**Laptop Price Prediction**

Download Datasets

*This data project has been used as a take-home assignment in the recruitment process for the data science positions at Allegro.*

**Assignment**

Your task is to define and train a machine learning model for predicting the price of a laptop (**buynow\_price** column in the dataset) based on its attributes. When testing and comparing your models, aim to minimize the RMSE measure.

**Data Description**

The dataset has already been randomly divided into the training, validation and test sets. It is stored in 3 files: **train\_dataset.json**, **val\_dataset.json** and **test\_dataset.json** respectively. Each file is JSON saved in [**orient=’columns’** format](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.to_json.html).

**Example how to load the data:**

>>> import pandas as pd

>>> dataset = pd.read\_json("public-dataset.json")

>>> dataset.columns

Index(['buynow\_price', 'graphic card type', 'communications', 'resolution (px)', 'CPU cores', 'RAM size', 'operating system', 'drive type', 'input devices', 'multimedia', 'RAM type', 'CPU clock speed (GHz)', 'CPU model', 'state', 'drive memory size (GB)', 'warranty', 'screen size'], dtype='object')

**Practicalities**

Prepare a model in Jupyter Notebook using Python. Only use the training data for training the model and check the model's performance on unseen data using the test dataset to make sure it does not overfit.

Ensure that the notebook reflects your thought process. It’s better to show all the approaches, not only the final one (e.g. if you tested several models, you can show all of them). The path to obtaining the final model should be clearly shown.