**Detecting Malicious Twitter Bots Using Machine Learning**

**Abstract:**

Undoubtedly, social media, such as Facebook and Twitter, constitute a major part of our everyday life due to the incredible possibilities they offer to their users. However, Twitter and generally [online social networks](https://www.sciencedirect.com/topics/computer-science/online-social-networks) (OSNs) are increasingly used by automated accounts, widely known as bots, due to their immense popularity across a wide range of user categories. Their main purpose is the dissemination of fake news, the promotion of specific ideas and products, the manipulation of the stock market and even the diffusion of sexually explicit material. Therefore, the early detection of bots in social media is quite essential. In this paper, two methods are introduced targeting this that are mainly based on [Natural Language Processing](https://www.sciencedirect.com/topics/engineering/natural-language-processing) (NLP) to distinguish legitimate users from bots. In the first method, a feature extraction approach is proposed for identifying accounts posting automated messages. After applying feature selection techniques and dealing with imbalanced datasets, the subset of features selected is fed in [machine learning algorithms](https://www.sciencedirect.com/topics/engineering/machine-learning-algorithm). In the second method, a deep learning architecture is proposed to identify whether tweets have been posted by real users or generated by bots. To the best of the authors’ knowledge, there is no prior work on using an attention mechanism for identifying bots. The introduced approaches have been evaluated over a series of experiments using two large real [Twitter datasets](https://www.sciencedirect.com/topics/computer-science/twitter-dataset) and demonstrate valuable advantages over other existing techniques targeting the identification of malicious users in social media.

**Existing system:**

Due to the vast array of opportunities they present, social media platforms like Facebook and Twitter have become deeply embedded in our daily lives. However, due to the widespread appeal of Twitter and other OSNs, automated accounts commonly referred to as bots are rapidly making use of them.The train data has many attributes. The required features are extracted using Spearman correlation method. Random Forest.  **Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset*.***

**Disadvantages**:

* Low security

**Proposed system:**

There are numerous characteristics to the train data. The necessary functionality is retrieved using the Spear-man correlation technique. Logistic Regression algorithm by Machine learning.Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Real-time data as displayed are used as the optimum learning model. Data are preprocessed and zero values are eliminated using pandas (tool for pre-processing). The dataset is trained and the test data set is the actual Twitter data.The output is in shape 0 or1. We developed an algorithm in our research that identifies Twitter bots. indicate Malicious URL and in above screen we can see URL prediction accuracy as 73%. we got ML accuracy of Logistic Regression is 74% Thus, word algorithms were used to real-time data and the Twitter bots have been detected effectively

**Advantages:**

* High security
* High efficiency

**SYSTEM REQUIREMENT:**

**HARDWARE REQUIREMENTS:**

# Processor - Intel i3(min)

* Speed - 1.1 Ghz
* RAM - 4GB(min)
* Hard Disk - 500 GB
* Key Board - Standard Windows Keyboard
* Mouse - Two or Three Button Mouse
* Monitor - SVGA

**SOFTWARE REQUIREMENTS:**

* Operating System - Windows10(min)
* Programming Language - Python 3.7