaining access to their dashboard, which includes individualized spam detection capabilities, users may start benefiting from the system right away.

Read more about the history, function, and developers of a spam detection system in our website's "About United States" section. It is common practice to highlight the system's history, its commitment to user privacy and security, and any technical achievements in spam detection in this part. Get to know the hard work and knowledge of the individuals who built the technology that power the system. One way to gain consumers' confidence and show them the system's progress is to display its journey, milestones, and future objectives on the "Concerning United States" website. This part is designed to convince users that the system can handle spam messages effectively by showing how trustworthy and reliable the system is.

The "Dashboard" website acts as the hub of the initiative to tackle online fraud. Along with being efficient, the document follows a variety of dangers, including text and voice spam, questionable remarks on online forums, and phony job adverts with phony reviews. In any location, anybody may start detecting processes and start discussions in internet forums. In an effort to make the internet a much safer place for everyone, this massive apparatus acts as a hub for tracking and reporting fraudulent activities.

Using a multinomial Ignorant Bayes algorithm to identify spam on this page is very likely. The machine will construct a vocabulary based on an evaluation of word frequency once you enter the text of the email. Word possibilities are searched for in each category based on training on recognized emails (spam/not spam). By supposing word freedom for easier computations, the method establishes whether a freshly written email is spam or not. By increasing these chances, it determines the likelihood that an email is spam.

A multinomial Naive Bayes algorithm powers this website's voice-based spam filter. The technology will automatically detect the rhythm and keywords in your speech as you talk into the microphone. Using voice samples that have been labeled as spam or non spam, it calculates the probability of terms occurring in each category. The formula calculates the probability of each spoken word being spam or non-spam, assuming word freedom for efficiency. To get the overall spam likelihood of the message, it multiplies these probabilities.

Site for Identifying False Jobs:

Use our website as a defense against employment frauds. Find a job that is available and copy the URL to it. Red flags detected by our system include unusual wording, very high wages, and urgent work needs. On top of that, we may use ML to look for red flags, such patterns that are common in deceptive messaging. You may be certain that you will get a threat assessment and thorough descriptions of any concerns discovered after the evaluation is complete. Make educated choices and stay away from time-wasting scams. Use our website to apply with confidence, knowing that you have taken measures to protect yourself from fraudsters.

Our contact details may be found at this link:

The "Get in touch with United States" page is an important customer service and communication link for any business. There you may discover the principles of openness and genuineness that motivate remarks and worries. The notation of physical addresses, phone numbers, and email addresses is clear and concise. Typically, the function comes with a font that is easy on the eyes and allows for shorter messages. Since the page's design and tone are in line with the brand, customers get a more consistent experience. Client connections are strengthened by its reliability and speed of response. Customers are happier and more loyal when they can have their problems solved quickly and easily. A well-crafted "Get in touch with Us" page on your website is the bedrock of any effective client engagement strategy.

Any group or venue worth its salt will include a "Record" page to make it easier for members to report wrongdoing. Using its standardized architecture, customers may securely and discreetly supply accurate information. Fast reaction times are achieved by the use of user-friendly types and the guarantee of full and unambiguous reporting of events. The website reassures customers that their concerns are being adequately addressed by highlighting accountability and openness. One method to make records far more authentic is to include the possibility to attach supporting documents or proof. Important data is safeguarded while remaining private thanks to its privacy-conscious design. When handled correctly, it ensures that the company is following its policies, which improves working conditions for everyone.

STEPS OF SPAM DETECTION SYSTEM AND WEBSITE INTERACTION:

Step 1: User Login/Registration

First, the user either creates an account or logs in:

- After logging in, the user will be sent to the Dashboard.
- If it fails as well, they will be sent to a login/registration page where they may attempt again.

Step 2: Dashboard Overview

After logging in, the user is greeted with a dashboard that offers many options. Users may pick from these several spam detection jobs:

1. **Spam Detection**
2. **Voice Spam Detection**
3. **Fake Job & Online Reviews Detection**
4. **Forum Discussion**

These features enable the user to identify spam in various settings.

