SMU®

Data Science Machine Learning DS7331.402

Amber Burnett
Lance Dacy
Shawn Jung
Jeremy Otsap

Lab 1

January 19, 2020

Business Understanding

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Purpose of the Dataset

Describe the purpose of the data set you selected (i.e., why was this data collected in the first place?).

Define and Measure the Outcomes

Describe how you would define and measure the outcomes from the dataset. That is, why is this data important and how do you know if you have mined useful knowledge from the dataset?

Effectiveness of a Prediction Algorithm

How would you measure the effectiveness of a good prediction algorithm? Be specific.

Data Understanding

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Data Description

Describe the meaning and type of data (scale, values, etc.) for each attribute in the data file.

Data Quality

Explain any missing values, duplicate data, and outliers.
Are those mistakes? How do you deal with these problems? Be specific.

Simple Statistics

Give simple, appropriate statistics (range, mode, mean, median, variance, counts, etc.) for the most important attributes and describe what they mean or if you found something interesting. Note: You can also use data from other sources for comparison. Explain the significance of the statistics run and why they are meaningful.

Visualization

Visualize the most important attributes appropriately (at least 5 attributes). Important: Provide an interpretation for each chart. Explain for each attribute why the chosen visualization is appropriate.

Explore Relationships

Explore relationships between attributes: Look at the attributes via scatter plots, correlation, cross-tabulation, group-wise averages, etc. as appropriate. Explain any interesting relationships.

Interesting Relationship

Identify and explain interesting relationships between features and the class you are trying to predict (i.e., relationships with variables and the target classification).

Other Features

Are there other features that could be added to the data or created from existing features? Which ones?

Appendix

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt.