THE UNIVERSITY OF HUDDERSFIELD

School of Computing and Engineering

ASSIGNMENT SPECIFICATION

Module	Details
Module Code	CHS2406-2425
Module Title	Data-driven Artificial Intelligence
Course Title/s	BSc (Hons) Computer Science, BSc (Hons) Computer Science with Games Programming, BSc (Hons) Computer Science with Cyber Security, BSc (Hons) Computer Science with Artificial Intelligence, BSc (Hons) Software Engineering, BSc (Hons) Computing, BSc (Hons) Applied Computing (Top-up)

Assessment	Weighting, Type and Contact Details
Title	Hotel Booking Prediction
Weighting	50%
Mode of working for assessment task	Individual This assessment task is to be completed on an individual basis and there should be no collusion or collaboration whilst working on and subsequently submitting this assignment.
Module Leader	Abirami Gunasekaran
Module Tutor/s	Abirami Gunasekaran

Submission	Submission and Feedback Details
Hand-out date	23/09/2024
How to submit your work.	Brightspace submission point
Submission date/s	**25/10/2024** by 12:00 noon – if you have any technical issues submitting your work, please contact the Module Leader as soon as possible.
Expected amount of independent time you should allocate to	20 hours

Submission	Submission and Feedback Details
complete this assessment	
Submission type and format	A jupyter notebook both in pdf and ipynb format.
Date by which your grade and feedback will be returned	**17/11/2025** Note: This is a maximum of three working weeks after the submission deadline.

Additional Guidance Information	Details
Your responsibility	It is your responsibility to read and understand the University regulations regarding conduct in assessment. Please pay special attention to the assessment regulations (section 10) on Academic Misconduct. In brief: ensure that you; 1. DO NOT use the work of another student - this includes students from previous years and other institutions, as well as current students on the module. 2. DO NOT make your work available or leave insecure, for other students to view or use. 3. Any examples provided by the module tutor should be appropriately referenced, as should examples from external sources. Further guidance can be found in the SCEN Academic Skills Resource and UoH Academic Integrity Resource module in Brightspace. If you experience difficulties with this assessment or with time management, please speak to the module tutor/s, your Personal Academic Tutor, or the School's Guidance Team. (sce.guidance@hud.ac.uk).

Additional Guidance Information	Details
Guidance on using AI:	 Level 2 – Some use Permitted Some use of AI tools is permitted in the research/early stages of this assignment but you must ensure that the work you submit is your own. If you use AI tools, you should acknowledge or reference this in your work. Use the Text reference builder to learn how to reference AI generated ideas. The sorts of questions to consider when using AI are: • Is it accurate? • Are the references genuine? • Has it reproduced bias?
Requesting a Late Submission	It is expected that you complete your assessments by the published deadlines. However, it is recognised that there can be unexpected circumstances which may affect you being able to do so. In such circumstances, you may submit a request for an extension. Extension applications must be submitted before the published assessment deadline has passed. There are two types of extension that you may request. You will be required to indicate which one you are applying for when you submit the request for Late Submission via MyHud/MyStudies. 1. Self-certified illness extension of up to 5 working days. • Evidence will not be required for this type of request, but you are limited to two self-certified extension requests in any academic year. 2. Extension request, of up to 10 working days. • This extension requires you to submit appropriate evidence in support of your request.

