Preprocessing Code

The Dataset acquired contained many duplicate values and Null values

To clear the Raw dataset we dropped the duplicated data from ‘English’ Attribute as for each English word there contained more than 1 Russian word. After preprocessing now it contains 1 Russian word for each English word.

import pandas as pd

data =pd.read\_csv('DataSetR.csv')

df = pd.DataFrame(data)

df\_unique\_english = df.drop\_duplicates(subset=['English'])

df\_unique\_english=df\_unique\_english.dropna()

df\_unique\_english.to\_csv('DataSetR.csv', index=False)