International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Optimizing SMS Spam Detection: Leveraging the Strength of a Voting Classifier Ensemble

[1]Manas Ranjan Bishi, [2]N Sardhak Manikanta, [3]G Hari Surya Bharadwaj, [4]P Siva Krishna Teja, [5]Dr G Rama Koteswara Rao

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: The paper addresses the challenge of SMS spam, which has seen a significant rise in recent years. This work proposes an ensemble learning` approach using Support Vector Machine (SVM), Naive Bayes, Extra Trees, and a Voting Classifier to enhance SMS spam detection. The system employs diverse machine learning algorithms, meticulously chosen and fine-tuned for optimal performance. The ensemble, centered on the Voting Classifier strengthened by a Random Forest Classifier, plays a crucial role in identifying spam messages. Evaluation is conducted using key metrics such as Accuracy, Precision, Recall, and F1-score to provide a comprehensive understanding of the model's effectiveness. Dataset exploration revealed unexpected dynamics, challenging initial assumptions. For instance, higher word counts were observed in spam messages, potentially reflecting strategic tactics employed by spammers. Additionally, the identification of over 6,000 duplicate texts within spam messages raises intriguing questions about spammers' methodologies. The development process incorporates meticulous data preprocessing steps like tokenization, lowercasing, and stop word removal. Rigorous training sessions with SVM, Naive Bayes, and Random Forest classifiers leverage their unique strengths, while the introduction of a voting ensemble method enhances the model's robustness and mitigates potential biases. The paper concludes by demonstrating the practical application of the SMS spam detector, achieving an accuracy of 94% through the combined application of various machine learning algorithms. This systematic and thoughtful approach positions the paper as a valuable contribution to the field of SMS spam detection, addressing real-world challenges in digital communication security.

Keywords: Phishing Offensives, Sophisticated, Machine Learning, Leverage, SMS Spamers.

1. Introduction

The SMS Spam Classifier project addresses the escalating issue of SMS spam inundating digital communication channels. In the realm of mobile communication, spam messages disrupt user experience, threaten privacy, and compromise security. Leveraging cutting-edge technology, the project employs machine learning (ML) and natural language processing (NLP) to construct a robust SMS spam classifier. Through sophisticated feature engineering, including word embeddings and Term Frequency-Inverse Document Frequency, a diverse array of ML algorithms and deep learning models are explored, ensuring adaptability to evolving spam tactics.

Real-time processing is a pivotal aspect of the project, swiftly categorizing incoming messages as spam or legitimate, reducing the risk of users encountering malicious content. The adaptive nature of the classifier, refined through continuous user feedback, enhances its accuracy and effectiveness over time. Alarming estimates by cybersecurity experts, such as Symantec, projecting a 30% increase in SMS spam in 2022, underscore the urgency of the issue. Phishing attempts, a prevalent form of SMS spam, pose serious threats to user security.

Reports from organizations like Kaspersky highlight the evolving sophistication of SMS spammers, necessitating ongoing research and development in SMS spam classification. The SMS Spam Classifier project not only aims to improve SMS communication but also exemplifies the practical application of AI in addressing real-world challenges.

In an environment where distinctions between legitimate and spam messages blur, the project serves as a crucial step toward fostering secure, efficient, and uninterrupted messaging. As the digital landscape evolves, the SMS Spam Classifier stands as a testament to the commitment to the safety and efficacy of digital communication, ensuring users can rely on their mobile channels without hesitation.

2. Literature Survey

Navaney, P., Dubey, G., & Rana, A. [1] "SMS Spam Filtering Using Supervised Machine Learning Algorithms," presented at the 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) in Noida, India. This work stands as a notable contribution in the field of SMS spam detection, focusing on the

[1,2,3,4,5]*Department of Computer Science and Information Technology, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, AP, India*
[1]*Email: manaslikegames@gmail.com*
[2]*Email: namana.sardhakmanikanta66@gmail.com*
[3]*Email: harisurya2500@gmail.com*
[4]*Email: sivakrishnatejapolisetty@gmail.com*
[5]*Email: drgrrao@kluniversity.in*

implementation of supervised machine learning algorithms to address the pervasive issue of SMS spam. By leveraging these algorithms, the research aims to refine and optimize spam filtering techniques within the realm of SMS communication.

The study delves into the exploration and application of supervised machine learning methodologies tailored explicitly for SMS spam filtering. Through meticulous analysis and experimentation, Navaney, Dubey, and Rana provide a comprehensive understanding of the efficacy and adaptability of supervised machine learning algorithms in combatting SMS spam.

Their work serves as a valuable resource for researchers, practitioners, and industry experts involved in mobile communication security. The methodologies and findings outlined in this paper contribute significantly to the ongoing pursuit of more robust and effective SMS spam filters, thereby enhancing the overall user experience and security in mobile communications.

Ubale, G., & Gaikwad, S. [2] "SMS Spam Detection Using TFIDF and Voting Classifier," presented at the 2022 International Mobile and Embedded Technology Conference (MECON) in Noida, India. This research paper focuses on SMS spam detection and employs TFIDF (Term Frequency-Inverse Document Frequency) coupled with a Voting Classifier. The study meticulously explores the practical implementation of these techniques, shedding light on their potential for enhancing the efficacy of existing spam detection mechanisms.

By leveraging TFIDF, the research underscores the pivotal role of feature extraction in distinguishing spam content from legitimate messages. Additionally, the integration of a Voting Classifier highlights the effectiveness of ensemble learning in bolstering the precision of spam detection models. This integration accentuates the importance of amalgamating diverse approaches for more robust and accurate spam identification systems.

Overall, the findings from Ubale and Gaikwad's investigation not only contribute to the ongoing refinement of spam detection algorithms in SMS communication but also emphasize the significance of leveraging feature extraction methods and ensemble techniques. Their insights offer promising pathways toward enhancing user experiences and fortifying security measures in mobile messaging platforms.

Subasi, A., Alzahrani, S., Aljuhani, A., & Aljedani, M. [3] "Comparison of Decision Tree Algorithms for Spam E-mail Filtering," presented at the 2018 1st International Conference on Computer Applications & Information Security (ICCAIS) in Riyadh, Saudi Arabia. This research undertook a meticulous comparative analysis, focusing on various decision tree algorithms employed in filtering spam emails. The primary objective was to discern the effectiveness and efficiency of these algorithms in combating the pervasive issue of email-based spam.

