**ATOS 4 group (Social media):**

**Andreas Mihaloianis, Momshad Alvee Dinuri, Chaitanya Agrawal**

**27.02.2015**

**Overview:**

The team has been working hard these two weeks to complete some very mundane but important task. We as a team had to read about 1000 tweets and label them one at a time. This task took some time, as we had to fetch the 1000 relevant tweets from twitter. Other tasks included having to clean the interface and getting the working version of the project live on azure as there are some simple syntactic changes that need to be made to make sure the project works everytime we export it from our local server.

**Summary of meetings held:**

**List of tasks completed and estimations:**

Around 1000 tweets have been manually labelled by the team as either good, bad or neutral. This data is to act as the training data for our algorithm. It was necessary to collect our own training data as similar datasets are not available online due to twitters terms and condition saying thats tweets cannot be stored in excess of 24 hours. The azure website has been updated with the latest working version of the code as many things work differently on azure compared to on our local web server. The interface has been cleaned and made consistent, the control-panel/settings page has been greatly improved to adjust its display according to the amount of data available etc. We are almost near to completing the project, we have the main infrasturcture setup and in the future we will add a few more addons to make it complete.

**Plan for next two weeks:**

We are going towards the end of the project and we almost have a functional app. From now on, we only add up some functionalities. Our main concern for the following 2 weeks will be working on the bag of words algorithm, which is an important stage of our work. We hope to have a working algorithm by the end of the next sprint. Also, we aim to have the latest, debugged application on the azure server so that our client can start giving us a more in depth feedback on it. This would help us track our development and maintain the good progress we are having at the moment.

**Individual Paragraphs:**

**Andreas:**

During the last 2 weeks I have been actively working in within the group’s activities. Our project starts to shape up as a whole, as we have identified the bag of words algorithm as the method to work out the sentiment analysis faster. Each of us (including myself) have tagged 300 tweets manually which added to the other tweets we manually tagged at the beggining of the project will constitute the training data of the algorithm. Also, I took care of making the interface of all the pages look the same which leads to a much user-friendlier UI. Nontheless, at the moment we are trying to make the application work on the azure server, because it behaves different than on the local machines, attempt on which all of us is engaged.

**Alvee:**

In the past two weeks I labelled 300+ tweets according to their sentiment. This data is to act as the training data for our bag of words algorithms. We will use these tweets to train the model and then cross-validate the model with test cases. I mainly focused on the bag of words algorithm, the algorithm is ideal for our applicaion but there are many difficulties in applying any natural language processing algorithm as many pre-processing cleaning steps need to be taken. Each of these string cleaning actions needed to be considered as the tweets contained a lot of mis-spellings. I also created a dummy untrained version of the algorithm to check the speed of execution and I am happy to say that we can expect instant respone with this new algorithm which our clients want, intead of the 30 seconds delay that we have using the DatumBox API.

**Chaitanya:**

I labeled nearly 300 tweets as good, bad or neutral to contribute to the dataset

for training our algorithm. There was a bug on the Control Panel/Settings

page that made the parameter buttons stick to each other. I fixed that bug by

introducing a HTML 5 closing break tag in the code. The emoticons rendering

task completed in the previous sprint was made more efficient by making use

of Twitter official emoticons rendering library, Twemoji and some JavaScript

code. Our application’s code currently behaves differently on Azure Cloud

Services server than a local web server and results in incorrect date/time,

hyperlink and emoticon rendering. There was another issue of tweets being

displayed as blank boxes on the Azure website which was fixed by inspecting

webpage elements in Google Chrome and making relevant syntax changes.

I also changed HTML 5 code of several webpages to make sure that the

design pattern was uniform and consistent throughout the application.