



Computer Science and Creative Technologies

Coursework Specification

Module Details

Module Code	UFCFMM-30-3
Module Title	Business Intelligence and Data Mining
Module Leader	Dr Kamran Soomro
Module Tutors	Dr Mahmoud Elbattah
Year	22-23
Component/Element Number	B
Total number of assessments for this module	2
Weighting	60%
Total Assignment Time	Approx 12 weeks
Element Description	Portfolio – Group Project

Dates

Date issued to students	31-01-2023
Date to be returned to students	06-2021
Submission Date	04-05-2023
Submission Place	Blackboard
Submission Time	14:00
Submission Notes	Each student should submit the final complete packages, including your notebook, data files, group contribution information, any presentations and README file. Specific instructions will be communicated in due time.

Feedback

Feedback provision will be	Written with interim feedback provided through Teams and in person.
-----------------------------------	---

Contents

Module Details	1
Dates.....	1
Feedback.....	1
Contents	2
Section 1: Overview of Assessment.....	2
Section 2: Task Specification	3
Task.....	3
Subtask1 Exploratory Data Analysis:	3
Subtask2 Customer Segmentation:	3
Subtask3 Build a Shopping Recommender System:.....	3
Subtask4 Evaluation:	3
Subtask 5 Sentiment Analysis:	4
Data	4
Comments	4
References	4
To get you started.....	4
Section 3: Deliverables	4
Section 4: Feedback mechanisms.....	5
Section 5: Marking Criteria	6

Section 1: Overview of Assessment

This assignment assesses the following module learning outcomes:

1. Investigate how data mining / machine learning is applied in a real-world business intelligence scenario – MO1.
2. Apply the concepts of machine learning and data mining in business intelligence applications and evaluate their effectiveness – MO2
3. Critically evaluate ethical, security and privacy-related considerations in business intelligence applications – MO3.

The assignment is worth **60%** of the overall mark for the module.

Broadly speaking, the assignment requires you to explore solutions to solve business intelligence and data mining problems. You will be given a dataset and a specific task. You will do research on recommender systems and related topics, identify the challenges and opportunities, implement, and evaluate your solutions. Furthermore, you will broaden your horizons with exploring cross-disciplinary approaches, which covers psychology and sociology.

The assignment is described in more detail in section 2.

This is a group assignment consisting of 3–4 members.

Working on this assignment will help you to practice problem solving skills, management skills, communication skills and team player skills. From business intelligence and data mining perspective, you will practice data processing/analysis, pattern discovery, as well as strategy identification and risk assessment.

If you have questions about this assignment, please use the Teams site to ask them.

Section 2: Task Specification

Task

Implement a shopping recommender system.

Subtask1 Exploratory Data Analysis:

Data exploration, pre-processing, and visualization.

In this task you will explore the data and try to understand it. For example, some of the things you can do are:

- Summarise the data and analyse its distribution. You can use techniques like scatter plots, box plots, histograms etc.
- Check for missing values and noise.
- Check if there's a timestamp data type that needs to be handled.
- Check for categorical data.
- Check if feature scaling is needed.
- Plot graphs or charts (or using other visualization techniques) to discover patterns.
- Check if other pre-processing steps are needed.

Subtask2 Customer Segmentation:

Customer behaviour refers to an individual's buying habits, including social trends, frequency patterns, and background factors influencing their decision to buy something. Businesses study customer behaviour to understand their target audience and create more-enticing products and service offers. A customer behaviour analysis is a qualitative and quantitative observation of how customers interact with your company. Customers should be segmented into buyer personas based on their common characteristics [1].

Subtask3 Build a Shopping Recommender System:

The purpose of recommender systems is to recommend products according to the user's preferences. A shopping recommender system should be able to identify the user's needs based on the user's shopping history and preferences. They may involve techniques such as collaborative filtering.

Subtask4 Evaluation:

Evaluate the recommender system that you have implemented using metrics such as Area Under the Curve (AUC), precision, recall, f1, etc [2]. You should carefully consider the various options available and choose appropriate evaluation metrics.

Marks will be given for critically interpreting the results. You should be able to justify the choice of evaluation metrics provide critical analysis.