Through rigorous evaluation and comparison of diverse decision tree models, the study aimed to offer insights into their relative strengths and weaknesses in mitigating spam. By dissecting the performance metrics and nuances of these algorithms, the research provides valuable benchmarks and guidelines for researchers and practitioners in the realm of email security. The findings serve as a fundamental reference point, facilitating the development and optimization of robust spam filtering mechanisms.

Subasi et al.'s meticulous comparative analysis not only contributes to understanding the intricacies of decision tree algorithms in email spam filtering but also offers a foundational resource for shaping future advancements in email security. These insights aid in refining and optimizing the efficacy of spam detection systems, catering to the evolving challenges posed by email-based spam.

Panigrahi, P. K. [4] "A Comparative Study of Supervised Machine Learning Techniques for Spam E-mail Filtering," presented at the 2012 Fourth International Conference on Computational Intelligence and Communication Networks in Mathura, India. This comprehensive study scrutinizes various methodologies, aiming to discern their individual efficiencies in combating the pervasive issue of email-based spam. By meticulously comparing and contrasting these supervised machine learning techniques, the research dissects their strengths and weaknesses. Through detailed performance evaluations and analysis, Panigrahi's study offers profound insights into the nuanced efficacy of each method. These insights provide a comprehensive understanding of the diverse approaches available for spam email detection, assisting researchers and practitioners in making informed decisions about suitable techniques tailored to their specific requirements.

Panigrahi's exhaustive comparative analysis serves as a crucial reference point, illuminating the landscape of supervised machine learning techniques for spam email filtering. The research findings contribute significantly to the field by outlining the distinctive attributes of various methodologies, enabling the refinement and development of more robust and adaptive spam detection systems.

3. Why Is This Research Important?

The research on the SMS Spam Classifier is crucial for various compelling reasons, emphasizing its significance in the digital landscape. Primarily, the study addresses the escalating issue of SMS spam, a pervasive problem disrupting the lives of individuals worldwide. The project's scope involves developing a sophisticated machine learning (ML) and natural language processing (NLP) system,

emphasizing accurate classification to swiftly distinguish between spam and legitimate content.

Adaptability is a key feature of the project, crucial for addressing the ever-evolving tactics of spammers. This ensures the SMS Spam Classifier remains effective in detecting emerging spam patterns, providing a dynamic defense mechanism. The research's importance is underscored by its commitment to enhancing the user experience, notably reducing the inconvenience caused by SMS spam. Through systematic filtering, the project aims to offer users a more seamless and efficient messaging experience.

The research places a strong emphasis on user security, recognizing it as a paramount concern in the digital age. The SMS Spam Classifier acts as a formidable defense against phishing attempts, malicious links, and various cyber threats associated with SMS spam. This proactive security measure contributes to shielding users from personal data breaches, financial fraud, and the dissemination of malware, significantly bolstering the safety of the digital environment.

The project's adaptability is a crucial attribute, ensuring consistent protection for users against ever-evolving spamming tactics. The research exemplifies the practical application of artificial intelligence (AI) in addressing real-world challenges, showcasing the potential of advanced AI technologies, including ML and NLP, in mitigating pervasive issues like SMS spam.

Ultimately, the SMS Spam Classifier project contributes to fostering a safer and more efficient digital messaging environment. By mitigating the impact of SMS spam, it lessens the susceptibility of the digital realm to spam-related inconveniences, fraudulent activities, and security breaches. This endeavor aligns with the overarching mission of making the digital world a more secure and trustworthy space for all individuals. In summary, the research is vital as it addresses a real and pervasive challenge in the digital age, enhances individuals' lives, and contributes to the safety of online interactions.

4. Proposed System

The proposed SMS spam classification system integrates a diverse set of machine learning algorithms, including Naive Bayes (NB), Support Vector Machine (SVM), and Extra Trees, meticulously tuned for optimal performance. The inclusion of a Voting Classifier, employing ensemble learning with SVM, NB, and Extra Trees, reinforced by a Random Forest Classifier, enhances the system's robustness in distinguishing between spam and legitimate SMS messages, ensuring adaptability to the evolving digital communication landscape.

The initial application of the Naive Bayes algorithm reveals commendable accuracy in differentiating between spam and legitimate messages, with message length identified as a critical feature. However, an intriguing nuance is uncovered where shorter messages with specific token characteristics may be misclassified as spam. Learning from this, an in-depth analysis of the learning curve suggests a balance between training and test set errors, prompting a strategy shift towards bias mitigation and the introduction of more meaningful features to enhance overall classification accuracy.

The Support Vector Machine (SVM) significantly contributes to high-precision SMS spam classification, excelling in identifying nuanced patterns in the data. Its efficacy in mitigating evolving SMS spam tactics enhances the system's overall performance. The introduction of the Extra Trees algorithm, an ensemble learning technique, further strengthens the classification system's accuracy and resilience, particularly in the presence of noisy data.

The culmination of these efforts is the Voting Classifier, combining SVM, NB, and Extra Trees, with the support of the Random Forest Classifier as the final estimator. This ensemble technique maximizes the strengths of individual algorithms, creating a comprehensive and reliable SMS spam detection.

To enhance the interpretability of the models, brief explanations are provided, shedding light on the decision-making processes. Moreover, adjustments are made to address the issue of misclassifying shorter messages as spam by exploring additional features and implementing strategies for bias mitigation.

Validation metrics such as precision, recall, and F1-score are emphasized for a more quantitative measure of the model's effectiveness.

The system is designed to be adaptable, with considerations for future updates and fine-tuning to handle emerging trends or new spam tactics.

