



# The 5<sup>th</sup> DeepDive Challenge

## Customer Churn Prediction

Mehr 1402



### Problem Statement

In this DeepDive challenge, we are going to investigate one of the most important challenges of businesses through a lens of data and artificial intelligence.

Customer churn occurs a company or business loses a customer. Customer churn rate is a critical metric because it is much less expensive to retain existing customers than to acquire new customers.

"Churn prediction" is a process where machine learning models are used to forecast which customers are at the highest risk of ending their patronage. Churn prediction is done based on customer behavior data, meaning machine learning models are trained to identify what behavioral patterns indicate higher chance of churn.

### Dataset Description

This dataset is randomly collected from an Iranian telecom company's database over a period of 12 months.

In the following tables, you will be informed about some characteristics of the dataset:

*Table 1 - Some important attributes of this multivariate dataset*

Number of records	Feature Type	Number of features	Any missing values
3150	Integer	13	No

*Table 2- Variable Information*

Variable	Value
Anonymous Customer ID	-
Call Failures	Number of call failures
Complains	Binary (0: No complaint, 1: complaint)
Subscription Length	Total months of subscription
Charge Amount	Ordinal attribute (0: lowest amount, 9: highest amount)
Seconds of Use	Total seconds of calls
Frequency of use	Total number of calls
Frequency of SMS	Total number of text messages
Distinct Called Numbers	Total number of distinct phone calls
Age Group	Ordinal attribute (1: younger age, 5: older age)
Tariff Plan	Binary (1: Pay as you go, 2: contractual)
Status	Binary (1: active, 2: non-active)
Churn	Binary (1: churn, 0: non-churn) - Class label
Customer Value	The calculated value of customer

## Task

The customer churn prediction problem can be viewed as a binary classification problem in the context of machine learning. So, the goal here is to train and evaluate binary classifiers. Here are some of the main steps you must consider doing:

1. EDA: Performing Exploratory Data Analysis before modeling can give you good insights and help you build better machine learning models.
2. Preprocessing: You must perform data cleaning and other preprocessing methods if needed.
3. Model building: Build various models using different algorithms.
4. Model evaluation: Evaluate the models you build using appropriate classification metrics such as *accuracy*, *precision*, *recall*, *f1*, and *confusion matrix*.
5. Model comparison: Compare the performance of the models.