David Kwon Han Lam

Abstract – CKME 136

Context

There is a saying that says a picture is worth a thousand words. In this data set I need to determine whether a given word matches the pictures shown. The theme I have selected is classification and regression analysis as this is a non-textual dataset.

Problem

Through this data set we can answer the following question

1. Which picture provides the strongest representation of the given word e.g. the greatest confidence level

This type of data is beneficial to help determine which picture best represents a word that can be used to promote a business or individual ad on their webpage, biography, business card etc.

Data

The image descriptions data set can be obtained from the website <https://www.figure-eight.com/data-for-everyone/>.

Techniques

To conduct classification and regression, I would be using python packages and R.

To present and showcase the data, I would be using tableau and other instruments learned throughout the courses