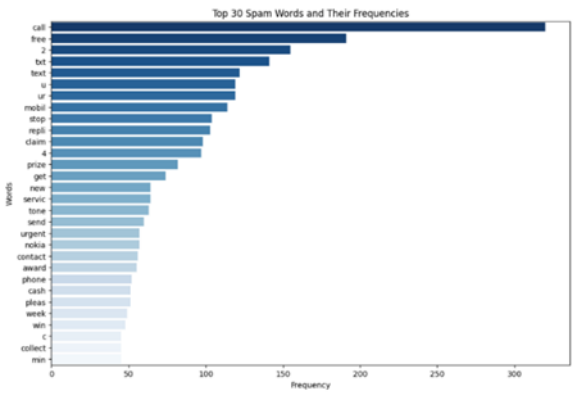

Fig.1. Process of SMS Spam Classification

4.1. Dataset

The SMS Spam Collection dataset serves as a valuable resource for the development of SMS spam classifiers, comprising a collection of SMS messages that have been meticulously curated for research purposes. Within this dataset, you'll find 5,574 SMS messages in English, each meticulously categorized as either "ham" (legitimate) or "spam."

Acquired from the UCI Machine Learning Repository and originally compiled by Almeida and Hidalgo [18], the public SMS dataset features 5,574 English messages categorized as legitimate (ham) or spam. Collected from diverse sources like the UK forum from the Grumble text website, NUS SMS Corpus (NSC), Caroline Tag's PhD Thesis, and SMS Spam Corpus v.0.1 Big.

Table. 1 Dataset Source Description

Source Description	Number of messages
Grumbletext Website: A collection of 425 manually extracted SMS spam messages.	425
NUS SMS Corpus (NSC): A subset of 3,375 randomly chosen ham messages.	3,375
Caroline Tag's PhD Thesis: A list of 450 SMS ham messages.	450
SMS Spam Corpus v.0.1 Big	1,324

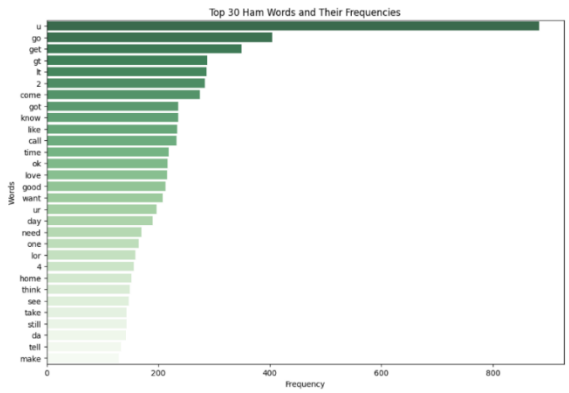

Fig.2. Top 30 Most Frequent Ham Words and Their Occurrences in the Dataset

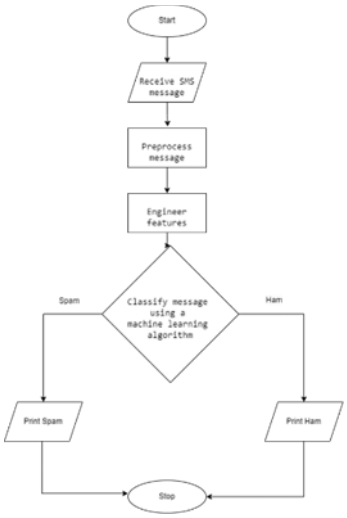

Fig.3. Top 30 Most Frequent Spam Words and Their Occurrences in the Dataset

Table.2 Dataset's Extremes: Maximum and Minimum Counts of Words and Sentences in Spam

	Number of words	Number of sentences
Min	2	1
Max	46	9

Table.3 Dataset's Extremes: Maximum and Minimum Counts of Words and Sentences in Ham

	Number of words	Number of sentences
Min	1	1
max	220	38

4.2. Process of SMS Spam classification

The process of SMS spam classification involves text preprocessing, feature extraction, model training using machine learning algorithms, and evaluation. Techniques such as natural language processing and ensemble methods contribute to the effective identification and filtering of spam messages from legitimate ones.

4.2.1. Receive SMS message

At the outset of the SMS classification process, the initial step is to receive the incoming SMS message. It marks the point at which the message is first accessed and enters the classification pipeline. Subsequent stages in the process, including preprocessing, feature extraction, and classification, depend on this crucial first step, making it the foundational element of the SMS message analysis and categorization procedure.

4.2.2. Preprocess message

Next, the stage involves preparing the SMS content for effective feature extraction. This step includes eliminating stop words, converting all text to lowercase, and stripping away punctuation marks. These actions streamline the text, ensuring subsequent analysis. By removing common words that hold little analytical value, standardizing the text to lowercase, and eliminating punctuation, the message becomes more refined. This preprocessing step enables clearer analysis and feature extraction, enhancing the accuracy and efficiency of downstream processes such as classification, sentiment analysis, or any text-related tasks within the domain of natural language processing.

4.2.3. Engineer features

Following preprocessing, features are derived from the message content. These features vary from basic aspects like word count or specific keywords to more intricate elements such as message sentiment or identified entities. The extraction process involves gathering crucial attributes that aid in the classification, enabling a comprehensive analysis of the SMS content for effective categorization.

4.2.4. Classify message using a ML algorithm.

The extracted features are employed to train a machine-learning algorithm for message classification. This process equips the algorithm to distinguish between spam and legitimate (ham) messages based on the learned patterns and attributes. By leveraging these features, the algorithm can efficiently categorize incoming SMS messages, aiding in the automated filtering and organization of messages for users.

4.2.5. Output prediction

After processing and analysis, the machine learning algorithm generates a prediction regarding the nature of the message, classifying it as either spam or ham. This prediction is the result of the algorithm's learned patterns and insights from the message's features, facilitating the automated determination of the message's legitimacy. Users can then benefit from accurate message categorization, helping to maintain inbox cleanliness and security against spam.

4.3. Feature Engineering

The For the SMS spam classifier, feature engineering involves lowercasing, tokenization, special character removal, eliminating stop words and punctuation, and applying stemming or lemmatization for text processing.

4.3.1. Conversion to Lowercase:

An initial and fundamental step in text preprocessing involves converting all text to lowercase. This standardizes the text representation, effectively eliminating potential issues stemming from variations in letter case. By doing so, the classifier can equate words with differing capitalization, thus enhancing its accuracy in SMS message classification.

4.3.2. Tokenization:

Tokenization is the practice of segmenting the text into individual tokens or words. This breakdown of text into its elemental components allows the classifier to inspect and process the content at a finer level. This serves as a cornerstone for extracting meaningful insights from SMS messages and forms the basis for subsequent analysis.

