# AXA Data Challenge

Aexandre Carton, Clément Fischer, Alexandre Guinaudeau

March 4, 2016

## 1 Introduction

In this report, we describe how we constructed a model to predict the number of incoming call to the AXA french center. The dataset on which the prediction is made consists in training data from the years 2011 and 2012, giving the number of calls depending on the date and a number of parameters such as the area of expertise of the call and the state of the call center. It comes up with another complementary set of data concerning weather information in France, which is supposed to have an influence over the number of accidents and thus incoming calls to AXA. We divided our work in – parts. First, the dataset has to be preprocessed to be used efficiently. Then we use this data to create features for our model. After this feature engineering step we train a model and use it to predict the number of calls asked in the submission file. We also evaluate our model using cross-validation techniques.

## 2 Preprocessing

The data comes in two types of csv files : meteo_2011.csv, meteo_2012.csv and train_2011_2012.csv. The main training dataset contains multiple informations on the incoming calls to AXA's centers in France. The number of incoming calls we have to predict, CSPL_RECEIVED_CALLS is one of the 86 columns of the file. For each value of DATE and ASS_ASSIGNMENT (the field of competence to which the call is assigned), the model has to predict the number of incoming calls for the three next days.

We used pandas library to read the files as databases objects. The main advantage of pandas' read_csv function is that it allows us to read only the columns we are interessted in, thus saving much computation time.

For the training dataset, we keep the columns 'DATE', 'ASS_ASSIGNMENT', 'CSPLRECEIVED_CALLS' and 'DAY_OFF'. We use this last column to eliminate data corresponding to non worked days in AXA. We then group the data by summing the number of calls having the same date and assignment values.

The meteo dataset consists in rows giving information at a given date for a some location - city and department (≃region) number - in France. From this, we extracted the number of french departements undergoing negative temperaures and number of them with rain.

# 3 Feature engineering