Classifying Crime & Terrorism Tweets

Samar Farzand Najmuddin

UNIVERSITY *of* STIRLING

MSc in Big Data

Project Aim: To collect tweets from twitter. Use different text classification techniques to label tweets as crime and terrorism related. From the volume of crime and terrorism tweets determine the safety of the city of London.

Supervised Learning Techniques Used :

- Support Vector Machines
- K Neighbors
- Logistic Regression
- Multinomial Naïve Bayes
- Decision Trees
- Random Forrest
- Ada Boost
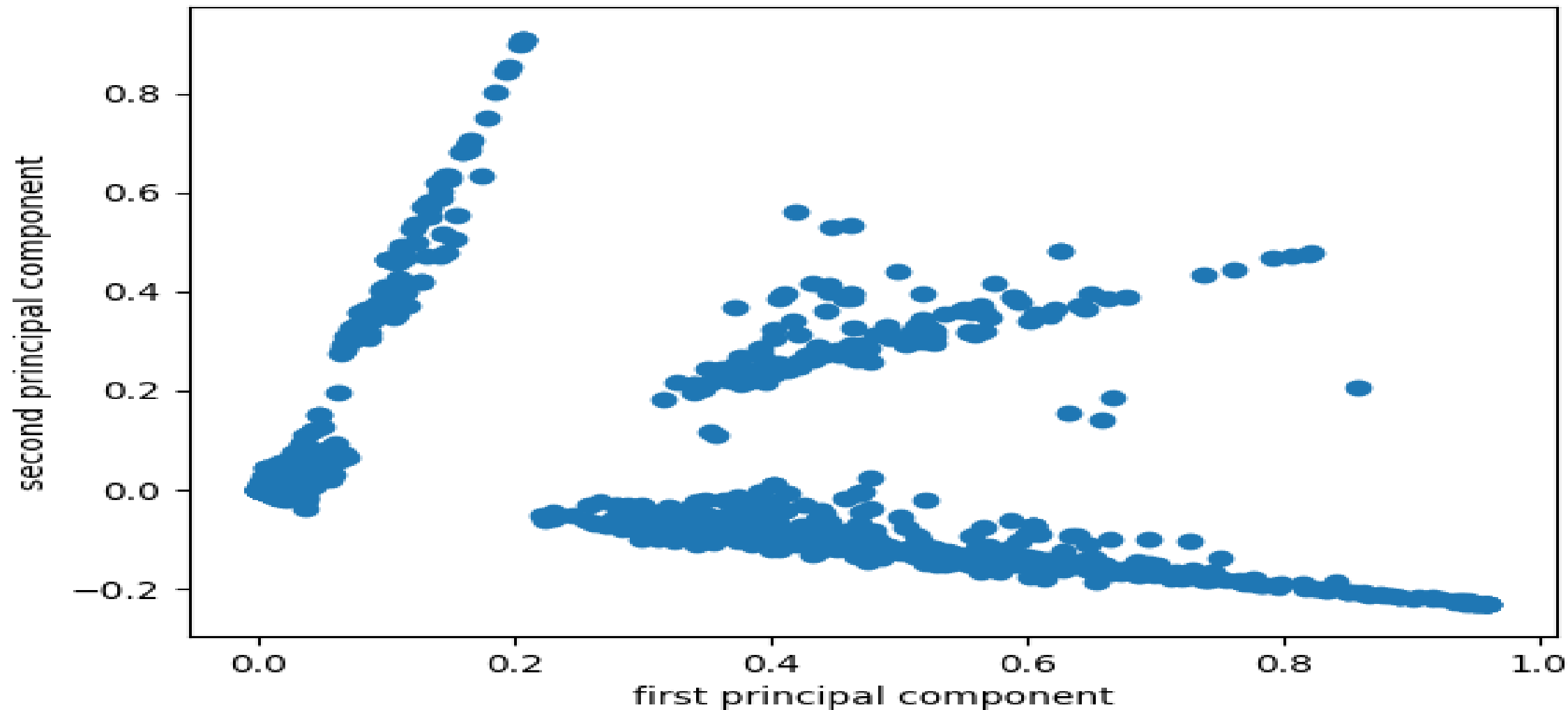- Neural Network Multilayer Perceptron

Unsupervised Learning Techniques Used:

- Latent Semantic Analysis
- Latent Dirichlet Allocation

Natural Language Analyses with NLTK

twitter

python

gensim — topic modelling for humans

scikit learn

matplotlib

Training of Model

Labelled data → Machine Learning Algorithm

Training Tweets → Feature Extractor → Features → Machine Learning Algorithm

Prediction of Model

Test Tweets → Feature Extractor → Features → Classifier Model → Labelled data

Plot of points against LSA principal components