4.3.3. Removal of Special Characters:

SMS messages frequently contain special characters, including symbols, emojis, and non-alphanumeric characters. The elimination of these special characters is critical to simplify the text data and focus on the textual content.

4.3.4. Elimination of Stop Words and Punctuation:

Common words, recognized as stop words [B] (e.g., "the," "and" "in"), and punctuation marks, while integral to language, can introduce redundancy in text classification. The removal of these elements reduces noise and streamlines the feature selection process. The classifier can then concentrate on the more informative words in the text.

4.3.5. Stemming:

The final preprocessing step entails reducing words to their root forms, accomplished through stemming. Methods aim to ensure that variations of words are treated as a unified entity.

4.4. Model Evaluation Metrics

The SMS spam classifier's performance is assessed using metrics precision, recall, and F1-score. Precision measures accurate spam predictions, recall gauges identified spam instances, and F1-score balances both. High values indicate robust spam classification.

4.4.1. Accuracy

Accuracy serves as a metric that assesses the proportion of SMS messages that are correctly classified out of the entire test dataset. It provides a holistic measure of the classifier's effectiveness in discerning spam and legitimate messages. While achieving a high level of accuracy is a desirable objective, it should be interpreted in conjunction with other pertinent metrics to ensure a comprehensive evaluation.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

4.4.2. Precision

Precision stands as a pivotal metric that measures the classifier's proficiency in accurately identifying spam messages among those it designates as spam. A higher precision score indicates that when the classifier labels a message as spam, there is a greater likelihood that it is

indeed a correct classification. This minimizes the chances of false alarms or misclassifying legitimate messages as spam.

$$\Pr ecision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

4.4.3. F1-Score

F1-score is a crucial metric for evaluating the SMS spam classifier, striking a balance between precision and recall. It considers both false positives and false negatives, providing a comprehensive assessment of the model's overall effectiveness in identifying spam while minimizing errors and maintaining a harmonious blend of precision and recall.

$$\text{F1-Score} = \frac{2 \text{ X Precision X Recall}}{\text{Precision+Recall}} \quad (3)$$

4.4.4. Recall

Recall is a pivotal metric assessing the SMS spam classifier's effectiveness in capturing all actual spam messages. A higher recall score signifies the model's proficiency in identifying a larger proportion of spam instances, minimizing the risk of overlooking potential threats and ensuring comprehensive spam detection.

$$\Re call = \frac{TP}{TP + FN} \quad (4)$$

5. VOTE CLASSIFIER

The vote classifier, a dynamic ensemble technique, stands as a cornerstone in the realm of SMS spam classification, especially when coupled with diverse classifiers such as the Support Vector Classifier (SVC), Extra Trees Classifier (ETC), Naive Bayes (NB), and Random Forest Classifier (RFC). This ensemble method orchestrates a symphony of individual classifiers, each contributing its unique strengths, to create a more potent and reliable predictive model. Support Vector Classifier (SVC) stands out for its prowess in detecting intricate patterns, proving invaluable for capturing nuanced features that signal spam messages. Its ability to identify complex relationships within the data enhances the classifier's precision in discerning subtle variations indicative of spam content. On the other hand, the Extra Trees Classifier (ETC) contributes robustness to the ensemble through a randomized decision-making process. This strategy not only mitigates overfitting but also elevates the model's generalization capabilities, enabling it to effectively navigate the diverse landscape of spam instances. Together, SVC and ETC create a formidable synergy, optimizing the SMS spam classifier's ability to accurately identify and filter out unwanted messages. The Support Vector Classifier (SVC) is a linchpin in the SMS spam classification ensemble, showcasing exceptional proficiency in discerning intricate patterns within the data. Its strength lies in its ability to identify complex relationships and dependencies, making it an asset in capturing nuanced features characteristic of spam messages. In the realm of SMS communication, where spammers continuously devise sophisticated tactics, SVC's capacity to unravel subtle variations ensures a heightened sensitivity to the diverse nature of spam content. This precision contributes significantly to the classifier's ability to make accurate predictions, crucial for distinguishing between legitimate messages and potential spam. Complementing SVC, the Extra Trees Classifier (ETC) introduces a critical element of robustness to the ensemble. By adopting a randomized decision-making process, ETC addresses the challenge of overfitting, a common pitfall in machine learning models. This randomized approach involves constructing multiple decision trees and selecting the most robust outcomes. The result is a more resilient model with enhanced generalization capabilities, allowing it to effectively handle a wide array of diverse spam instances. The synergy between SVC and ETC within the ensemble not only improves the overall accuracy of the SMS spam classifier but also fortifies its adaptability, ensuring reliable performance in the face of evolving spamming techniques and patterns. The vote classifier's essence lies in harmonizing diverse strengths. Utilizing a democratic voting mechanism, each classifier contributes its perspective on SMS message classification. This ensures the final prediction isn't unduly swayed by any single model's idiosyncrasies but rather reflects a consensus among multiple algorithms. This collaborative approach amplifies overall accuracy and robustness, creating a more reliable and adaptive classification system capable of navigating the intricacies of varied spam instances in SMS communications.

This ensemble approach significantly fortifies the SMS spam classifier against the multifaceted challenges posed by spam messages. By combining the discerning eye of SVC, the robustness of ETC, the probabilistic modeling of NB, and the decision-making diversity of RFC, the vote classifier creates a comprehensive and adaptable defense against the ever-evolving tactics employed by spammers.

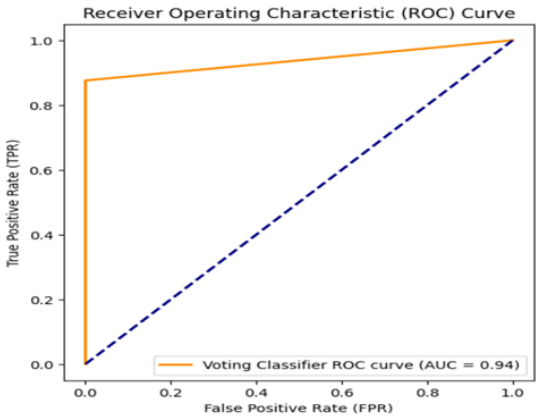

Fig.4. ROC Curve of Voting Classifier

The Receiver Operating Characteristic (ROC) curve is a graphical representation of the performance of a binary classifier, in this case, the Voting Classifier. It is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. TPR is also known as sensitivity or recall, representing the proportion of actual positive instances correctly identified, while FPR is the ratio of falsely identified negative instances. The resulting ROC curve is visualized using Matplotlib. The orange line represents the Voting Classifier's ROC curve, with its AUC annotated on the plot. The dashed navy line represents a random classifier, serving as a baseline. The curve's deviation from the diagonal line signifies the classifier's effectiveness, with a curve closer to the top-left corner indicating superior performance.

