

TEA DATA SCIENCE CODING CHALLENGE

Design a machine learning system which preprocesses given input data and predicts the future statistics. The dataset consists of user statistics about a mobile application which can be reached via this [link](#). The machine learning system should be responsible for:

- Prediction of the number of users for the next 6 hours using the latest 24 hour's data. (There will be 6 predicted points which represent data for every predicted hour.)
- Total number of users will be observed 24 hours later than the latest data point using the latest 5 day's data. (The exact date is 30.12.18 09:00)

Requirements

- Preprocessing and predicting models should be in a pipelined system. Sci-kit learn library could be used for creating pipelines.
- Necessary metrics should be returned by the system. There should be at least 2 metrics for the monitoring (Ex: R2 score, accuracy). Metric selection should be done by the candidate.
- Total duration and the memory usage of the overall system should be returned by the system.

Rules

- Challenge documents should be sent via Github link.
- Challenge source code should be provided with necessary documentation. (How to run the system etc.)
- Preprocessing and predictions can be done by any Python or R.
- Model training source codes should be included in the project.