



Introduction to Regression Project

Project Deliverables

You will be required to submit:

- A GitHub repository with your project written in Python or R.

Instructions

Problem Statement

Mobile carrier Megaline has found out that many of their subscribers use legacy plans. They want to develop a model that would analyze subscribers' behavior and recommend one of Megaline's newer plans: Smart or Ultra.

You have access to behavior data about subscribers who have already switched to the new plans (from the project for the Statistical Data Analysis course). For this classification task, you need to develop a model that will pick the right plan. Since you've already performed the data preprocessing step, you can move straight to creating the model.

Develop a model with the highest possible accuracy. In this project, the threshold for accuracy is 0.75. Check the accuracy using the test dataset.

1. Open and look through the data file.
2. Split the source data into a training set, a validation set, and a test set.
3. Investigate the quality of different models by changing hyperparameters. Briefly describe the findings of the study.
4. Check the quality of the model using the test set.
5. Additional task: sanity check the model. This data is more complex than what you're used to working with, so it's not an easy task. We'll take a closer look at it later.

Dataset

- Dataset URL (CSV File): <https://bit.ly/UsersBehaviourTelco>

Data description

Every observation in the dataset contains monthly behavior information about one user. The information given is as follows:

- **calls** — number of calls,
- **minutes** — total call duration in minutes,
- **messages** — number of text messages,
- **mb_used** — Internet traffic used in MB,
- **is_ultra** — plan for the current month (Ultra - 1, Smart - 0).

Project evaluation

We've put together the evaluation criteria for the project. Read this carefully before moving on to the task.

- Here's what the reviewers will look at when reviewing your project:
- How did you look into data after downloading?
- Have you correctly split the data into train, validation, and test sets?
- How have you chosen the sets' sizes?
- Did you evaluate the quality of the models correctly?
- What models and hyperparameters did you use?
- What are your findings?
- Did you test the models correctly?
- What is your accuracy score?
- Have you stuck to the project structure and kept the code neat?

You can use the CRISP-DM methodology that we've used in the past, whose approach will help you answer the above questions.