6. WORK DONE

In the comprehensive exploration and development of the SMS spam classifier dataset, a fascinating revelation emerged, unraveling intricate dynamics within the realm of spam and legitimate messages. Contrary to initial expectations, spam messages exhibited a higher word count. This unexpected finding suggested a strategic maneuver by spammers, employing verbosity as a means to establish credibility and manipulate recipients. In stark contrast, legitimate messages prioritized clarity and conciseness, adhering to a communication style that eschewed unnecessary embellishments.

Delving deeper into the dataset, a surprising discovery surfaced: a significant presence of over 6,000 duplicate texts within the realm of spam messages, surpassing occurrences observed in legitimate messages. This raised intriguing questions about the intent behind such duplications and hinted at potential tactics employed by spammers to saturate message volumes or enhance the perceived importance of their content.

In navigating the intricacies of SMS spam classification, the developed classifier adopted a systematic methodology, steering clear of advanced. The journey commenced with meticulous data preprocessing, involving essential steps such as tokenization, lowercasing, and the removal of stop words. These preparatory measures aimed to streamline the text data, facilitating subsequent feature extraction for effective classification.

Moving through subsequent stages, the classifier underwent rigorous model training, leveraging diverse classifiers including Support Vector Machines (SVM), Naive Bayes (NB), and Random Forest Classifier (RFC). Each classifier brought forth unique strengths to the overarching classification process. SVM, with its proficiency in capturing intricate patterns, NB's adeptness in probabilistic modeling, and RFC's decision-making diversity, collectively contributed to a well-rounded and robust classification framework.

To fortify the classifier's resilience and enhance its decision-making capacity, a voting ensemble method was introduced. This ensemble comprised SVM, NB, RFC, and other classifiers, collectively contributing to the final decision-making process. The voting classifier aimed to harness the collective intelligence of diverse models, potentially mitigating biases and enhancing overall performance by aggregating insights from multiple perspectives.

In the final stages of the SMS spam classifier's development, its practical utility shone through as it adeptly categorized new SMS messages as either spam or legitimate based on the extracted features. This classification methodology proved to be efficient and accurate in distinguishing between spam and legitimate messages, showcasing its potential for real-world applications.

In conclusion, the exploration and development of the SMS spam classifier underscored the significance of a systematic and thoughtful approach. By eschewing overly complex techniques in favor of a robust ensemble method, the classifier showcased its adaptability and efficacy in the face of evolving spam tactics. The integration of diverse classifiers, augmented by the introduction of a vote classifier, underscored the importance of collective intelligence in enhancing the classifier's performance for practical, real-world applications in SMS spam detection.

7. RESULTS

Table 4. Results acquired from Algorithms.

Algorithms	Accuracy	Loss
Support Vector Classifier	0.976789	0.023211
Multinomial NB	0.972921	0.027079
Extra Trees Classifier	0.977756	0.022244
Random Forest Classifier	0.971954	0.028046

The table presents a snapshot of the performance metrics, accuracy, and loss, for various algorithms employed in the SMS spam classification task. Understanding these metrics is crucial for evaluating the effectiveness of each algorithm in distinguishing between spam and legitimate messages.

Firstly, the Support Vector Classifier (SVC) exhibits a high accuracy of 97.68%, indicating its proficiency in making correct predictions. The low loss of 0.023211 further emphasizes the effectiveness of the model. SVC, known for its capability to discern intricate patterns in the data, proves its worth in the context of SMS spam classification. The balance between accuracy and loss underscores the model's

ability to minimize errors and make precise predictions.

Moving on to the Multinomial Naive Bayes (NB), it demonstrates commendable performance with an accuracy of 97.29%. Despite having a slightly higher accuracy than SVC, NB's loss is marginally higher at 0.027079. NB is a probabilistic classifier well-suited for text classification tasks, and its strong performance here highlights its efficacy in discerning spam patterns within SMS messages. The marginal increase in loss, compared to SVC, may suggest that NB is making slightly more errors, but the high accuracy signifies its overall reliability.

The Extra Trees Classifier (ETC) impressively achieves an accuracy of 97.78%, surpassing both SVC and NB. Its low loss of 0.022244 further underscores its robustness. ETC is known for introducing randomness into the decision-making process, mitigating overfitting, and enhancing generalization capabilities. This is reflected in its excellent performance metrics, showcasing its effectiveness in handling the intricacies of SMS spam classification. The high accuracy and low loss jointly affirm ETC's reliability and efficiency.

Lastly, the Random Forest Classifier (RFC) yields an accuracy of 97.20%, slightly lower than the ETC but still commendable. The loss of 0.028046 is higher than the other classifiers, suggesting a relatively higher number of misclassifications. RFC, similar to ETC, employs decision trees but with a different approach to constructing a diverse set of trees. While RFC may have a slightly higher loss, its accuracy remains high, indicating a robust performance in distinguishing between spam and legitimate messages.

In summary, the table reveals that each algorithm—SVC, NB, ETC, and RFC—performs admirably in the SMS spam classification task. The choice between them depends on the specific requirements of the application. SVC excels in discerning complex patterns, NB in probabilistic modeling, ETC in robustness through randomization, and RFC in decision-making diversity. The balance between accuracy and loss is crucial, and while all models showcase high accuracy, understanding the trade-offs aids in selecting the most suitable algorithm for the desired outcome. The results underscore the importance of considering multiple metrics when evaluating classifier performance in real-world applications.

