TF-IDF classification based Multinomial Naïve Bayes model for spam filtering configuration manual

Alan Chavez
MSc Cybersecurity
National College of Ireland
Dublin, Ireland
X19137516@student.ncirl.ie

Configuration manual

In order to replicate the Multinomial Naïve Bayes model proposed is necessary install the following software.

- Download SMS Spam Collection Data Sat from Machine learning Repository https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection
- Python 3.8.3
- Anaconda 64 – Bit
    - Create an environment dedicated to use, select python 3.6
    - Intall Jupyter Notebook
    - Install the following python libraries:
        - Sklearn Library
        - Texbob Library
        - Matplotlib Library
        - Pandas Library
        - Sklearn Library
        - Nltk Library

    - Import the following libraries:
        - import matplotlib.pyplot as plt
        - import csv
        - from textblob import TextBlob
        - import pandas as pd
        - import sklearn
        - import pickle
        - import numpy as np
        - import nltk
        - from nltk.corpus import stopword

Libraries imported in jupyter notebook:

```
import sklearn

import pickle

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

import pickle

from textblob import TextBlob

import pandas as pd

import nltk

from nltk.corpus import stopwords

import string

from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer

from sklearn.naive_bayes import MultinomialNB

from sklearn.tree import DecisionTreeClassifier

from sklearn.metrics import classification_report, f1_score, accuracy_score, confusion_matrix

from textblob import TextBlob

from sklearn.model_selection import RandomizedSearchCV,learning_curve, GridSearchCV,
train_test_split, cross_val_score, StratifiedKFold

from sklearn.pipeline import Pipeline
```

Finally, The python code is included into the .zip file

The code base was obtained from https://radimrehurek.com/data_science_python/, properly referred in TF-IDF classification based Multinomial Naïve Bayes model for spam filtering, in addition some improvements were realized to achieve a better performance.