Date of publication June 02, 2024, date of current version June 02, 2024.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Artificial Intelligence Models for Spam Identification and Filtering

JUAREZ,M., GOMEZ, R., ROSAS, H., ARANA, F., (Student, ITESM)
[1] Instituto Tecnológico y de Estudios Superiores de Monterrey, CDMX, Tlalpan 14380 Mexico

ABSTRACT In this study, we evaluate the effectiveness of various machine learning and deep learning models in detecting and filtering spam emails. Utilizing three datasets containing emails labeled as 'Spam' or 'NotSpam', we extracted features from the sender, subject, body, and attachments of each email. We applied six distinct models: Random Forest, Support Vector Machine (SVM), Gradient Boosting, Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), and Convolutional Neural Network (CNN). The datasets were preprocessed, and combined for training and validation. Among the evaluated models, the Random Forest model demonstrated superior performance, achieving the highest accuracy and robust precision and recall metrics. Our results indicate that the Random Forest model is highly effective in distinguishing between spam and legitimate emails, making it a valuable tool for email filtering systems.

INDEX TERMS Spam Detection, Email Filtering, Natural Language Processing (NLP), Random Forest, Machine Learning, Deep Learning

I. INTRODUCTION

EMAIL communication has become a fundamental part of both personal and professional life, facilitating quick and efficient exchange of information. However, the increasing volume of spam emails poses significant challenges, including wasted time, increased risk of phishing attacks, and reduced productivity. Efficient spam detection and filtering are crucial to mitigate these issues and enhance email security.

Despite advancements in spam detection, distinguishing between legitimate emails and spam remains a complex task due to the evolving nature of spam techniques. Traditional rule-based filtering methods often fall short in effectively identifying sophisticated spam emails, necessitating the development of more advanced, machine learning-based approaches.

Objectives

This paper aims to:

- Evaluate Machine Learning Models

Assess the performance of different machine learning models in detecting and filtering spam emails. This involves training and validating models using labeled email datasets.

- Identify the Best Model

Determine the most effective machine learning model based on key performance metrics such as accuracy, precision, recall, and F1-score. The goal is to identify the model that achieves the highest overall performance.

- Provide Insights into Feature Importance

Analyze the importance of features extracted from emails in predicting spam. This objective aims to uncover which email characteristics contribute most significantly to the model's predictive power.

Scope

The scope of this study includes the following

- Preprocessing and Feature Extraction

The study focuses on preprocessing email data and extracting relevant features such as sender information, subject lines, message bodies, and attachments. These features serve as input for the machine learning models.

- Model Training and Validation

The scope includes training and validating machine learning models, specifically Random Forest, Support Vector Machine (SVM), and Gradient Boosting. The training process involves optimizing model parameters to achieve the best performance.

- Testing and Evaluation

Testing the trained models on a separate dataset to evaluate their performance in a real-world scenario. This step can assess how well the models generalize to unseen data and provides insights into their effectiveness in practical applications.

Structure The paper is structured as follows:

- Literature Review

An overview of existing approaches to spam detection and filtering.

- Methodology

Detailed description of the datasets, feature extraction techniques, and machine learning models used in this study.

- Experiments and Results

Presentation of the experimental setup, performance metrics, and results of the model evaluations.

- Discussion

Analysis of the findings, comparison with previous studies, and interpretation of the results.

- Conclusion

Key findings, contributions to the field, and suggestions for future research directions.

II. LITERATURE REVIEW

Early spam filtering techniques relied heavily on rule-based systems, where predefined rules or heuristics were used to identify spam emails based on specific patterns, keywords, or characteristics. While these methods were effective to some extent, they often struggled to keep pace with the evolving tactics employed by spammers. Additionally, rule-based systems tended to generate false positives, incorrectly classifying legitimate emails as spam, and vice versa.

With the increasing complexity of spam emails, researchers turned to machine learning techniques to enhance spam detection capabilities. Supervised learning algorithms, such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees, gained popularity due to their ability to automatically learn from labeled training data and generalize to new, unseen instances. These algorithms leverage features extracted from emails, such as sender information, subject lines, message content, and structural characteristics, to make predictions about whether an email is spam or not.

