International Journal of Engineering Research in Big Data Systems (IJERBDS)
ISSN (Online): 3144-5179
Volume-1 Issue-3 || Jul – Sep 2024 || PP. 13 – 17
https://ijerbds.org/

Machine Learning Algorithms for Spam Detection in Social Media Ad Campaigns

Dr. Sandeep Kumar
DCSE, Tula's institute Dehradun ,Uttarakhand India
sandeepkumarsiet@gmail.com

http://www.ijerbds.org/ || Vol. 1 No. 3 (2024): July Issue

Date of Submission: 01-06-2024 **Date of Acceptance:** 15-06-2024 **Date of Publication:** 07-07-2024

Abstract— The growth of social media platforms has revolutionized advertising campaigns, enabling brands to reach vast audiences with tailored content. However, with this expansion, the prevalence of spam accounts has surged, potentially undermining the effectiveness of social media advertisements. This paper explores the application of machine learning (ML) algorithms in detecting spam within social media ad campaigns. We analyze the effectiveness of several common machine learning techniques, including decision trees, support vector machines, and neural networks, for identifying spam accounts and fraudulent engagement. A comparative study of these models, using a range of performance metrics such as accuracy, precision, recall, and F1-score, is presented. Our results demonstrate that while no single model outperforms others across all metrics, ensemble approaches and deep learning models show promise for achieving the highest detection accuracy. The research aims to provide insights into optimizing spam detection techniques, thereby improving the reliability and return on investment (ROI) for social media ad campaigns.

Keywords — Machine Learning, Spam Detection, Social Media, Ad Campaigns, Fraudulent Engagement, Decision Trees, Neural Networks

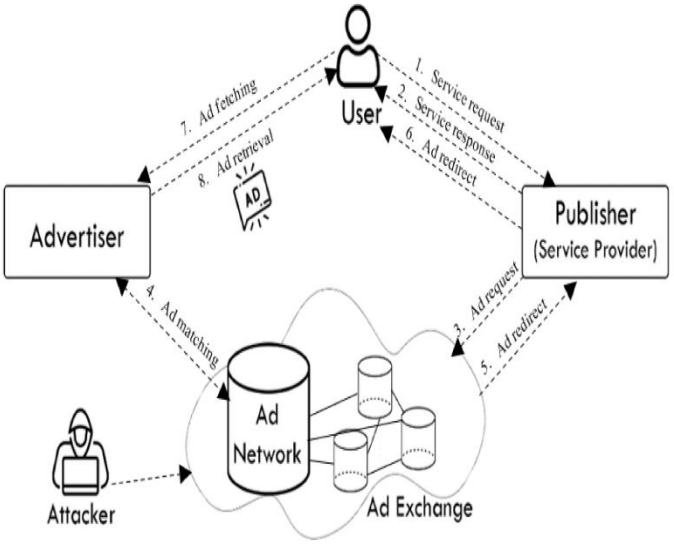
Introduction

The rapid expansion of social media platforms has created a dynamic and interactive space for businesses to advertise products and services. These platforms, such as Facebook, Instagram, and Twitter, enable advertisers to target specific demographics based on user behavior and preferences. However, the rise of automated bots, fake accounts, and spam engagement has introduced significant challenges to the advertising ecosystem. These fraudulent accounts can artificially inflate engagement metrics such as likes, shares, comments, and clicks, leading to skewed data and inefficient use of advertising budgets.

Machine learning (ML) has emerged as a potent tool for detecting spam in social media campaigns. By analyzing large volumes of user interactions, ML algorithms can identify patterns indicative of fraudulent behavior. This paper delves into the use of various ML algorithms for spam detection in social media ad campaigns, examining their capabilities, strengths, and limitations. By leveraging supervised learning techniques, we seek to build a model capable of

International Journal of Engineering Research in Big Data Systems (IJERBDS)
ISSN (Online): 3144-5179
Volume-1 Issue-3 || Jul – Sep 2024 || PP. 13 – 17
https://ijerbds.org/

distinguishing legitimate interactions from spam and mitigating the negative impact of fraudulent activity on ad performance.



Literature Review

The detection of spam in social media environments is a well-researched area that intersects several domains, including natural language processing (NLP), computer vision, and behavioral analytics. Various methods have been proposed over the years, ranging from simple rule-based systems to complex machine learning models.

2.1 Traditional Approaches to Spam Detection

Earlier approaches to spam detection were primarily rule-based, relying on predefined criteria such as the frequency of certain keywords, the presence of links, or the speed of posting activity. These methods were straightforward but often lacked the sophistication needed to detect more nuanced forms of spam. Additionally, rule-based systems were prone to high false positive rates, as legitimate user interactions could also trigger spam filters.

