PRIFYSGOL BANGOR UNIVERSITY

Ysgol Busnes Bangor
Bangor Business School

BANGOR BUSINESS SCHOOL

ABJ-4061 / DATA SCIENCE

Assignment: Project on Data Analysis

The due date for the assignment is **7th May 2025** at 23:55.

- This assignment contributes to 50% of the overall module grade.
- Assignments are to be submitted through Turnitin in Blackboard. You can submit your assignment multiple times up until the deadline. Once the submission date has passed, the system will not allow re-submission.
- Students may request an extension to the stipulated due date. To request an extension, you will need to log into My Bangor https://my.bangor.ac.uk/en/ and locate the request centre from the drop-down list of the Online Services tab.
- Where an assignment is late without an agreed extension, the following rules will apply without prejudice:
    a. Submitted up to 7 days after the original submission date - Assignment will be capped at the lowest possible passing grade, which at 50 marks.
    b. Submitted after 7 days without a valid extension – Assignment will be zero graded.

  Before submitting your assignment, please ensure that it is properly referenced to guard against accusations of malpractice. Guidance on referencing can be found at Teaching & Learning Support Resources Gateway *and* Referencing for Bangor Business School.
- You are reminded not to copy material from any sources without properly referencing it, as this constitutes plagiarism. Plagiarism is defined as using without acknowledgement another person's words or ideas and submitting them for assessment as though it were one's own work. This includes copying materials from the internet, unfair use of generative artificial intelligence software such as Chat GPT, rewriting published material without acknowledging the source and the translation of materials using unauthorised methods or essay mills. Cases of plagiarism will be referred to the Business School Academic Integrity Officer for investigation and may be subject to a deduction of marks, which can result in an overall mark of 0%.

  You are reminded that the use of such services to generate work in substitute for your original contributions contravenes Bangor's Academic Integrity policy. Any detected attempt to use such tools will also result in a referral to the Business School Academic Integrity Officer. Academic Integrity Procedure: https://www.bangor.ac.uk/regulations/procs/proc05.php.en.

COURSEWORK DETAILS

Please direct all your enquiries on the project to Dr Heather He. This coursework consists of five tasks. You must complete ALL of them. This assignment is about data analysis using R and Power BI. Therefore, all data import/output, manipulation, exploration, analysis, and presentation must be conducted using R & Power BI. You will lose credit if, for example, you use Python to perform any of the tasks.

Task 1: Crafting a Tailored Cover Letter for a Business Data Analyst Position [20%]

The purpose of this task is to help you develop skills in tailoring your experiences and skills to meet job requirements, a crucial element in improving employability. You will practise this by researching a real job posting for a Business Data Analyst in the UK and crafting a compelling cover letter. You're required to:

- **Research a Job Posting**: Visit https://uk.indeed.com/ and search for a Business Data Analyst position in the UK. Choose a job posting that matches your interests and skills. Copy and paste the job title, company name, and full job description into your report.
- **Write a Cover Letter (Maximum 400 Words).** Your cover letter should:
    - Begin with a brief introduction stating the role you are applying for and explaining why you are interested in it.
    - Highlight specific experiences, skills, and achievements from your academic or personal background that directly relate to the job requirements. For example, if the role requires data visualisation skills, mention relevant projects where you used tools like Power BI or similar.
    - Conclude with a professional closing, reiterating your enthusiasm for the role and how you can add value to the company.

For Task 2 - 4, please load the data from a MySQL Database that is stored on your own machine into R session.

For Task 5, please load the data stored in a csv format into PowerBI.

Task 2: Understanding Market Campaign Data [20%]

You are a newly employed data scientist at BangorTelco. The first question you have been asked to address is: 'Is there a way to determine in advance which customers are likely to accept the offer in the marketing campaign?'

The IT team provide you with access to the company database, which contains a table (BangorTelco_MarketCampaign) containing data on over 20,000 customers, randomly selected from a larger database of previous customers, including whether they have accepted the offer in the last campaign. The data in this table are:

ID	Customer's unique identifier
Year_Birth	Customer's birth year
Education	Customer's education level
MaritalStatus	Customer's marital status
Income	Customer's yearly household income
Recency	Number of days since customer's last purchase
NumWebPurchases	Number of purchases made through the company's website
NumStorePurchases	Number of purchases made directly in stores
NumWebVisitsMonth	Number of visits to company's website in the last month
Response	1 if customer accepted the offer in the last campaign, 0 otherwise

Your goal is to create a decision tree which can predict class membership of the Response variable. The sales team will then use your tree's rules to determine whether a customer is likely to accept the offer in the next campaign. Clearly, the better the tree prediction, the more marketing costs can be reduced.

N.B. You may encounter certain attributes that contain abnormal data points (e.g., Year_Birth entries earlier than 1900 appear to be input errors). Please handle these abnormal data points with careful consideration to ensure the reliability of your model.

Task 3: k Nearest Neighbours [20%]

Management was impressed with your work on Decision Trees. They have heard it may be possible to infer the probability a person is likely to accept the offer from whether they have similar characteristics to other known customers who have already left.

Your goal is to build the best k Nearest Neighbours model you can to predict the probability a given customer will accept the offer.

Task 4: Clustering [20%]

Previously, management have been quite directed about what they were looking for and what they wanted to predict. Now, they are asking you what kind of things can be 'discovered' from the data. In particular they are interested in whether there are any kind of natural groupings that exist within their customers.

Your goal is to find and explain any ONE natural grouping you find within the data. You only need to concentrate on finding one way to group users, and then explain that grouping in business terms.

Task 5: Building a Data Science Dashboard [20%]

Management was impressed with the models you have built so far!

They are also starting to think about deployment. At a recent conference on Data Science and Big Data, management attended a seminar about Microsoft Power BI, and they think this would be the perfect deployment approach for their models.

For this task, you only need to implement the tree and KNN models.

Management wants you to build a Data Science Dashboard. As a minimum, for the tree and KNN models, the dashboard will load the models and for each model, it will show visualizations of the model and its fitness for purpose. The dashboard would then allow a user to input the required details and receive a prediction from each of the underlying models.

Beyond this required functionality, you may enhance the dashboard as you see fit.

Instructions

Submit to Blackboard:

- **A report in .PDF format produced from knitr.** Your report should contain your responses to Tasks 1-4. You must submit the report to the Turnitin "Report Submission" tab on Blackboard. The report must be produced from knitr; failure to deliver this will result in reduction to your marks. There are two approaches you could take to produce the PDF report:
    - You could either knit your R Markdown file into a HTML file and then save it to a PDF file, or
    - You could directly knit your R Markdown file into a PDF file.
- **The Power BI dashboard (.pbix file)**. The dashboard should be submitted to the "Dashboard Submission" tab on Blackboard. Failure to submit this file will result in reduction to your marks. Note that the Power BI software (free license) is available via University Digital Services - Microsoft Office 365.
- **An R Markdown file in .RMD format containing your codes.** The markdown file should be submitted to the "R Markdown Submission" tab on Blackboard. Failure to submit this file will result in reduction to your marks.
- Please name your files using your student ID number, e.g. 500123456.PDF, 500123456.pbix, and 500123456.RMD. You could include author name inside your report/dashboard/R Markdown file, but **do not include your name in your file names.**
- You are allowed to test submit your report before the due date. You can use Turnitin to check your assignment for plagiarism before you submit your final version. The last submission prior to the deadline will be treated as the final submission and will be the copy that is assessed by the marker.