MODULE HANDBOOK 2023-2024

Module Title: Data Mining and Predictive Modelling

Module Code:CT119-3-2

Year /Level:2

Credits:3

Co – requisite or pre-requisite module(s):Probability and Statistics

School:School of Computing

Semester:4

Contents

1.0 Introduction

Welcome to **Data Mining and Predictive Modelling**

This module introduces the students to data mining methods and models, including association rules, clustering, and predictive models. Students will be guided to work with datasets and apply their newly acquired data mining expertise to solving real problems using large, real-world data sets. The students will be exposed to employing data mining tools in the market to enhance their technical knowledge.

2.0 Module Team

Module Leader: Dr.Preethi Subramanian | preethi@apu.edu.my

Module Team

(a) Mr.Mafas Raheem | raheem@apu.edu.my

(b)

3.0 Module Learning Outcomes

Upon successful completion of this module, you will be able to:

CLO 1	Explain the basic concepts of data mining, modelling and analytical challenges of interpreting & presenting data (C2, PLO1)
CLO 2	Apply data mining techniques to produce a solution (C3, PLO2)
CLO 3	Perform analysis on the various data mining techniques for solving specific problems in real-world scenarios (A5, PLO9)

4.0 How will I learn on this module?

In this module you will attend a range of classes as well as studying independently and preparing for assessments. The plan below describes the work you will need to do to be successful in this module.

Week #	Weekly learning Outcomes	Topic Coverage	Hours	In class Learning Activities	Independent Learning Activities
1	Explain the tasks and functionalities of data mining Exhibit familiarity with the data mining software	Intro to Data Mining and Methodologies	2	Intro to the module and topic. Introduction to the data mining software used Lab1_Identification of data mining problem	Complete installation of the software and try the basic tasks

2	Discuss the need for exploratory data analysis Perform EDA on a given dataset using statistical and graphical methods	Exploratory Data Analysis	2	Lecture on EDA concepts. A simple Q&A will be conducted in class Briefing on the assessments Lab2_Graphical EDA	Students can try practice datasets for EDA
3	Explain the different forms of data preprocessing Apply the different types of data pre processing appropriately	Data Pre-processing	2	Lecture on concepts of data preprocessing. Case studies and identification of open source datasets Lab3_Statistical data exploration	Students can find a dataset for the assignment
4	Discuss the various data reduction strategies Apply data reduction on a given dataset	Data Pre-processing	2	Lecture on the various data reduction methods. Confirm the datasets chosen for the assignment by discussing with the lecturer Lab4_Data Preprocessing	Students can try to preprocess the assignment dataset
5	Explain about the different data mining methodologies	Data Mining Methodologies	2	Lecture on the data mining methodologies and the phases Quiz	Students can read on the methodologies and challenges in data mining projects

	Compare and contrast between the different methodologies				
6	Briefly summarize the methods used in data mining	Data Mining Methods	2	Lecture on data mining methods Progress Check 1	Students need to prepare for the progress check
7	Explain the concept of a predictive model and its potential benefit Apply tree-based modeling for classification and interpret the outcomes	Predictive Models	4	Lecture on tree based modeling. Demo on decision trees Lab 5_Tree based models	Students can read the given materials and text books about tree based models Practice on classification datasets
8	Differentiate between linear and logistic regression models Apply regression-based modeling to predict a dependent variable and interpret the outcomes	Predictive Models	4	Lecture on regression models Demo on Linear regression Lab 6_Regression Models	Students can read the given materials and text books about regression models Practice on prediction datasets
9	Explain the concept of a neural network and its benefits Apply NN based modeling to predict a	Predictive Models	4	Lecture on NN models Demo on NN and HPNN Lab 7_NN models	Students can read the given materials and text books about neural network models

