

**Parallel Distributing Computing**

**Group Name: Dynamic Developers**

**Team Members:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Student Id** | **Class Id** |
| Muhammad Ghous Sarwar | 59455 | 100048 |
| Tahreem Feroz |  |  |
| Syed Fasih |  |  |
| Ismail Javery | 60152 | 100048 |

**Project Name: Find Sentiment of Tweet on Cluster**

**Introduction:**

* Twitter.com is a popular microblogging website.
* Each tweet is 140 characters in length.
* Tweets are frequently used to express a tweeters emotion on a particular subject.
* There are firms which poll twitter for analyzing sentiment on a particular topic.
* The challenge is to summarize the overall sentiment on a specific topic.

**Problem Statement:**

* Creating a huge data set for training, testing, and evaluation purposes partially using distributed computing.
* Evaluation of any give topic by using our trained algorithm.

(negative, positive, and neutral)

**Methodology:**

**Distribution of Tweets:**

This could be done in two different ways:

* Distribution of words: where the words are distributed on different machines. In this case, the final polarity of a tweet cannot be decided until the polarity of individual words has been received.
* Distribution of tweets: in this method, (complete) tweets are distributed on different machines. In this case, polarity of a tweet can be calculated on a machine.

**Steps Required:**

* Pre-process the tweets. Pre-processing includes but is not limited to stemming, passing through a stop-list, replacing URLs with a keyword URL, replacing user names etc.
* Perform tokenization in order to get words, which be uni-grams, bi-grams or tri-grams.
* Distributing either (complete) tweets or words of a tweet on different machines as discussed in Section 1
* Application of SentiWordNet to find the polarity of a tweet