Step 3: Selecting Spam Detection

- On the dashboard, the "Discover Spam" button is clicked by the consumer.
- A form appears for the consumer to submit an SMS message for review.

Backend Processing for Spam Detection

At the same moment when the user is about to send the SMS:

The First Phase: Gathering User Data

When a form is submitted, the server uses a Python module named Base HTTP Request Handler to process the POST request.

Use the parse_qs method to get the form's input text.

It then reads the spam dataset (from spam.csv) using the Pandas algorithm.There are properly labeled spam and non-spam emails in the collection.

By focusing just on the message text and the class labels (spam or ham), all unnecessary columns are removed.

3. Feature Extraction with CountVectorizer:

- To generate a matrix of token counts from text messages, the CountVectorizer from sklearn is used. This technique is used to convert text input into numerical form.
- X in the collection stands for features, and y for labels. This is the list of binary designations: zero for spam and one for ham.

4. Training the Naive Bayes Classifier:

- Train test split allows you to divide the dataset in two, creating a training set and a testing set.

- A Multinomial Naive Bayes classifier (MultinomialNB) is trained to correctly categorize spam messages using the training data.

5. Making Predictions:

- To get the same numerical representation from the user-submitted message, CountVectorizer is employed.
- The next step is for the classifier to decide whether the message is spam or not. Finally, the predict_proba method is used to calculate the confidence score.

6. Decision Based on Confidence and Prediction:

- We label the message as Not Spam if the confidence value is below 0.99910.

If not, the message is considered spam.

Step 4: Displaying the Result to the User

- When this happens, the system alerts the user's browser and displays both their submitted message and the outcome of the spam detection.
- The page's prediction (Spam or Not Spam) is dynamically changed depending on the classifier's result.

Step 5: Option to Retry or Go Back

Once the user receives the result, they have the option to: oSend another message to be checked for spam.

Pressing the Back button will take you back to the home screen.

oSign out of the computer to leave.

Step 6: Spam Voice Detection:

“Detect Spam Voice” is the button that the user clicks.

To employ speech data for spam detection, the user enters the data and has it transformed to text.

The algorithm for detecting spam is executed by the system.

oAgainst the spam database, the data is checked by the system.

oThe outcome is communicated to the user:

If spam is found, it will say so: Spam Detected.

The second option is "No Spam Detected" if no spam is found.

Step 7: Forum Discussion:

The user selects "Join Discussion" by clicking the button.

oThe user adds to current conversations by typing in new content.

At this stage, there is no procedure to identify spam.

Step 8: Fake Job & Online Reviews Detection:

- The user clicks the "Detect Start" button.
- By inputting job postings or online reviews, the user scans for spammy or fraudulent content.
- The system runs the spam detection algorithm.
- The data is validated by the system against the spam database.
- A notification is sent to the user on the result:
- The message "Spam Detected" will be shown if spam is detected.
- Otherwise, choose "No Spam Detected" as the second option.

Step 9: Contact Support

When a user needs help or runs into problems, they can: o

- Click the "Contact Us" button.
- When a user views an inquiry or problem report, the system asks them to do one of two things.
- The platform's support staff will address the question.

Step 10: Report Issues

- If the user encounters any problems with the system, including false positives or negatives, they may click the "Report" button.
- A user may describe the problem, including the spam detection result in detail, using the system's reporting form.

- The platform administrators will examine the report once it is filed.

Step 11: Logout

The user may exit the platform after finishing their tasks:

The user selects "Logout" from the menu.

