# Sentiment Analysis using Rapid Miner

**Introduction:**

This project is designed to help organizations and individuals to get better and relevant suggestions of travel destinations, so that individuals can correctly decide their travel destination of their choice and prepare accordingly. The suggestions are based on useful reviews given on websites by the users. This project will analyze the review comments provided by travelers on websites and then provide a summary of the textual content that is present on the website.

**Method**:

In this project, we collected data from websites through web scraping and then populating it in a CSV file. This data can be used by the Rapid Miner Tool to perform Topic Analysis and Sentiment Analysis on the travel destinations and come up with the ones that have highest positive rating. EXTRA PLUGIN FOR SENTIMENT ANALYSIS.

a. ***Sampling Design***: This project considers only a segment of population that has travelled to a place and then written a review about it on popular social media. We have taken a convenient sample of about 300 reviews by travelers. This kind of sampling is actually a random set of data and it does not reflect the sentiments of all those who have travelled to that place recently.

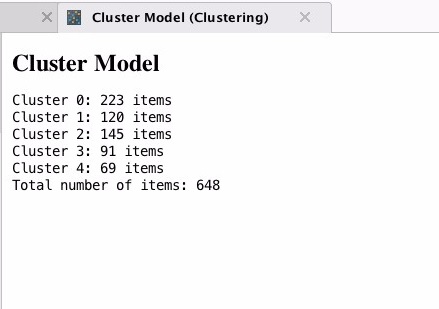
b. ***Data Collection***: There are a lot of social media sites, travel blogs and travel websites that have a lot of data about travel destinations. We scraped our data from the <https://tripadvisor.com>. The data was scraped using a tool called “importio”. It is an online tool that gives us data that has been scraped from the URL of a website that we provide to the tool. We need to select the required columns from among the available list of columns. To ensure quality, we made sure that multiple people did web scraping and then the process was repeated with different data sets until we achieved consistent results. The Data was mostly textual in nature. We initially tried to scrape it using python but there were some issues and inconsistencies. So we finally decided to use a tool for scraping the data from the web.

c. ***Data Analysis***: This scraped data was provided to the Rapid Miner tool for getting insights. Since we were performing sentiment analysis, Topic Analysis which is also a part of sentiment analysis was also undertaken and the results of both of them could be seen in the charts and graphs. The results are shown in screen shots below.

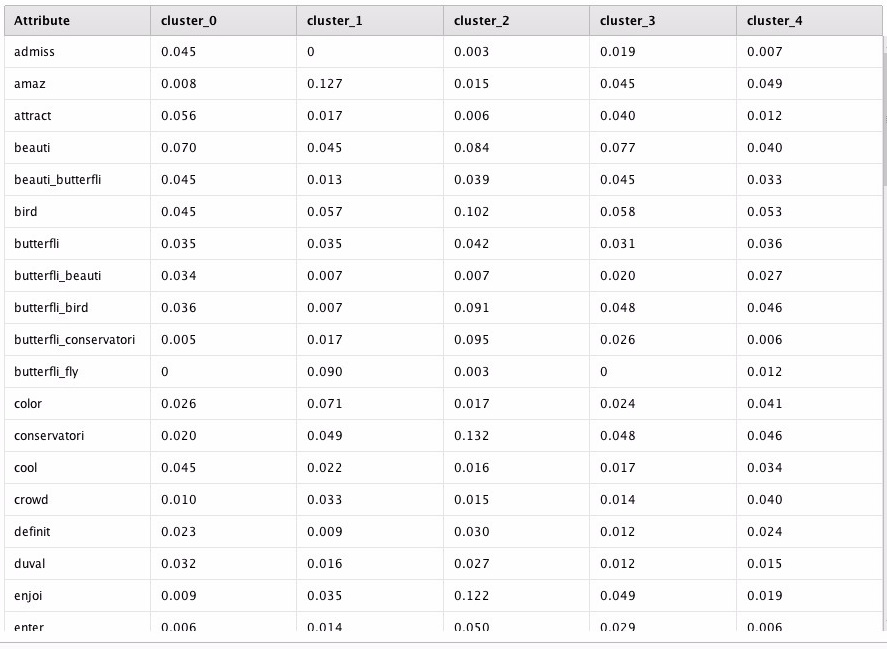
**Results**:

