Company: INSPECTORIO

Team: Data Science

Date: 2021-07-01

Introduction:

On this page, we aim to describe about the project scenario and expected outcomes which are used for ML/DS candidates with the purpose of coding test.

Project Scenario:

In the manufacturing industry, the risk of every product needs to be monitored. On this project, a ML/DS candidate is required to build a ML/DL model that can classify a product as LOW or HIGH risk. Each sample (or row) in the provided data corresponds to a product, contain all information related to the product along with its label (column name ‘label’).

Datasets:

* train.csv: used for training
* test.csv: used for testing

Expected Outcomes:

* Define appropriate metrics for this problem.
* Visualize and describe useful insights from data.
* Show your skill in feature engineering. It’s important to note that the types of features include categorical, numerical, text.
* Build ML/DL models to boost your chosen metrics and show your best performance. To be plus, you could implement several types of algorithms, including machine learning, deep learning, … to show your implementing skills.
* Draft some conclusion about this project.
* Propose several methods to boost the performance.

Submission Files:

* A jupyter notebook to show all your implementation and description.
* A csv file to contain your prediction on the testing dataset.
* Your pre-trained model and related others.
* A report to document your result and finding.