BIRZEIT UNIVERSITY

Electrical and Computer Engineering Department

ENCS434 Artificial Intelligence, First Semester, 2021-2022

Machine Learning Project

Due: January 10, 2022

Automatic Tweet Spam Detection

Twitter spam is unwanted content manifesting in many ways! Including bulk messages, profanity, insults, hate speech, malicious links, and fraudulent reviews. Lets tackle this problem by building a classifier to detect when a tweet is "Quality" content or "Spam"!

1. *What is Spam?*

Spam is defined as the tweets that are posted by known fake twitter accounts that are:

- Politically Motivated
- Automatically generated content
- Meaningless content
- Click Bait

For more information about Spam tweet detection please read the following papers:

- Unsupervised collective-base d framework for dynamic retraining of supervised real-time spam tweets detection model. Mahdi Washha, Aziz Qaroush, Manel Mezghani, Florence Sedes.
- Kabakus, A. T. , & Kara, R. (2017). A survey of spam detection methods on twitter. International Journal of Advanced Computer Science and Applications.
- Wu, T., Wen, S., Xiang, Y., & Zhou, W. (2017). Twitter spam detection: Survey of new approaches and comparative study. Computers and Security, 76 . doi: 10.1016/ j.cose.2017.11.013 .

2. Problem Statement

The goal is to develop an automated solution that filters out the spam tweets based on extracted features from the tweet content and other attributes.

3. Data set:
You will train your system using the attached data set.

The data set contains the following fields:

- Tweet
  This is the text that was tweeted
- following
  The number of people the account that tweeted is following
- followers
  The number of people following the account that tweeted
- actions
  The total number of favorites, replies, and retweets of said tweet
- is_retweet
  Binary [0,1] value: If 0 its not a retweet, if 1 it is a retweet
- location
  The self-written location provided by the user on their profile, May not exist, be "Unkown", and is NOT standardized! ex. could be ("NY", "New York", "Upper East Side", Etc!)
- Type
  Either Quality or Spam

5. Submissions: Please submit the following:
1. **Report**:

- Describe in details your formalization of the problem including the stages of your solution, selected features, and results.
- describe in details how you designed each feature.
- The results you obtained including evaluation method, measures, and comparing t other related work.

2. **Source Code:** Include all the source code you developed or extended from the program. These need to be submitted only electronically (no hardcopies of the code). The running program needs also to be submitted electronically.

3. **Demo:** You will be asked to demo your work to your instructor. For that you need to be able to work with your program, introduce minor modifications and defend your choices.

Honor Policy: All are required to adhere to the University honor policy and violations will be dealt with according to university regulations.

Good Luck