DEGREE: MSc Data Analytics
Module: Predictive Analytics and Machine Learning using Python

Assignment Title: Data analysis and machine learning with Python for real-world problems: From data to informed decision making
Assignment Type: Report
Word Limit: 2000-3000 words (+/- 250)
Weighting: 50%
Issue Date: 25/07/2024
Submission Date: 25/09/2024
Feedback Date: 16/10/2024

Plagiarism:
When submitting work for assessment, students should be aware of the InterActive/Canvas guidance and regulations in concerning plagiarism. All submissions should be your own, original work. You must submit an electronic copy of your work. Your submission will be electronically checked.
Learner declaration
I certify that the work submitted for this assignment is my own and research sources are fully acknowledged.
Student signature: Date:
Harvard Referencing:
The Harvard Referencing System must be used. The Wikipedia, UKEssays.com or similar websites must not be used or referenced in your work.
Learning Outcomes:
LO1. Demonstrate the understanding of basic concepts of dealing with different types of data – ordinal, categorical, encoding along with collecting, storing and making it ready for processing.
LO2. Explain the various components of predictive analytics, with the models for regression, classification and clustering to analyze real-life business problems.
LO3. Implement various models and work on a project life cycle from end to end to solve an analytical problem which translates into a business problem solution using machine learning and artificial intelligence.
Assessment Criteria: Weighting 50%
2000-3000 words
Introduction
This assignment aligns with the unit's objectives of introducing learners to predictive analytics, machine learning, and artificial intelligence concepts, and developing their ability to apply these techniques to solve real-world business or social problems using Python.
Tasks:
Each student is required to develop a comprehensive written report on the topic of

Data analysis and machine learning with Python for real-world problems:
From data to informed decision making. The report should include the following elements:

- An introduction that justifies the significance of predictive analytics and machine learning for data-driven decision making (10 pts.).
- The main text that focuses on the following tasks:

Task 1. Select a data set that reflects a real-world business/social problem and identify all the data types in it. Check whether data cleaning and/or encoding is needed. Conduct an exploratory data analysis (related to LO1, 25 pts.).

Task 2. Select two or more machine learning models appropriate for the business/social problem in question. For each algorithm, write a python code and clearly describe each step of it (related to LO2 and LO3, 25 pts.).

Task 3. Using appropriate metrics, evaluate your machine learning models and visualize their results. Compare the outcomes of different algorithms and recommend the best performing one (related to LO1 - LO3, 25 pts.).

- A conclusion that summarizes your findings and their implications for the business/social problem in question (10 pts.).
- A list of references (5 pts.).

The Harvard Referencing System must be used. The Wikipedia, UKEssays.com or similar websites must not be used or referenced in your work. Your essay should be 2000-3000 (+/- 250) words, excluding Title, Table of Contents, Bibliography and Appendices. For more information please refer to the BSBI Essays Guide.