Sheffield Hallam University

College of Business, Technology and Engineering (BTE)

Programming Concepts and Practice

Academic Year: 2021/2022, Semester 2

ASSIGNMENT 2

Module Leader: Dr. Bayode Ogunleye		Level: 7
Module: Programming Concepts and Practice		Module Code: 55-706555
Assignment Title: PCP Assignment 2		
Academic Year: 2021/2022		
This is an individual task. There should be no collusion or collaboration whilst working on and subsequently submitting the assignment.		
Individual	Weighting: 40%	Wordcount: 1500
Submission date/time: **17-May-2022, 14:59**	Blackboard submission: Yes Turnitin submission: Yes	Format: source code, digital media, report.
Planned feedback date: **07-June-2022**	Mode of feedback: Written and verbal	In-module retrieval available: No

Module Learning Outcomes

- LO1: select appropriate programming techniques and data structures to develop effective software implementations of relatively complex systems using an appropriate programming language
- LO2: apply relevant program design strategies to the implementation of software applications using that programming language.
- LO3: design and implement well-engineered, domain specific software using that programming language

1. Introduction

You have been asked by SHU Data Analytics team to explore different programming and analytics techniques to analyse and evaluate student performance and thus, build predictive model to make informed decisions. For this purpose, you will make use of programming concepts such as use of custom module, function definitions, file processing and exception handling, use of scientific computing, data analysis, data visualization and machine learning libraries (such as numpy, pandas, matplotlib, and scikitlearn) in the implementation.

2. Dataset

The dataset for this assignment can be downloaded from the PCP Blackboard module site or the link below. Please study the dataset in terms of size, data type and variables.

- Students performance dataset (available on Blackboard & UCI repository).

3. Assignment Key Tasks

The following tasks are to be performed in this assignment:

i). Exploratory Data Analysis

You are required to write codes to check if there's any data missing (if yes, apply an appropriate cleaning technique). Is there any other data preprocessing you need to conduct? If yes, write codes for this purpose. In addition, the module should have methods that perform descriptive statistical analysis of the dataset (such as mean, median, standard deviation, variance, minimum, maximum, skewness and kurtosis): choose a range of the variables of your interest, find their frequencies and dependencies through bar plots, grouped bar plots, pie-charts, etc. Draw conclusions.

ii). Regression

Split the dataset on training and testing sets. Build Random Forest Regression model to predict a final year grade (G3). Evaluate your model using the test dataset. Plot the feature importance graph. Estimate mean square error and accuracy. Comment on your results.

iii). Classification

Based on the final grade, divide the students into 3 categories. For example, poor achieving students (low), average achieving (medium), well achieving (high). Investigate class imbalance problem by producing the plot of the class distribution. If there is presence of class imbalance problem, use at least 2 techniques to balance the class distribution (Algorithm or Sampling technique). Build **three** classification models (Support vector machine, Random Forest classifier and Multi-Layer Perceptron Neural Networks). Evaluate your models using test dataset and provide the confusion matrix for all models. Report and compare performance of the models in terms of accuracy, precision, recall and F1-Score. Draw conclusions and provide recommendations. Please provide justification for chosen methods.

4. Requirement

This assignment is an individual piece of work, and your submission must be in the form of modules (.py files) or Jupyter Notebook file. Tutors should be able to open and run your modules on a standard campus computer.

I. You are required to submit at least two python files. One of that should implement **a custom module** (.py files)**.** An example can be custom module (saved as EDA.py) which consists of functions defined to perform the descriptive statistics. Seconded by main.ipynb file which consists of EDA module imported (to perform exploratory data analysis), regression and classification solutions.

II. You are required to submit a report (Ms word or PDF file). The report should provide justifications for your analysis of the solution, design decisions and pseudocodes. It should explain the relationships between the modules. A good report should be based on evidence with critical analysis of the implemented system. In addition, your report should include a reflection section of your experience while executing this project. The reflection should detail what went well or not and lesson learnt. What would you do differently if you have another chance to execute this project again?

III. You are required to submit a video recording demonstrating the programming and machine learning concepts you have adopted. **Kindly explain the concepts in detail.**

IV. Any evidence of collusion/plagiarism will be penalised if appropriate! If there is some doubt about the authenticity of a particular piece of work, then the person submitting it will be expected to defend such work, including reasons for the programming decisions

taken. You must document with references any use of libraries or existing code in your report.

V. Appropriate use of variable names for clearer understanding is desirable

VI. Adequate commenting of your codes for easier understanding during grading is also desirable.

5. Submission Process

- Your assignment should be submitted electronically through the module's Blackboard site as a single ZIP file that contains **all your source codes, video demonstration and report**.
- In addition, a copy of your report should be submitted on Turnitin.
- If your video is longer than 15 minutes, we will stop watching it during grading at exactly 15th minute.
- Please check your upload to ensure you have submitted the correct files successfully as any issues will not be considered after the deadline.
- Kindly provide an explanation in your report on how to execute your application.
- You must also check your report's similarity score using Turnitin on the Blackboard before final submission. Please do not submit any report with similarity score higher than 20%. Otherwise, you will be penalized for plagiarism or collusion.
- Your assignment must be submitted on **Tuesday, 17th May 2022 by 2:59 pm.**
- Note that late submission will attract penalty. The penalty is capping of your mark to 50%.

