Advanced (Business) Data Analytics

ASSIGNMENT 1

Summary

- Type: Project report
- Learning Objectives Assessed: 1, 2, 3
- Due Date: 26 April 2022 13:00
- Deliverable: A written report submitted via TurnItIn and a Jupyter Notebook file
- Weight: 50%

This assignment is an individual assignment. The aim is to provide experience in the steps involved with creating, evaluating, improving classification models, and finally presenting and interpreting the model in a business report. You are strongly encouraged to commence this assignment by the end of the third week of the semester, and you should progress thoughtfully through the steps. Hasty decisions made early in the design process may result in much more work later.

Feel free to discuss concepts and ideas with peers, but remember your submission must be your work. Be careful not to allow anyone to copy your work.

Specification

Bush Fire Hazard Reduction (BFHR) burning is a low intensity, controlled fire crucial to reducing the fuel load (vegetation) on a certain section of land and therefore reducing the potential impacts of a bush fire on life, property, and the environment. Currently private landowners are permitted to conduct BFHR if the correct standards are followed, under the approval of the Rural Fire Service (RFS). The burning of vegetation can be potentially hazardous and ultimately the private landowners are responsible for any fire that burnt out of control and may be liable for the damages it causes.

Bushfires have caused significant damage to the Australian lives, land and wildlife and therefore ensuring that safe and well-planned reduction efforts will help to reduce the risk of future disasters. To ensure safe burning is carried out, the RFS provides a range of standards to consider before lighting a fire. However, the season and weather conditions consideration are proving more difficult to plan due to Australia's weather conditions becoming more erratic, especially in recent times. This consideration factors in the some of the following conditions:

a) Selecting the right time and season.
    i. Generally, burning in late spring (after vegetation has dried during winter rain) is usually avoided because of the potential for re-ignition during summer when rainfall is low, and conditions are hot and dry.
b) Temperature
    i. Ideally temperatures should be less than 25°C for low intensity burning.

c) Wind speed and direction
    i. Low intensity burns are best carried out in wind conditions less than 15 km/h as measured in the open. The direction of the wind affects the direction in which the fire develops as well as how fast it progresses.

As a result, the Australian Rural Fire Service has requested you to propose a classification model, that can predict if a Bush Fire Hazard Reduction burn can be performed on a certain day given the previous day's weather results. Such a model is valuable to the RFS in providing them accurate information when approving reduction burn proposals from landowners. Your model will play a crucial role in helping to protect Australian lives, land, and wildlife, whilst making the jobs of the RFS more efficient to allow them more time to help fight fires.

Dataset

The data provided is a range of season and weather condition data gathered by the RFS over the last 10 years. The provided dataset contains 134692 records and 24 inputs, ordered by date (from July 2013 to March 2022). The classification goal is to predict if the reduction burning can be permitted (yes/no) on the following day (Burn_Tomorrow variable).

There are 4 types of input variables and only 1 target/label/special variable:

A) INPUT - Location data:

1. City_Name (type: categorical)
2. City_State (type: categorical)

B) INPUT - Related to Time and Seasonal data:

1. Date (type: numeric)
2. Season: Current season in area (type: categorical)
3. Climate: Climate of the area (type: categorical)
4. Minimum_Temperature: Daily minimum temperature (°C) (type: numerical)
5. Maximum_Temperature: Daily maximum temperature (°C) (type: numerical)
6. Temperature_AM: Daily average temperature (°C) in the morning (type: numerical)
7. Temperature_PM: Daily average temperature (°C) in the evening (type: numerical)

C) INPUT - Related to wind data:

8. Wind_Direction_AM: Daily average wind direction in the morning (type: categorical)
9. Wind_Direction_PM: Daily average wind direction in the evening (type: categorical)
10. Wind_Speed_AM: Daily average wind speed (kph) in the morning (type: numerical)
11. Wind_Speed_PM: Daily average wind speed (kph) in the evening (type: numerical)
12. Max_WindGust_Direction: Daily maximum wind gust direction (type: categorical)
13. Max_WindGust_Speed: Daily maximum wind gust speed (kph) (type: numerical)

D) INPUT - Other weather conditions:

1. Significant_Rainfall: Was there significant rainfall (type: categorical)
   Note: Significant rain is defined as >= 1mm.
2. Rainfall_Amount: Daily amount (mm) of rain (type: numerical)
3. Humidity_AM: Daily average humidity (%) in the morning (type: numerical)
4. Humidity_PM: Daily average humidity (%) in the evening (type: numerical)
5. Atmospheric_Pressure_AM: Daily average atmospheric pressure (hpa) in the morning (type: numerical)
6. Atmospheric_Pressure_PM: Daily average atmospheric pressure (hpa) in the evening (type: numerical)
7. CloudCover_Oktas_AM: Fraction of sky obscured by cloud in the morning (type: numerical)
8. CloudCover_Oktas_PM: Fraction of sky obscured by cloud in the evening (type: numerical)

   Note: This Oktas are a unit of eighths. For example, a partially cloudy day may have 3 Oktas referring to there being 3/8's of the sky covered.

