Requirements Analysis

* Environment/IDE – Eclipse
* Machine Learning Algorithm – Naïve-Bayes Algorithm, Random Forest, etc.
* Platform- Java
* Operating System- windows 10 64-bit

Function Specification

* The function of the system is to categorize the tweets from the logged in user into Spam & Non-Spam categories
* The relevancy of the tweet/post should be checked by the application and then segregated.
* If there are a high number of spam tweets from a particular user then that user shall be blocked.
* Label based logic will be applied for categorization.

External Interface Specifications

1. User Interface

* The user interface will be the Twitter site itself.
* It will show the user the segregated tweets into spam and non-spam parts and the main feed shall consist for the relevant tweets.
* It will consist of a screen which will provide the user with relevant information.

1. Communication Protocol

* Communication interface will communicate the result acquired by the application to the user, this can be done after the user has logged-in into the account and then the application shall run in the background providing with the expected result.
* The screen shall also show all the default information that can be seen in any normal twitter account.

1. Hardware Interface

* In this spam detection project, there isn’t much of specific hardware requirement so any kind of machine capable of Web and Java is applicable.

1. Database Backend

* The training dataset will be a csv file which will all the nominal features of a tweet as parameters like time, date, no. of characters, etc.
* The algorithms will train the application on this dataset and provide a classification model to be used over the actual real-time logged-in data.
* The dataset will try to achieve maximum accuracy by training the ML algorithms over relevant data.

Technical specifications

1. Hardware Details :

* Ram : 8GB
* GPU : Graphics Card with basic HTML/PHP support
* Hard Disk Storage : 500 GB
* CPU : 2.5GHz

1. Operating System :

* Microsoft Windows/Linux

1. Performance :

* Processing time required by the system to process the input and and provide with the categorized tweets.
* Handle large amount of data generated from the huge number of tweets/posts.
* Using the FPR, TPR and F-measure attributes for statistical analysis.

1. Programming Languages and technologies :

* Java
* Eclipse/ IntelijIDEA/ Eclipse Che

1. Versions of components used :

* Microsoft Windows 10 - 64-bit
* Java (latest version)
* Eclipse Neon