

Fall 2018 Term Project

## Smart and Connected Health

The term project for INFO 4670/5810 requires that you apply what you have learned in this course to analysis a dataset retrieved from National Science Foundation (NSF). The dataset is about the NSF awards on Smart and Connected Health (SCH). We have worked on a paper about this – the published paper at the 6<sup>th</sup> International Conference for Smart Health (<http://2018.icschconf.org/>). Background information, or the NSF program's call for proposal can be found here: <https://www.nsf.gov/pubs/2018/nsf18541/nsf18541.pdf>.

This project used a similar dataset as the attached paper, but with different sizes. The attached paper only analyzed 146 projects. The dataset for our term project includes 295 records (projects, or lines in .csv table). So, I will expect different results even for the same analysis as the paper.

### Analysis Requirements

You are required to use both Excel and RapidMiner for this project to perform the analysis specified below. Your final submission should be a report that answers the following questions or meets the following requirements.

1. What are the 10 states that received the most funding (the amount in dollars) from NSF on Smart and Connected Health? Use a histogram to present your results.
2. What are the five organizations that have the most funding? What are the five organizations that have the most funded projects? Use a Pie chart to present your results
3. How is Texas doing in terms of Smart and Connected Health based on this dataset?
4. Perform an association rules analysis to see whether you can find any patterns from this dataset. Report your results. This can be analysis of one column values or a combination of multiple column values. For example, just the column 'Abstract' or combination of column 'AwardedAmountToDate' and column 'OrganizationState'.
5.
  - a) Perform a general text analysis similar to your assignment 4 on the Abstract attribute. The processes include: tokenization, transformation (to lower cases), and stop word filtering.
  - b) Then perform a K-Means clustering analysis on the results from question 5a. The analysis will be similar as you did for assignment 5. Report the 10 clusters in a table listing the first 20 terms for each cluster in your report.

### Report Requirements

You need to submit a term project report that answers the above five questions. The content of the report should answer each question in a separate section and a final section about any comments. Below is a possible outline of the term project content:

**Section One:** Distribution of NSF SCH Funds among States

**Section Two:** SCH Organizations

**Section Three:** How is Texas Doing on SCH?

**Section Four:** An Association Analysis of NSF SCH Projects

**Section Five:** A Clustering Analysis of NSF SCH Projects

**Section Six:** Comments and Conclusions

### Detailed Instructions

Respond to the questions. The report should be formatted as follows:

- Use APA 6<sup>th</sup> Edition for citations and references and any point not addressed in these

instructions.

- For all text, use 12-point Cambria (preferred) or 12-point Times Roman font
- Text should be double spaced within paragraphs and double spaced between paragraphs
- Label the graphs, charts and images.
- Spell-check the document
- Do not include a title page or an abstract page
- Do not include References in the page count
- If you use section headings in the paper, use 12-point or larger type.
- 1-inch margins
- At top right on page 1, type three lines: your full name, course number, and semester and year like this:

Jane Doe  
INFO 5810  
Spring 2016

- Put the title information at the top of the first page, and then begin the body of your paper
- Create a one-line header at top right, beginning on page 2, with your last name and page number, like this:

Doe/p. 2

- The resulting document should be a single file in MS Word or PDF. The preferred format to is a .docx file
- Use the following convention for naming your files [Lastname][AssignmentName] like this: DoeTermproject.docx/pdf

**Submit your Term Project to the Assignment link available in Canvas available in Assignments by **December 10<sup>th</sup>, 2018 by 11:59 PM**. Please follow the naming convention when you upload your assignment files: yourLastname\_Termproject.docx.**