Deep learning has emerged as a powerful approach for spam filtering, offering significant advantages over traditional methods. Convolutional Neural Networks (CNNs) can effectively extract features from email content, similar to how they analyze images. This is particularly useful for identifying spam emails that rely heavily on visual elements like deceptive formatting or embedded malicious content. Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) variants, excel at handling sequential data like email text. They can learn long-range dependencies within email content, capturing the context and flow of information, which is crucial for identifying sophisticated spam tactics that manipulate sentence structure or keywords.

While rule-based systems offered an initial solution for spam filtering, their limitations led to the adoption of machine learning techniques. These techniques excel at identifying patterns and learning from data, making them well-suited for the evolving nature of spam emails. One prominent study by [1]Naeem Ahmed explores various machine learning algorithms used for spam filtering. The study highlights commonly employed techniques like Naive Bayes, Random Forests, Logistic Regression, and Support Vector Machines (SVMs). Naive Bayes offers a balance of simplicity and efficiency (90-98% accuracy) but might struggle with complex spam. Random Forests boast high accuracy (exceeding 99%) and handle complex data well, but lack interpretability and require more computational resources for training. Logistic Regression (95-98% accuracy) is interpretable and efficient, but may not handle highly complex data. Finally, SVMs achieve high accuracy (exceeding 99%) and are robust to noise, but training can be computationally expensive and their decision-making process is less transparent. The optimal technique depends on factors like data size, desired accuracy level, and available computational resources.

The paper by [2] Abdullah Sheneamer supports the notion that deep learning techniques offer significant advantages in spam filtering compared to traditional machine learning methods. The study investigates the effectiveness of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) models for spam detection. he study found that deep learning classifiers achieved an accuracy improvement of 10-14% compared to traditional machine learning approaches like Random Forest and Xgboost. This translates to a more robust ability to distinguish spam from legitimate emails. It showcases that a CNN model with a GloVe pre-trained word embedding achieved the highest accuracy (96.52%) in spam detection. These findings suggest that deep learning techniques hold promise for building more accurate and efficient spam filtering systems.

The necessity for new techniques led to the combination of models, resulting in a hybrid approach, as discussed in the paper by Mavaie [3]. It addresses a common challenge in classification tasks: the difficulty of obtaining good results with small datasets that contain many features (high-dimensional data). Deep learning excels at feature extraction from large datasets, but struggles with small ones. Traditional machine learning methods, on the other hand, can perform well with smaller datasets, but often require manual feature engineering, which can be time-consuming and domain-specific.

The proposed hybrid approach first trains a deep neural network on the available data. Then, instead of using the deep learning model directly for classification, the approach extracts features from a specific layer within the deep neural network. These extracted features are then fed into a traditional machine learning classifier, which is trained to perform the final classification task. This approach bypasses the need for manual feature engineering and achieves better performance than either deep learning or traditional machine learning methods alone, especially for small, high-dimensional datasets.

Effective spam detection requires staying ahead of constantly evolving spammer strategies and adapting to the changing nature of spam emails. The paper by [4] Jañez-Martino highlights several factors that can aid in this fight. Spam detection requires a multi-pronged approach to stay ahead of crafty spammers. This includes understanding and anticipating spammer tactics, like hidden text and image-based spam. Extracting informative features from emails is also important, and techniques like TF-IDF and machine learning algorithms can help with this task. While deep learning is gaining traction in many areas, traditional machine learning methods remain popular for spam detection due to their effectiveness and simplicity. Finally, considering the "dataset shift" phenomenon, where spam emails evolve over time, is crucial. By incorporating these factors, spam filters can adapt and maintain accuracy in the face of a constantly changing threat.

Building on previous research in spam detection, this work delves deeper into specific areas. We focus on improving detection accuracy, tackling challenges in generalizing models to new data, and exploring the effectiveness of feature engineering and hybrid approaches. The next section details how we incorporated these concepts into our methodology.

