

Programming Language and Libraries versions

The code requires Python version 3.10, Pip and all libraries mentioned below.

Programming Language used:

Python –version 3.10

Libraries used:

```
pandas==1.4.4  
category_encoders  
xgboost==1.6.2  
imblearn==0.9.1  
sklearn==0.0  
joblib
```

Package Dependencies/Installation

The steps to install and load the Python, Pip and the libraries is mentioned in the notebook. Please follow the links provided in the jupyter notebook file for assistance on how to do them.

Running the code and generating predictions

Restart and Run all the cells in the jupyter notebook which goes through the complete process as shown below:

1. Reading the data
2. Defining the pipeline and associated code for each stage in the pipeline
3. Fitting the pipeline on the given data
4. Making predictions on the test data
5. Persisting the results in a csv file

For more information on what each step of the process does, refer to the text/markdown before the code cells.

NOTE: For more information on reasoning/justification for the preprocessing and modelling in the code please refer to the report.