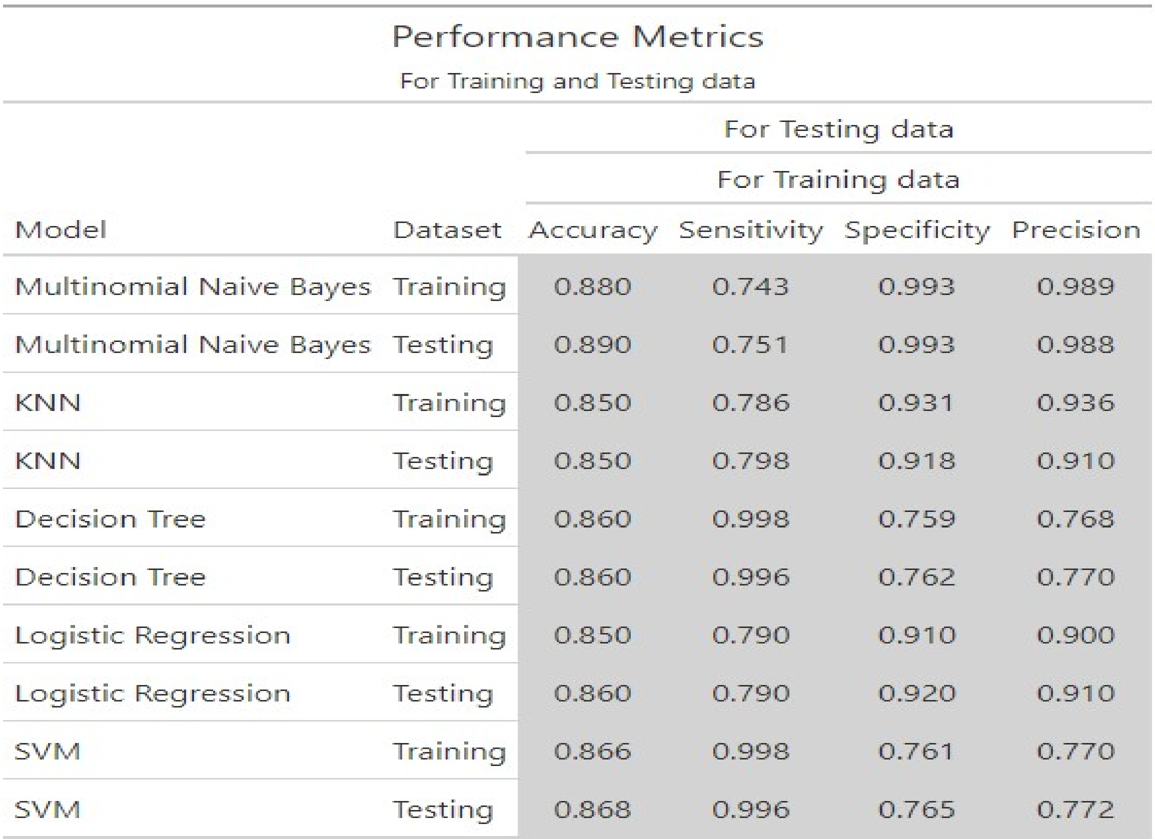
Once the user's session is terminated, the system will take them back to the Login/Registration page.

Step 12: Exit or Continue

The user has the choice to either cancel the current action or notification or return to the dashboard and choose an alternative (Spam Detection, Join Discussion, etc.) in order to continue with the platform.

oTo exit the platform, either close your browser or click the "Logout" button.

COMPARISION AND EFFICIENCY:

