CSC8499 - Project and Dissertation for MSc in Advanced Computer Science

Twitter Spam Detection

Interim report

By
Kailash Balachandiran
220243160

1. Introduction

The goal of the Twitter spam detection project is to develop a method for quickly and accurately identifying and removing spam content from the social networking site. It uses several ML models, NLP strategies, and an intuitive web application [1]. Flask is used to build the backend, and PyMySQL is utilized to connect to databases and securely store user login data. The user interface for interacting with the spam detection system is the web application. It has an easy-to-use front end made using HTML and CSS that enables users to enter Twitter account information or tweet content for spam analysis. The system employs ML models that were developed using instances of both spam and valid content from a labelled dataset [2]. Each account or tweet is given a spam chance score based on features gleaned from user accounts, tweet content, engagement patterns, and network relationships. Communication between the front-end and ML models is facilitated using the adaptable web framework Flask. Users are given the option to submit queries and get instant spam detection results. PyMySQL makes ensuring that data is persistent and that user credentials are securely stored in MySQL databases, protecting user privacy [1]. A complete system for detecting Twitter spam is created by combining ML models, NLP approaches, Flask, PyMySQL, and HTML/CSS. The web application offers users a user-friendly interface, improving the functionality and security of the Twitter network.

2. Aim and Objectives

This project aims to develop a reliable Twitter spam detection system [4]. To filter out spam, it examines user accounts, tweet content, interaction patterns, and network connections using state-of-the-art machine learning algorithms and natural language processing methods [3]. Additionally, the project places a lot of emphasis on creating an easy-to-use online application for quick access to the spam detection system, which improves Twitter user experience. The goals are to provide effective spam identification and filtration while also enhancing the platform's anti-spam environment.

- Develop and train machine learning models that can classify tweets and user accounts as authentic or spam, continually enhancing the models to consider new spamming tactics.
- Apply natural language processing techniques to extract relevant features from tweet content and identify spam-related patterns such as unwanted ads and malicious links.
- Construct a user-friendly online application with HTML/CSS and Flask that enables users to enter Twitter account details or tweet content and receive real-time spam detection results.
- Ensure data security and persistence by using PyMySQL and SQL databases to safely store and retrieve user login information.
- Taking user suggestions into account and making the system more capable of handling a lot of spam detection requests at once.

By fulfilling these goals, the project hopes to provide a safer and more enjoyable user experience by offering a practical and straightforward method for identifying and reducing spam on Twitter.

3. Overview of Progress

3.1 Programming:

To train the system for Twitter spam identification, Multiple machine-learning models and techniques have been incorporated., To do this, code had to be written to pre-process and modify the data, and models had to be trained on a labelled dataset. To create and improve the machine learning models, the implementation makes use of Python libraries like sci-kit-learn and TensorFlow. To construct the web application's backend logic, Flask is used as the web framework. To do this, technology had to be written to manage user requests, connect with machine learning models, and deliver real-time spam detection findings. The front end of the web application was designed using HTML and CSS to be both aesthetically beautiful and user-friendly [6].

3.2 Research & Analysis:

There has been a lot of research done to comprehend the difficulties and intricacies of Twitter spam identification [1]. To understand more about various machine learning algorithms, natural language processing techniques, and network analysis methodologies for spam identification on social media platforms, an assessment has been done on the body of current literature, research papers, and industry best practices. Also, performed analysis and trials to find characteristics and trends that are telling of spam content on Twitter [2]. To do this, numerous feature extraction strategies were investigated, and the effectiveness of those techniques was assessed using pertinent metrics [1]. To ensure the system's accuracy and dependability, the research phase has helped identify the best methods for spam identification.

Natural language processing techniques have been applied to analyze tweets' textual content [5]. This includes sentiment analysis, subject modelling, and spotting spam-related trends like overuse of hashtags, mentioning URLs, or using certain terms frequently [2].

To improve the system's adaptability and efficacy, Twitter user behaviors and spam patterns from the real world have been examined [3]. Monitoring spamming strategies, seeing new spam trends, and putting this knowledge into the spam detection algorithms were required for this [4]. The system's capacity to identify and remove spam content has been constantly improved.

Since deep learning methods have been developed, researchers have considered employing neural networks for Twitter spam detection [3]. This includes the use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to automatically extract valuable information from tweet content and improve spam detection accuracy [6].

Additionally, research has concentrated on creating real-time spam detection algorithms that are capable of accurately identifying and filtering spam content as it appears on Twitter [1]. These analyses entail tracking, evaluating, and modifying the most recent spamming strategies to modify detection methods.

Overall, in terms of programming, research, and analysis, the project has advanced significantly. An efficient and user-friendly system for Twitter spam detection has been developed because of the use of machine learning models, the construction of the web application, and the research and analysis that have been done.

4. Project Plan

Month	April 2023				
Week	1	2	3	4	5
	Dissertation topic selection			Project Initiation	
Month	May 2023				
Week	1	2	3	4	
	Research and Requirements Gathering		Ethical form approval	Data collection and preparation	
Month	June 2023				
Week	1	2	3	4	5
	Interim report	Machine Learning Model Development		Natural Language Processing Techniques	
Month	July 2023				
Week	1	2	3	4	
	Web Application Development		Evaluation and Refinement		
Month	August 2023				
Week	1	2	3	4	5
	Documentation and Reporting			Final Demonstration	

Table 1

Table 1 project plan explains the key actions and tasks required in creating the Twitter spam detection system. It includes analysis, data preparation, creation of machine learning models, creation of web applications, integration, testing, and evaluation, as well as documentation, deployment, and maintenance. The project's goals will be met, and a reliable, approachable spam detection solution will be delivered thanks to the plan's systematic methodology.

References

[1] bkenar, Sepideh & Haghi Kashani, Mostafa & Akbari, Mohammad & Mahdipour, Ebrahim. (2020). Twitter Spam Detection: A Systematic Review. Available at: https://arxiv.org/abs/2011.14754

[2] Wu, Tingmin & Liu, Shigang & Zhang, Jun & Xiang, Yang. (2017). Twitter spam detection is based on deep learning. 1-8. 10.1145/3014812.3014815. Available at: https://www.researchgate.net/publication/312428491_Twitter_spam_detection_based_on_deep_learning

[3] Kabakus, Abdullah Talha & Kara, Resul. (2017). A Survey of Spam Detection Methods on Twitter. International Journal of Advanced Computer Science and Applications. 8. 10.14569/IJACSA.2017.080305. Available at: https://www.researchgate.net/publication/315966273_A_Survey_of_Spam_Detection_Methods_on_Twitter

[4] Anisha P Rodrigues, Roshan Fernandes, Aakash A, Abhishek B, Adarsh Shetty, Atul K, Kuruva Lakshmanna, and R. Mahammad Shafi. (2022). Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques. Available at: https://www.hindawi.com/journals/cin/2022/5211949/#related-articles

[5] Sun, Nan & Lin, Guanjun & Qiu, Junyang & Rimba, Paul. (2020). Near real-time twitter spam detection with machine learning techniques. International Journal of Computers and Applications. 44. 1-11. 10.1080/1206212X.2020.1751387. Available at: https://www.researchgate.net/publication/340708941_Near_real-time_twitter_spam_detection_with_machine_learning_techniques

[6] Zulfikar Alom, Barbara Carminati, Elena Ferrari, A deep learning model for Twitter spam detection, Online Social Networks and Media, Volume 18, 2020, 100079, ISSN 2468-6964, Available at: https://doi.org/10.1016/j.osnem.2020.100079. (https://www.sciencedirect.com/science/article/pii/S2468696420300203)