E) OUTPUT - Desired Target:

- Burn_Tomorrow: Indicates whether a Bushfire Hazard Reduction burn is possible on the following day (type: binary)

Deliverables

Your reports should include the following parts:

- Introduction and data exploration and clustering
- Model building: you can develop your initial model with DT. Then try alternative approaches, such as ANN, NB and KNN.
- Model evaluation: evaluate your initial model and alternative model using various approaches.
- Conclusion and summary: Include those results that are most significant for your strategy development and recommendations and justify them.

You may decide to have four main sections in your report, including *a) introduction, b) model building, c) model evaluation, and d) conclusion & summary*. Alternatively, you may decide to combine part b) and part c) and have three sections, including *1) introduction, 2) model building and evaluation, 3) conclusion and summary*. Both approaches are accepted.

It is up to you to decide what proportion of your report goes to each part. You may include tables, charts, or tables of your analysis and models. At the end of your analysis, your Jupyter Notebook should be uploaded along with your report.

The consistency of your Jupyter Notebook will be checked with the results in your report. You **do not need** to provide the **screenshots** of your Jupyter Notebook, as the marker can observe them from your file. Consider the following points for designing your process:

- You need to create only one Notebook with as many cells and outputs that are needed.
- You should not modify "A1_Data.xlsx" file before importing it in your Notebook.
- All of your analysis should be done after importing "A1_Data.xlsx" in Jupyter Notebook, not Excel, or any other analytical tool.
- Your Jupyter Notebook should start with loading "A1_Data.xlsx" file from your desktop.

Formatting and professionalism

The project report is to be written to a professional standard. This requires a formal writing style – do not use dot points - and adopt a professional tone. Given the report's nature, you may choose to write this essay in the first person. The report must be consistent with the University's policies on academic integrity, plagiarism, and consequences as noted below. The report should be typed (in Times Roman 12-point font or larger, single-spaced) and the Word Count should be 1500 words (+/- 10%) in total length. The Word Count excludes the title page, tables, footnotes and references (if required). The word limit must be observed or the assessment will be affected as noted in the Rubric. No appendices are to be provided.

Submission

Submissions are to be submitted via the two following links in the Individual Assignment: Classification folder found in the Assessment Section of the Blackboard course site: Written report to be submitted through the TurnItIn link, the Jupyter Notebook to be submitted by the Assignment Submission link.

Acceptable submission formats for the written report are Microsoft Word and PDF formats, and a Jupyter Notebook file. The files MUST be named in the format of StudentLastName_StudentID.pdf (or a. docx or .doc extension). If your ID is 41724593 and your surname is Mory, the name of your files would be Mory_41724593.pdf. The written assignment file should not be zipped.

Plagiarism

It is understandable that students talk with each other regularly and discuss problems and potential solutions. However, it is expected that the submitted assignment is a unique document – all parts of the assignment are to be completed solely by the individual student. In cases where an assignment is perceived to not be a unique work, a loss of marks and other implications can result. For further information about academic integrity, plagiarism, and consequences, please visit:

http://ppl.app.uq.edu.au/content/3.60.04-student-integrity-and-misconduct.

Administrative Requirements

Consultation sessions

To ensure that an equal and sufficient amount of time is allocated for every student who attends consultation sessions regarding the practical aspects, the average consultation time (during busy consultation times) will be limited to 5 minutes per student. The main aim of this restriction during busy periods is to ensure student equity and minimize waiting time. However, in circumstances where no other students are waiting, longer consultation times will be provided.

Submission Date

11:50 AM 26th Apr 2022.

For each calendar day (i.e., including Saturdays and Sundays) or part thereof after the submission deadline, a penalty of 10% of the total possible assignment marks will be deducted until the assignment is submitted.

Deadline extensions

An extension to the assignment deadline will only be considered for legitimate reasons and with supporting documentation (e.g., medical certificate). A request for an extension is assessed by the Assessment, Examinations & Misconducts Coordinator. You may discuss your situation with your course coordinator, but you still need to make a formal extension request using the form identified on the Electronic Course Profile for this course. Extensions will not be granted where the school is not satisfied; you took reasonable measures to avoid the circumstances that contributed to you not submitting by the due date. The following are not grounds for an extension:

- holiday arrangements (including overseas travel)
- misreading a due date
- social and leisure events
- moving house
- the pressure of work/competing deadlines
- computer issues

Please refer to the Electronic Course Profile for this course for more detail.