

Hands-on exercise - Data Science

Navikenz - Version 1.4

Objective

- To test the hands on experience, expertise of the candidate for areas within statistics/data science
- This exercise is focused on the AI/ML engineer -> Data Scientist track, with specialisation in statistical models, data and analytics
- This exercise will require you to analyze, design and train a model to predict the number of predicted bike rentals for a specific day/hour of day

Instructions - Building the prediction model [MANDATORY]

- Download the Seoul bike rentals dataset from the following URL: <https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>
- The dataset has a collection of recorded bike rental counts for about 8000 records - the URL contains further details about the dataset itself
- We need you to perform EDA on the dataset, and design/build predictive models to return the expected number of bike rentals for a specific day (based on a selected attribute set - day of week, weather conditions, etc)
- The models need to be implemented using at least 2 algorithms - you can use any algorithms/models you are familiar with
- Compare the performance of the selected models and recommend which approach works best for this dataset
- Describe all the observations on the data, alternatives considered, selected option(s) and actions done for each stage in the implementation within the notebook itself
- If any of the algorithms/models you have used is a 'black-box', you'll also need to demonstrate that the model *has learnt the correct selection criteria* when being trained
- We will be evaluating the results on the decisions taken, options considered and the overall accuracy of the predicted results - we will test the model with queries based on selected input attributes

Expectations

- We expect to be given a notebook (and or supporting code/instructions) that includes both code and output
- You can use any technology/framework/algorithms/models that you are familiar with to build the answers, refer to online sites for documentation, etc while you are implementing the exercise
- We expect to be given:
 1. The URL of a git repository (or Colab notebook, etc) which you created/used for the exercise (the URL should be valid (and we need to have read access to it) till we examine/run the code)

2. The git or Colab repo should contain a Readme.md that describes instructions (if any) to be carried out before executing the cells in the notebook, and the notebook itself (instructions can be within the notebook)
 3. The notebook itself should be self-explanatory, and contain the steps you followed to design/build/test the models - and/or choices made
 4. We will run a few tests manually towards the end of the notebook - these tests will not be published/shared online
 5. Plagiarism rules apply for your submission
- Based on our evaluation, we may reach out to you for a next round, where we may ask you to take us through the choices/decisions made, and or include a selected 'change' scenario - we will expect this change scenario to be finished *during the session, with your screen shared*
 - There is no limit on the amount of time you can take to finish the submission - however, the time taken by you to finish the submission (from the time we sent this to you) **is a factor** (albeit a smaller factor) in the overall evaluation

All the best.