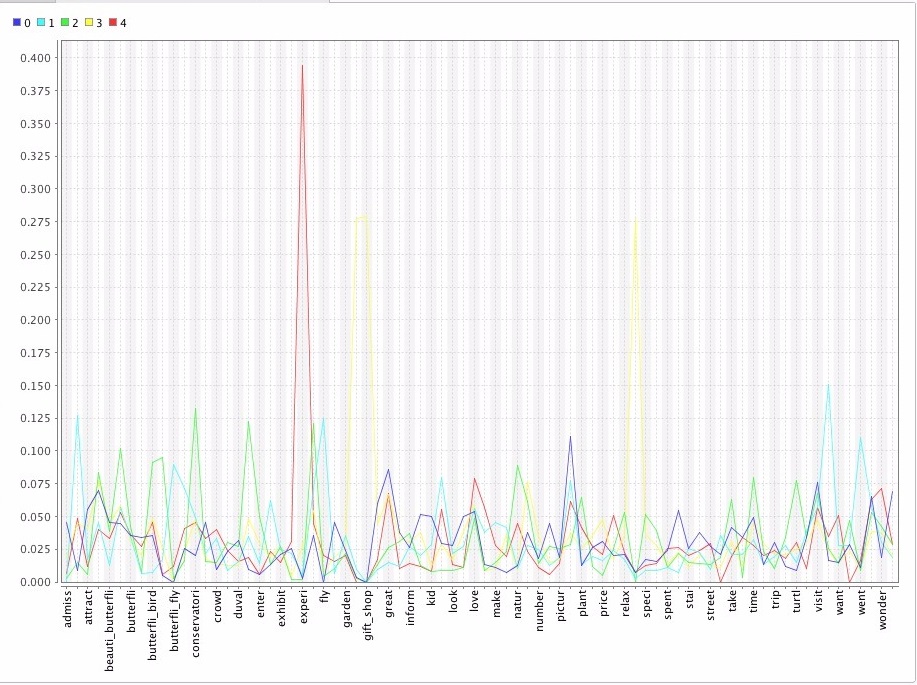
We observed that the Rapid miner tool was able to perform Topic Analysis and Sentiment Analysis on the input Data obtained after web scraping. After the Topic Analysis is performed, the data words are divided into clusters of most often repeated words. We can also see the Centroid Table which shows the frequency of each attribute. After the completion of sentiment analysis, the comments are analyzed and the polarity of each comment is calculated along with the polarity co-efficient of the comment.

