

BPP Coursework Cover Sheet

Please use the table below as your cover sheet for the 1st page of the submission. The sheet should be before the cover/title page of your submission.

Programme	MSc Management with Data Analytics
Module name	Applied Modelling and Visualisation
Schedule Term	
Student Reference Number (SRN)	
Report/Assignment Title	Applied Modelling and Visualisation
Date of Submission <i>(Please attach the confirmation of any extension received)</i>	
<u>Declaration of Original Work:</u>	
I hereby declare that I have read and understood BPP's regulations on plagiarism and that this is my original work, researched, undertaken, completed and submitted in accordance with the requirements of BPP School of Business and Technology.	
The word count, excluding contents table, bibliography and appendices, is ____ words.	
Student Reference Number:	Date:
By submitting this coursework you agree to all rules and regulations of BPP regarding assessments and awards for programmes. Please note, submission is your declaration you are fit to sit.	
BPP University reserves the right to use all submitted work for educational purposes and may request that work be published for a wider audience.	
BPP School of Business and Technology	

MSc Management with Data Analytics

Applied Modelling and Visualisation

Coursework Assessment Brief

Submission mode: **Turnitin online access**

1. Assessment Brief

This assessment brief gives you an overview of the formative and summative assessments that are part of this module. The learning outcomes below will be tested in the assessment contained in this brief.

1.1. Module Learning Outcomes (LOs)

ILO 1: Formulate innovative data driven solutions to commercial problems

ILO2: Critically evaluate the use of algorithms and model when developing analytical solutions

ILO3: Critically appraise the concepts, tools and techniques for data visualisation

1.2. Assessment Overview

A written report, with a code submission included in an appendix or included within a single zip submission. See instructions at the end on how to create a zip file containing your work.

Please note: ensure you read the general assessment guidance at the end of this document.

2. Assessment Structure/Guidance

Every year billions of pounds are lost to credit card fraud. While fraud is a substantial cost to the financial system, so too is the cost of detecting credit card fraud. In a competitive market, banks need to balance the cost of fraud against the impact on customer experience of onerous controls. In other words, false positives in the fraud detection process can lead to customer transactions being erroneously declined. This highlights the need for effective detection systems with low rates of error.

Your task, as a data analyst working for a large financial services advisory firm, is to ***analyse a set of transactions, develop the respective feature inputs and create an appropriate, scalable model for predicting potential fraudulent transactions.***

Formative Submission

Your formative submission will be a written report (at most 1000 words) that should attempt tasks 1 and 2 as described in the summative submission and select **one** relevant analytical model models to classify whether a card transaction is potentially fraudulent or not, and critically analyse the model, as described in task 3. You should **include a code appendix** that performs the associated tasks.

Summative Submission

You are provided with a set of 100,000 card transactions* – recorded on 13th and 14th October 2020. Each transaction feature is listed:

- Transaction ID
- Date
- Time
- Type of Card – Visa, MasterCard

- Entry Mode – Tap, PIN
- Amount
- Type of Transaction – Online, POS, ATM
- Merchant Group
- Transaction Country
- Shipping Address
- Billing Address
- Gender of Cardholder
- Age of Cardholder
- Issuing Bank

Your submission should be a written report (at most 3000 words) that describes how modelling and visualisation could be applied to reducing the rate of fraud in a financial context. You must also **include a code appendix** that can perform the following tasks:

1. Provide a rationale of the steps taken during each step of the Extract, Transform and Load (ETL) phase of the project, discussing any ambiguities, assumptions, and anomalies in the provided data and how you should deal with them (ILO1).
2. Explain the justification for performing Exploratory Data Analysis (EDA) and the use of appropriate descriptive statistics and visualisations to understand the results of that analysis, and critically analyse how the EDA process will guide your selection of analytical models (ILO3).
3. Select **two** relevant analytical models to classify whether a card transaction is potentially fraudulent or not (ILO2). Critically analyse the strengths and limitations of each model with references to the relevant literature. You should choose from the following models:
 - i. Logistic regression,
 - ii. Decision Tree Classifier
 - iii. Bagging Classifier
 - iv. Random forest classifier
 - v. AdaBoost Classifier
 - vi. XGBoost
 - vii. Artificial neural network
 - viii. Another appropriate state-of-the-art algorithm
4. Provide a **critical evaluation of each model** selected in the previous task by using your test data set (ILO2).
 - a. Including an explanation of your chosen loss function.
 - b. A short discussion of the accuracy metrics.
 - c. Cast the accuracy metric, the number of correct predictions, and the number of incorrect prediction results of all the models to a table to allow for comparison.
 - d. Based on your findings, make a critically justified recommendation for the use of one model for reducing the rate of fraud.

5. **Communicate your findings** (ILO3). Provide several graphical outputs (with commentary) such as a ***correlation matrix, a heat map or confusion matrix*** of your results, in order to illustrate your analytical outputs in a visual manner.
6. A code appendix that performs the following tasks (ILO1):
- i. Import, clean and prepare the data for analysis, ensuring the relevant test, validation and training sets are prepared
 - ii. Perform Exploratory Data Analysis with appropriate visualisations
 - iii. Train and test the two analytical models you selected in task 3
 - iv. Evaluate the two models based on your choice of loss function
 - v. Produce appropriate visualisations of your results

Research and referencing

Your report should include a list of references used to develop the report and research to support the suggested approach. The list should use only the *Harvard Referencing System* as highlighted in the *General Assessment Guidance* section of this document. All the figures/tables used in the report must have captions and, wherever needed, properly referenced, and explained in your submission.

Creating a zip file

- Step 1: Locate the report file **and** Python notebook (or a folder that contains them both). Select both the files (use Ctrl+click to select both files), or the folder.
- Step 2: Once you have selected them, right-click on them, select “Send to”, and then select Compressed (zipped) folder.
 - A new zipped folder with the same name is created in the same location.
 - Check that it worked correctly: double click on the compressed file to open it, you should see it contains both files or the folder with the files inside it.
- Step 3: Upload the zipped file (containing both of your submission files) onto the submission link. Done!