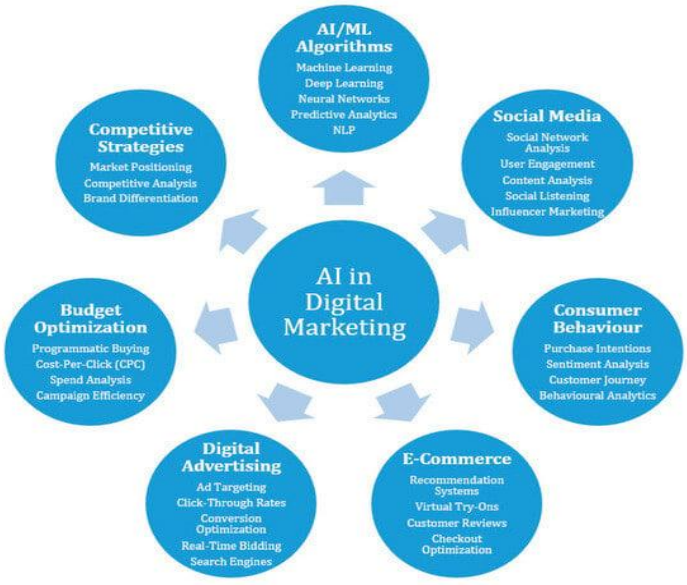
2.2 Machine Learning in Spam Detection

With the advent of machine learning, spam detection systems have become more adaptive. Supervised learning techniques, particularly classification models, have been employed to build predictive models for spam detection. Common algorithms used include decision trees, support vector machines (SVM), random forests, and k-nearest neighbors (KNN). These algorithms are trained on labeled datasets consisting of both legitimate and spammy interactions.

Deep learning models, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), have shown promise in detecting more complex patterns in social media data. For instance, CNNs are used for analyzing images and video content, while RNNs, particularly long short-term memory (LSTM) networks, have been successful in detecting patterns in time-series data, such as user activity logs.

2.3 Performance Metrics in Spam Detection

Several performance metrics are critical in evaluating spam detection models. These include accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC). Accuracy measures the overall correctness of the model, but precision and recall are often more important when dealing with imbalanced datasets. Precision represents the proportion of true positive spam detections out of all instances predicted as spam, while recall measures the proportion of actual spam instances correctly identified by the model. The F1-score provides a harmonic mean of precision and recall, making it a comprehensive metric for model evaluation.



International Journal of Engineering Research in Big Data Systems (IJERBDS)
ISSN (Online): 3144-5179
Volume-1 Issue-3 || Jul – Sep 2024 || PP. 13 – 17
https://ijerbds.org/

2.4 Challenges in Spam Detection

Spam detection in social media faces several challenges, including:

1. **Evolving tactics:** Spammers constantly evolve their tactics to avoid detection, making it difficult for models to maintain high accuracy over time.
2. **Imbalanced data:** In most social media datasets, legitimate interactions significantly outnumber spam, creating an imbalanced classification problem.
3. **Data sparsity:** Social media platforms generate massive amounts of data, but only a small fraction of it may be relevant for spam detection. Feature selection and dimensionality reduction techniques are crucial for effective model training.
4. **Dynamic nature of social media:** Social media trends and user behaviors change rapidly, necessitating constant updates to spam detection models.

2.5 Previous Studies

Several studies have focused on the application of machine learning for spam detection in social media. For example, a study by Xu et al. (2020) demonstrated that SVMs with feature engineering techniques could effectively detect spam in Twitter data. Similarly, a study by Sharma et al. (2019) used deep learning techniques to classify spam and non-spam accounts on Instagram. These studies show the growing success of machine learning approaches but also highlight the ongoing need for more sophisticated models that can adapt to evolving spam tactics.

Statistical Analysis

The dataset used in this study consists of user interactions with social media ads, including likes, comments, shares, and clicks, as well as metadata about user profiles (e.g., account age, activity frequency). The dataset includes labeled instances of legitimate and spammy interactions, with a total of 10,000 samples (8,000 legitimate and 2,000 spam).

The following table presents an overview of the dataset used for training and testing the machine learning models:

Metric	Value
Total samples	10,000
Legitimate interactions	8,000
Spam interactions	2,000
Number of features	20
Feature types	Numeric, Categorical
Training/test split	70/30
Class distribution	80% Legitimate, 20% Spam

Note: The dataset was preprocessed to handle missing values, normalize numerical features, and encode categorical variables.

