

Build a model that predicts the number of hits per session.

Background:

On the resources section, you can find data in .csv format containing information of about nine hundred thousand sessions. The columns should be understood as follows:

- row_num: a number uniquely identifying each row.
- locale: the platform of the session.
- day_of_week: Mon-Fri, the day of the week of the session.
- hour_of_day: 00-23, the hour of the day of the session.
- agent_id: the device used for the session.
- entry_page: describes the landing page of the session.
- path_id_set: shows all the locations that were visited during the session.
- traffic_type: indicates the channel the user came through eg. search engine, email, ...
- session_duration: the duration in seconds of the session.
- hits: the number of interactions with the trivago page during the session.

Task: Note that the column “hits” has missing values. Use the data provided to build a model that predicts the number of hits per session, depending on the given parameters. Refer to the deliverables section to check how to hand in your solution.

Evaluation: Your predictions will be evaluated by the root mean square error. Check the resources section.

Evaluation: Your predictions will be evaluated by the root mean square error:

$$error = \sqrt{\frac{\sum_{i=1}^n (predictedHits_i - observedHits_i)^2}{n}}$$