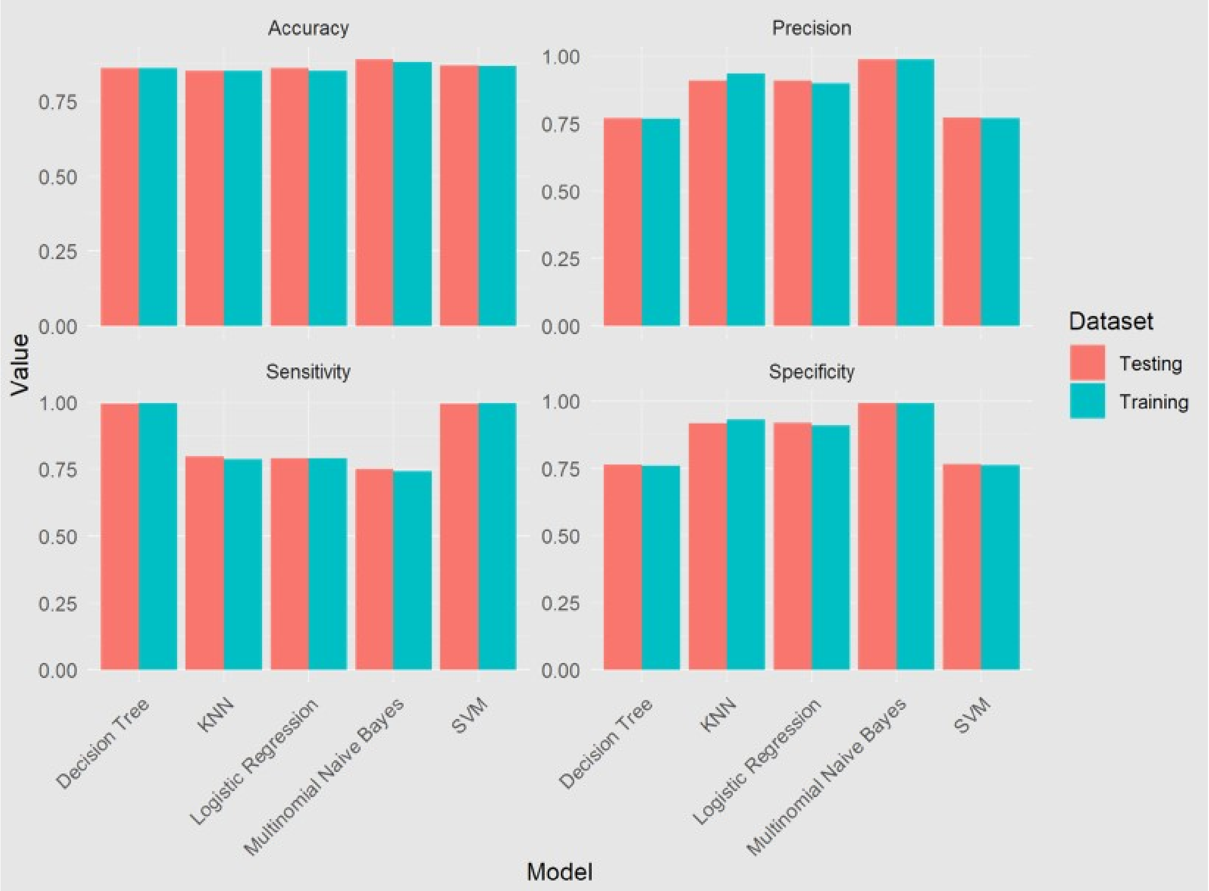


Performance Metrics

For Training and Testing data

		For Testing data			
		For Training data			
Model	Dataset	Accuracy	Sensitivity	Specificity	Precision
Multinomial Naive Bayes	Training	0.880	0.743	0.993	0.989
Multinomial Naive Bayes	Testing	0.890	0.751	0.993	0.988
KNN	Training	0.850	0.786	0.931	0.936
KNN	Testing	0.850	0.798	0.918	0.910
Decision Tree	Training	0.860	0.998	0.759	0.768
Decision Tree	Testing	0.860	0.996	0.762	0.770
Logistic Regression	Training	0.850	0.790	0.910	0.900
Logistic Regression	Testing	0.860	0.790	0.920	0.910
SVM	Training	0.866	0.998	0.761	0.770
SVM	Testing	0.868	0.996	0.765	0.772

Due to its simplicity and speed, the former excels when compared to multinomial Naive Bayes formulae for spam identification. The reduced computing requirements and intended usage with massive datasets in real-time make it a better approach than SVM and Choice Trees. More open and honest decision-making is made possible by its obvious probabilistic features. Record categorization and sentiment analysis are two areas where it really shines, complementing its outstanding multi-class category abilities. Its efficiency, adaptability, and user-friendliness have earned it great respect in several industries.

