

Internship Task

Please complete **only one** of the technical tasks presented below:

Task 1: Fraud Detection Model

Task 2: Recommendation System

and submit it [here](#).

The deadline for submission is 23:59 May 23 UTC+4.

Task 1: Fraud Detection Model

Create a fraud detection model based on the data in [fraud_profile.csv](#) and [fraud_behavior.csv](#) files. The goal of this task is to check data science skills in data preparation and modeling.

Data description:

`fraud_profile.csv` contains user signup and profile data

- user: unique user_id-s
- fraud: dependent variable (1 -fraud, 0 -non-fraud)
- email_domain: encoded email domain
- ip_asn: encoded signup ASN
- signup_datetime: user signup time
- gender: encoded user gender
- marital: user marital status
- employment: encoded user employment status
- pet: the list of user's pets

`fraud_behavior.csv` contains user clickstream behavior

- user: unique user_id-s
- click_time: the time when the user clicked on an activity
- length_of_activity_in_minutes: the expected completion time of the activity in minutes
- promised_reward: the reward promised for the activity
- ip_asn: user's device ip-asn
- postback_time: activity ending time, if this is null then the user terminated the activity and there is no activity_status and gained_reward. To calculate the user activity duration, you can subtract click_time from postback_time.

- activity_status: status of the activity, only 'successful' indicates a full success for the activity
- gained_reward: reward gained by the user for the completion of the activity

Task 2: Recommendation System

Create a recommendation system based on the data in [recsys_data.csv](#) file. The goal is to recommend surveys with the highest completion probability to users (up to 3 surveys to each user). The columns are:

- user - the id of the user
- survey - the id of the survey
- status - whether the completion of the survey was successful or not (1 - successful, 0 - unsuccessful)
- date - the date of the survey completion attempt.

Note: Surveys have a limited lifetime but exact lifetime information is not available (it can be 20 minutes, a week, a month, etc.). If you can create a recommendation system that pays more attention to newer surveys, it would be better. You can disregard this part if it makes the task harder for you.