Vrije Universiteit Brussel

Proposal bachelor thesis

Title: Analyzing commuter satisfaction through Twitter

Promotor: Y-M De Hauwere

Advisor: Maarten Deville, Peter Vrancx

Includes preparation course: Yes

Context

The performance and punctuality of the Belgian Train Services (NMBS/SNCB) have been the topic of many complaints in recent years. While trains can run late for different reasons, the effect for commuters is the same: poor service. As of the 14th of december, a new network plan will be in effect that has a major impact on existing lines. The end result should be a more reliable service, and thus more satisfied users. Whether or not this will be the case remains to be seen and is something we will investigate in this project.
Data science has recently emerged as a tool to provide new insights in problems that are otherwise hard to analyse. By mining social media, researchers have studied the effects of disasters, predicted election results and optimized public transport systems. In order to measure user satisfaction with the train services, we will turn to freely available Twitter data. This data should provide insights into commuter satisfaction, as unsatisfied users tend to turn to social networks to ventilate their frustrations. As of October 2013, the NMBS/SNCB even has a full-time Twitter team to respond to user complaints, ensuring a high availability of data.

Proposal bachelor thesis

In this bachelor thesis you will employ data science techniques to analyze Tweets that are related to NMBS/SNCB. The idea is to obtain as much insights about the train performance as possible and then use these insights to compare data from both the old network plan (in effect up till 13th of december) with data from the new network plan (in effect from 14th of december). This provide us with some insights to whether or not the new plan has a positive impact on the commuter satisfaction.
More precise, your project will consist out of multiple tasks:

- Fetch Tweets based on hashtags related to NMBS/SNCB from the Twitter API
- Perform sentiment analysis on these Tweets and evaluate the relation between average train delay and tweet sentiment.

- Gather contextual data such as Tweet location, weather information and train information from NMBS/SNCB to see how these relate to the amount of Tweets
- Examine and visualize patterns in the twitter data. This will consist of studying relations between time, location and sentiment.
- Compare insights from the old and the new network plan to evaluate its impact on commuter satisfaction.

Preparatory course bachelor thesis

This thesis requires at least basic knowledge about data science and data visualisation techniques. Therefore, students will be required to read specific chapters from the Python for Data Analysis book (http://shop.oreilly.com/product/0636920023784.do). The student will complement these reading assignments by working through the materials of an online Coursera Data science course. The student's knowledge will be tested by individual assignments offered throughout the Coursera course, as well as through a final presentation given to the assistants of this project.