

CHAPTER 1

INTRODUCTION

1.1. Background

Massive tweets loaded in twitter every second. User must know whether the tweet is negative or positive. This mean to avoid harmful tweets such as racism and intolerance tweets. The way of knowing such of all those things are by clustered the document of tweets via hashtag.

Using the calculation of *Tf-Idf* algorithm for data training and testing to get the rank result of each data. The data then continue calculated by the implementation of *k-Nearest Neighbour* (k-NN) to classify negativity on the document.

By the classification which had found out using *k-Nearest Neighbour* (k-NN), user will know which one is the negative tweet or positive.

1.2. Problem Formulation

1. What hashtag is used for data sampling?
2. Which trending region is used for data sampling?
3. How many data will be taken?
4. How accurate will be the classification?

1.3. Scope

To separate and to classify the negative and positive one and to find out how toxic social media will be.

1.4. Objective

This project classified the tweets from twitter by looking from the lowest score on each tweet and compared with data training.