**Project on Predictive analytics**

For this project, you will need to collect data upon which you can run predictive analytics.

There are several publicly available datasets you can search for on the internet. Example: [https://www.kaggle.com](https://www.kaggle.com/). It could be any dataset that can be useful for your analysis.

Your dataset must have at least 4 predictor variables (can be more) and an outcome variable (can be a categorical or a numerical column) and at least 200 records.

Using this dataset, you will need to create a model, to predict the outcome variable of your dataset using Frontline solver.

You can choose to either run a classification (logistic or classification tree) or a numerical prediction algorithm (Multiple Linear Regression and Regression Tree) depending on the outcome variable of your Dataset.

If you have any questions you can communicate with me about your project via email or during zoom office hours.

For this project you have 2 submissions (1) Project Proposal (2) Project Report

**Submission 1 (Project Proposal) (Due on April 12, 2021, Monday)**

Your Project proposal submission must include a 1) word document and 2) excel file and 3) project participation declaration document

1) Word document must provide details about

* Introduction about your project (What it is about and how this predictive analytics project can be useful)
* Literature: Look for previous literature on this topic (for example if your data is about sports, look for research studies published on sports analytics and write about it. You can search for research papers/articles on the Milner Library website)
* Data source: How the data was obtained, what is the data source, what kind of data does the source website host and cite the sources.
* Describe the outcome variable and the predictor variables- describe in detail the data present in the variables, whether numerical or categorical, what does the data represent, How each variable important for your study, etc.
* What you are planning to do with your dataset. - Write your plan about applying data visualization, data cleaning, and what are you trying to predict with this dataset ( I need only the plan in your words, results of applying the analytic techniques and visualization techniques must be provided in part 2)
* The predictive analytics method you are planning to use and why you chose a particular method.

2) Excel file with the Data set that you plan to work on.

Grading Rubric

|  |  |
| --- | --- |
| Introduction - Gives a clear, relevant, and appropriate hook about the topic | 5 |
| Depth of research - Well researched, gave appropriate background and put together in own words to support the topic | 5 |
| Data - Found Appropriate Data | 10 |
| Description of variables of the study | 5 |
| Structure of the Project Document and citations- Well-formatted and easy flow of the content with appropriate headings, subheadings, cited previous research | 5 |
| Plan for analysis - Has a clear idea about what to do with the dataset |  |

**Submission 2 (Project Report) (Due on April 20,2021 Tuesday)**

Your Project Report submission must include 1) Project Proposal word document

1) Word document is an extension of the previous submission (In addition to the previous submission in the project report it must also include)

* Introduction about your project (What it is about and how this predictive analytics project can be useful)
* How the data was obtained
* Describe the outcome variable and the predictor variables.
* What you are planning to do with your dataset.
* Explain the predictive analytics algorithm/method you are planning to use and why you chose a particular method.
* Explain the results obtained after applying data visualization, data cleaning, and data preparation methods you applied
* Provide your Model after running predictive analytics
* Explanation of the Results: Model Fit indices and Determination of Goodness of fit
* Conclusion

2) Excel file with the Analysis and results you obtained with the dataset.

Grading Rubric

|  |  |
| --- | --- |
| Explanation of results from data visualization, data cleaning, and data preparation methods | 10 |
| Predictive analytics model creation | 5 |
| Explanation of Model Fit indices | 10 |
| Structure - Well formatted and easy flow of the content with appropriate headings, subheadings | 5 |
| Conclusion - Gives a suggestion, prediction, question, or an opinion | 5 |