6. Suggested Structure of the Report:

i. Cover page with your name, student number and title

ii. Introduction which contains a short description of the context & method.

iii. Answers on the stated questions should be **well discussed.** For example, approach, and result/findings.

iv. All evaluative results should be presented in a table (screenshots of Python result in Appendix)

v. All plots, figures and graphs must be numbered and clearly labelled.

vi. Provide conclusion and recommendations.

vii. Please include the reference, especially, to all libraries used in the programming. You

are required to use the APA for referencing.

7. References

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

Zohair, L. M. A. (2019). Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education, 16(1), 1-18.

Zeineddine, H., Braendle, U., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. Computers & Electrical Engineering, 89, 106903.

Vultureanu-Albişi, A., & Bădică, C. (2021). Improving Students' Performance by Interpretable Explanations using Ensemble Tree-Based Approaches. In 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 215-220). IEEE.

8. Assessment Criteria

This assignment will be assessed through the report, testing of implementation and video demonstration of the submitted codes using the data files submitted. The video demo should demonstrate how your solution meets the assessment criteria. In general, the coursework will be assessed against the Learning Outcomes (LOs) using a set of assessment criteria. This set of assessment criteria allows assessing how successful you have met the LOs. In order to ensure consistent use of the relevant criteria, the assessment criteria are summarised in the following assessment matrix and grid. This is an indicator of how the marks will scale across each category of the learning outcomes it covers.

Table 1: Assessment Matrix

Assessment 2 Criteria	Marks	Learning Outcome Covered
Clear understanding of relevant programming concepts like use of custom module, functions, parameters, and argument for EDA solution	10%	LO1, LO2, LO3
Use of python (data science) libraries to solve regression problem	10%	LO3
Use of python (data science) libraries to solve class Imbalance & classification problem	15%	LO3
Knowledge of subject/Interpretations/Video Demonstration/Report/Deliverable	5%	LO2, LO3

Note that the University's common grading descriptor in Table 3 will be used to determine marks within each area shown in Table 3. The marking scheme embeds the concept of extended work by rewarding only the highest marks to those who demonstrate evidence of independent investigation, learning, and thought. Thus, to achieve top grades, you will need to go beyond the materials presented in lectures and labs and undertake some of your own research (i.e. read and discuss related materials).

Table 2: Assessment Marking Grid

Fail (<50%)	Pass (50-59)	Merit (60-69)	Distinction (70% +)
Understanding of relevant programming concepts for EDA solution (/10)			
No evidence of understanding and use of parameter and argument passing. Nothing is submitted.	Evidence of clear and consistent understanding of the function definition, parameters, and argument passing. Evidence of practical solution.	Very good and appropriate definition of functions, parameters and argument passing.	Exceptional understanding and creative use of programming solutions.
Use of python (data science) libraries to solve regression problem (/10)			
No evidence of the use of Python libraries such as numpy or pandas. Not able to apply appropriate python libraries. No submission	Clear and good evidence of the use and application of python libraries (such as numPy, pandas, matplotib and scikit-learn) with some correct and expected outputs. But some minor issues with outputs.	Very good understanding and good implementation using python libraries to implement some of the functionality of the system with justifications and correct outputs. Program executes and produces expected	Excellent understanding and implementation of relevant python libraries, such as numPy , pandas, matplotib and scikit-learn with outstanding results. excellent user interaction through GUI, etc.
Use of python (data science) libraries to solve class Imbalance & classification problem (/15)			
No evidence of the use of Python libraries such as numpy or pandas. Not able to apply appropriate python libraries. No submission	Clear and good evidence of the use and application of python libraries (such as numPy, pandas, matplotib and scikit-learn) with some correct and expected outputs. But some minor issues with outputs.	Very good understanding and good implementation using python libraries to implement some of the functionality of the system with justifications and correct outputs. Program executes and produces expected	Excellent understanding and implementation of relevant python libraries, such as numPy , pandas, matplotib and scikit-learn with outstanding results. excellent user interaction through GUI, etc.
Knowledge of subject/Interpretation/Video Demonstration/Report/Deliverable (/5)			
Report lacking good structure, no personal reflection, no description of the deliverable or explanation and justification of decisions. Poor use of language. No/Poor video demonstration.	Good structure, evidence of personal reflection on what went well or not. Good recommendations were made. Good justification for design and implementation decisions. Good use of language. Good video demonstration.	Very good structure, evidence of personal reflection on what went well or not. Very good recommendations were made. Very good justification for design and implementation decisions. Very good video demonstration. Good use of language.	Excellent structure, excellent personal reflection on what went well or not. Excellent recommendations were made. Good justification for design and implementation decisions. Excellent video demonstration. Excellent use of language. Evidence of innovation in the deliverable, e.g. excellent user interaction through UI, etc.

