PROJECT REPORT

For the final phase of the project, you are required to submit a report containing details of the work done, results, analysis, accomplishments, and key learning. The report is due on the last day of class – April 29.

We are most interested in your Big Data strategy, results, and analysis. You should be able to explain each of these areas clearly in your report. Please do not include source code as part of the report. The maximum page limit for this report is 10 pages.

This report will be part of a Turnitin assignment, which means that it will check your report for plagiarism and copying. So, please do not lift sentences or large blocks of text from online or other sources. Please write your own report.

The final report should contain the following parts:

I. Design:

Design document should contain your proposed design of the solution. At the minimum, it should contain the following:

- Summary of problem definition
  - Focus on explaining what you want to do in the project, any assumptions, and limitations.

- Description of input data
  - In the design document, you have to provide a summary of your data – for example, data format, attributes, and metadata. Please do not include data inside the report.

- Explanation of your algorithm and pseudocode

- Explanation of your Big Data strategy
  - Which Big Data strategy did you use and how does it make sense in the overall picture. If you used multiple technologies, list the project phases where you used each.

- Create a data flow diagram for your application.
  - An example of DFD for MapReduce is here:
  http://creately.com/diagram/example/h21wfdxq2/MapReduce

  Similarly, you can create a data flow diagram for your machine learning/data analysis strategy.

- Details of how your application handles bad or missing data and is your strategy robust i.e. can it recover from errors. Similarly, if you use machine learning, how do you handle overfitting.

II. Analysis of Results:

In this section, you will present your final results and analyze them. Following are certain key points:

- You should be able to summarize your results well. This could be in the form of tables, graphs, plots, or other visualization tools.

- You should be able to validate your results. For example, you can compute the accuracy of your model on the test dataset, or use cross-validation on the training dataset, or you can show that there is a correlation between positive review and star rating, or that there is a correlation between positive sentiment and stock price. Try to come up with numerical results.

- If you results are below expectation, explain probable reasons.

III. Conclusion

In this section, you will present your conclusions. Following are key points:

- Explain how using Big Data helped you with this project? Explain how using Big Data helped you arrive at a better/faster/more efficient solution.

- Describe what you learned in this project.

- Describe how your technique/strategy can be improved.

IV. Role of Each Team Member

In this section, you will describe the role of each team member.

V. References

Please cite all references that you used. These could be websites, online tutorials, papers, etc.