**Spam detection on Twitter**

The project based on detecting spammers on Twitter platform. This project used classification approach for predicting the spam or non-spam. Scripts written in python programming language.

**Prerequisites**

* Install PyCharm IDE from internet. It’s open-source IDE.
* Install Anaconda environment from internet – set Python version – 3.5, you may use latest.
* Libraries and packages used – Pandas, Pandas as pd, Seaborn, Matplotlib, NLTK, WordCloud, Numpy, Math, PrettyTable

**Installing**

* Install PyCharm
* Create interpreter as conda environment and install all required libraries and packages
* Install libraries using pip install and always install libraries in conda environment

**Running the tests**

The folder provides different excel dataset files and python scripts.

Major 6 files provide for classification model – 3 files for cleaning the dataset and 3 files for classification results. Ignore files such as .idea, graphs, venv

Environment files also provided if have knowledge about setting up on PyCharm. Otherwise install all requires libraries and packages.

First run dataset cleaning script, then run classification script.

1. **Honeypot\_Classification.py** – 2009 dataset classification script
2. **Honeypot\_cleaning.py** – 2009 dataset pre-processed script
3. **Honeypot\_New\_Claassification.**py – 2019 dataset classification script
4. **Honeypot\_New\_Cleaning.py** – 2019 dataset pre-processed script
5. **SPD\_Classification.py** – SPD dataset classification script
6. **SPD\_cleaning.py** – SPD dataset cleaning script