A Network Tour of Data Science
Project Description

Gaétan Ramet
235934
gaetan.ramet@epfl.ch

Benjamin Schloesing
227866
benjamin.schloesing@epfl.ch

Yuan Yao
267363
yuan.yao@epfl.ch

November 28, 2016

1 Project Insight

The goal of this project is to predict the gender of twitter users based on various features, especially on the tweets they post. We will try to analyze how well do words in tweets can define the gender of a user, and also see which words are the most important for gender prediction.

Our project would consist in :

- The use of a Kaggle data set as a basic training and testing data : "https://www.kaggle.com/crowdflower/twitter-user-gender-classification"

- Creation of another database of twitter users

- Extraction of relevant features

- Maybe apply the same algorithm on Facebook data, in order to observe if the results differ, due to difference of behavior depending on the social network.