MGSC 5126 (Fall 2020)
Group Project Guideline
Cape Breton University
Shannon School of Business

Project requirements

For the group project, you can pick a topic that is relevant and interesting to your group. The idea needs to be embraced by all the team members.
You will be expected to choose a project that leverages your knowledge of data mining methods in real-world applications.
The project should discover or leverage interesting relationships within a significant amount of data.

Possible interesting topics

A short list of possible interesting topics are as follows:

- Mining Educational Data to Predict Student's Academic Performance
- Analyzing Hashtags and Tweets
- Early Stage Diabetics Risk Prediction
- Comparing the Performance of Decision Trees and Neural Networks for Classification of Breast Cancer
- Human Activity Recognition Using Smartphones
- House Price Prediction (i.e. in Cape Breton Island)
- Detecting Parkinson's disease
- Data Mining Analytics to Minimize Logistics Cost
- Using a classification method to Predict and Forecast Transportation Demands (i.e. in Maritime)
- Implement a collaborative filtering technique on certain basket/item data (from Ebay or Amazon, for instance)
- Movie Recommendation System
- Using Data Mining to Determine Causal Associations Between Drugs and Condition
- Understanding Correlations between Product Reviews and Ratings

Example of External Free Data Sources

1. UC Irvine Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php)
2. Information Is Beautiful (https://informationisbeautiful.net)
3. BEST DATA SOURCES FOR BUSINESS ANALYSIS (https://www.analyticsinsight.net/best-data-sources-for-business-analysis)

Deliverables

1. Executive Summary
    a. The executive summary should capture briefly the questions you addressed and your key findings. Also briefly explain datasets. Summarize with your main suggestions for how to use these results in decision making.

2. Research paper (5-8 pages)
    a. What problems you specifically addressed, including details in technical as well as business terms

   b. The process you followed: Preprocessing steps, techniques used.
   c. Interesting results. Include the visualization of the result, as well as the specific result based on your performance measures (e.g., the actual association rule, classification accuracy, confusion matrix, etc.)
   d. Conclusions
3. A presentation (10 -15 minutes)
4. Source Code
5. Datasets

Timeline

To ensure progress towards the completion of your project, your team will meet with me during certain milestones, listed below. Those meetings are mandatory for full grade on the project.

Sep 18	**Building your team**	**Grade**
Oct 7	**Project proposal (max 2 pages) due** Submit a document that describes your initial research on the topic	**2%**
Oct 19	**Executive Summary** Draft paper (Introduction), Methods to be used, Experimental setup (Dataset, Training/Testing setup method, performance measures)	**2%**
Nov 2	**First revision due** First complete paper	**3%**
Nov 20	**Second revision due** (revised paper based on the professor's comments)	**3%**
Nov 30	**Final research paper due.**	**10%**
Nov 30, Dec 4	**Final presentations.**	**5%**

Final Project Evaluation

Make sure you do the following:

- Complete implementation for your Data mining technique.
- Compile your code successfully.
- Understand your technique and code.
- Implement a user interface (or easy use menu) covering the main data-mining primitives for user interaction.
- Execute your code successfully.
- Obtain real or synthetic data based.
- Obtain interesting results on your data.
- Submit a complete research paper of your work.
- Presentations showing the objectives, motivations, techniques used, how to use your system, data, and results (10-15 minutes).

The research paper should include the following:

- An abstraction that specifies your objectives, motivations and results of the work.
- Introduction including your motivations.
- Literature review describing objectives and motivations behind your project, the main data mining algorithm or technique used in the project, the data (how you collected and preprocessed your data or your resources)
- Discussion with different points of comparison and experimental setup.
- Extension to the work already done with a clear point of view
- Results of your comparison, extensions or discoveries (references, and results showing sample runs from your program.
- Conclusion about your work and results.
- State your references with a correct citation (15-20 at least).

Final Project Evaluation Criteria:
Section# Group#

Criteria	**Strong**	**V. Good**	**Good**	**Weak**
1. Title correctly reflect the subject of your paper.				
2. An abstraction that specifies your objectives of the survey.				
3. Introduction with clear motivations.				
4. Literature review is covering all algorithms under comparison.				
5. Showing an understanding of the chosen algorithms and application				
6. Code is running fully correct				
7. Data sets are enough and correctly preprocessed.				
8. Points of comparison or extension.				
9. Experimental set up is clear and illustrates your point of view.				
10. Results are clear and prove your points of view.				
11. Add up to the current research.				
12. References are enough and in the correct format.				
13. Correct English (Grammar, spelling, ….)				
14. Paper is of quality that is suitable for publication.				
15. Presentation is clear and organized between the group members.				
16. Group Work				