3. Analysing the Sentiments of Tweets

You are going to work on the same tweets dataset of the previous lab. In this task, you are asked to analyse the sentiments of the tweets. For example tweet A: “I like Nokia” has a positive sentiment, a tweet B: “I hate latex” has a negative sentiment, and a tweet C: “Trump visited Germany” has a neutral sentiment. The expected outcome is to classify a given tweet into positive, negative, or neutral depending on its content. The steps to follow for tweet sentiment classification are:

1. The dataset is already labeled. Identify the meaning of each label.
2. Define the attributes that represent the tweets (class label must be excluded).
3. Use Weka to classify the tuples with Decision Tree (J48) and Bayesian classifier. Try different evaluation methods, what do you observe?
4. Reduce your features only to positive and negative words. The lists of positive and negative words are available on the course web page.
5. Do you see any difference in the results? Explain your findings.
6. Optional: how the algorithms behave with a much larger training set? To answer this question, use your dataset as a test set and take the training set from the course web page. The training set contains over 1 million tweets.

**Deliverable:** Write a report of 1-2 pages about your findings explaining (1) how did you construct the attributes, (2) A comparison between the two models with different settings and evaluation strategies, (3) the effect of using only positive and negative words as attributes.

**Deadline:** 14/12/2017 at 2pm.

---

2. [https://www.cs.uic.edu/liub/FBS/sentiment-analysis.html](https://www.cs.uic.edu/liub/FBS/sentiment-analysis.html)