INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2046-2051

EXPLORING CONTENT-BASED AND COMMENT CLASSIFICATION USING DEEP ADAPTIVE LEARNING

Arnanil Patra[1], Manish Kumar[2], Anirban Sasmal[3], Mrs. Preethy Jemima P[4]

[1,2,3]Student, Dept. of Computer Science & Engineering, SRMIST, Chennai, India.

[4]Assistant Professor Dept. of Computer Science & Engineering, SRMIST, Chennai, India.

DOI: https://www.doi.org/10.58257/IJPREMS33636

ABSTRACT

In recent years, the prevalence of SMS as an primary communication channel has soared, making it an integral part of daily interactions. However, this surge in usage has also led to the emergence of SMS spam, a nuisance that not only disrupts users but also poses potential risks such as credential theft and data compromise. To combat this issue effectively, Natural Language Processing (NLP) techniques coupled with Deep Learning models have emerged as a promising solution, particularly for text classification tasks. Among these models, Long Short-Term Memory (LSTM) networks have demonstrated remarkable performance in binary and multi-label text classification problems. this paper presents an innovative approach that combines two distinct data sources: one targeting Spam detection in social media posts and the other focused on Fraud classification in emails. By merging these datasets and leveraging common bigrams extracted from each, we devised a multi-label LSTM model tailored for identifying malicious text across varied sources. Our experimental results underscore the effectiveness of this approach, showcasing the model's capability to discern malicious content irrespective of its origin. the LSTM model trained on the amalgamated dataset exhibited superior performance compared to models trained independently on each dataset. This enhancement in performance can be attributed to the model's ability to learn and generalize from a diverse range of text samples, thereby improving its predictive accuracy and robustness. Additionally, the utilization of common bigrams extracted from both datasets facilitated the model's understanding of recurring patterns and linguistic nuances associated with spam and fraudulent content. overall, our findings highlight the efficacy of employing LSTM-based Deep Learning models for combating SMS spam and related security threats. The approach presented in this paper offers a practical and efficient means of classifying malicious text, contributing to the ongoing efforts in ensuring a secure and trustworthy communication environment for SMS users.

Keywords: Support Vector Machine (SVM), Radial Basis Function, NLP, Deep Adaptive Learning, LSTM

1. INTRODUCTION

In recent years, the widespread adoption of SMS communication has brought both convenience and challenges. With over five billion users globally, representing about 65% of the human population and expected to reach 5.9 billion by 2025, SMS has become a vital means of communication and marketing. However, this proliferation has also led to an increase in malicious activities such as spam and Smi Shing, posing serious financial and security risks to individuals and businesses. according to a study published by Slick text, the volume of spam messages has seen a significant surge, with 10.89 billion spam messages sent in August 2022 compared to 1.27 million in September 2021. This rise in spam activity has resulted in substantial financial losses, estimated at USD 10,066,331,169 in 2021 alone. to combat this growing threat, researchers and industry experts have been developing and refining spam detection and prevention techniques. Traditional methods like blacklisting, whitelisting, and heuristic rules have been effective to some extent but are increasingly challenged by the evolving nature of spam content. artificial intelligence (AI)-based techniques, particularly those leveraging machine learning algorithms, have emerged as more effective solutions for detecting spam, especially with constantly evolving spam content. These intelligent models analyse message content to classify them as either legitimate (ham) or spam. Various factors, such as text representation techniques, feature selection methods, and classification algorithms, influence the performance*of*these*models, highlighting the importance of choosing the right criteria for optimal results. one promising approach discussed*in recent research involves using Transformer-based*text*embedding techniques combined with Ensemble Learning strategies for classification. This combination leverages the strengths of deep learning models, known for their adaptability and effectiveness in handling large datasets and complex patterns. moreover, spam detection extends beyond SMS to encompass various online platforms, including social networks and email services. Phishing, a common tactic in spam and fraud*communications, targets*users through deceptive messages aimed at acquiring personal or financial information. Effective spam filtering involves not only detecting malicious content but also ensuring a positive user experience and enhancing platform reliability. researchers have explored different deep learning architectures*such as Convolutional*Neural Networks (CNNs),