**Topic Analysis Clusters:**

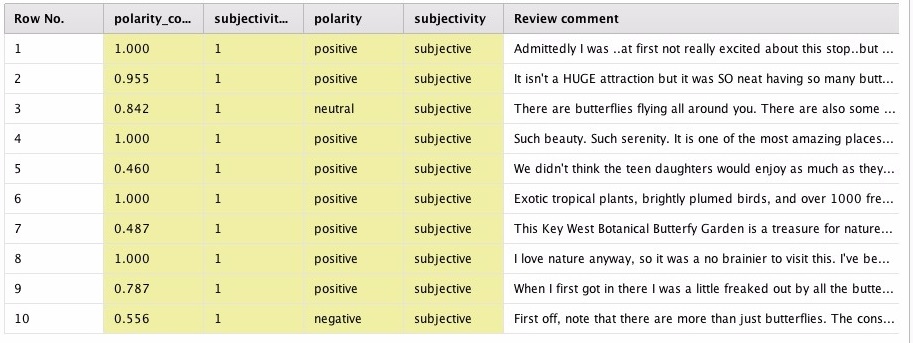


**Topic Analysis Centroid Table:**

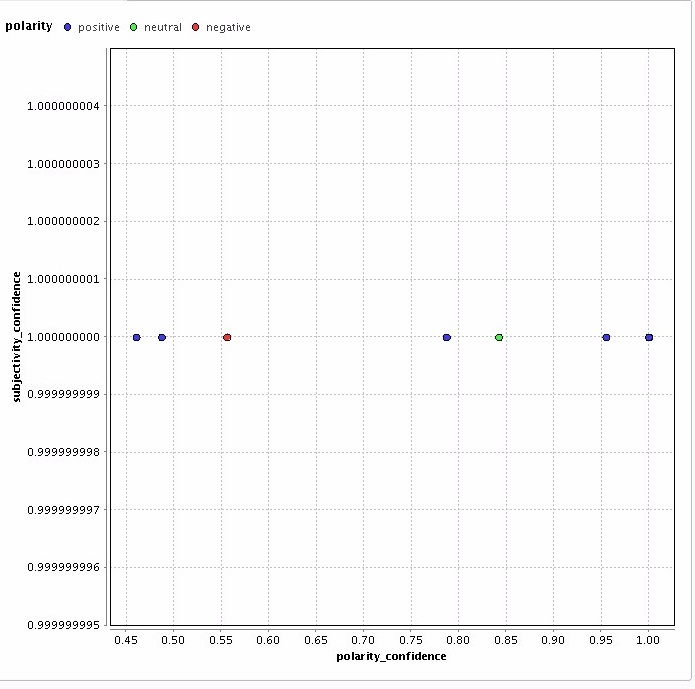
**Topic Analysis Output Visualization:**

****

**Sentiment Analysis Data:**

****

**Sentiment Analysis Charts:**

****

**Conclusion & Limitations**:

This is the final section of your paper and will take the results and provide interpretations and recommendations based on the results. This section must at least contain the following subsections (or the information in these subsections):

a. *Summary*: Since the “Results” section included all the analysis of your data, here you summarize the key findings that form the basis of your recommendations. The most important findings should be summarized here and should tie in well with the management question you started your report with.

b. *Recommendations*: This brief section should develop some very clearly stated. recommendations based on your research findings. This will be evaluated based on how well your research findings answered the management question at the start of your report. There should be something that is clear and actionable at this point (even if it is clear and specific direction for an additional research study). For a good grade, make sure your recommendations are clear, specific, and linked to the management question/dilemma.

c. *Limitations*: This is an important part of your report for this class. Given the time constraints on doing this project, there should be several things that could have or should have been done better. Here is where you get to lay out all the problems with your research project. What are all the things you should have done differently? How did these “flaws” impact your results? It may be better to simply number these limitations (just as I have done in this outline) and give a brief subhead before explaining how this limitation may have impacted your results. Don’t arbitrarily put things in here that couldn’t possibly have had an impact on your particular study. However, the more you are able to identify possible sources of error and bias, and explain how it could have affected your results, the better your grade will be.

**References**:

Include a detailed bibliography. Please follow a consistent format. I am partial to the APA style (because that is what I am used to) both for citations within your report and the format of the references. However, you can use your preferred style (e.g., Chicago Manual) as long as it is consistent throughout the paper and through all the references.

* <https://www.tripadvisor.com/>
* [https://dash.import.io](https://dash.import.io/)

**Appendices:**

Include details here that would have disrupted the flow of the paper. So, you can include things like full sample, tools screenshots, etc.

Sample Data set that was scraped from the TravelAdvisor.com:



The List of tools that were used for the implementation of this project:

1. Rapid Miner
2. ImportIO.
3. KNIME

**CIS 8045 – Report Evaluation Sheet**

Group: Unstructured Engineers

Semester:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Poor  1 | 2 | 3 | 4 | Excellent  5 |
| **Intro and Background Research**  • Clearly defined problem statement  • Clear link from mgt dilemma to research question  • Thorough analysis of surrounding issues  • Flowed well and led to hypotheses/objective  • Was appropriately cited & referenced |  |  |  |  |  |
| **Results and Analysis**  • Were appropriate for data  • Were well explained  • Focused on substantive outcomes  • Were thorough and complete |  |  |  |  |  |
| **Discussion & Conclusions**  • Summarized substantive issues  • Completely discussed limitations  • Discussed business implications well  • Clearly addressed research questions  • Clear and actionable outcomes |  |  |  |  |  |
| **Overall**  • Was well written (grammar, spelling)  • Had a professional appearance  • Was well organized and structured |  |  |  |  |  |

Comments:

Final Report Grade: