

Recruitment task for Data Science

We would like to learn about your technical abilities to approach machine learning tasks. Therefore, we would like you to prepare a solution in the open-source technology (preferably Python) to the following predictive task.

Dataset:

<https://drive.google.com/open?id=1kfrqft4Pw2PUTHF00KAUNF19Y25uX90R>

Dataset contains current and past loans granted by the company. Variables description is available at the above link.

1. Prepare a model(s) that would predict probability of default at the moment of loan application. Default client is defined as one with *loan_status* variable taking on the following levels:
 - a. charged off
 - b. default
 - c. does not meet the credit policy. Status: Charged Off
 - d. late (31-120 days)
2. Present no more than 3 competing predictive models and select one of them. Explain criteria for the selection.
3. Calculate models' performance metrics. If necessary, make appropriate variables' pre-processing.
4. Share with us a programming code with comments and the presentation of main results (max. 4 slides).
5. Find and describe interesting multivariate relationship in the data (1 slide).

We are interested in your approach, the challenges of data, business & analytical problems you faced and how you have approached them or how you would approach them on the project if you were given more time. We want to know your analytical steps with their rationale. Most important parts are: (1) data pre-processing steps and feature engineering, (2) performance evaluation & selection of predictive models, (3) next steps you would perform if given more time.

Additional hints:

1. If you lack computing power, sample dataset.
2. Nasty hacks are allowed as long as you describe why you need them and what would be a correct approach.
3. We want rationale, not numbers - we won't rank candidates based on score they achieve.
4. Visual side of presentation is secondary.