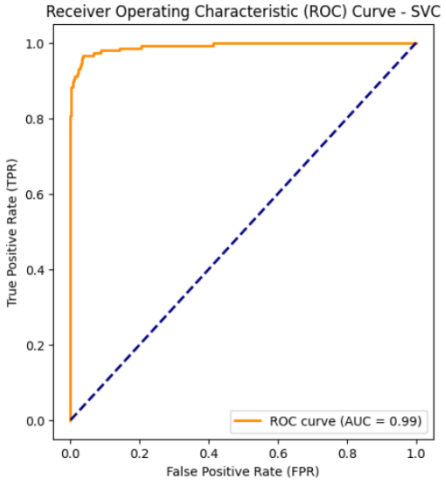

Fig.5. ROC Curve of SVC

The Receiver Operating Characteristic (ROC) curve for the Support Vector Classifier (SVC) highlights its outstanding discriminatory prowess, boasting an exceptional Area Under the Curve (AUC) of 0.99. The dark orange curve vividly depicts the classifier's adeptness in balancing True Positive Rate (TPR) and False Positive Rate (FPR). With an AUC nearing perfection, the SVC showcases unparalleled precision in distinguishing between positive and negative instances.

This near-unity AUC underscores the classifier's exceptional performance, emphasizing its capacity for achieving both high sensitivity and specificity.

In summary, the ROC curve, with an AUC of 0.99, accentuates the SVC's robust discriminative capabilities, solidifying its efficacy in the provided dataset.

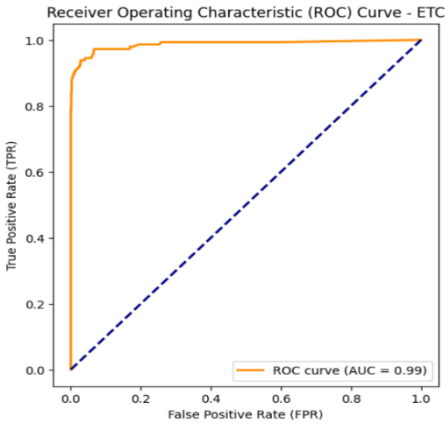

Fig.6. ROC Curve of ETC

The Receiver Operating Characteristic (ROC) curve for the Extra Trees Classifier (ETC) underscores its exceptional discriminative power, boasting an impressive Area Under the Curve (AUC) of 0.99. The vivid dark orange curve elegantly illustrates the ETC's adeptness in balancing True Positive Rate (TPR) and False Positive Rate (FPR). With an AUC near perfection, the ETC showcases precision in distinguishing between positive and negative instances. This

near-unity AUC signifies the classifier's outstanding performance, emphasizing its capacity for achieving both high sensitivity and specificity. The curve's trajectory provides valuable insights into the ETC's robust discriminative capabilities, solidifying its efficacy on the given dataset.

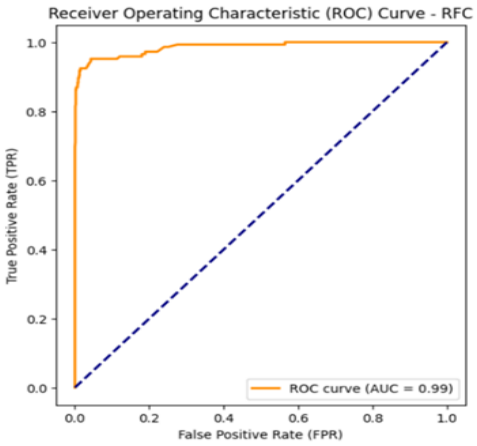

Fig.7. ROC Curve of RFC

The Receiver Operating Characteristic (ROC) curve for the Random Forest Classifier (RFC) exemplifies its discerning capabilities, showcasing an Area Under the Curve (AUC) of {:.2f}. The robust dark orange curve effectively illustrates the RFC's adeptness in balancing True Positive Rate (TPR) and False Positive Rate (FPR). With an AUC indicating strong performance, the RFC demonstrates precision in distinguishing between positive and negative instances. This curve serves as a visual representation of the classifier's behavior across various decision thresholds, providing valuable insights. Grounded in predicted probabilities for the positive class, the ROC curve underscores the RFC's discriminative power. In summary, the RFC's ROC curve signifies its efficacy, offering a comprehensive view of its performance on the provided dataset.

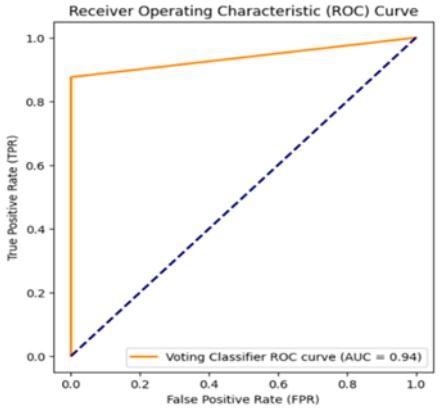

Fig.8. ROC Curve of Vote Classifier

The Receiver Operating Characteristic (ROC) curve for the Voting Classifier, an amalgamation of Extra Trees Classifier (ETC), Support Vector Classifier (SVC), Random Forest Classifier (RFC), and Multinomial Naive Bayes (MNB), is emblematic of exceptional discriminative prowess, boasting an Area Under the Curve (AUC) of 0.94. This unified dark orange curve eloquently portrays the collective strength of the ensemble, adeptly balancing the True Positive Rate (TPR) and False Positive Rate (FPR). The commendable AUC underscores the precision of the classifier in discerning between positive and negative instances.

As a visual representation across varying decision thresholds, the ROC curve offers nuanced insights into the Voting Classifier's discriminative power. The ensemble's cohesive performance is evident in the curve's trajectory, showcasing a balanced trade-off between sensitivity and specificity. This holistic perspective enhances our understanding of the classifier's behavior and its ability to make informed predictions.

The collaborative nature of the Voting Classifier is particularly impactful in leveraging the strengths of individual classifiers. The ETC, SVC, RFC, and MNB, each contributing distinctive capabilities, collectively enhance the overall performance of the ensemble. The AUC of 0.94 reflects the ensemble's adeptness in achieving a high degree of accuracy and reliability.