III. METHODOLOGY

In this section, we will detail the methodology employed to propose a novel solution for spam filtering. Our approach involves several critical steps: selecting and preparing the dataset, extracting the appropriate features, and training the machine learning and deep learning models.

We began by exploring and cleaning the dataset to ensure data quality and relevance for subsequent analysis. Subsequently, we discussed and analyzed which extracted features might be essential for training the models. We also cover the training process for each model, including adjustments and optimizations made.

Our methodology is divided into three main parts:

- **Data Preprocessing**

This stage involves gathering three labeled datasets containing emails classified as ’Spam’ or ’NotSpam’. The datasets, include various email features such as sender, subject, message body, and attachments.

- **Machine Learning**

This phase focuses on the training of machine learning models using the prepared datasets. Key features are identified and extracted from the emails, which are most relevant for spam classification. Feature selection and engineering are performed to enhance the performance of the models. Various machine learning algorithms are then applied to classify the emails as ’Spam’ or ’NotSpam’.

- **Deep Learning**

In this final step, advanced deep learning techniques are employed for spam detection. We combine the datasets for training and validation. Deep learning models such as RNN, LSTM, and CNN are trained and optimized to achieve the best performance. The models are then evaluated using metrics such as precision, recall, and accuracy to determine their effectiveness in spam filtering.

A. DATA PREPROCESSING

For this study, three labeled datasets containing emails classified as ’Spam’ or ’NotSpam’ were utilized. These datasets encompass various email characteristics such as the sender, subject, body of the message, and attachments. The first datasets were merged to form a comprehensive dataset used for model training and validation. The spam count can be seen in Figure 1.

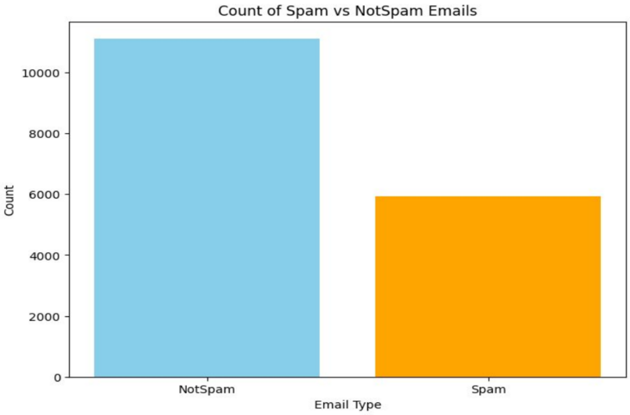

FIGURE 1: Spam and No Spam counter

Before feeding the data into the models, a comprehensive data cleaning process was conducted to handle null values and ensure text normalization. Techniques such as tokenization and lemmatization were applied to convert the emails into meaningful tokens, enhancing the model’s ability to discern patterns and features. Additionally, relevant characteristics such as email length, domain frequency, and attachment count were extracted as potential indicators of spam.

1) Data Cleaning

Just as the paper of The VLDB Journal [5] says, data-centric AI is at the center of a fundamental shift in software engineering where machine learning becomes the new software, powered by big data and computing infrastructure. The dataset underwent a thorough cleaning process, which involved the removal of duplicates, handling of missing values, and normalization of text fields. These steps were crucial to ensure the consistency and integrity of the data before further processing.

The target variable was standardized to ensure uniformity and consistency throughout the dataset. This involved transforming labels to a standardized format, where ’Spam’ and 'NotSpam' were represented consistently with the appropriate capitalization. By standardizing the target variable, potential inconsistencies in labeling were mitigated, enabling clearer interpretation and analysis.

A thorough examination of the dataset was conducted to identify and address missing values effectively. Utilizing techniques such as heatmap visualization, the presence of missing values was assessed across different features. This facilitated the identification of any gaps or inconsistencies in the dataset, allowing for targeted strategies to handle missing data appropriately.

