

Industrial project description: Churn Prediction

The chosen role is **analyst**.

1 Planning the industrial research project

1. Goal of the project. (**Expected development result.**) — The primary goal of this project is to develop a machine learning model to predict customer churn, enabling proactive retention strategies. The expected result is an accurate predictive model that can identify customers at high risk of leaving.
2. Applied problem solved in the project. (**How will the result be used?**) — The result will be used by the company's customer retention team to target at-risk customers with personalized offers or interventions, improving overall customer loyalty and reducing churn rates.
3. Description of historical measured data. (**Formats and timing.**) — The historical data includes customer demographics, transaction history, usage patterns, and support interactions. The data is structured in tabular form and collected over the past 2-3 years at monthly intervals.
4. Quality criteria. (**How is the quality of the obtained result measured, what is in the report?**) — The quality of the result will be measured using metrics such as accuracy, precision, recall, and the area under the ROC curve (AUC). The final report will include these performance metrics along with insights from feature importance analysis.
5. Project feasibility. (**How to show that the project is feasible, list of possible risks.**) — Feasibility is demonstrated through a pilot study on a sample dataset. Potential risks include data quality issues (e.g., missing or inconsistent data), model overfitting, and changes in customer behavior that reduce model accuracy over time.
6. Conditions necessary for successful project implementation. (**Organization of work.**) — Access to a high-quality dataset with relevant features is crucial. Collaboration between data scientists, domain experts, and the customer retention team is also essential for defining useful features and evaluating model results.

7. Solution methods. (**Procedure libraries.**) — The solution will involve using machine learning techniques, specifically logistic regression, decision trees, or ensemble methods. Hypotheses about factors influencing churn will be tested, and optimal probability models will be developed using Python libraries such as scikit-learn and XGBoost.

2 Research or development?

Analyst: The impact of this research lies in its ability to provide actionable insights for reducing customer churn, which is a critical issue for many industries. The model can help optimize customer retention strategies, ultimately increasing profitability and customer satisfaction.

Expert: (**How long will the model be used? What will replace it in the future?**) The model is expected to be used for 1-2 years, after which it may need re-training or replacement due to changes in customer behavior or market conditions. Future advancements may include deep learning models or integration with real-time data for instant churn prediction.