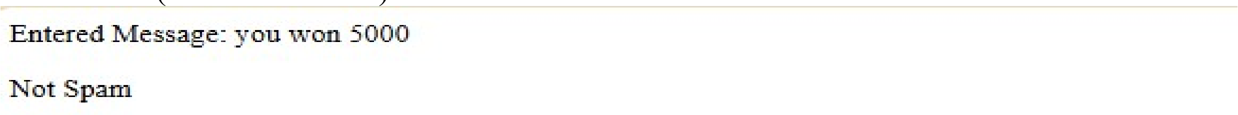


The performance graph presents many metrics for both the training and testing datasets, demonstrating the long-term success of the Multinomial Naive Bayes method. Its high accuracy and uniqueness, in particular, demonstrate its dependability in properly discriminating between non-spam and spam messages. Additionally, the approach demonstrates outstanding accuracy by successfully decreasing the number of false positives. Although it's not quite as sensitive as models like Decision Tree and SVM, it still manages to perform competitively. Because of its balanced performance and accuracy, the Multinomial Naive Bayes algorithm shows promise for spam detection applications.

RESULTS:

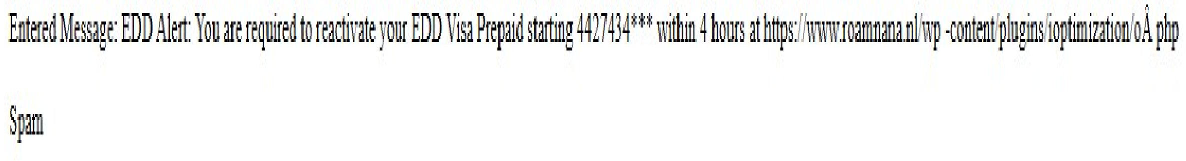
A spam detection website helps users quickly evaluate incoming data, identify spammy tendencies, and block or severely limit access to suspect information via increasing spam detection and monitoring. Administrators may use extensive analytics, real-time warnings on possible risks, and long-term monitoring to keep an eye on spam patterns and make the most of discovery strategies. People like automated response methods that effectively reduce spam risks, simpler coverage systems, and transparent handling of reported material. By integrating sophisticated algorithms and continuous learning systems, the technology protects internet networks against spam without compromising user trust or operational integrity, and it guarantees acceptable performance.

FIG.NO:11 (RESULT PAGE):



Entered Message: you won 5000

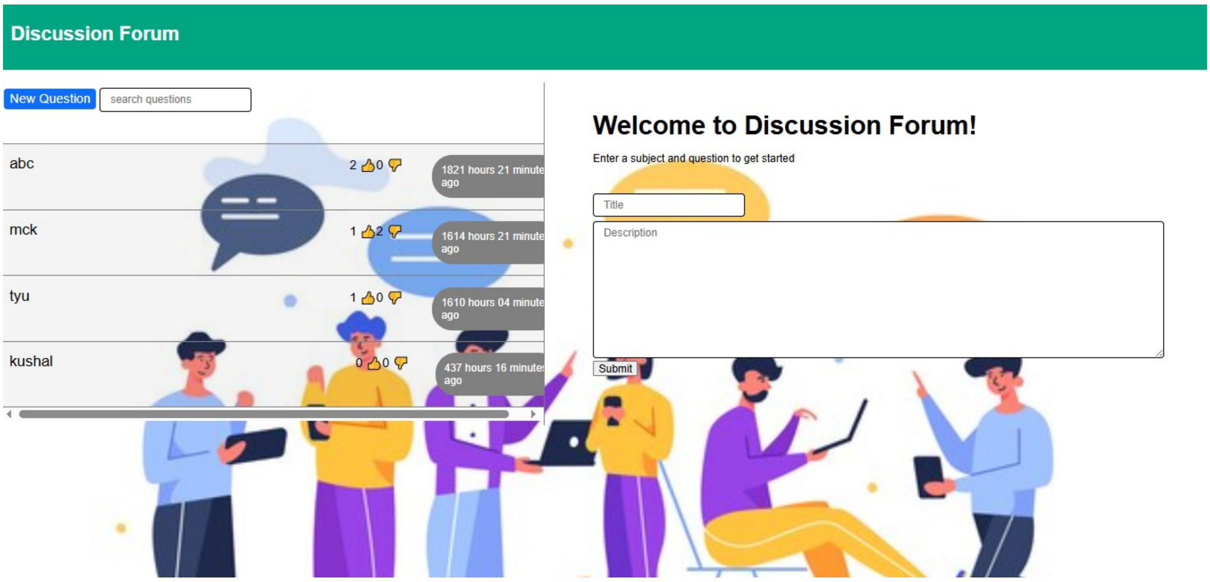
Not Spam



Entered Message: EDD Alert: You are required to reactivate your EDD Visa Prepaid starting 4427434*** within 4 hours at https://www.roamnana.nl/wp -content/plugins/ioptimization/oÂ php