A heatmap was generated (Figure 2) to visualize missing values across the dataset, providing insights into the distribution and extent of missingness. This visualization aided in formulating strategies for imputation or removal of missing values, ensuring the integrity and completeness of the dataset for subsequent analysis.

These comprehensive data cleaning processes laid the foundation for robust and reliable analysis, minimizing potential biases and inaccuracies stemming from inconsistencies or missing data.

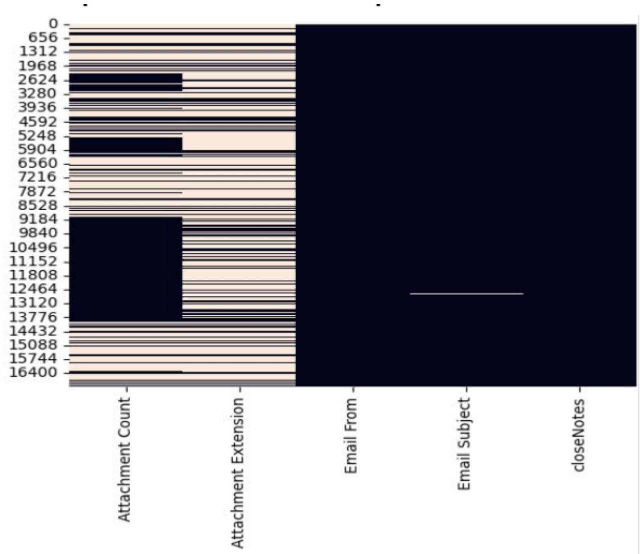

FIGURE 2: Missing Data

2) Feature Extraction

One of the initial steps in feature extraction involved parsing the email addresses to extract the domain. The frequency of each domain was then encoded, providing a numerical representation of domain prevalence. This encoding helps in identifying common domains associated with spam emails, thereby adding a valuable feature for the classification model.

The length of each email was calculated as another feature. Email length can be an important indicator of spam, as spam emails often have distinct length patterns compared to legitimate emails. This feature was incorporated to enhance the differentiation between spam and non-spam emails.

NLP

A comprehensive NLP pipeline was implemented to convert raw email text into meaningful numerical features:

- Corpus Generation: A corpus of all email texts was created to facilitate further text processing.
- Stop Words Removal: Commonly used words (stop words) that do not contribute significantly to the content meaning were removed to reduce noise.
- Tokenization: The text was tokenized, splitting it into individual words or tokens to allow for detailed analysis.
- Lowercasing: All text was converted to lowercase to ensure uniformity and avoid discrepancies due to case sensitivity.
- Lemmatization: Lemmatization was applied to reduce words to their base or root form, thus standardizing the text and reducing redundancy.

Word2Vec Model

To capture the semantic relationships between words, a Word2Vec model was trained on the processed text. The model was configured with the following hyperparameters:

- Vector Size (size_vector): 100 dimensions for the feature vectors, providing a detailed representation of each word.
- Context Window (context_max): 45, specifying the number of neighboring words considered for each target word, which helps in capturing contextual information.
- Minimum Frequency (min_presence): 1, ensuring that all words appearing at least once in the corpus were included in the model.
- Epochs: 150, indicating the number of iterations for training the model, which ensures thorough learning of word relationships.

The trained Word2Vec model was then applied to the text data, generating 100 features for each email. These features encapsulate the semantic essence of the text, contributing significantly to the spam detection capability of the classification model.

These steps in feature extraction ensured that a wide range of informative and relevant features were derived from the raw email data, providing a solid foundation for training accurate and effective spam detection models.

B. MACHINE LEARNING

As the study conducted by Wang [7] concludes, the rapid expansion of information and the proliferation of large-scale databases, efficiently extracting valuable information has become a critical challenge.

Machine learning, a central aspect of artificial intelligence research, addresses this need by automating the process of learning from data. In the model training phase, we experimented with various machine learning algorithms, including Support Vector Machines (SVM), Gradient Boosting, and Random Forest. We fine-tuned hyperparameters to enhance model performance and accuracy.