Table 3: Level 7 RUBRIC for Grading

Class	Mark range	CG %	General Characteristics
DISTINCTION (Excellent)	93 - 100	96	Exceptional breadth and depth of knowledge and understanding evidenced by own independent insight and critical awareness of relevant literature and concepts at the forefront of the discipline; evidence of extensive and appropriate independent inquiry operating with advanced concepts, methods and techniques to solve problems in unfamiliar contexts; Cogent arguments and explanations are consistently provided using a range of media demonstrating an ability to communicate effectively in a variety of formats using a sophisticated level of the English language in an eloquent and professional manner to both technical and non-technical audiences; a sustained academic approach to all aspects of the tasks is evidenced; academic work extends boundaries of the disciplines and is beyond expectation of the level and **may achieve or be very close to publishable or commercial standard**.
	85 - 92	89	Excellent knowledge and understanding **evidenced by some clear independent insight and critical awareness of relevant concepts some of which are at the forefront of the discipline**; evidence of appropriate independent inquiry operating with core concepts, methods and techniques to solve complex problems in mostly familiar contexts; Arguments and explanations are provided that is well-supported by the literature and in some cases uses a range of media demonstrating an ability to communicate effectively in a limited number of formats using own style that is suited to both technical and non-technical audiences; a sustained academic approach to most aspects of the tasks is evidenced; **one or more aspects of the academic work is beyond the prescribed range** and evidences a competent understanding of all of the relevant taught content.
	78 - 84	81	
	70 - 77	74	
MERIT (Good)	67 - 69	68	Very good knowledge and understanding is evidenced as the student **is typically able to independently relate taught facts/concepts together some of which are at the forefront of the discipline**; evidence of some competent independent inquiry operating with core concepts, methods and techniques to solve familiar problems; Arguments and explanations are provided that are typically supported by the literature and in some cases may challenge some received wisdoms; competently uses all taught media and communication methods to communicate effectively in a familiar settings; an academically rigorous approach applied to some aspects of the tasks is evidenced; some beyond the prescribed range, may rely on set sources to advance work/direct arguments; demonstrates autonomy in approach to learning.
	64 - 66	65	
	60 - 63	62	

PASS (Satisfactory)	57 - 59	58	Good knowledge and understanding of the area of study **balanced towards the descriptive rather than critical or analytical and mostly confined to concepts that are not at the forefront of the discipline**; evidence of some independent reading and research to advance work and inform arguments and approaches; Arguments and explanations are limited in range and depth although some are adequately supported by the literature albeit descriptively rather than critically; competently uses at least one taught media and communication method to communicate appropriately in familiar settings; although the approach applied to some aspects of the tasks may lack academic rigour, there are some clear areas of competence within the prescribed range. Relies on set sources to advance work/direct arguments and communicated in a way which shows clarity, but structure may not always be coherent.
	54 - 56	55	
	50 - 53	52	
FAIL (Insufficient)	40 - 49	45	**Knowledge and understanding is marginally insufficient as the student is typically only able to deal with terminology, basic facts and concepts**; Adequate knowledge of concepts within the prescribed range but fails to add meaningful detail or make sufficient links between concepts and facts to adequately solve problems posed by the assessment; some ability to independently select and evaluate reading/research however there is a strong reliance on set sources and to provide descriptive and unsubstantiated arguments/methods; communication/presentation is competent in places and at a threshold level as it fails to demonstrate clarity and focus; inability to adequately define problems and make reasoned judgements; The general approach to tasks lacks rigor and where there is competence and rigor, it is not sustained.
	30 - 39	35	
	20 - 29	25	**Knowledge and understanding is highly insufficient as the student is unable to evidence any meaningful understanding of two or more taught concepts or methods**; very limited evidence of reading and research to advance work; inadequate technical and practical skills as the student is unable to use and apply such skills to address problems or make judgements; limited or lack of understanding of the boundaries of the discipline and does not question received wisdom; approach to learning lacks autonomy and approach to tasks is not sustained; inability to communicate coherently.
	10-19	15	
	1-9	5	
ZERO	0	0	Work of no merit OR absent, work not submitted, penalty in some misconduct cases.

****

All work must be your own. If evidence of collusion/copying is found, then such collusion will be penalised, severely if appropriate! If there is some doubt about the authenticity of a particular piece of work, then the person submitting it will be expected to give a detailed explanation of such work, including reasons for the programming decisions taken.

****

You must submit **a demonstration video (not more than 15 minutes)** demonstrating your running prototype system and explaining how it meets the assessment criteria. Please look at the marking criteria and prepare your demonstration accordingly. In case you face any challenge in video recording, please run your system, and provide screencast explaining how your system satisfy the assessment criteria,

****

WARNING

All work must be yours. If evidence of collusion/copying/plagiarism is found, then such collusion will be penalised, severely if appropriate! If there is some doubt about the authenticity of a particular piece of work, then the person submitting it will be expected to give a detailed explanation of such work, including reasons for the programming decisions taken