

**Machine Learning Project Description**

**Student: Akos Redey**

**Batch:** Machine Learning Course March – May 2021

**Date: 5. May 2021**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Title: Churn Prediction of Telecom Customers**

**Goal**: I plan to use the Kaggle Telecom Users Dataset to predict churn of telecom customers

**Involves:** Exploratory Data Analysis (EDA), Preprocessing and Feature engineering, Modelling, Interpretation of selected model,

**Dataset:** https://www.kaggle.com/radmirzosimov/telecom-users-dataset

**Description**: The idea is to have a prototype notebook/code with the following functionalities:

* Get dataset from [Kaggle](https://www.kaggle.com/datasets)
* Store the raw data locally.
* Clean and preprocess the data, perform EDA on it (descriptive stats, visualizations)
* Run the encoding of categorical and scaling of numerical variables
* Build the initial models and compare performance, do a hypothesis test on them
* Perform hyperparameter tuning and compare the 2-3 best performing models
* Select final model and make predictions, intepret the results (try explainable AI)
* Save model artifacts and the model itself; perform version control with git

**Possible extensions:**

* Explore Plotly library to design interactive visualizations
* Build an ml app with functions, classes
* Build an API with Flask or Django

**Work Packages:**

* Sketch how the data is going to be stored.
* Sketch the ML process workflow
* Build the project structure with data, scripts and modules
* Implement the functionalities listed in Description