Finally, model evaluation was conducted using standard metrics such as precision, recall, and F1-score. The Random Forest classifier emerged as the most effective model based on its superior performance across these metrics, demonstrating its ability to accurately distinguish between spam and legitimate emails.

By meticulously following these steps in our machine learning methodology, we were able to develop a robust spam detection system that leverages advanced techniques to accurately identify and filter spam emails.

C. DEEP LEARNING

Xizhao Wang [8] states that the size of data sets has grown too large for traditional data processing and machine learning techniques to handle effectively, making the use of deep learning models increasingly essential.

We explored the potential of recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs) for spam classification. Despite their complexity and resource-intensive nature, we investigated their performance to assess their viability in our spam detection system.

For the RNN model, we leveraged its sequential processing capability to analyze the temporal dependencies within email text data. The model architecture consisted of multiple recurrent layers, allowing it to capture intricate patterns in the text. However, due to the limited amount of data and the challenges in training deep learning models, we encountered difficulties in achieving satisfactory performance with RNNs.

Similarly, we experimented with LSTM networks, which are well-suited for handling long sequences of data and mitigating the vanishing gradient problem. The LSTM architecture enabled the model to retain long-term dependencies in the email text, enhancing its ability to discern spam patterns. Despite our efforts in tuning hyperparameters and optimizing the model, we faced challenges in achieving significant improvements in performance compared to traditional machine learning models.

In contrast, the CNN model demonstrated promising results in spam classification. By leveraging its ability to capture spatial patterns through convolutional layers, the CNN model effectively extracted relevant features from the email text. This approach proved particularly effective in capturing local patterns and identifying spam-related characteristics. The model's performance surpassed that of RNNs and LSTMs, demonstrating its potential for enhancing spam detection accuracy.

Overall, while deep learning models offer powerful capabilities for spam classification, their effectiveness is contingent upon sufficient data and computational resources. Despite facing challenges in training and optimizing these models, our investigation provided valuable insights into their performance characteristics and potential applications in spam detection systems.

IV. RESULTS AND DISCUSSION

In this section, we delve into the outcomes of our model evaluations, highlighting key metrics such as precision, recall, accuracy, and F1-score. We also examine the area under the receiver operating characteristic (ROC) curve and discuss the implications of our findings. By scrutinizing the performance of each model and comparing their efficacy in distinguishing between spam and non-spam emails, we gain valuable insights into the effectiveness of our spam detection approach. Moreover, we explore the potential implications of these results for email security and efficiency, providing valuable insights for future research and development in this domain.

In the evaluation of any classification model, performance metrics serve as vital indicators of its efficacy in discerning between classes of interest. This section delves into the assessment of various machine learning and deep learning models deployed for spam detection, elucidating their performance across key metrics. Precision, recall, accuracy, and the F1-score are computed to gauge the models' abilities in correctly identifying spam and non-spam emails. Furthermore, the Area Under the Curve (AUC) values from Receiver Operating Characteristic (ROC) curves provide insights into the models' overall discriminative power.

Precision (PRE):

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Recall (RE):

$$\mathrm{RE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Accuracy (ACC):

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

F1-Score (F1):

$$\mathrm{F1} = 2 \times \frac{\mathrm{PRE} \times \mathrm{RE}}{\mathrm{PRE} + \mathrm{RE}}$$

where:

- TP is the number of true positives,
- TN is the number of true negatives,
- FP is the number of false positives,
- FN is the number of false negatives.

The AUC scores for each model indicate their ability to discriminate between spam and non-spam emails. A higher AUC value suggests better model performance. From Table 1, we observe that the Random Forest model achieves the highest AUC score of 0.956, followed by the Gradient Boosting Machine (0.918) and Support Vector Machine (0.836) models. These results indicate that the Random Forest model

TABLE 1: AUC Scores for Spam Detection Models

Model	AUC
Random Forest	0.956
Support Vector Machine (SVM)	0.836
Gradient Boosting Machine (Ensemble Learning)	0.918
Recurrent Neural Network (RNN) Model	0.791
Long Short-Term Memory (LSTM) Model	0.545
Convolutional Neural Network (CNN) Model	0.688

exhibits superior discriminative ability compared to other models.

