

Industrial project description (Banking credit scoring)

Answer the question to outline your project. Choose one of the roles: an Expert or an **Analyst**.

1 Planning the industrial research project

Before planning the research, the analyst and (**expert**) discuss the key issues. After the long dash — our remarks.

1. Goal of the project. (**Expected development result.**) — The goal of the project is a system capable of assessing the creditworthiness of the client with great accuracy.
2. Applied problem solved in the project. (**How will the result be used?**) — The applied task is to process all available data about the client, analyze them, which will allow to establish their quality and the possibility of their use in the system, develop a machine learning model that, based on the available characteristics of the client, is able to assess his creditworthiness. The result is illustrated in the form of graphs describing the accuracy of the model on customers who have true information on their credit history over time. And graphs are also needed to illustrate the actual and simulated proportion of customers who are not creditworthy.
3. Description of historical measured data. (**Formats and timing.**) — User data is presented in the form of different sources: credit history (table), registration data (table), his transactions (logs), loan information (term, amount). All sources contain both numerical, categorical and textual information
4. Quality criteria. (**How is the quality of the obtained result measured, what is in the report?**) — The desired target is the probability of default for the 12th month of the loan. If the loan was defaulted on, then the target is 1. The quality criterion is ROC_AUC.
5. Project feasibility. (**How to show that the project is feasible, list of possible risks.**) — To assess feasibility, it is necessary to evaluate the available technologies and tools for creating a model. You need to make sure that you have access to the necessary data and computing resources. It is necessary to assess the economic costs of the project and compare it with the potential profit from the project. It is also

necessary to assess how possible it is to implement this project into the banking process. Risks in the implementation of the project may be in poor-quality customer data, as well as in the absence of the customer's consent to their processing. There are operational risks associated with errors in the development/implementation of the system in the banking process. There is a risk of losing customers, as the new model may reduce the number of customers being credited, thereby potentially increasing the number of customers from competitors.

6. Conditions necessary for successful project implementation. (**Organization of work.**) — A sufficient number of qualified developers are needed who are able to develop a model, process data, and implement the system into the banking process. In order for the project to be completed, it is necessary that the new approach surpasses the existing baseline, otherwise the project will be irrelevant. It is also necessary to approve this project from a part of the business, since the new process may reduce the flow of loans issued.
7. Solution methods. (**Procedure libraries.**) — This problem can be solved by using a machine learning model, such as boosting. The simplest solution is to build a boost on all the available numerical and categorical data that we have. However, to improve the accuracy of the model, unstructured data, such as user transactions, can also be processed representing a log. You can also use LLM to process text data: user description, chat history. You can process the data by obtaining useful data for the model, for example, the client's credit load, his salary, etc.

2 Research or development?

In other words, novelty or technological advancement?

Analyst: What impact will the research have on the field of knowledge? How useful will it be?

The research project will reveal how accurately the client's credit profile can now be assessed based on his data. Perhaps this study will reveal the knowledge that some of the available data can well describe the client for the bank in terms of risk. For the bank, this study is useful because it will potentially reduce the share of losses associated with non-repayment of the loan due to new methods of assessing the client.

Expert: (**How long will the model be used? What will replace it in the future?**)