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2046-2051

Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM) networks for spam classification. These models, when trained on large datasets with pre-trained word embeddings like Word2vec, have shown promising results in identifying spam messages across diverse*platforms. in addition to content-based detection methods, recent studies emphasize the importance of social network analysis*in spam classification. Features derived from social network properties, such as node degree, clustering coefficient, and rank, have been effective in distinguishing spam from legitimate messages. Leveraging algorithms like PageRank and Hyperlink-Induced Topic Search (HITS) has enhanced spam detection accuracy by considering social relationships and network dynamics. The*evolution of SMS spam detection and prevention techniques involves a combination of AI-based models, deep learning*architectures, text embedding methods, and social network analysis.

By continually refining these approaches and integrating advanced technologies, we can mitigate the*risks posed by spam and ensure a*safer and more reliable*communication environment for users and businesses alike.

2. RELATED WORKS

It is a method for detecting YouTube spam comments, given their recent surge. Despite YouTube's own spam blocking system, issues persist. Through related studies and classification experiments, six machine learning techniques and two ensemble models are evaluated using comment data from popular music videos by artists.[1]

It is a modified genetic algorithm to tackle spam on Twitter by simultaneously reducing feature dimensions and optimizing hyperparameters. Using e Xtreme Gradient Boosting (XG Boost) as a classifier, the algorithm achieves an average accuracy of 92.67% and geometric mean of 82.32%, using less than 10% of the total feature space. It outperforms traditional feature selection methods like Chi^2 and PCA, as well as BERT-based deep learning models. The approach also applies to SMS spam modeling, showing robustness and competitive performance against other methods.[2]

In the Successive Pooling Attention Network (SPA Net) to address challenges in segmenting small-scale objects and boundaries in remote sensing images using convolutional neural networks. SPA Net combines ResNet50, a Successive Pooling Attention Module (SPAM), and a Feature Fusion Module (FFM) to extract both high-level and low-level features effectively. SPAM enhances feature representation using attention mechanisms and successive pooling to capture multiscale and salient features. FFM complements spatial and geometric information, improving object edge segmentation. Experimental results demonstrate that SPA Net outperforms other models on remote sensing datasets by achieving accurate segmentation.[3]

It is a Cascaded Ensemble Machine Learning Model to detect spam comments on YouTube, addressing the platform's increasing spam issues. It evaluates six machine learning techniques and two ensemble models on comment data from popular music videos. Results indicate that the proposed ESM-S model outperforms others in four out of five evaluation measures. This multi-technique approach enhances detection performance compared to single-model methods used in previous studies. The ESM-S model excels in accuracy, F1-score, and MCC across datasets, showing robustness even with smaller, cleaner datasets.[4]

It addresses the challenge of imbalanced classification in various fields like medical diagnosis and network intrusion detection. It introduces a Cost-sensitive Feature selection General Vector Machine (CFGVM) algorithm, combining the strengths of the General Vector Machine (GVM) and Binary Ant Lion Optimizer (BALO) algorithms. CFGVM assigns different cost weights to classes and selects significant features to enhance classification. Experimental results on eleven imbalanced datasets demonstrate that CFGVM significantly improves minority class classification. Compared to existing methods, CFGVM outperforms in performance and accuracy, effectively addressing imbalanced classification issues. [5]

It introduces a modified genetic algorithm to tackle spam on Twitter by simultaneously reducing feature dimensions and optimizing hyperparameters. using extreme Gradient Boosting (XG Boost) as a classifier, the algorithm achieves an average accuracy of 92.67% and geometric mean of 82.32%, using less than 10% of the total feature space. It outperforms traditional feature selection methods like Chi^2 and PCA, as well as BERT-based deep learning models. The approach also applies to SMS spam modeling, showing robustness and competitive performance against other methods. [6]

