

REPORT

Lab6.Pandas Data Cleaning Part-II

In this lab we have learned various data cleaning techniques with the help of using pandas and scikit library.

Scikit library has some encoders which does the pre-processing or data cleaning part very easy in few steps.

Label Encoder, One Hot Encoder, MinMaxScaler, Binarizer, Imputer.

dedupe and fuzzywuzzy these packages help to de-duplicate or Entity Resolution and String Matching dedupe library helps to identify duplicate or repeated string.

To import those libraries use (!pip install fuzzywuzzy) then from fuzzywuzzy import fuzz, (!pip install dedupe) then import dedupe