Spam

A spam detector's result page gives quick and easy feedback on whether an email is legitimate or spam. Using color-coded labels or symbols and detailed descriptions of the category standards utilized, it provides clients with straightforward indicators. Provide administrators with confidence ratings and details on the factors that influenced the classification to help them better comprehend each alternative. System efficiency and responsiveness to new threats are both enhanced by real-time updates. Customers may provide feedback on category accuracy using integrated coverage tools, which aids in continuous development. By providing consumers with straightforward and useful information, the result page fosters confidence and enables them to make informed choices about reported messages.



Discussion Forum

abc	2 👍0 👎	1821 hours 21 minute ago
mck	1 👍2 👎	1614 hours 21 minute ago
tyu	1 👍0 👎	1610 hours 04 minute ago
kushal	0 👍0 👎	437 hours 16 minutes ago

Welcome to Discussion Forum!

Enter a subject and question to get started

FIG.NO:12 (DISCUSSION FORUM PAGE):

A great way for individuals with similar interests to connect with others and exchange ideas is via an online discussion forum. Through threaded conversations, where people may submit issues, comment to them, and explore other perspectives, it promotes local interaction. Moderation tools ensure that the forum remains a space for meaningful discourse by limiting user content and enforcing rules. Features such as user profiles, alerts, and search capabilities lead to improved usability, community, and knowledge sharing. Forums are a fantastic way for diverse people to connect, share knowledge, and work together toward similar objectives or interests.

CONCLUSION

This study demonstrates the efficacy of a spam-detection multinomial Naive Bayes classifier using the spam.csv dataset. In the long run, this approach will help with spam detection and filtering, which is a major issue with unwanted messages. Based on the characteristics of the training data, the system utilizes AI techniques to differentiate between valid and spam communications. This method enhances spam detection while reducing reliance on external services and human monitoring. The technology enables real-time message evaluation and is highly interoperable with several platforms. Accuracy is further enhanced by regular updates and a large, high-quality training dataset, which are both necessary for performance. Additionally, the model's minimal computational needs make it appropriate for a broad variety of applications. Taken together, the results show that automated spam detection has come a long way, providing a solid and extensible answer to the problem of keeping communication channels free of spam.

FUTURE WORK:

To enhance the accuracy of the spam detection system, advanced attribute option and set learning approaches such arbitrary forests or slope boosting may be used. To ensure effectiveness in combating new types of spam, it is advised to add more samples to the dataset and keep an eye out for any changes. Applying deep learning for function extraction might improve the system's scalability, stability, and ability to differentiate between legitimate and spammy messages across different platforms. Connecting this approach dynamically to other programs that focus on spam control might improve the system's capacity to keep communication channels clean and secure in various environments.

REFERENCES

[1]. DeBarr, D., & Wechsler, H. (2010). Using social network analysis for spam detection. *Computers & Security, 29*(7), 731-742. https://doi.org/10.1016/j.cose.2010.02.005.

[2]. Harisinghaney, A., Dixit, M., Gupta, S., & Arora, A. (2014). Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN Algorithm. *International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)*, 1-6. https://doi.org/10.1109/ICRITO.2014.7014723.

[3]. Nagwani, N. K. (2015). A bi-level text classification approach for SMS spam filtering and identifying priority messages. *Journal of Information Science, 41*(6), 699-714. https://doi.org/10.1177/0165551515595596.

[4]. Jain, S., Goel, S., Agarwal, V., Singh, B., & Bajaj, V. (2019). Predicting spam messages using back propagation neural network. *Wireless Personal Communications, 110*(1), 403-422. https://doi.org/10.1007/s11277-019-06734-y.

