**Dell Technologies - Hackathon**

**Problem to Solve & Context**

**Background**

Pricing any product accurately is an important aspect for every business. As the technology advances, new configurations/models are being built to satisfy the customer's needs. However, predicting the price of different configuration based on the pricing of each components is always a challenging task.

An enhanced performance prediction model would benefit our business by helping us accurately predict the price of the new laptop/desktop configuration.

**Problem Statement**

The provided datasetconsists of Laptop/Desktop configurations and its pricing. Every component has different models. Different laptop/ desktop configurations are created by combining various components and its models.

Predict the price of the systems having new configuration offerings as provided in the dataset.

**Evaluation Criteria**

| **Criteria** | **Details** | **Scoring** |
| --- | --- | --- |
| **Accuracy of predicted Price** | How does the algorithm perform against the pricing in test set? | MAPE score  Average (ABS (Actual –Predicted)/ Actual) |

**Data Description**

There are 3 datasets available along with a data dictionary

* Train set
* Validation set
* Test set
* Dictionary

File Descriptions:

1. Train set – It contains the laptop/desktop configuration along with pricing to be used for model training
2. Validation set- It contains the laptop/desktop configuration along with pricing to be used of model performance validation
3. Test set - It contains the laptop/desktop configuration for which price needs to be predicted
4. Dictionary – It contains the order of each component

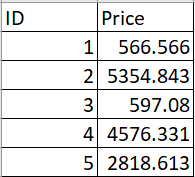
The data folder contains CSV flat files at the following level of granularity:

1. CPU
2. RAM
3. Storage
4. GPU
5. ScreenSize
6. ScreenType
7. TouchType
8. Form
9. Region
10. Price - Dependent variable

**Submission Instructions**

1. Total Number of submissions per team = 1
2. Submission will be through the following Google Form: <https://forms.gle/bJU4NPSRkDhyEc6z9>
3. Only **.csv files** with a single sheet (as in the prescribed format) will be accepted for submissions
4. The file name should be the team name
5. No. of entries should match with the expected record count for the problem
6. Submissions are expected to have ONLY two columns with case-sensitive column names

Problem 1A – Columns ***ID*** and ***Price*** *,* filename**- *Prediction\_1A***



1. The ***ID*** column should have the exact number of records and same values as in Test set
2. Keep the row order of the Test file as is, otherwise, your score will be inaccurate.
3. Submit a single prediction for each row.
4. The Prediction column values of should be in Number format