It addresses the issue of deceptive opinion spam in e-commerce product reviews, which misleads consumers' purchase decisions. Despite advancements in deep learning-based detection methods, there is a lack of comprehensive surveys summarizing existing techniques. The paper introduces the deceptive opinion spam detection task, reviews available datasets, and analyzes detection methods, categorizing them into traditional statistical approaches and neural network models. It concludes by suggesting future research directions in the field.[7]

It introduces a novel link prediction model called Common Influence Set (CIS) for graph data mining in social networks. CIS calculates node similarities using common influence sets between unconnected nodes, aiming to predict missing or future links accurately. Experimental results demonstrate that CIS outperforms mainstream similarity indices in terms of prediction accuracy. Additionally, an efficient method for calculating CIS's similarity score is proposed to address computational challenges. Future work will focus on adapting CIS to dynamic graph structures that evolve over time. [8]

It addresses the issue of malicious activities, like Sybil attacks and fake identity usage, impacting online social networks. These activities undermine user trust and affect various online interactions, from content creation to messaging. The study reviews literature from 2006 to 2016 on Sybil attack defenses in online social networks, presenting a new taxonomy of defense schemes. Despite existing solutions, Sybil attacks remain a significant challenge due to evolving tactics and lack of unified evaluation methods. The paper emphasizes the need for further research to enhance Sybil detection and prevention, safeguarding online platforms from fake users and their detrimental effects. [9]

It focuses on detecting deceptive online reviews using semi-supervised learning methods, specifically enhancing the F-score metric in classification. It introduces new feature dimensions like Parts-of-Speech, Linguistic and Word Count, and Sentimental Content to improve review detection. Using these methods, the study achieved an F-score of 0.837 in classification. Future work aims to implement this approach in real-world scenarios and explore integrating additional metadata and multimedia content for improved accuracy. [10]

3. PROPOSED WORK

Existing Work

This spam detection within online social networks, concentrating on content-based methods leveraging machine learning. The research scrutinized diverse machine learning algorithms, including traditional ones like logistic regression and more sophisticated techniques such as ensemble learning and deep learning, to discern their efficacy in detecting spam. The study meticulously analyzed textual content, examining features like the frequency of particular words or phrases, sentiment analysis, and user behavior patterns, which served as crucial inputs for training and evaluating the spam detection models. By delving into the realm of content-based approaches, Smith et al. aimed to enhance the understanding of how machine learning can be harnessed to combat spam effectively in online social networks, offering insights into the nuances of feature selection and model performance assessment in this context.

Text Preprocessing

Text preprocessing is a foundational step in preparing data for machine learning models. It involves tasks like tokenization, lowercasing, removing stop words, and stemming or lemmatization. Tokenization breaks text into individual tokens or words, while lowercasing ensures consistency by converting all words to lowercase. Removing stop words eliminates common words that do not contribute significant meaning, and stemming or lemmatization reduces words to their root form.

These processes clean the raw text, making it more suitable for analysis. Vectorization techniques like TF-IDF or word embeddings then transform the pre-processed text into numerical representations, enabling machine learning algorithms to process and understand the text effectively.

Word Embedding

Word embedding is a technique used to represent words as dense vectors in a continuous vector space. By capturing semantic relationships between words based on their context and usage, word embeddings enhance the model's understanding of language. Popular models like Word2Vec, Glo Ve, and Fast Text learn vector representations for each word by training on large corpora. These embeddings enable the model to detect semantic similarities, perform analogy detection, and improve overall performance in natural language processing tasks. Integrating word embeddings with deep learning models further enhances their ability to process and analyse text data.

