**Project “Twitter Data Stream Sentiment Analysis”**

The following report covers our minor project in big data analysist. Detailing the challenges faced and the progress made over the span of several weeks.

The main objective we set out to accomplish was performing sentiment analysis on the content of tweets, and to present this in several directly understandable manners. This task can be split into three distinct sections: gathering and sorting the data, performing analysis on the data and lastly portraying this analysis in an useful medium.

(soort achtige afbeelding)

**Table of contents:**

* **Motivation**
* **In-depth project setup**
  + **Obtaining the data**
  + **Analysing the data**
  + **Visualisation of analysis**
* **Results**
* **Addendum**
  + **Similar projects**
  + **Alternate possibilities**

**Objective synopsis: (general description)**

* obtaining the data.
* analysis.
* Visualisation of the analysis.

**Motivation:**

Description of the use of what we did  
Description of made choices? Vs include in in-depth description

**In-depth project setup:**

**Obtaining the data:**

We have chosen to concurrently use two methods of data gathering, i.e. already created data sets obtained from a database, along with live streamed data directly from twitter itself.

**Data sources: what data do you need?/where would you find it**  
Mined data from twitter, gotten directly from twitter through the usage of an api.  
Text, coupled with several features (i.e. Likes, retweets, reactions, user data) saved in a to be determined format.

**Converting obtained data:**

Tools used to obtain data + convert said data into a usable format:  
Used environment / packages  
from-to formats

**Analysing the data:**

* Divide the data into positive - negative tweets
* Give the overall opinion of the public (i.e. twitter users) of the topic
* Give the popularity of given subject

Tools used to sentiment analyse the data  
used environment / packages

**Visualisation of analysis:**

Description of used manners of visualisation

Detail on tools used:  
used environment / packages

**Results:**

Couple of graphs + part of tweet data set?

**Addendum:**

**Similar project / methods:**

Multiple project setup to do twitter analysis

Mainly business focused models renting their services for r&d purposes

**Alternate methods we could have used:**

Spark streaming