Supervised learning using logistic regression and classification trees.

Assignment Instructions for Fundamentals in Business Analytics:

Your term Assignment is to analyse an existing data set and prepare a report for business decision or policy makers. The modelling context is limited to supervised learning and specifically data that support analysis by logistic regression and classification trees – where the dependent variable is 'Yes' or 'No'.

Finding a data set
Enter the following search string into your browser
https://archive.ics.uci.edu/datasets?Task=Classification&NumInstances=100-1000&NumAttributes=10-100&skip=20&take=10&sort=desc&orderBy=NumHits&search=&Area=Business&Area=Life+Science&Area=Social&Types=Multivariate

This gives 52 datasets from which to choose from. Your group may select any one of these sets for the analysis.

For many of you this will be your first application of Business Analytics. However, I expect everyone has completed a report with some data analysis and whether calling the exercise analytics or market research, or economic analysis the format is standard.
Guidelines: #pages and words give you an idea of the size of the report, considering that some sections will have graphs and tables and others will not.

Instructions:

1. You are expected to **work in teams of three**. These have been assigned randomly in LMS.
2. Use machine learning principles such as test data, K-fold for estimation and other relevant techniques.
3. Visualize the data and the results using R.
4. Your report should include the following sections (see over for more details):
    a. An executive summary
    b. Introduction: Briefly describe the problem and the data you will use to solve it.
    c. Explanation of ML techniques being used: Describe the techniques you will use for predictive analytics, including logistic regression and classification trees.
    d. Descriptive analysis and visualization: Conduct a descriptive analysis of the data and visualize the results.
    e. Presentation of results and interpretation: Present the results of your analysis and interpret what they mean.
5. Each team should upload their report as a pdf and their R script as a **text file (*.txt).**
6. The report should be written in clear and concise language, and should be no longer than 6 pages, *excluding the appendix and references.*
7. The assignment will be graded based on the quality of the analysis, the clarity of the report, and the completeness of the code (see rubric)

Submission Guidelines:

1. The assignment is due on [insert due date].
2. **Submit your report as a pdf** and your **R script as a text file (*.txt).**
3. Submit your files through the LMS system.

Good luck with the assignment! If you have any questions, please do not hesitate to contact me.
Regenerate response

Business Analytics Report

For many of you this will be your first application of Business Analytics. However, I expect everyone has completed a report with some data analysis and whether calling the exercise analytics or market research, or economic analysis the format is standard.
Guidelines: #pages and words give you an idea of the size of the report, considering that some sections will have graphs and tables and others will not.

Title

Executive Summary. Single page summary showing: the purpose of the report; key descriptive statistics; main result and inferences; recommendations and conclusion. You may choose to keep to a traditional report format or present the Exec Summary as an infographic (i.e. more like a single page poster). *single page limit*

Table of Contents

1. Introduction. What is the business context (as you see it[1]) of the analysis? What is the purpose of your analysis (i.e. the aim of your report)? *½ page ~ 200 words*

2. Method discuss your chosen method (logistic regression for classifying binary outcomes and classification trees) from a theoretical perspective. This is not meant to be a textbook treatment, but a discussion on the appropriateness of your model to the context and the assumptions of the model. You should also discuss your method of segmenting the data into training and testing sets. *1 pages ~ 400 words*

3. Descriptive Statistics and Preliminary Correlation Analysis Summary statistics for the dependent and independent variables. Examination of correlations where appropriate. Crosstabs where are appropriate. Charts (bar graphs, histograms, line and scatter graphs) go here. *Charts and tables 2+ pages ~ 250 words*

4. Analytics Present your regression and decision tree models and their performance on the test data set. Choose a best model and draw appropriate inferences for business decisions. *Charts and tables 1-2 pages ~ 300 words*

5. Recommendations & Conclusions Briefly outline the implications of the model and possible business decisions (campaigns, risk management strategies, customer screening in the finance sector). *½ page ~ 250 words*

6. Reference List Follow APA or Harvard Style but be consistent.

7. **Appendices** We will try to avoid appendices. Only use an appendix for full results for a model of which the main results are tabled in the report. Example maybe that you listed a summary of fit statistics (MSE, R^2, AIC or BIC) for many models and the appendix contains a table that lists the full output for each model. ALL TABLES in the **APPENDIX must be of interest and REFERRED** to in the in the main body of text.

Please see LMS Assessments for data and rubric

SPARK Plus is used to reflect on your involvement and that of your team mates.
https://www.uwa.edu.au/students/-/media/Project/UWA/UWA/Students/Docs/Faculties/Business/SPARK-Guide-for-Students-2017.pdf

[1] You will be accessing secondary data for data science repositories. Often the purpose of the data is to test an algorithm or to demonstrate some practical skills. You need to see through this and discuss the problem from the original context (i.e. why would institution want to know your results and what decisions will be influenced by the outcome of your analysis). In short, what is the business context?