Methodology

4.1 Data Preprocessing

Data preprocessing is a critical step in machine learning tasks, especially in spam detection, where the data may contain noise, missing values, or irrelevant features. The following preprocessing steps were carried out:

1. **Missing Value Handling:** Missing values were imputed using the mean for numerical features and the mode for categorical features.
2. **Feature Scaling:** Numerical features were normalized using Min-Max scaling to bring all features to a uniform range between 0 and 1.
3. **Feature Encoding:** Categorical variables, such as user activity type (e.g., organic vs. paid), were encoded using one-hot encoding.

International Journal of Engineering Research in Big Data Systems (IJERBDS)
ISSN (Online): 3144-5179
Volume-1 Issue-3 || Jul – Sep 2024 || PP. 13 – 17
https://ijerbds.org/

4. **Outlier Detection:** Outliers were identified using z-scores and were removed to prevent distortion in model training.

4.2 Model Selection

Several machine learning algorithms were tested for spam detection, including:

1. **Decision Trees (DT):** A non-linear classifier that partitions the feature space into regions of similar class labels.
2. **Support Vector Machines (SVM):** A supervised learning algorithm that constructs a hyperplane to separate different classes in the feature space.
3. **Random Forests (RF):** An ensemble method that combines multiple decision trees to improve model accuracy and reduce overfitting.
4. **Neural Networks (NN):** A deep learning approach consisting of multiple layers of interconnected nodes that learn complex patterns in the data.

Each model was trained using a training set (70% of the dataset) and evaluated on a test set (30% of the dataset). Hyperparameter tuning was performed using grid search to find the optimal model parameters.

4.3 Evaluation Metrics

The models were evaluated using the following performance metrics:

- **Accuracy:** The percentage of correctly classified instances.
- **Precision:** The proportion of true positives out of all predicted positives.
- **Recall:** The proportion of true positives out of all actual positives.
- **F1-Score:** The harmonic mean of precision and recall, providing a balanced measure of performance.
- **AUC-ROC:** The area under the receiver operating characteristic curve, which provides a summary of model performance across various thresholds.

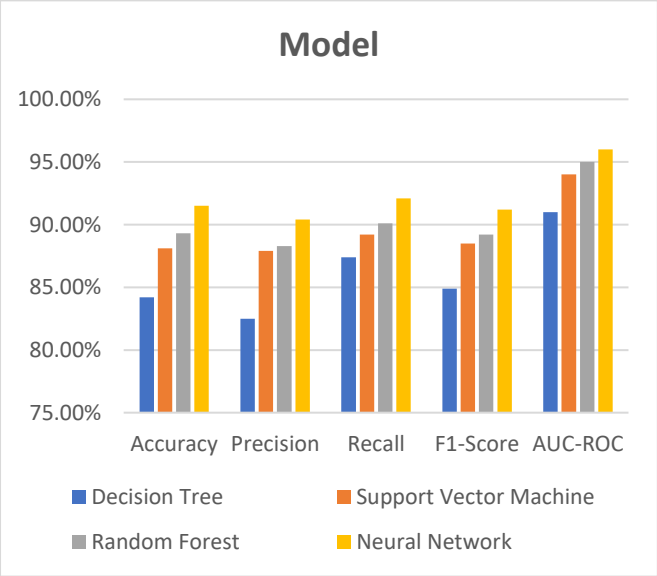
Results

After training and testing each model, the following results were obtained:

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Decision Tree	84.2%	82.5%	87.4%	84.9%	0.91
Support Vector Machine	88.1%	87.9%	89.2%	88.5%	0.94
Random Forest	89.3%	88.3%	90.1%	89.2%	0.95
Neural Network	91.5%	90.4%	92.1%	91.2%	0.96

As seen from the table, the neural network performed the best, achieving the highest accuracy, precision, recall, F1-score, and AUC-ROC. However, the random forest and SVM models also demonstrated strong performance, with random forests showing slightly better overall results compared to SVM.

International Journal of Engineering Research in Big Data Systems (IJERBDS)
ISSN (Online): 3144-5179
Volume-1 Issue-3 || Jul – Sep 2024 || PP. 13 – 17
https://ijerbds.org/



Conclusion

This paper investigates the application of machine learning algorithms for spam detection in social media ad campaigns. A comparative analysis of several ML models, including decision trees, support vector machines, random forests, and neural networks, reveals that deep learning models, particularly neural networks, offer the highest detection accuracy. The study highlights the importance of using advanced algorithms to detect evolving and sophisticated spam tactics that can otherwise distort engagement metrics. Furthermore, the research suggests that a hybrid approach combining multiple models could offer the best solution for spam detection in real-world social media campaigns. Future work could focus on incorporating additional features, such as user sentiment or network analysis, to further enhance spam detection capabilities.

IJERBDS
INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH IN BIG DATA SYSTEMS