Additional Guidance Information	Details
	The maximum extension that can grant is 10 working days. Accepted grounds for an extension • Serious short-term illness or accident (of a nature which in employment would result in a health-related absence); • Evidence of a long-term health condition worsening; Emerging mental health condition, or worsening of an existing mental health condition; • Bereavement. If you are unable to submit work within the maximum late submission period of 10 days, contact the School's Guidance Team. (sce.guidance@hud.ac.uk), as you may need to submit a claim for Extenuating Circumstances (ECs).
Extenuating Circumstances (ECs)	An EC claim is appropriate in exceptional circumstances, when an extension is not sufficient due to the nature of the request. You can access the EC claim form on the Registry website; where you can also find out more about the process. You will need to submit independent, verifiable evidence for your claim to be considered. Once your EC claim has been reviewed you will get an EC outcome email from Registry. If you are unsure what it means or what you need to do next, please speak to the Student Support Office – Room SJ1/01 An approved EC will extend the submission date to the next assessment period (e.g July resit period).
Late Submission (No ECs approved)	Late submission, up to 5 working days, of the assessment submission deadline, will result in your grade being capped to a maximum of a pass mark.

Additional Guidance Information	Details
	Submission after this period, without an approved extension, will result in a 0% grade for this assessment component.
Tutor Referral available	YES
Resources	• Please note: you can access free Office365 software and you have 1 Tb of free storage space available on Microsoft's OneDrive – Guidance on downloading Office 365.

Assignment 1: Hotel Booking Prediction

1. Assignment Aims

The assignment is designed to assess your ability to work with noisy and inconsistent real-world data in the context of a practical machine learning classification problem. Its purpose is to provide you with an opportunity to showcase your practical skills in developing a real-world machine learning application.

2. Learning Outcomes:

- Analyse thoroughly current and emerging data-driven AI concepts and techniques and discuss their limitations.
    - Understand the issues associated in a realistic dataset, compared to an academic dataset.
- Critically evaluate the suitability and effectiveness of data-driven techniques in particular application areas
    - Demonstrate the skills and techniques required to convert noisy, inconsistent data into high-quality data required for good training and learning of a machine learning model.
    - Train a machine learning model on the processed data and produced acceptable prediction accuracy (more than 70%) of the model in a realistic setting.

3. Assessment Brief

You've been provided with a dataset containing booking information for two renowned hotels in Portugal: Resort hotel and City hotel. The hospitality industry is exceptionally time sensitive. Unlike businesses dealing in physical goods, hotels cannot sell their services tomorrow if they cannot do so today. Consequently, hotels strive to maintain a consistent stream of bookings throughout the year. However, a significant challenge in the hotel business is booking cancellations, as they result in substantial financial losses.

Your task is to work with a dataset spanning from July 2015 to August 2017, encompassing all the bookings made at these two hotels, and build a machine learning classifier that predicts whether a booking will be cancelled.

Given the realistic nature of this problem and the dataset's origin from the hotels' SQL database, inconsistencies and data quality issues are expected. Furthermore, the dataset includes over 30 features and more than 100,000 records. More information about the dataset could be access in the following paper.

https://www.sciencedirect.com/science/article/pii/S2352340918315191

Your responsibilities include:

Data Pre-processing:

- Gain a comprehensive understanding of the dataset through summarisation and basic data checks. At minimum, you should examine data types and identify missing values.
- Employ suitable techniques to address missing or inconsistent values, with clear justifications for your choices. Handle the outliers as well.
- Do the necessary data type conversions of columns.

Exploratory Data Analysis (EDA):

- Extract valuable insights from the dataset that could benefit hotel management, including the following at minimum:
    - Calculating cancellation percentages for City and Resort hotels.
    - Identifying the most frequently ordered meal types.
    - Determining the number of returning guests.
    - Finding the most booked room types.
    - Exploring correlations between room types and cancellations.

  For each of these tasks, choose a suitable type of visualisation covered in the practical sessions, such as:

    - Bar chart
    - Pie chart
    - Line chart
    - Heat map

Feature Engineering:

At minimum, you should:

- Clearly explain any decisions to drop or convert features from one format to another.
- Select a subset of relevant features (e.g. more informative) from a larger set of features to improve model performance.
- Consider binning as a method to simplify certain features and reduce variability.
- Evaluate whether scaling is necessary for your model.

