

Report for Advanced Analytics in

Business

Cheung Wai Chun, r0817438

David Badajkov, r0604517

Ana Maria Giraldo Vargas, r0822450

Sonia Rocio Socadagui Casas, r0823960

Marcela Lopez Viveros, r0773141

April 28, 2021

Contents

Assignment 1
Feature engineering
Missing values
Exploratory data analysis
Model building
Interpretation
Reflection
Assignment 2
Assignment 3
Assignment 4

Assignment 1

Feature engineering

The dataset for Assignment 1 consists of 55463 observations and 78 features. The number of features. It is easy to observe that there are a lot of missing values and categorical or date features in the dataset. In this section, we would discuss the strategies used in handling such problems.

Missing values

As shown in figure 1, there are 55 feature which contain missing values, and 24 out of them contain more than 80% of missing values. For those features with a high proportion of missing values (more than 50%), missing values are treated as an extra category. By treating missing values as an extra category, the information of the non-missing entries of those features can be retained and learnt by the model, whereas removal of those features may lead to a loss in information or pattern.

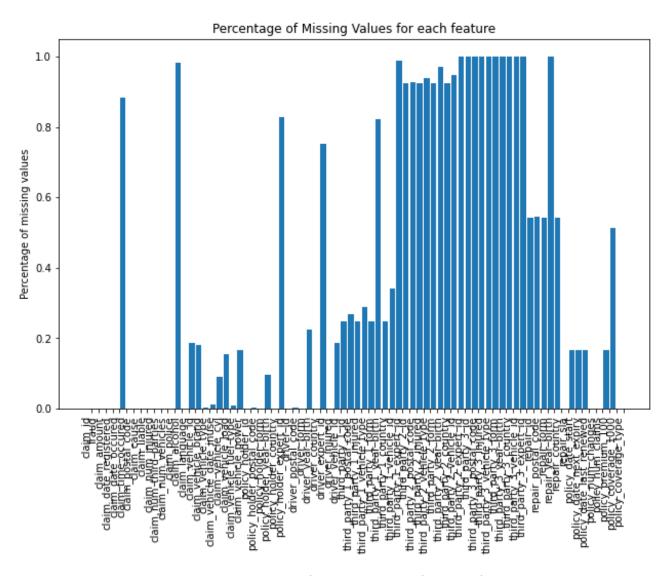


Figure 1: Proportions of missing values for each feature

Exploratory data analysis

Model building

Interpretation

Reflection

Assignment 2

Assignment 3

Assignment 4