ABN·AMRO

ABN AMRO Data Talent Programme Data Scientist

As part of the application process for the job position Data Scientist, we would like you to work on a realistic case and present your work. This document should explain everything you need to know to get started.

At ABN AMRO we use R, Python, MATLAB, SAS, Scala and other technologies. You are free to choose any technology, but make sure that we can run your code or otherwise replicate your work.

Take-home case

Case description

ABN AMRO relationship managers for small and medium enterprises (SMEs) have a large client portfolio. This makes it difficult to keep track of their clients. As a result, we mostly serve SMEs in a reactive manner. ABN AMRO wants to help relationship managers to better serve their clients. This might include automatic recommendations: which clients to contact, so our service becomes proactive instead of reactive.

A common client request is a (renewal on their) business loan. As part of the MVP, we want to use machine learning to suggest to our relationship managers which customers are likely to apply for (renewal of) a credit, and when they are likely to do so. The goal is a solution to help relationship managers to serve their customers more proactively.

Data

Along with this assignment we send you two datasets that contain a selection of customer data. The data you receive covers a time period of 32 months in total (January 2014 until August 2016). The first dataset "customers" contains a subset of our internal customer data like credit volumes, debit volumes, number of transactions, etcetera.

The second dataset "credit_applications" contains information regarding historic credit applications. The "credit_application" field indicates that in the given month (see column "yearmonth") has a value of 1 if a client in that month applied for credit with ABN AMRO and otherwise it has the value 0. Field "nr_credit_applications" indicated how many times a client applied for credit in a given month.

Assignment

You are requested to build a model that predicts which clients will be applying for a loan. You can base your prediction on any subset of months, as long as you explain your decision. As a bonus, you can additionally build a model to predict WHEN those customers will apply for a loan.
As part of your interview, we ask you to present your results. Prepare for 40 minutes of interactive presentation (15 minutes of slides and demo, 25 minutes of discussion). Make sure to include the following elements:

- Description of, and arguments for, your experimental set-up;
- How you handled missing values, and why you did so;
- The metrics you optimized for, and why;
- Recommendations for further improvement of the model (e.g. feature engineering, extra data);
- Recommendations to management for how best to use your predictions.

Please add any intermediate results (such as tables, statistics, visualizations) to the appendix.

Evaluation

Your presentation will be evaluated on the following criteria:

- Your understanding of the problem, from a business and analytical perspective;
- The process you followed and experimental design (data processing, model selection, etc);
- Your recommendations for further research and to management;
- The way you present your outcomes.

For any questions regarding the case, please get in touch with your hiring manager.

We look forward to your presentation. Good luck!

Attachments to the case:

- *credit_applications.csv*
- *customers.csv*