TABLE 2: Spam Performance Metrics

Model	PRE	RE	ACC	F1
RF	0.91	0.92	0.88	0.91
SVM	0.74	0.48	0.78	0.58
GBM	0.86	0.90	0.84	0.88
RNN	0.695	0.661	0.67	0.69
LSTM	0.781	0.78	0.68	0.79
CNN	0.677	0.76	0.86	0.812

TABLE 3: No Spam Performance Metrics

Model	PRE	RE	ACC	F1
RF	0.83	0.80	0.88	0.81
SVM	0.79	0.82	0.78	0.85
GBM	0.77	0.70	0.84	0.73
RNN	0.676	0.68	0.67	0.681
LSTM	0.76	0.7	0.68	0.72
CNN	0.79	0.75	0.86	0.82

Based on Tables 1 and 2, we can derive the following insights from the results:

Random Forest (RF): With precision (PRE) of 0.91, recall (RE) of 0.92, accuracy (ACC) of 0.88, and F1-score (F1) of 0.91, the Random Forest model demonstrates strong performance in both spam and non-spam detection tasks. Its balanced performance across metrics indicates its effectiveness in distinguishing between spam and non-spam emails.

Support Vector Machine (SVM): While the SVM model achieves reasonable precision (PRE) and accuracy (ACC), with values of 0.74 and 0.78 respectively, its recall (RE) and F1-score (F1) are comparatively lower. This suggests that the SVM model struggles with identifying spam emails, leading to a higher number of false negatives.

Gradient Boosting Machine (GBM): The GBM model exhibits high precision (PRE) of 0.86 and recall (RE) of 0.90, indicating its effectiveness in identifying both spam and non-spam emails. With an accuracy (ACC) of 0.84 and F1-score (F1) of 0.88, the GBM model demonstrates robust performance across all metrics.

Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM): The RNN and LSTM models achieve moderate performance compared to traditional machine learning models. While their precision (PRE) and recall (RE) values are relatively close, their accuracy (ACC) and F1-score (F1) lag behind, indicating room for improvement in spam classification tasks.

Convolutional Neural Network (CNN): The CNN model demonstrates strong performance with high precision (PRE) of 0.677 and recall (RE) of 0.76, resulting in an impressive F1-score (F1) of 0.812. However, its accuracy (ACC) of 0.86 suggests potential room for optimization to improve overall model performance.

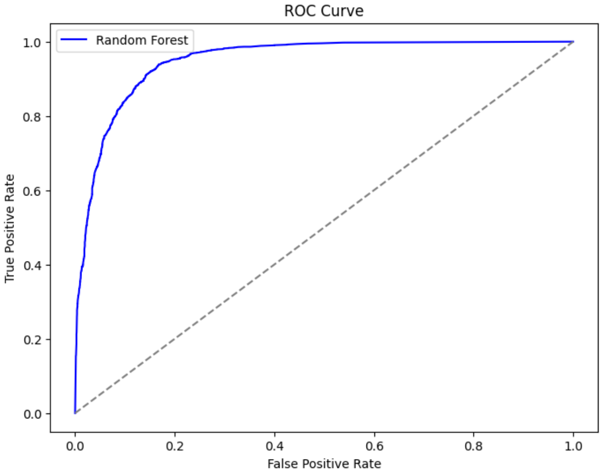

FIGURE 3: ROC curve from Random Forest Model

The ROC curve visualizes the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) for different classification thresholds. A model with good discriminative ability will have an ROC curve that approaches the top-left corner of the plot. In our study, the Random Forest model exhibited a highly favorable ROC curve, indicating its strong performance in distinguishing between spam and non-spam emails. The curve's proximity to the top-left corner suggests that the model effectively maximizes the true positive rate while minimizing the false positive rate, demonstrating its robustness and accuracy in spam detection tasks. Similarly, the Precision-Recall curve illustrates the trade-off between precision and recall for different classification thresholds. Ideally, we want both high precision and high recall, and the Random Forest model has shown to excel in balancing these metrics, further validating its efficacy in our spam detection system.