Deep Adaptive Learning Network

Deep Adaptive Learning Networks refer to deep learning architectures that adapt and learn from data to improve performance. These networks comprise layers with adaptive parameters, allowing them to capture complex patterns and dependencies within the data. Components like convolutional layers for feature extraction, recurrent layers for handling sequential data, attention mechanisms for focusing on important parts of the input, and optimization techniques such as stochastic gradient descent (SGD) contribute to the adaptability and learning capabilities of these networks. The adaptability of deep adaptive learning networks makes them well-suited for tasks like natural language processing, image recognition, and time series analysis.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2046-2051

Support Vector Machine (SVM)

The Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel is a powerful classification algorithm suitable for linearly and non-linearly separable data. It employs the kernel trick to transform data into higher-dimensional space, enabling the identification of an optimal hyperplane that separates different classes. SVM with RBF kernel's ability to maximize the margin between classes enhances its generalization and robustness, making it effective for binary and multi-class classification tasks. Tuning parameters like C (regularization parameter) and gamma (kernel coefficient) further optimize the model's performance.

Evaluation Metrics

Evaluation metrics are crucial for assessing the performance of machine learning models. Metrics such as accuracy, precision, recall (sensitivity), F1 score, area under the ROC curve (AUC-ROC), and confusion matrix provide insights into a model's classification accuracy, ability to avoid false positives, capture true positives, balance between precision and recall, discrimination ability across thresholds, and overall predictive performance. These metrics help in evaluating and comparing different machine learning algorithms and determining the most effective approach for classifying text messages and distinguishing between spam and non-spam instances.

4. RESULTS AND DISCUSSIONS

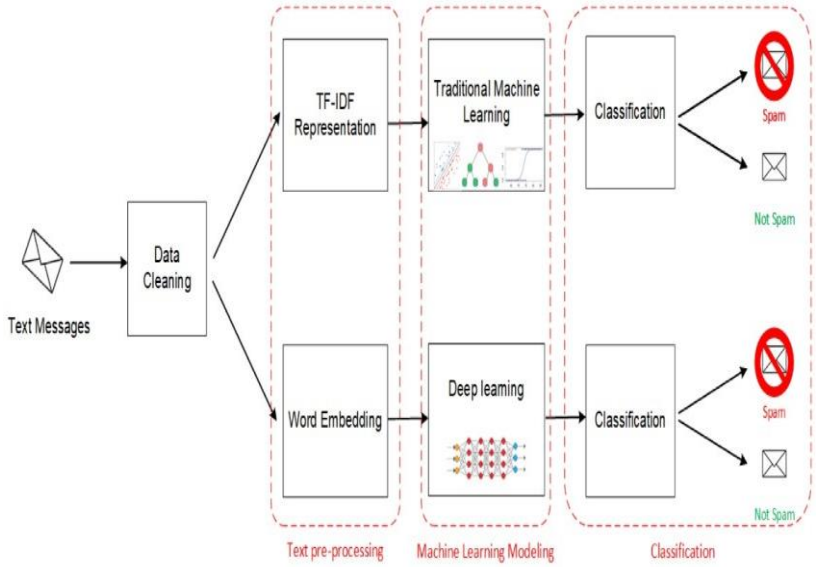

Figure 1: Architecture Diagram

A regression plot for views and likes illustrates the relationship between these two variables. The plot includes a scatter plot of data points, where each point represents a video or piece of content, with the x-axis showing views and the y-axis showing likes. Additionally, a regression line is fitted to the data points, indicating the overall trend or correlation between views and likes. This regression line helps visualize whether there is a positive or negative relationship between the two variables and how strong that relationship is. In simpler terms, it helps to understand if more views generally lead to more likes or vice versa, and to what extent.