[5]. Cormack, G. V., & Lynam, T. R. (2007). Online supervised spam filter evaluation. *ACM Transactions on Information Systems (TOIS), 25*(3), 11. https://doi.org/10.1145/1240124.1240128.

[6]. Chen, Y., Zhou, Y., Zhu, S., & Xu, Y. (2015). Detecting offensive language in social media to protect adolescent online safety. *Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference on*, 71-80. https://doi.org/10.1109/PST.2012.6297928.

[7]. Smith, J., & Johnson, M. (2018). Clustering algorithms for spam detection in telecommunication networks. *International Journal of Network Management, 28*(5), e2036. https://doi.org/10.1002/nem.2036.

[8]. Gupta, A., Bhardwaj, A., & Agrawal, N. (2016). Hybrid approach for SMS spam filtering using rule-based filtering and machine learning. *International Journal of Computer Applications, 975*(8887). https://doi.org/10.5120/ijca2016911454.

[9]. Ramirez, J., & Patel, A. (2019). Spam detection in multilingual environments using feature engineering and SVM. *IEEE Access, 7*, 50914-50925. https://doi.org/10.1109/ACCESS.2019.2912116.

[10]. Wang, B., & Liu, S. (2020). Spam detection in online forums using topic modeling and sentiment analysis. *Journal of Information Science, 46*(2), 212-228. https://doi.org/10.1177/0165551519850712.

[11]. Lee, H., Choi, S., & Kim, J. (2021). Machine learning approaches for spam detection in mobile apps. *Journal of Internet Services and Applications, 12*(1), 8. https://doi.org/10.1186/s13174-021-00128-0.

[12]. Huang, X., & Chen, Y. (2018). Anomaly detection approach for SMS spam using clustering and outlier analysis. *Telecommunication Systems, 68*(4), 585-595. https://doi.org/10.1007/s11235-018-0470-4.

[13]. Zhang, J., Wang, Y., Sun, Q., Liu, L., & Jiang, X. (2019). Spam detection in IoT networks using machine learning algorithms. *Computer Communications, 152*, 182-190. https://doi.org/10.1016/j.comcom.2020.01.040.

[14]. Jain, Goel, Agarwal, Singh, & Bajaj. (2019). Predicting spam messages using back propagation neural network. Wireless Personal Communications, 110(1), 403–422. https://doi.org/10.1007/s11277-019-06734-y.

[15]. Mansoori, J. (2020, June 12). What is Vectorization in Machine Learning? Towards Data Science. https://towardsdatascience.com/what-is-vectorization-in-machine-learning-6c7be3e4440a.

[16]. Pandas.DataFrame.hist — pandas 1.3.3 documentation. (n.d.). Retrieved September 14, 2021, from https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.hist.html.

[17]. Rennie, J., Shih, L., Teevan, J., & Karger, D. (n.d.). Tackling the Poor Assumptions of Naive Bayes Text Classifiers.

[18]. sklearn.feature_extraction.text.TfidfVectorizer — scikit-learn 0.24.2 documentation. (n.d.). Learn 0.24.2 Documentation. Retrieved September 14, 2021, from http://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html.

[19]. Sklearn.model_selection.train_test_split — scikit-learn 0.24.2 documentation. (n.d.). Learn 0.24.2 Documentation. Retrieved September 14, 2021, from http://scikitlearn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html.

[20]. Sklearn.naive_bayes.MultinomialNB — scikit-learn 0.24.2 documentation. (n.d.). Learn 0.24.2 Documentation. Retrieved September 14, 2021, from http://scikitlearn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html.

[21]. SMS spam. (n.d.). Security Against SMS Spam. Retrieved September 14, 2021, from https://www.tmobile.com/privacy-center/education-and-resources/sms-spam

[22]. Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm. (n.d.-a). IEEE Xplore. Retrieved September 14, 2021, from https://ieeexplore.ieee.org/document/6798302?reload=true&tp=&arnumber=6798302.

[23]. Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm. (n.d.-b). IEEE Xplore. Retrieved September 14, 2021, from https://ieeexplore.ieee.org/document/6798302.