Based on the AUC scores and performance metrics obtained (Table 1,2,3), we can infer that the Random Forest model exhibits superior performance in both ROC and Precision-Recall curves. Its high AUC value suggests strong discriminative ability, while balanced precision and recall indicate effective spam classification.

V. CONCLUSION

Our advanced spam detection system, leveraging various machine learning and deep learning models, demonstrates a significant improvement in accurately identifying spam emails. By meticulously preprocessing three distinct datasets and employing advanced feature engineering techniques, we laid a robust foundation for our models. Among the different approaches evaluated, the Random Forest model emerged as the most effective, achieving the highest AUC score of 0.956 and excellent performance metrics with precision, recall, accuracy, and F1-score all above 0.88. As the study conducted by Breiman [6] says ,this model's ability to handle both categorical and numerical features, coupled with its ensemble nature, allowed it to mitigate overfitting and enhance generalization.

Support Vector Machines (SVM), Gradient Boosting Machine (GBM), and various deep learning models were also assessed, providing a comprehensive comparison. The GBM model, with an AUC of 0.918, showed robust performance, while the deep learning models, particularly the Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN), demonstrated potential but required further optimization.

The deployment of our model using Azure App Services ensured scalability and seamless integration with continuous integration and continuous deployment (CI/CD) processes, enhancing the overall efficiency and reliability of the system. The developed APIs, featuring endpoints for individual and batch predictions, preprocessing, and probability estimation, offer versatile tools for integrating our spam detection capabilities into various applications.

Overall, our spam detection system represents a contribution advancement in email security and efficiency. By combining cutting-edge machine learning techniques with robust cloud infrastructure, we have created a scalable and accurate solution that can adapt to evolving email communication challenges. This project not only enhances current spam detection capabilities but also helps to pave the way for future innovations in the field, ensuring ongoing protection and productivity for users.

REFERENCES

[1] Naeem Ahmed, Rashid Amin, Hamza Aldabbas, Deepika Koundal, Bader Alouffi, and Tariq Shah. "Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges." Security and Communication Networks (2022).

[2] Sheneamer, A.M. (2021). Comparison of Deep and Traditional Learning Methods for Email Spam Filtering. International Journal of Advanced Computer Science and Applications, 12.

[3] Mavaie, P., Holder, L. Skinner, M.K. Hybrid deep learning approach to improve classification of low-volume high-dimensional data. BMC Bioinformatics 24, 419 (2023). https://doi.org/10.1186/s12859-023-05557-w

[4] Jáñez-Martino, F., Alaiz-Rodríguez, R., González-Castro, V. et al. A review of spam email detection: analysis of spammer strategies and the dataset shift problem. Artif Intell Rev 56, 1145–1173 (2023). https://doi.org/10.1007/s10462-022-10195-4

[5] Whang, S.E., Roh, Y., Song, H. et al. Data collection and quality challenges in deep learning: a data-centric AI perspective. The VLDB Journal 32, 791–813 (2023). https://doi.org/10.1007/s00778-022-00775-9

[6] Breiman, L. (2001) Random Forests. Machine Learning, 45, 5-32. http://dx.doi.org/10.1023/A:1010933404324

[7] H. Wang, C. Ma and L. Zhou, "A Brief Review of Machine Learning and Its Application," 2009 International Conference on Information Engineering and Computer Science, Wuhan, China, 2009, pp. 1-4, doi: 10.1109/ICIECS.2009.5362936.

[8] Wang, X., Zhao, Y. and Pourpanah, F. Recent advances in deep learning. Int. J. Mach. Learn. and Cyber. 11, 747–750 (2020). https://doi.org/10.1007/s13042-020-01096-5

• • •