Classifier Training:

Train a basic classifier on the pre-processed data, such as a decision tree or random forest. While all of the steps you followed previously should improve classification accuracy, for this assignment there are no strict requirement on excellent accuracy. Your model will be deemed accuracy as long as it attains more than 70% accuracy. Ensure you use stratified train-test splitting to maintain consistent class distribution in both datasets.

Feature Importance:

Determine which features carry more significance in your model's learning process.

Your work will not only involve building a predictive model but also ensuring that the data is cleaned, transformed, and explored to maximise the model's accuracy and value to hotel management.

Submission format:

The Jupyter notebook file must be submitted in both PDF and IPYNB formats. Do not submit any other formats, and do not attempt to compress the files. Doing either will result in marks being deducted for presentation.

4. Marking Scheme

Data Pre-processing: (25%)

- Missing values: (10%)
- Inconsistent values: (10%)
    - Outliers' analysis via Box-plot visualisation
- Column data type conversion (5%)

Exploratory Data Analysis: (25%)

- 5 Visualisations x 5% = 25%

Feature Engineering: (20%)

- Feature selection (5%)
- Binning (5%)
- Encoding (5%)
- Scaling (5%)

Classifier Training (20%)

- Data Splitting (5%)
- Model Training (10%)
- Model Evaluation (5%)

Feature Importance: (10%)

- Visualise and discuss important features (10%)

5. Grading Rubric

	HONOURS (FHEQ LEVEL 6)
90 +	**Outstanding demonstration of independent scholarly achievement and critical evaluation**. • well-structured assessment that comprehensively addresses the module learning outcomes and specific criteria • critical evaluation is evident through systematic, relevant, and comprehensive coverage of content • skilfully communicated in a style appropriate to the assessment brief • very limited areas for improvement

	HONOURS (FHEQ LEVEL 6)
80 +	**Exceptional demonstration of independent scholarly achievement and critical evaluation**. • well-structured assessment that addresses the learning outcomes and specific criteria for the module • critical evaluation is evident through systematic, relevant and comprehensive coverage of content • skilfully communicated in a style appropriate to the assessment brief
70 +	**Excellent demonstration of independent scholarly achievement and critical evaluation.** • well-structured assessment that addresses the learning outcomes and specific criteria for the module • critical evaluation is evident through systematic and relevant coverage of content • skilfully communicated in a style appropriate to the assessment brief
60 +	**Very good demonstration of independent scholarly achievement. and critical evaluation** • well-structured assessment that addresses the learning outcomes and specific criteria for the module • critical evaluation is evident through relevant coverage of content • clearly communicated in a style appropriate to the assessment brief
50 +	**Good demonstration of independent scholarly achievement and critical evaluation**. • fairly well-structured assessment that addresses the learning outcomes and specific criteria for the module • some critical evaluation is evident through coverage of content • good communication in a style appropriate to the assessment brief
40 +	**Adequate demonstration of independent scholarly achievement and critical evaluation.** • adequately structured assessment that addresses the learning outcomes and specific criteria for the module • some critical evaluation is evident through coverage of content which is also descriptive • communicates in a style appropriate to the assessment brief
30 +	**Limited demonstration of independent scholarly achievement and critical evaluation t.** • poorly structured assessment that does not completely address the learning outcomes and specific criteria for the module • work is descriptive in its coverage of the content • Poor communication that does not use a style appropriate to the assessment brief
20 +	**Minimal demonstration of independent scholarly achievement and critical evaluation.** • poorly structured assessment that only address a small part of the module learning outcomes and specific criteria for the module • work is descriptive in its coverage of the content, and in places may be inadequate • poor communication that does not use a style appropriate to the assessment brief

	HONOURS (FHEQ LEVEL 6)
10+	• poorly structured assessment that does not address the learning outcomes and specific criteria for the module • coverage of the content is inadequate or incomplete • poor communication that does not use a style appropriate to the assessment brief
0+	Poorly structured assessment that does not address at all the learning outcomes and specific criteria for the module