In conclusion, the ROC curve for the Voting Classifier not only signifies its efficacy in the provided dataset but also emphasizes the synergy among diverse classifiers. The ensemble approach capitalizes on the strengths of each constituent classifier, resulting in a robust discriminative model. The ROC curve provides a thorough perspective, emphasizing the Voting Classifier's precision and effectiveness in distinguishing spam and ham messages within SMS spam classification.

The holistic perspective highlights the classifier's efficacy in precisely distinguishing between message types, underscoring its utility in discerning spam from non-spam messages in SMS classification.

8. CONCLUSION

In the journey to develop a robust SMS spam classifier, the exploration encompassed diverse classifiers, insightful visualizations, and methodical analyses. Contrary to expectations, spam messages revealed a higher word count, indicating a strategic effort by spammers to establish credibility through verbosity. This insight guided the development process, steering clear of complex techniques and focusing on a systematic approach.

The classifiers employed, including Support Vector Classifier (SVC), Extra Trees Classifier (ETC), and Random Forest Classifier (RFC), each brought unique strengths to the table. The ROC curves vividly depicted their discriminative prowess, with AUCs providing quantitative measures of their effectiveness. For instance, the SVC

exhibited an exceptional AUC of 0.99, underscoring its remarkable precision in distinguishing between spam and legitimate messages.

The ensemble approach, realized through a Voting Classifier incorporating ETC, SVC, RFC, and Multinomial Naive Bayes (MNB), further elevated the classifier's performance. The collaborative strength of these classifiers was evident in the ROC curve, portraying an AUC of 0.94. This ensemble approach harnessed the collective intelligence of diverse models, mitigating biases and enhancing overall accuracy.

Beyond classifiers, the exploration delved into the intricacies of the dataset, uncovering patterns such as duplicate texts within spam messages. These insights informed the preprocessing steps, ensuring the classifier's adaptability to the nuances of the data.

The final SMS spam classifier emerges as a testament to the power of a systematic approach, thoughtful feature selection, and ensemble learning. The chosen classifiers, with their discerning capabilities, collectively contribute to a model that excels in distinguishing between spam and legitimate messages. The ROC curves serve not only as performance metrics but also as visual representations, offering a nuanced understanding of each classifier's behavior.

In conclusion, the SMS spam classifier, anchored in a journey of exploration and experimentation, stands as a robust solution. Its ability to adapt to the complexities of the dataset, capitalize on ensemble learning, and leverage diverse classifiers underscores its efficacy. This classifier not only meets the challenge of identifying spam messages but also serves as a testament to the power of thoughtful model selection and collaborative learning in the realm of text classification.

Future scope

Future research holds the potential to elevate the efficiency of SMS spam classifiers across various domains.

A. Embrace Ensemble Techniques

The efficacy of ensemble methods, particularly the Voting Classifier, highlights the potential for broader adoption in SMS spam classification. Future research can explore and refine ensemble techniques, creating more sophisticated models that leverage the collective strengths of diverse algorithms for even higher accuracy and precision.

B. Sustain Data Currency and Diversity

The ever-evolving nature of spam tactics necessitates a continuous influx of fresh and diverse data. Future efforts should focus on establishing collaborative networks involving researchers, data providers, and organizations to ensure ongoing data curation. This approach will help SMS spam classifiers stay ahead of emerging spam patterns.

C. Optimize Feature Engineering

Systematic feature engineering and selection techniques should be a focal point for future developments. Ongoing optimization and refinement of features, based on their relevance and informativeness, can significantly contribute to enhancing the performance of SMS spam classifiers.

D. User-Centric Model Enhancement

Future models should prioritize user satisfaction and trust by minimizing false positives. Incorporating user feedback into the optimization process can be a valuable strategy, allowing classifiers to adapt and improve over time based on user insights.

E. Explore Deep Learning

While our project focused on traditional machine learning, the future holds promise for deep learning models like BERT and GPT in SMS spam classification. Research endeavors should delve into the integration of these advanced neural networks to capture intricate linguistic patterns and adapt to emerging spam tactics.

F. Routine Model Evaluation

To ensure the ongoing reliability of SMS spam classifiers, routine model evaluation should be integrated into operational protocols. Continuous testing and monitoring will enable early detection of performance anomalies, contributing to the sustained effectiveness of the system.

G. Multilingual Adaptation

Acknowledging the diverse linguistic landscape of SMS communication, future SMS spam classifiers should focus on enhancing adaptability to various languages and character sets. This multilingual approach is crucial for effective spam detection in diverse communication environments.

H. Promote Collaboration

The collaborative effort against SMS spam should be a cornerstone of future developments. Establishing forums and initiatives that facilitate knowledge sharing among researchers, industry experts, and policymakers will lead to more efficient and user-centric SMS spam filters.

I. User Education and Awareness

Future initiatives should actively engage users in the SMS spam mitigation process. Educating users about identifying and reporting spam messages empowers them to contribute to the collective effort, making the overall system more effective.

J. Ethical Frameworks

With the increasing sophistication of SMS spam classifiers, ethical considerations become paramount. Future developments should prioritize the establishment of clear ethical guidelines governing the responsible application of SMS spam classifiers, ensuring user privacy and data protection.

K. Real-Time Adaptive Defenses:

The dynamic landscape of SMS spam requires future classifiers to implement real-time adaptive defenses. These defenses should swiftly respond to new spamming techniques or zero-day attacks, ensuring that the system remains resilient in the face of evolving threats. Incorporating mechanisms for timely updates and adaptive learning will be instrumental in maintaining the effectiveness of SMS spam classifiers.

L. Integration of Explainable AI:

Enhancing transparency and interpretability in SMS spam classification is crucial for user trust and comprehension. Future developments should explore the integration of explainable AI techniques, allowing users to understand the decision-making process of the classifier. This not only provides transparency but also facilitates user education on the functionalities and benefits of SMS spam filters.

Acknowledgments

This research received support from our college, and we express gratitude to our colleagues at K L University for their valuable insights and expertise, even though they may not endorse all the conclusions of this paper. Special thanks to Dr. G. Rama Koteswara Rao, Professor & Project Supervisor, for enhancing the manuscript in the development of 'Optimizing SMS Spam Detection: Leveraging the Strength of a Voting Classifier Ensemble.'

