COMP4431 Data Mining

Assignment 2

Megan Hoeksema

Introduction

* Intent of the application
* Dataset to be used, including source
* Use case

Dataset Analysis

* Define variables
* Define labels

Inputs

* No user interaction is required.

Proposed Libraries

* Libraries
  + Numpy
  + Pandas
  + Matplotlib.pyplot
  + Sklearn.datasets
  + Seaborn
* Library source
  + Numpy: <https://numpy.org/doc/stable/index.html>
  + Pandas: <https://pandas.pydata.org/>
  + Matplotlib.pyplot: <https://matplotlib.org/>
  + Sklearn.datasets: <https://scikit-learn.org/stable/>
  + Seaborn: https://seaborn.pydata.org/

Proposed Solution

* Functional description

Proposed Outputs

* Define application outputs

Proposed Visualization

* Define visualization of outputs

Conclusions

* Analysis of results
  + Exploring, visualizing, and making calculations are the main goals of this assignment. The outcome of this assignment should be to understand the capabilities of python in exploring and understanding a dataset as well as evaluating and displaying visualizations representative of the data.