What the main point when writing about data preprocessing?

* Definition of and mathematic formular of methods
* Purpose of using methods.

4.2 Data Preprocessing

One-Hot Encoding was use for categorical features. One-Hot Encoding creates a separate binary dimension for each category within a feature. Each dimension takes a value of 1 for its corresponding category and 0 for all others. Then we use Z-score normalization to normalize data. Z-score is our choice since the method is more outliner-Resistant. Mathematically, Z-score was calculated as followed.

x' = (x - μ) / σ

where μ is mean and σ is standard deviation of data.

By the end, the dataset was divided into train and test data with a ration of 70:30, where 30% was kept aside for testing model’s generalizability. All of data preprocessing step are performs on pandas Dataframe.