Marking Guide/ Assessment Criteria

Modelling and Visualisation		0-39% Fail	40-49% Low Fail	50-59% Pass	60-69% Merit	70-79%	80-100% Distinction
30%	Formulate data-driven solutions (ILO1) Attach code appendix that performs the outlined tasks and discuss ambiguities, assumptions and anomalies.	<p>Notebook fails to execute, fails to display the options, or halts during execution.</p> <p>Inadequate and often implicit knowledge base with some omissions and/or lack of theory <i>relating to the use of ETL processes</i>. No discussion of ambiguities, assumptions or anomalies.</p> <p>Notebook fails to produce any outputs which can be used to communicate your findings</p>	<p>Notebook correctly loads the input data file into a Python data structure. No comments are given on the method used. Notebook uses a package to conduct EDA, as well as fraud likelihood prediction with the outputs of the appropriate model outcomes and metrics but with no explanation or comments.</p> <p>Weak and often implicit knowledge base with some omissions and/or lack of theory <i>of the use of modelling and visualisation for a data project (and relevant code libraries)</i></p> <p><i>Notebook correctly uses a package to produce communication tools but does not contain any explanation or commentary.</i></p>	<p>Notebook correctly loads the input data file into a Python data structure. Comments are given on the approach taken.</p> <p>Notebook correctly handles duplicate values as well as EDA. Comments are given.</p> <p>The script achieves prediction for the fraud likelihood and also correctly outputs appropriate model outcomes and metrics with reasonable level of commentary and explanation.</p> <p>Notebook correctly uses a package to produce communication tools, with reasonable explanations and comments.</p>	<p>Notebook correctly loads the input data file into a Python data structure. Comments and explanations are given with detail on the extract phase of the project.</p> <p>Notebook handles duplicate values, missing values as well as descriptive statistics explaining the steps taken to reach the results.</p> <p>Notebook also achieves prediction for the fraud likelihood with good explanation and comments about the method used. There are model evaluation metrics outputted alongside predictions.</p> <p>Notebook correctly uses a package to produce communication tools with good explanation and comments about the method used.</p>	<p>Notebook correctly loads the input data file into a Python data structure. The comments provided cover technical details of the extract phase of the project, demonstrating extensive knowledge on dataframe imports.</p> <p>Notebook handles duplicate values, missing values and explains in detail the steps taken to reach the results.</p> <p>Correctly uses a package to achieve prediction for the fraud likelihood and outputs the appropriate model outcomes and metrics.</p> <p>Explanations are detailed and profound.</p> <p>Notebook correctly uses a package to produce communication tools, with very detailed explanation and comments about the model output and your chosen method of communication conveys this.</p>	<p>Notebook correctly loads the input data file into a Python data structure in a modular fashion. The comments provided cover exceptional technical details of the extract phase of the project, demonstrating extensive knowledge on dataframe imports and their peculiarities.</p> <p>Notebook handles duplicate values, handles missing values, correctly uses a package to achieve prediction for the fraud likelihood and outputs the appropriate model outcomes, metrics as well as an example of the prediction in action for a new mock transaction entry. Comments provided are profound in detail.</p> <p>Explain in detail the steps taken to reach the results with further explanation of methods to expand the steps taken or process followed.</p> <p>Also explains rationale behind the methods used.</p> <p>Notebook correctly uses a package to produce communication tools with very detailed explanation and comments about the method used including examples of similar practices and suggestions to further enhance the communication of results.</p>
30%	Critically evaluate the use of models, analysing the strengths and weaknesses (ILO2)	<p>Inadequate and often implicit knowledge base with some omissions and/or lack of theory <i>relating to the use of programming for predictive modelling</i>. No explanation of loss function, accuracy metrics, or recommendation of model for fraud reduction.</p>	<p>Weak and often implicit knowledge base with some omissions and/or lack of theory <i>relating to the use of programming for predictive modelling</i>. Weak explanation of loss function, accuracy metrics, or recommendation of model for fraud reduction.</p>	<p>Satisfactory knowledge base that begins to explore and analyse the theory relating to the use of programming for predictive modelling. Satisfactory explanation of loss functions, accuracy metrics and comparative strengths of models based on ability to reduce fraud rate drawing on the academic literature.</p>	<p>Good knowledge base that explores and analyses the theory relating to the use of programming for predictive modelling. Good explanation of loss functions, accuracy metrics and comparative strengths of models based on ability to reduce fraud rate drawing on the academic literature with originality and autonomy.</p>	<p>Excellent knowledge base that explores and analyses the theory relating to the use of programming for predictive modelling. Excellent explanation of loss functions, accuracy metrics and comparative strengths of models based on ability to reduce fraud rate drawing on the academic literature with considerable originality and autonomy.</p>	<p>Outstanding knowledge base that explores and analyses the theory relating to the use of programming for predictive modelling. Excellent explanation of loss functions, accuracy metrics and comparative strengths of models based on ability to reduce fraud rate drawing on the academic literature with outstanding originality and autonomy at the cutting edge of current scholarship.</p>

