Data Mining Honors Project Final Report

## Motivation and objectives

The project compares UT Arlington professors based on their outcomes in the Student Feedback Surveys (SFS) conducted at the end of each semester and their ratings on the Professor Rating Site RateMyProfessors.com, winner of two People's Voice Webby Awards for 2015. Based on the data in both datasets, UT Arlington professors are classified into five groups in three different stages and the results of all three stages are combined and compared to give a fuller picture of the efficacy of UT Arlington Professors.

## Data mining/analysis tasks tackled

The chief data mining task was clustering performed using a k means clustering algorithm, but other tasks were performed as follows:

1. Downloading appropriate webpages from the websites
   1. Download the ratemyprofessor webpages for UT Arlington professors
   2. Download the UTA SFS webpages
2. Mining the web pages downloaded for rating information
   1. Convert the tables from the UTA SFS webpage into an internal data structure
   2. Convert the ratemyprofessor webpage information into an internal data structure
3. Converting the collected data into classes
4. Plotting the results of the clustering on a graph and producing tables showing the results.

## Datasets collected and Used

* + - 1. UT Arlington Student Feedback Survey
      2. Rate My Professor.com data for UT Arlington Professors

## Design of methods

All methods are part of one of four classes:

1. The Cluster class, which is the main class, containing the implementation of the clustering algorithm and controlling the flow of the program.
2. The Professor class, which stores details about the Professors collected from the dataset.
3. The Rating class, which stores information about a Professor’s rating from the RateMyProfessor dataset.
4. The Point class, which stores information about a Professor’s rating from the Student Feedback Survey dataset.

## Implementation of methods

There are three major methods, all belonging to the Cluster class. They are clusterRMP(), clusterSFS(), and clusterRMPandSFS(). These classes implement a k means clustering algorithm to cluster the data into five clusters.

Supporting methods are parseSFS() and parseRMP() which parse the data from the SFS and RMP datasets respectively. All other methods in the four classes merely support the functions of these methods.

## Results and Evaluation

The results show that the data in the UT Arlington SFS and RateMyProfessor datasets have graphs that are roughly alike when the mean of the clusters are plotted against the number of professors in that cluster. Both datasets show the majority of professors belong to the second largest cluster. However, while the clusters for the SFS dataset range from 63 to 96 percent of the maximum score, the clusters for the RMP dataset range from 38 to 90 percent of the maximum score.

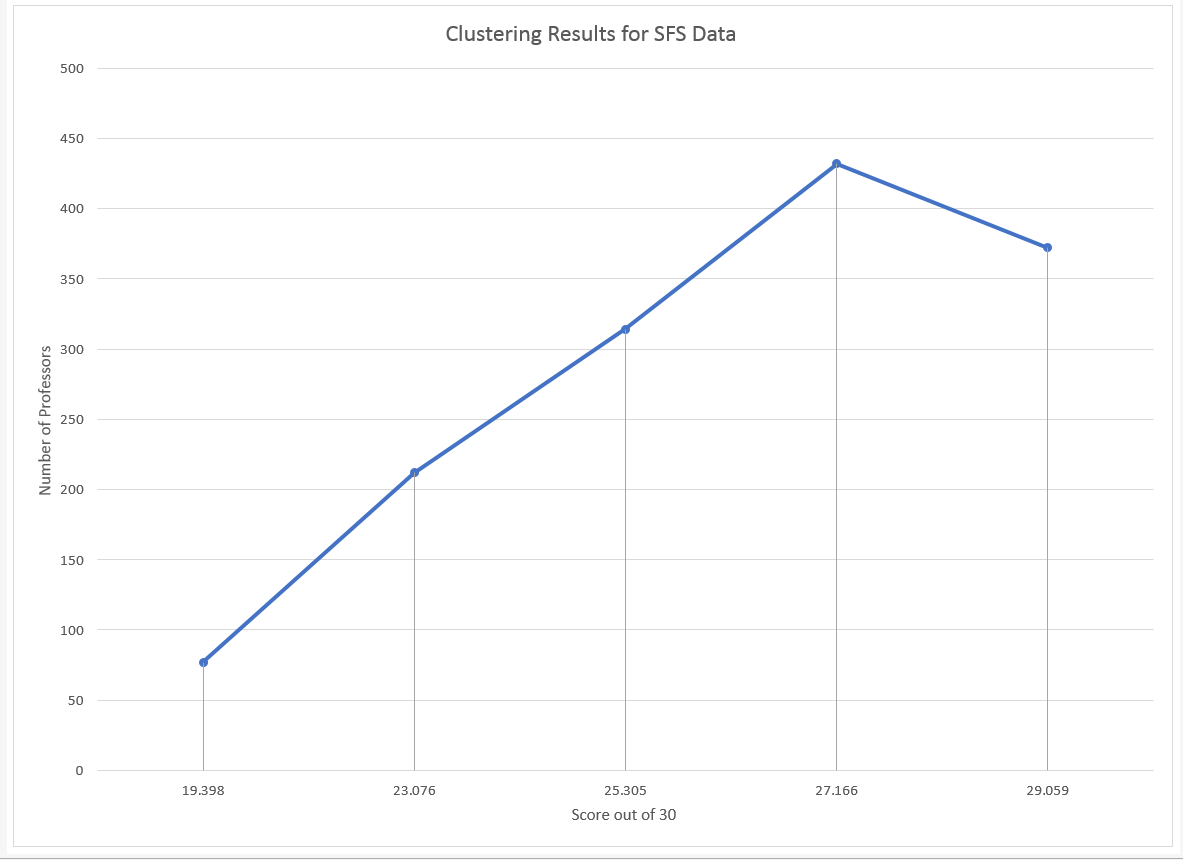
This greater range in scores for the RMP dataset suggests that scores for measures used by RateMyProfessor.com produces a different kind of result from those used by the UT Arlington SFS survey. In other words, the less descriptive terms like "clarity" used in the RMP dataset produces a greater range of responses than the more descriptive prompt: "The instructor used teaching methods that facilitated my learning" used by the UT Arlington SFS dataset.