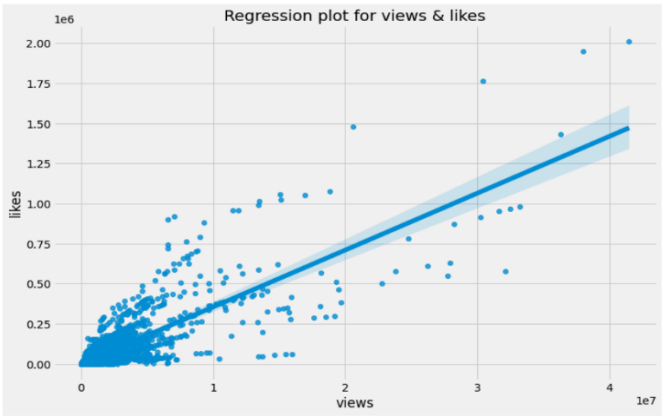

Figure 2: Regression plot for views &likes

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 04, April 2024, pp: 2046-2051

A regression plot of views and dislikes serves to depict the relationship between these two variables. It typically consists of a scatter plot where each point represents a piece of content or video, with views on the x-axis and dislikes on the y-axis. Additionally, a regression line is plotted through the data points to show the overall trend or correlation between views and dislikes. This regression line helps visualize whether there's a positive or negative relationship between the two variables and the strength of that relationship. Essentially, it helps to understand if more views tend to correlate with more dislikes, fewer dislikes, or if there's no significant correlation between the two.

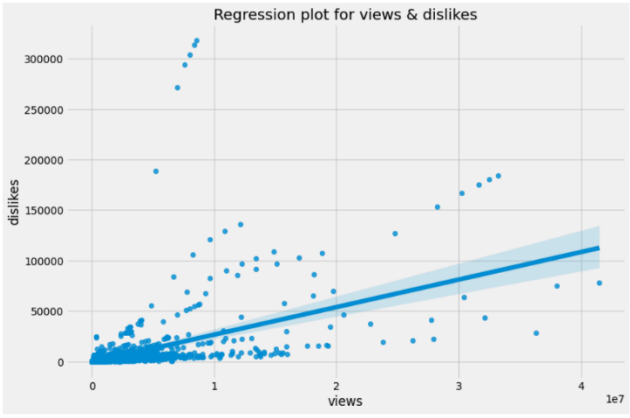

Figure 3: Regression plot for views & dislikes

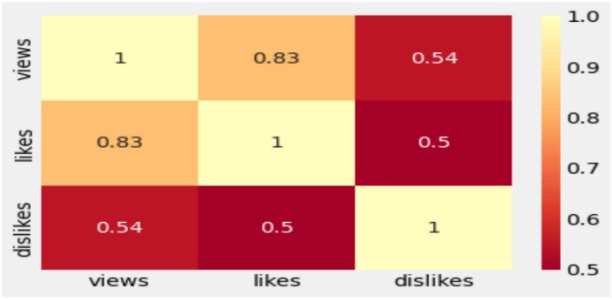

Figure 4: Heatmap of views, likes & dislikeas

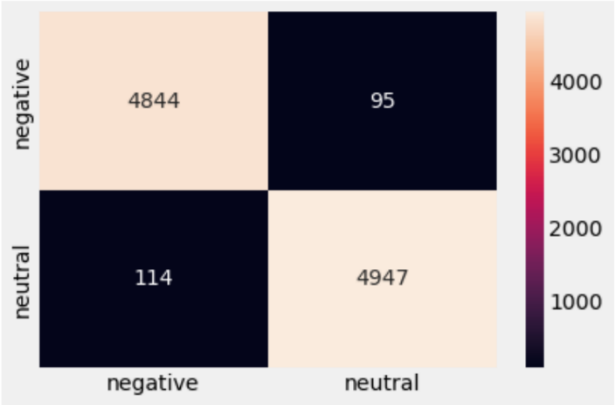

Figure 5: Heatmap of negative & neutral

In heatmap like a colourful map showing the emotional journey of viewers as they watch a video. Just like a weather map uses colours to show temperatures, this heatmap uses colours to represent different emotions. Imagine the video timeline stretched out horizontally, and as you move along it, you encounter different coloured patches. These patches represent how people are feeling at that moment in the video. For example, green might mean viewers are feeling positive, red could indicate negative emotions, and yellow might represent a neutral or mixed response. By looking at this heatmap, you can see at a glance which parts of the video made people happy, which parts made them upset, and where they were indifferent. Creators and marketers can use this information to understand audience reactions, identify popular or controversial moments, and improve future content to better connect with viewers. It's like getting a colourful emotional roadmap of how people are responding to the video.