30%	Critically using and appraising data visualisation techniques (ILO3)	<p>Inadequate and often implicit knowledge base with some omissions and/or lack of theory <i>relating to the use of EDA, descriptive statistics and data visualisation. There are no data visualisations, neither in the notebook nor the report.</i></p> <p><i>The student did not explain the justification for performing EDA, did not present appropriate descriptive statistics and has not explained how EDA guides model selection.</i></p>	<p>Weak and often implicit knowledge base with some omissions and/or lack of theory <i>relating to the use of data visualisation. There isn't sufficient evidence of useful data visualisations, neither in the notebook nor the report.</i></p> <p><i>There is weak explanation for performing EDA, coming up with appropriate descriptive statistics and how EDA guides model selection.</i></p>	<p>Satisfactory knowledge base that begins to explore and analyse the theory relating to the use of data visualisation.</p> <p>The student has presented several appropriate data visualisations, communicating insights visually both in the report and the notebook.</p> <p>There is satisfactory explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.</p>	<p>Good knowledge base that explores and analyses the theory relating to the use of data visualisation.</p> <p>The student has presented several appropriate data visualisations, communicating insights visually both in the report and the notebook.</p> <p>There is good explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.</p>	<p>Excellent knowledge base that explores and analyses the theory relating to the use of data visualisation techniques.</p> <p>The student has presented several high quality data visualisations, excellently communicating insights visually both in the report and the notebook.</p> <p>There is excellent explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.</p>	<p>Outstanding knowledge base that explores and analyses the theory relating to the use of data visualisation.</p> <p>The student has presented several outstanding data visualisations, excellently communicating insights visually both in the report and the notebook.</p> <p>There is outstanding explanation for performing EDA, appropriate descriptive statistics and how EDA guides model selection.</p> <p>There are examples of data visualisation techniques at the cutting edge of industry.</p>
10%	Academic Research and Referencing Skills	<p>Inadequate critical analysis or evaluation with some difficulties. Largely imitative and descriptive. Some difficulty with structuring the line of logical argument and accuracy in expression of argument.</p> <p>Inadequate references and notes but may contain inconsistencies, errors or omissions.</p>	<p>Limited critical analysis and/or evaluation with reflection and broad evidence-based critique. Solid structure or argument including line of logical reasoning and accuracy in expression of argument.</p> <p>Limited and full and appropriate references and notes with minor or insignificant errors</p>	<p>Satisfactory critical analysis and/or evaluation. Good reflection and solid, well-reasoned judgements forming from evidence-based critique. Consistent logical structure of argument including the line of reasoning and accuracy in expression of argument.</p> <p>Satisfactory with precise, full and appropriate references and notes.</p>	<p>Good critical analysis and/or evaluation skills. Demonstrates intellectual originality and imagination</p> <p>Good with precise, full and appropriate references and notes at a high standard.</p>	<p>Excellent critical analysis and/or evaluation skills. Demonstrates intellectual originality, integrity, coherence and imagination.</p> <p>Excellent with precise, full and appropriate references and notes at near-publishing standard.</p>	<p>Outstanding critical analysis and/or evaluation. Demonstrates intellectual originality, integrity, coherence, creativity and imagination working consistently in the higher cognitive domains to a professional standard.</p> <p>Outstanding with precise, full and appropriate references and notes at publishing standard.</p>

General Assessment Guidance

- You are required to submit all elements of your assessment via **Turnitin online access**. Only submissions made via the specified mode will be accepted and hard copies or any other digital form of submissions (like via email or pen drive etc.) **will not be accepted**.
- For coursework, the submission word limit is notionally **3000** words.
- **Please ensure your student registration number is on your front cover sheet**
- You are required to achieve minimum **50%** to **pass** this module and must address all learning outcomes.
- You are required to use only Harvard Referencing System in your submission. Any content which is already published by other author(s) and is not referenced will be considered as a case of plagiarism.
- *You can find further information on Harvard Referencing in the online library on the Hub. You can use the following link to access this information:*
<http://bpp.libguides.com/Home/StudySupport>
- BPP University has a strict policy regarding authenticity of assessments. In proven instances of plagiarism or collusion students will go through the malpractice process. You are advised to read the rules and regulations regarding plagiarism and collusion in the GARs and MOPPs which are available on Hub in the Academic registry section.
- You should include a completed copy of the **Assignment Cover sheet**. Any submission without this completed Assignment Cover sheet may be considered invalid and not marked.
- Please note that any work submitted in an appendices is for information only and is NOT marked. You should ensure you hit all requirements and learning outcomes within the word count set.