Final task. Group 3.

Deadline: February 3, 23:59

- Files should be submitted in Jupiter-readable format .ipynb.
- All cells should not produce any errors.
- Code should be clear, important comments are necessary. Clarity and transparency will be assessed, not only correct answers.
- Answer should be visible. It means that when grader executes cell he should see answer clearly as cell output.

Problem [**Precision-recall curve**]

Read data_gr3.csv file. Column «Correct» contains true values, column «Prediction» contains predicted probabilities p for dependent variable y to belong to class 1. Define class predictions $\hat{y}$ by

$$\hat{y}(p) = \begin{cases} 1, if\ p \geq p_0 \\ 0, if\ p < p_0 \end{cases}$$

Here $p_0 \in [0,1]$ is specified threshold.

a) Build parametric curve $x = recall(p_0), y = precision(p_0), p_0 = 0, 0.01, 0.02, \ldots .0.99, 1$. It means output graph – points (x, y) which you've got for different p_0. Here precision and recall are corresponding metrics calculated for $\hat{y}(p_0)$ and $y\_correct$. You can use sklearn to calculate metrics.

b) On the same graph plot precision-recall curve using sklearn https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_curve.html