The result of treating both datasets together is less significant due to a marked difference in the number of professors in the RMP dataset compared to the SFS dataset. The gap between the smallest and largest clusters remains the same as that of the RMP results, but the look of the graph changes and largest number of professors belong to the highest cluster. This is likely due to the fact that a large amount of SFS data points belong in that cluster.

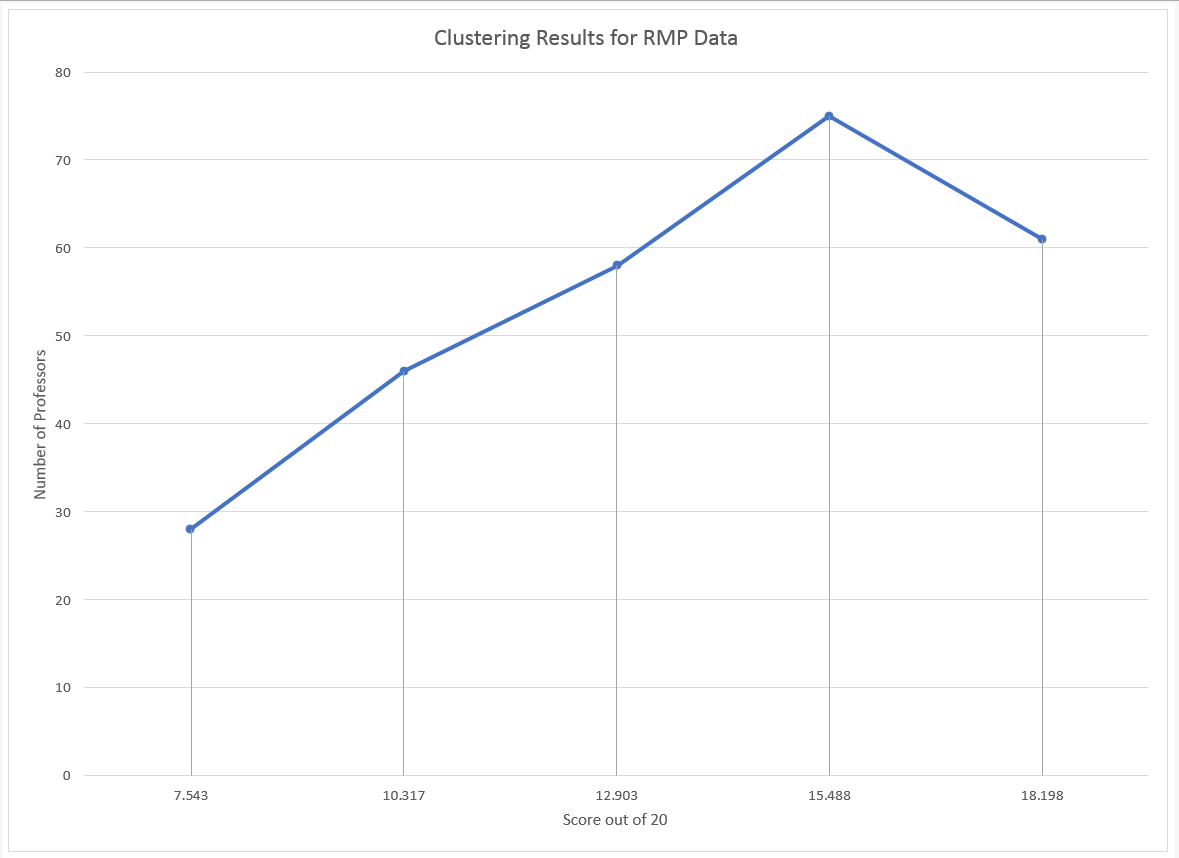
## Presentation/Visualization of the Outcome