Subtask 5 Sentiment Analysis:

Sentiment analysis is a powerful technique that can allow companies to computationally analyse customer reviews to understand the overall sentiment customers are expressing. This is a bonus task in which you can use sentiment analysis to extract additional intelligence about the dataset.

Data

The dataset consists of several files containing various attributes. You are expected to go through each file and understand and explore the dataset. There is a data schema provided with the dataset for your convenience. Most of the fields should be self-explanatory but if you do not understand something ask your tutor.

Comments

The intended end-users for this recommender system should be the customers.

References

[1] <https://blog.hubspot.com/service/customer-behavior-analysis>

[2] Schröder, G., Thiele, M. and Lehner, W., 2011, October. Setting goals and choosing metrics for recommender system evaluations. In UCERSTI2 workshop at the 5th ACM conference on recommender systems, Chicago, USA (Vol. 23, p. 53).

To get you started

- Do some research on the topics: e-commerce data analysis, purchasing behaviour, recommendation agent, unplanned purchase, user satisfaction (or related topics).
- Download the datasets from blackboard/assignments.
- Get familiar with the attributes and values of the dataset.
- Carry out the tasks.

Section 3: Deliverables

The final submission should consist of a zip file in the format *Group.zip* where *Group* is your group number according to Blackboard.

There will be only one submission per group.

The zip file should contain your notebook (.ipynb) file, any associated data files (.csv), a README file containing any special instructions for your code, as well as your group contribution document and presentation. **You may present directly from your notebook file if you so wish.** However, if you create a separate presentation, you **MUST** submit it along with the rest of your work.

The **notebook** should clearly list all the subtasks that you have performed and provide a narrative explaining your solution to each subtask, including justification and description followed by your interpretation of the results and any critical analysis. Well-explained and well-justified solutions will gain more marks.

Furthermore, you are also expected to conduct your own research on tools and techniques and cite them in your narrative.

The **group contribution document** is a rating of each member's contributions from 1–10, to be mutually agreed upon by the whole group.

You should also include a **README** file with any specific instructions on how to run your code.

AS A GROUP PROJECT, THE 5-DAY GRACE PERIOD DOES NOT APPLY TO THIS ASSESSMENT.

Section 4: Feedback mechanisms

Summative feedback will be communicated in written form Blackboard. Interim feedback will be communicated via the Teams site.

Section 5: Marking Criteria

Group Work (70%)	Completeness (60%)	0-35	36-45	46-55	56-65	66-75	76-100
		No tasks are fully completed.	Only Subtask 1 task is fully completed, the rest are either not or partially completed. The possible approaches and aspects were poorly explored.	Subtasks 1 and 2 completed fully, and the rest are either missing or partially completed. The possible approaches and aspects were not well explored.	Subtasks 1, 2 and 3 are completed fully, and the rest are either missing or partially completed. The possible approaches and aspects were explored to some extent.	Subtasks 1, 2, 3, and 4 are fully completed. The possible approaches and aspects were well explored.	All subtasks are fully completed. The possible approaches and aspects were thoroughly explored. The required tasks were completed with high standard.
	Clarity of narrative (10%)	Description of tasks is incomplete and unclear.	Description and justification of tasks is minimal with no critical insight and no citations.	Description and justification of tasks is adequate with some critical insight and minimal citations.	Description and justification of tasks is well-explained with fair amount of critical insight with good citations.	Description and justification of tasks is excellent with excellent critical insight.	Description and justification of tasks is outstanding with deep critical insight.
	Presentation (20%)	Poorly presented, misleading or lacking relevance.	The main parts of the contents were presented logically and clearly. Part of arguments are not supported by the contents.	All the contents were presented logically and clearly. Most of the arguments are supported by the contents.	All the contents were presented logically and clearly. All of the arguments are supported by the contents.	All the contents were presented logically and clearly. A strong connection is identified between the contents.	Outstanding presentation with novel ideas and critical insights.
	Group Behaviour (10%)	Team members never coordinates with other members.	Team members coordinate with other members minimally.	Team member coordinate with other members occasionally.	Team members coordinate with other members often.	Team members coordinate with other members often.	Team members coordinate with other members very often.
Individual Contribution (30%)	All group members must agree on the individual contribution of each member.						