Q. Create a model to extract data from invoices and categorise raw data into a readable

format.

NER – named entity recognition

We have tried checking , and scanning the pdf files of invoices , pick out entities , invoice number , date and details

Step 1: prepare basic training data – adding entities in training data

Step 2: load necessary packages and libraries

Step 3: create empty model – for feeding raw data , and enable ‘en’ language

Step 4: feed data to NLP model pipeline (language detection , entity recognition)

Step 5: training data and collect losses

Run each jupyter notebook one cell by next

Outputs are in jupyter noteboom cell itself