Data Mining Project Brief.

Phase 1.

Please select 3 team members for this class. To get the complete score of Phase 1, please take the following steps:

Report structure
The name of document: CSCI-6401-01_P1_your team name_your UNH ID
Team name (Please avoid choosing any name that may offend a group of people or be immoral)
Team member names and UNH emails: The head of your team and 2 other teammates
Please write two paragraphs to describe
A. yourself
B. why and how you selected these teammates and the leading person

Caution
Please submit the report by the deadline (all the team members should submit individually).
Please submit your report with extension of .pdf.
Please use Writing Center services if you need.

Phase 2.

Please submit your second phase report by considering the following points:

1. Your report should include the name of all team members, their UNH emails, your team name.

2. Provide a clear research question with the length of less than 3 lines.

3. Support the merit of answering your research question using literature review or other valid references (min 2, max 5).

4. In a separate section of your report describe the selected dataset that you want to work with. Your descriptions should also include the accessibility of the selected dataset, data collection methods, and information about the datatype, etc.

5. Please clearly describe your plan [steps, data mining model/algorithm] for solving the research question using the selected dataset.

Phase 3.

Please submit your third phase report by considering the following points:

1. Your report should include the name of all team members, their UNH emails, your team name.

2. Please introduce your selected data set and research question.

3. Please put a list of your related reviews. This list should include the title of paper, the authors names and affiliation, the publication date, and the name of publisher.

4. Please critically review the related literature from different perspectives, such as the selected data set, the number of sample records, their research questions, their data mining techniques, the performance metrics, and the highest quantitative performance outcome [short paragraph for each literature] .

6. The head of each team should submit the team report in a format of a .pdf file.

Phase 4.

Please submit your fourth phase report by considering the following points:

1. Your report should include the name of all team members, their UNH emails, your team name.

2. Please introduce your selected data set and research question.

3. Please put a list of the exploration techniques, which you used in this work. You can find an introductory list of data exploration techniques in https://www.saedsayad.com/data_mining_map.htm.

4. Please describe your data explorations from different perspectives using varied visualization techniques such as tables and charts. Finally, you should conclude your data exploration in a paragraph, which describes your findings based on the data exploration.

5. The head of each team should launch a private GitHub repository - add all team members, I (shivanjali.khare@gmail.com) and TA (Tangella, Nikhil Teja 181fa04465@gmail.com) as collaborators. Following this you should upload the source code of your analysis and the pdf file in the repository. Please make sure you put the repository address on your report.

7. The head of each team should submit the team report in a format of a .pdf file on Canvas.

Phase 5.

Please submit your fifth phase report by considering the following points:

1. Your report should include the name of all team members, their UNH emails, your team name.

2. Please introduce your selected data set and research question.

3. Please put a list of the data mining techniques, which you used in this report. You can find an introductory list of modeling / data mining techniques in https://www.saedsayad.com/data_mining_map.htm.

4. Most data mining techniques have both model parameters and hyperparamters, which optimize the selected technique for a particular problem. Please separately list all parameters/hyperparameters of your data mining techniques. Also, it is a good idea to provide a brief description of a hardware that you used to perform your experiments.

5. Please describe the outcomes of your data mining techniques from different perspectives using varied performance metrics. Your report should include various visualization techniques such as tables and charts. Finally, you should conclude your data modeling in a paragraph, which describes how well you answered your research question.

6. The head of each team should upload the source code of your modeling part and the pdf file in your Github repository - same repo as before. Make sure you put the repository address on your report. Your code should separately use training and testing data.

7. The head of each team should submit the team report in a format of a .pdf file on Canvas.

Phase 6.

Please submit your sixth phase report by considering the following points:

1. Your report should include the name of all team members, their UNH emails, your team name.

2. Introduce your selected data set and research question.

3. Put a list of the data mining techniques, which you used in this report.

4. Separately list all parameters/hyperparameters of your data mining techniques.

5. Separately list the techniques that you used to optimize the values of your parameters/hyperparameters.

6. Describe how your optimization techniques enhanced your data mining techniques outcomes from different perspectives and varied performance metrics. Your report should include various visualization techniques such as tables and charts. Finally, you should conclude your optimization step in a paragraph, which describes how you improved your previous answer to your research question.

7. The head of each team should upload the source code of your optimization part and the pdf file in the repository. Please make sure you put the repository address on your report. Your code should separately use training and testing data.

8. The head of each team should submit the team report in a format of a .pdf file on Canvas.

Phase 7.

You should upload you presentation file before your presentation time on Canvas to receive the points of presentation in this course. Please read the suggested instructions for this on-ground presentation.

Presentation Template (Recommended)

1. Introduction (1-slide) – Introduce your topic.

2. Research Question & scope (1-slide) – Clearly define research question & scope.

3. Motivation (1-slide) – Why did you choose this topic?
This model will be useful for which domain?
How beneficial is it to the society?
4. Literature Review – Present similar works in this area along with their drawbacks/limitations – should not be more than 1 slide – keep it simple, concise, for e.g., you can use tables/charts.

5. Methodology (no slide limit) – Discuss your dataset
Present the steps you followed to train the model
What decisions you took?
What parameters did you set? And why those?

6. Results (no slide limit) – Present your results (e.g., visual charts/tables)

7. Conclusion (1-slide) – Present your conclusion.

Overall work.
Connect your results with your research question.
Answer your research question.

8.Limitation (1-slide) – Discuss few limitations of your work.

9. Future work (1-slide) – Present some ideas based on your work.
How you can improve your work.

Phase 8.

Your academic paper must include the following sections:

- Title (0.25 points).
- Emails and affiliation of the authors (0.25 points).
- Abstract (1.75 points).
- Introduction (2 points).
- Related work [include atleast 5] (1.75 points).
- The proposed method (4 points).
- The experimental results (4 points).
- Discussion (3 points).
- Conclusion and future work(3 points).
- Appendix for link to the GitHub repository (0.25 points).
- References (0.25 points).
- Proofreading with an email from Writing Center (1 points).