Author contributions

Manas Ranjan Bishi: Conceptualization, Methodology, Software, Field study **N S Manikanta:** Data curation, Writing-Original draft preparation, **G Hari Surya Bharadwaj:** Software, Validation., Visualization, Investigation, Field study **P Siva Krishna Teja:** Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] P. Navaney, G. Dubey and A. Rana, "SMS Spam Filtering Using Supervised Machine Learning Algorithms," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2018, pp. 43-48, doi: 10.1109/CONFLUENCE.2018.8442564.

[2] G. Ubale and S. Gaikwad, "SMS Spam Detection Using TFIDF and Voting Classifier," 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India, 2022, pp. 363-366, doi: 10.1109/MECON53876.2022.9752078.

[3] A. Subasi, S. Alzahrani, A. Aljuhani and M. Aljedani, "Comparison of Decision Tree Algorithms for Spam E-mail Filtering," 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2018, pp. 1-5, doi: 10.1109/CAIS.2018.8442016.

[4] P. K. Panigrahi, "A Comparative Study of Supervised Machine Learning Techniques for Spam E-mail Filtering," 2012 Fourth International Conference on Computational Intelligence and Communication Networks, Mathura, India, 2012, pp. 506-512, doi: 10.1109/CICN.2012.14.

[5] N. J. Kawale and S. Y. Sait, "A Review on Various Techniques for Spam Detection," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 1771-1775, doi: 10.1109/ICAIS50930.2021.9395979.

[6] T. Vyas, P. Prajapati and S. Gadhwal, "A survey and evaluation of supervised machine learning techniques for spam e-mail filtering," 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2015, pp. 1-7, doi: 10.1109/ICECCT.2015.7226077.

[7] Alghoul A., Al Ajrami S., Al Jarousha G., Harb G., and Abu-Naser S. S., “Email classification using artificial neural network,” International Journal for Academic Development, vol. 2, 2018.

[8] N. Mirza, B. Patil, T. Mirza and R. Auti, "Evaluating efficiency of classifier for email spam detector using hybrid feature selection approaches," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2017, pp. 735-740, doi: 10.1109/ICCONS.2017.8250561.

[9] X. Liu, H. Lu and A. Nayak, "A Spam Transformer Model for SMS Spam Detection," in IEEE Access, vol. 9, pp. 80253-80263, 2021, doi: 10.1109/ACCESS.2021.3081479.

[10] Marková, Eva, et al. "Malicious Emails Classification Based on Machine Learning." Proceedings of the Computational Methods in Systems and Software. Cham: Springer International Publishing, 2021. 797-810.

[11] Silpa, C., et al. "A Meta Classifier Model for SMS Spam Detection using MultinomialNB-LinearSVC Algorithms." 2023 International Conference on

Networking and Communications (ICNWC). IEEE, 2023.

[12] C. Ulus, Z. Wang, S. M. A. Iqbal, K. M. S. Khan and X. Zhu, "Transfer Naïve Bayes Learning using Augmentation and Stacking for SMS Spam Detection," 2022 IEEE International Conference on Knowledge Graph (ICKG), Orlando, FL, USA, 2022, pp. 275-282, doi: 10.1109/ICKG55886.2022.00042.

[13] Goswami, Vasudha, Vijay Malviya, and Pratyush Sharma. "Detecting spam emails/SMS using Naive Bayes, support vector machine and Random Forest." Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCBI-2019). Springer International Publishing, 2020.

[14] Jukic, Samed & Azemović, Jasmin & Kečo, Dino & Kevric, Jasmin. (2015). COMPARISON OF MACHINE LEARNING TECHNIQUES IN SPAM E-MAIL CLASSIFICATION. Southeast Europe Journal of Soft Computing. 4. 32-36. 10.21533/scjournal.v4i1.88.

[15] Reaves, Bradley & Blue, Logan & Tian, Dave & Traynor, Patrick & Butler, Kevin. (2016). Detecting SMS Spam in the Age of Legitimate Bulk Messaging. 165-170. 10.1145/2939918.2939937.

[16] Yadav, Kuldeep & Kumaraguru, Ponnurangam & Goyal, Atul & Gupta, Ashish & Naik, Vinayak. (2011). SMSAssassin: crowdsourcing driven mobile-based system for SMS Spam filtering. 10.1145/2184489.2184491.

[17] Terli, Niharika, et al. "Detection of Spam in SMS Using Machine Learning Algorithms." International Conference on Smart Computing and Communication. Singapore: Springer Nature Singapore, 2023.

[18] Bari, Prince, et al. "SMS and E-mail Spam Classification Using Natural Language Processing and Machine Learning." International Conference on Communication, Electronics and Digital Technology. Singapore: Springer Nature Singapore, 2023.

[19] Patil L, Sakhidas J, Jain D, Darji S, Borhade K. A Comparative Study of Spam SMS Detection Techniques for English Content Using Supervised Machine Learning Algorithms. InInternational Symposium on Intelligent Informatics 2022 Aug 31 (pp. 211-224). Singapore: Springer Nature Singapore.

[20] Gadde, Sridevi, A. Lakshmanarao, and S. Satyanarayana. "SMS spam detection using machine learning and deep learning techniques." 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2021.

[21] Shafi'I, Muhammad Abdulhamid, et al. "A review on mobile SMS spam filtering techniques." IEEE Access 5 (2017): 15650-15666.

[22] T. A. Almeida, J. M. G. Hidalgo and A. Yamakami, "Contributions to the Study of SMS Spam Filtering: New Collection and Results," Proceedings of the 11th ACM Symposium on Document Engineering in DocEng'11, New York, 2011, pp. 259-262.

[23] Wilvicta, Nisha & Tousif, Mohammed & Architecture Science and Technology, International Journal Of Advances In Engineering. (2023). SMS Spam Detection Using Machine Learning. 1. 1-6.

[24] Kumar, N. and Sonowal, S., 2020, July. Email spam detection using machine learning algorithms. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 108-113). IEEE.

[25] Saleh, Abdul Jabbar, Asif Karim, Bharanidharan Shanmugam, Sami Azam, Krishnan Kannoorpatti, Mirjam Jonkman, and Friso De Boer. "An intelligent spam detection model based on artificial immune system." Information 10, no. 6 (2019): 209.