EAS 508: Statistical Learning and Data Mining-I : Project Details

The team project requires students to integrate the concepts described in this course to address a problem through data mining. The project should use techniques presented in the course to a large dataset. The main focus of the project is to develop a clear data mining logic which integrates multiple techniques. That is, the project should not be the application of all techniques individually with a series of results, but instead should only show results which contribute to the final conclusions. The project will result in both a team presentation and paper. The teams will design their own project, with the topics likely falling (although not required to) under one of following categories: (i) Medical Informatics, Clinical Data Science, and mHealth, (ii) Energy Science and Renewable Energy, (iii) Bioinformatics, (iv) Weather and Climate, (v) Finance, (vi) Education, (vii) Operational Workflow Management, (viii) Computational Social Science, (ix) Security and Fraud Detection, and (x) Sports Analytics.

The final presentations will be given in class at the end of the semester. Each team presentation will be 5 minutes. The quality of the presentation style will contribute to the project grade. The final paper should be 8 – 10 pages in length. It is required to include an introduction which lays out the problem and the challenges that the project is addressing, a description of the data, a description of the analysis methodology developed and applied, results, and a conclusion section which clearly summarizes the accomplishments and any remaining challenges.

The following is the schedule for the project:

- Group formulation: each group will be made up of five (5) members. You are allowed to select your teammates. By **September 22**, email Jie He (jhe53@buffalo.edu) with your team arrangement. If you do not select a team, you will be randomly assigned to a group. Additionally, if your team has fewer than 5 members, additional members will be randomly assigned to your group. One email from the team is sufficient, each member does not need to send an individual email.
- Proposal submission: by **October 1**, submit your project idea on UBLearns. This should be in the form of a short (less than 1 page) proposal, with the sections of problem definition, anticipated approach to address problem and anticipated data. By **October 8**, each group will receive feedback on their proposed project and any required changes. The clearer and more specific the proposal is will allow for more meaningful feedback.
- Presentation 1: In class on **November 2 and November 4**, each group will provide a three (3) minute presentation defining their problem and providing an update of the current status, using one slide. That slide must be submitted by **November 1**. The quality of the presentation, including style, will comprise part of the final project grade.
- Presentation 2: The final 5 minute presentations will be given in class on **December 2, 7 and 9**. The presentation slides must be submitted on UBLearns by **December 1**, regardless of the presentation day. The slides submitted on December 1 will be used in the presentation, and no further modifications are allowed after submission. The presentation should focus on the impact of the results / conclusions, and do not need to cover everything done during the project.
- Final paper: The paper is due by **December 10**. It is to be submitted on UBLearns.

The grade for the project will be determined as follows:

Technical Merit, Scientific Rigor, and Impact (40%). This includes the quality of the analysis, the challenge of the problem, the quality of interpretations, and the potential impact as defined in the conclusion.

Paper (30%). This grade is based on the quality of the paper, including clarity of discussing results and interpretations, clarity in defining the problem, appropriateness of references, quality of figures, etc. Grammar will be considered in the grading. Plagiarism detection software will be used to check each paper.

Presentation (30%, with Presentation 1 making up 10% and Presentation 2 making up 20%). This will be based on the quality of the presentation, with the expectation being a professional style presentation like would be seen at a conference. A penalty will be applied for any team going beyond the allotted time.