Human Performance Analytics Project

MSSA - 60530 Spring 2024

One of the main goals of Human Performance Analytics is the creation of and extraction of insights from a real performance dataset. The only requirement for the project topics is that there should be an aspect of individual contribution at some level of the analysis. If the main level of the analysis relates to team performance that is not a problem once it is linkable back to players at some level. Some potential questions to be answered by the analysis would be:

1. How to provide adequate ratings for players?
2. What factors contribute to winning/losing in a given sport? (Game prediction can be framed in this way if so interested)
3. What factors lead to injury?
4. What factors lead to success/failure on a given play?
5. What is the optimal strategy in a given sport?

Many other questions are possible, please feel free to think up your own and discuss further with me.

Teams

You may work in teams of 1-5 for this project. It is expected that larger teams carry out a more detailed analysis.

Initial Meeting

Once you have decided upon a potential project, found a suitable dataset and carried out an initial investigation of the data, I intend to have a meeting with each group in advance of submission of the first project report. Please select a slot during office hours when you are ready to discuss the project. If none of the office hour slots are suitable then please send an email to arrange a time.

Dataset

Several potential datasets for analysis are provided in a document on the Sakai page for the class. I am completely open to the use of other datasets so please feel free to use others that you identify and we can discuss at the 1-1/group meeting.

Project Deliverables:

This project will account for 55% of the final grade for this course split amongst the following deliverables:

1. **Short report** outlining problem, dataset and the project plan due on 04/10/2024 (15% of final grade)
2. **In-class presentation** delivered in the final week of class presenting an executive summary of the project and your findings. (10% of final grade)
3. **Final detailed report** of project and process with R-code, due on last day of class. (30% of final grade)

For both the short report and final report you will need to submit the report itself along with the dataset and r-code used in the project. Whereas for the project presentation you will need to submit the slides which are used.

The different aspects of the project are described below. Recommend page lengths regarding text are included but please treat this as merely a guideline and feel free to write as much or as little as is necessary.

Problem Outline Short Report:

The short report should provide a solid introduction to the project you are working on, the dataset you have chosen, the framing for the problem, the proposed solution, and a review of previous work done in this area. This report should contain the title of the project and the authors who are part of the project.

For the project outline I will be looking for:

1. Clarity of problem statement and proposed solution
2. Quality of the data overview, summary statistics, challenges in the dataset and appropriate data visualization.
3. Novelty of problem - Is this an interesting or real-world problem? Are the results of this project likely to have a real-world outcome?
4. Short review of previous work done in this area/on this problem.

Format:
Your problem outline report should be 1-3 pages in length. This does not include code and/or graphs. Please include the following sections:

1. Motivation - What is the problem you are considering and why is this an interesting problem? Has this problem been attempted before?
2. Problem framing - What is the proposed solution to the problem?
3. Data overview - What is the dataset you are going to use to solve this problem? Describe the dataset and its characteristics.

4. Contribution (Does not count towards limit) - Which team members contributed to each area of the project.
5. Bibliography (Does not count towards limit) - Citations for the references used.
6. R-code (Does not count towards limit) - R-script with code used for initial data analysis.

Please submit the short report on Canvas or via email to mbarron2@nd.edu by 6:55 pm on 04/10/2024. The report should be either a word document or a pdf file. The r-code should be sent as a separate r-script. The dataset used should also be submitted, the best method for this is likely to share the dataset via google drive.

Presentation:

The in-class presentations will be an executive summary of the problem, the solution and findings. This should be treated as a high-level overview and does not need to go into extreme detail on the methods used. In industry you will often present work to those who are less technical than you are and therefore you should attempt to describe the problem, briefly describe how it was solved and then discuss the insights gained.

The in-class presentations will be approximately 10 minutes in length.

The grading for the presentation will be based on the following factors:
1. Clarity of presentation.
2. Accurate conveying of problem and solution.
3. Insights drawn from the results.

Final Report

The final report should provide a comprehensive overview of the project and the findings that were generated from the project. This report should be readable and interpretable by someone who has no prior knowledge of your analysis.

The grading for the final report will be based on the following factors:
1. Clarity of presentation and explanation of project findings and results.
2. Technical quality of the work - Does the project make sense? Was the application of the methods reasonable?
3. Significance of the problem - Was this a real and interesting problem? Were the authors able to extract novel insights from the analysis?
4. Interpretation of the modelling process.
5. Analysis of results and insights that can be drawn from the results.
6. Plan for implementation of the results.

7. Next steps - what would you do differently, what would be the process for improving the modelling results.
8. Peer grading (Teams only) – For those who work in teams part of the grade given to students will be determined by their team members. Each team member will be submit a grade for each of the members of their team.

Format

Your final report should be approximately 3-5 pages in length and not more than 6. Feel free to include additional graphs and figures that you believe provide supporting information as an appendix if you go over five pages in length. The below is a rough guide to the format of the final report but feel free to modify it as you feel best suitable and feel free to discuss it with me in advance of final submission:

- Introduction (0.5-1 pages) - Outline the problem, its significance, and your motivation for pursuing the problem. Describe the inputs and outputs to the project in detail. An abstract paragraph outlining the entire project can also be included.
- Related work (0.5 pages) - What previous work has been done in this area and by whom? Remember to cite your sources. How is your approach different/an improvement on previous attempts to solve the problem?
- Data Description (0.5 - 1 pages) - Describe the dataset used for the problem, the pre-processing steps taken during the modelling process. How many samples and variables were used, how were these split amongst the training, test and validation sets. Try to include some summary graphs of your dataset.
- Methods (0.5 - 1 pages) - Provide a one to two paragraph description of the method used, its applicability to/suitability for the problem along with the methods strengths and limitations. I am looking to see your understanding of the method here.
- Results (1-1.5 pages) - Describe your application of the method to your problem. What were the hyper parameters chosen? How was the model fit? What was the output of the method? How did the method perform on the training, test and validation sets? What metrics were used to judge the method and provide a description of them (e.g. accuracy, sensitivity, AUC, etc.)? For classification problems you should include a confusion matrix of model accuracy while for regression problems you should provide an estimate of the error such as MSE. Present any challenges or issues in the project or model fitting process.
- Actions (1-2 pages) – Describe the plan for the implementation of the results in a real world context. This should include change management considerations of the results. Include suitable visualizations to support your conclusions.
- Conclusion and Future work (0.25-0.5 pages) - Provide a summary of the report and the key points from the analysis. What would be the next steps in this project given more time and resources?
- Contributions (Does not count towards limit) - What work was done by each of the team members on the project?
- Bibliography (Does not count towards limit) - List of sources used for the project.
- R-code (Does not count towards limit) - This should be provided separately as an r-script and should allow the findings and graphs in your project report to be recreated.

Your font size should not be smaller than 10pt for the final report. The final report can be submitted as a word document or pdf if you so wish. The r-code used for the project should be submitted as a separate r-script.

Please submit your final report via canvas or email to mbarron2@nd.edu by 1pm on the last day of class.

Citations:

For the project reports please include appropriate citations of sources which you make use of. These should be included in the report with references at appropriate points in the text where the information is used and a bibliography should be included at the end of the report listing the sources used.

Some citation tools can be found at:
https://libguides.library.nd.edu/business-general/citing-sources

A guide to citing sources can be found here:
https://libguides.library.nd.edu/plagiarism