Tania Soutonglang

CS 584-02

April 2, 2023

Interim Report - Disneyland Rating Estimator Based on Reviews

I am choosing to change the topic of my project. After reviewing what I would need to do, I don't have the skills or the test subjects to replicate the original paper's model since they used motion capture to record the data. Instead, I would like to create a model that would take reviews from Disneyland and estimate a rating between 1 to 5 stars. With this, I would be able to create a model and presentation on time.

I will be using a dataset from Kaggle and then pre-processing the data by lowercasing the sentences, tokenizing each word, stemming each word, and removing stop words. Afterward, I will create a Naive Bayes classifier that will take the counts of each word and then find the probability of the word based on each rating. After training the data, I would create the test cases and also have a presentable model that can also be used.

So far, I have already finished the pre-processing for training the model, this upcoming week I plan to finish making the Naive Bayes model. If this no longer works in the frame of what the final project is supposed to be, I would appreciate it if you can reach out to me.