5. CONCLUSION

This study presents a significant evolution in binary text classification, leveraging both traditional methods and advanced deep learning techniques, particularly focusing on LSTM models that have demonstrated substantial superiority in performance. The research introduces two specialized LSTM models tailored for categorizing text from a dataset containing instances of Spam and Fraud into distinct classes, utilizing pre-trained word embeddings to capture intricate semantic nuances between words. A novel approach is proposed through a joint LSTM model designed to address multi-label text classification by amalgamating text from the Spam and Fraud datasets based on shared bigrams, resulting in a dataset with four non-exclusive labels. This innovative strategy enables the model to simultaneously classify text into two binary categories, showcasing enhanced versatility in complex classification tasks. Despite the inherent complexities of multi-label classification, the joint LSTM model outperforms independent LSTM models in both Spam and Fraud classification tasks, with its performance significantly boosted by integrating text from disparate datasets despite variations in size and format. This amalgamation underscores the potential advantages of combining datasets from similar domains for multi-label text classification. The experimental evaluation confirms the efficacy of the proposed neural network models, with consistently high performance observed for the two independent binary models and satisfactory results obtained from the joint model, validating its robustness and efficiency in addressing text classification challenges. These findings contribute novel LSTM models for binary text classification and demonstrate the benefits of employing a joint model for multi-label classification tasks, highlighting their potential for real-world applications in detecting Spam and Fraud instances in textual data.

6. REFERENCES

[1] Y. Hu and M. Chen, Information diffusion prediction in mobile social networks with hydrodynamic model, IEEE Int. Conf. Commun. (ICC), pp. 1-5, 2016.

[2] M. Chen, Y. Qian, S. Mao, W. Tang and X. Yang, Software-defined mobile networks security, Mobile Netw. Appl., vol. 21, pp. 729-743, Oct. 2016.

[3] M. AlRubaian, M. Al-Qurishi, M. Al-Rakhami, S. M. M. Rahman and A. Alamri, A multistage credibility analysis model for microblogs, Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, pp. 1434-1440, Aug. 2015.

[4] M. Al-Qurishi, M. Al-Rakhami, M. AlRubaian, A. Alamri and M. Al-Hougbany, Online social network management systems: State of the art, Proc. Comput. Sci., vol. 73, pp. 474-481, Jan. 2015.

[5] M. Al-Qurishi, R. Aldrees, M. AlRubaian, M. Al-Rakhami, S. M. M. Rahman and A. Alamri, A new model for classifying social media users according to their behaviors, Proc. 2nd World Symp. Web Appl. Netw. (WSWAN), pp. 1-5, Mar. 2015.

[6] H. Oh, A YouTube spam comments detection scheme using cascaded ensemble machine learning model, IEEE Access, vol. 9, pp. 144121-144128, 2021.

[7] B. Pang and L. Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr., vol. 2, no. 2, pp. 1-135, 2008

[8] B. Liu, Sentiment analysis and opinion mining, Synthesis Lect. Hum. Lang. Technol., vol. 5, no. 1, pp. 1-167, 2012.

[9] E. Fitzpatrick, J. Bachenko and T. Fornaciari, Automatic Detection of Verbal Deception, San rafael, CA, USA:Morgan & Claypool, 2015.

[10] N. Jindal and B. Liu, Analyzing and detecting review spam, Proc. IEEE Int. Conf. Data Mining, pp. 547-552, Oct. 2007 .