The following graphs and tables provide are a pictorial representation of the results:

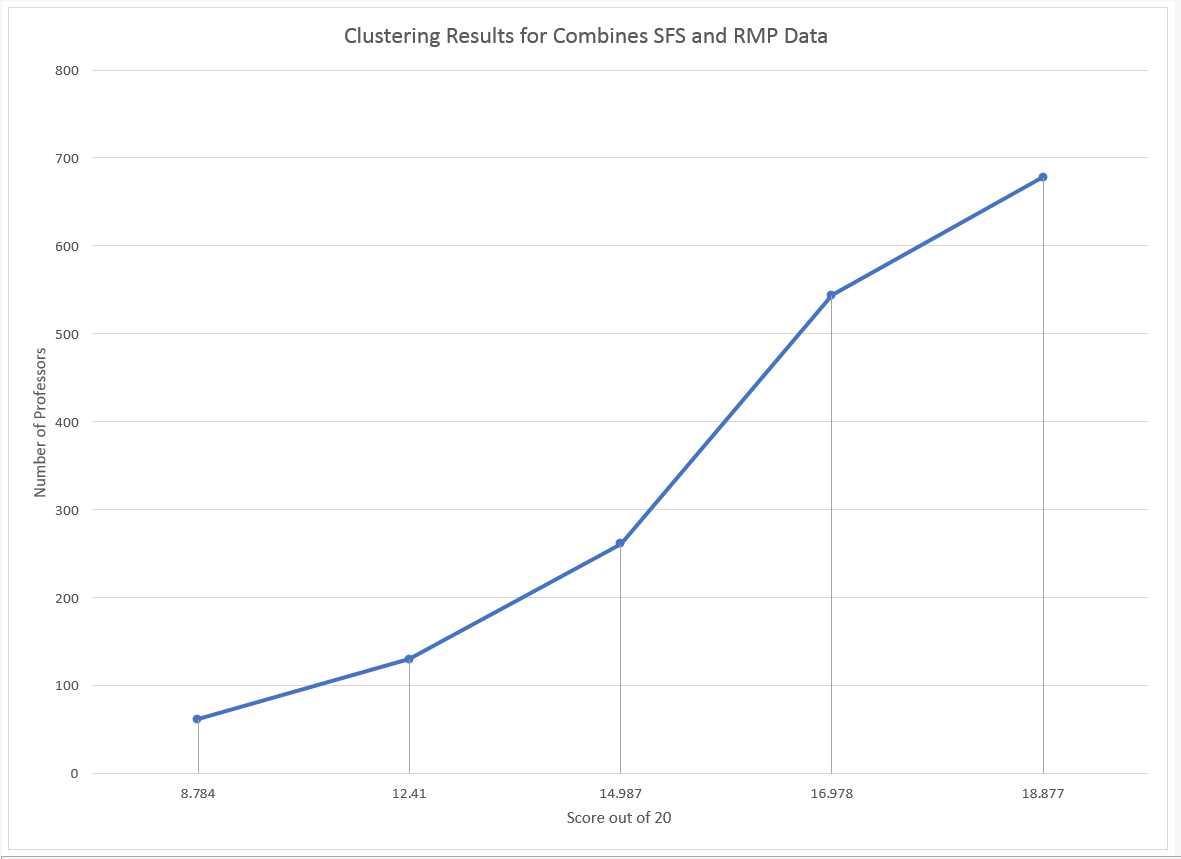
|  |  |  |
| --- | --- | --- |
| Total SFS Score | Professors | Percentage |
| 19.398 | 77 | 5.4726368 |
| 23.076 | 212 | 15.06752 |
| 25.305 | 314 | 22.316986 |
| 27.166 | 432 | 30.10453 |
| 29.059 | 372 | 25.118163 |



|  |  |  |
| --- | --- | --- |
| Total RMP Score | Professors | Percentage |
| 7.543 | 28 | 10.447761 |
| 10.317 | 46 | 17.164179 |
| 12.903 | 58 | 21.641791 |
| 15.488 | 75 | 27.985075 |
| 18.198 | 61 | 5.7655955 |



|  |  |  |
| --- | --- | --- |
| Total RMP + SFS Score | Professors | Percentage |
| 8.784 | 62 | 3.7014925 |
| 12.41 | 130 | 7.761194 |
| 14.987 | 261 | 15.58209 |
| 16.978 | 544 | 32.477612 |
| 18.877 | 678 | 40.477612 |



## URL of your project website

The website for the project can be found at: https://sites.google.com/site/tracyogunihonr/home