Shreeman Gautam

April 6, 2022

Peer Review

Group 2

The main contribution of this project is the evaluation of clustering algorithms based on

the given data. Evaluation, specifically, rests on drawbacks, benefits and accuracy of each

algorithm. It would be nice if the terms drawback, benefit and accuracy could be defined loosely

for each algorithm. The project idea is novel because categorical values are converted to

numerical values(using probability distributions, wassterstein distance) and as is known,

clustering is usually done with numerical values. At first glance, especially with 31.8 million

records, the plan might seem unreasonable, but as mentioned, the algorithms were tested on all

the records and further, GPUs will be involved in the future to expedite the process. The idea of

the project is to evaluate clustering and so far, the report has shown how the different clustering

algorithms work in regards to the data without specifying what the drawbacks, benefits and

accuracy of each algorithm are currently. I hope that the final report will make the evaluation

clear.

Group 22

The main contribution of the project is to do textual analysis on 30 reports from the Daily Utah Chronicle. Textual analysis, as explained by the paper, involves running the Misra Gries algorithm on each report and finding what the most common words in each report are to better paint a picture of the report in a highly efficient way, and to box each report into a category. As far as novelty goes, I cannot say anything because even though I am unaware of textual analysis being done in reports published by the U, it might exist for all I know. The plan is very reasonable because the idea is to build on the intermediate report on a larger scale. Misra Gries is a fast algorithm, and I would assume that the reports are not too big in size. Current results show textual analysis in one report and like the authors envision in their idea, the result shows the 3 most words used in the report. The results of the rest of the reports remains to be seen but I am sure the authors can do it feasibly.