					Practice on prediction datasets
10	Explain the different types of clustering algorithms Apply segmentation analysis on a given dataset	Clustering and Segmentation Analysis	4	Class Test: 2 hours Lecture on clustering and a gamification activity on clustering and segment profile Lab 8_Clustering	Prepare for the class test Apply the clustering principles on a given dataset
11	Explain how association analysis works on a given dataset	Association Analysis	4	Lecture on association analysis Demo on a transaction dataset Lab 9_Association	Apply association analysis on a given dataset and practice
12	Implement hybrid models such as semi-supervised models for a given dataset Apply ensemble models on a process flow with multiple predictive models	Hybrid Models	4	Demo on hybrid models Lab 10_Hybrid models	Apply hybrid models on a practice dataset
13	Solve the given data mining problem and discuss the results	Case Study	3	Explain the case study and the dataset given. Students are divided into groups	Brainstorming and implement the data mining concepts learned

				to solve the problem	
14	Solve the given data mining problem and discuss the results	Case Study	3	Explain the case study and the dataset given. Students are divided into groups to solve the problem	Brainstorming and implement the data mining concepts learned

5.0 Attendance and Absence

Attendance for all formal teaching is compulsory. On some occasions, for instance illness, your absence may be unavoidable. All absences must be notified to the relevant lecturer or Office immediately. Please refer to your Programme Handbook for contact details. The University needs to satisfy itself that you are engaged in your studies and will monitor your attendance at regular intervals. This is a particular requirement for international students but applies equally to all students. Details of when and how this will be undertaken will be given to you at orientation briefing.

6.0 Learning Resources

Provide a list of resources used by the module; book lists; key journals; software to be used.

Essential Readings:

Kotu,V., Deshpande, B (2019). Data Science: Concepts and practice. Cambridge, MA : Morgan Kaufmann Publishers. ISBN: 9780128147627.

Yang, X.S. (2019). Introduction to algorithms for data mining and machine learning. London, United Kingdom ; San Diego, CA, United States : Academic Press. ISBN: 9780128172179.

Sarma, K.S. (2017). Predictive Modelling with SAS Enterprise Miner. SAS Institute. ISBN: 9781635260380

Further Readings:

Key Journals:

Special Requirement (e.g: software, nursery, computer lab, simulation room, etc):SAS Enterprise Miner

7.0 Assessments

This module is assessed by:

Assessment Summary						
Form of Assessment	**Description**	**Duration (hour(s)**	**Hand out Date**	**Hand in Date**	**Marks Allocation**	**CLOs Assessed**
Final Assessment	Assignment		Week 2	Week 15	60%	CLO2, CLO3
Continuous Assessment	Test	2		Week 10	40%	CLO1

To pass a module, you must attempt every element of assessment and achieve at least 20% for each element and 50% in the module overall.

If you fail to pass the module, you will be required to re-sit any assessment components for which you did not pass. *Details on Regulation and Policies are available at* https://apiit.atlassian.net/wiki/spaces/RR/overview

8.0 Submission and Feedback

How do I submit my assessments?
Each module will have a different set of assessments and submission dates/times as stated in Section 7 above. It is your responsibility to be aware of the deadlines and to meet them.

You will be required to submit written assignments through Turnitin via Moodle. All submissions are automatically time-stamped by the system. No submission will be accepted by the system after the deadline.

Failure to meet any of the given deadline will be treated as a non-submission and a Grade Point 0 (zero) will be awarded for that component. This may result in failing the overall module. The only exceptions to these rules apply where a valid claim for extenuating circumstances can be made and is approved.

There may be occasions when you are unable to submit or undertake a piece of assessment due to circumstances beyond your control. APIIT has put in place a procedure for dealing with such extenuating circumstances. You can find more information in the Student Guide to Extenuating Circumstances which is available in the Policies and Procedures Space on APU's Knowledge Base , https://apiit.atlassian.net/wiki/spaces/PP/pages/2152267840/Extenuating+Circumstancess

How do I get feedback on my work?
You will normally receive feedback on all assessments, other than examinations, within 20 working days following the date of submission.
APIIT aims to release feedback within set weeks so that you can have as much of your feedback at once making the process less stressful.
Feedback will vary between modules; Feedback may also be received by forums, interviews, and individual feedback sessions. In the case of written coursework that is marked electronically, the lecturer will add commentary and feedback directly onto the electronic version of your written assignment, usually using a tool called GradeMark, which is integrated into Turnitin. You receive this feedback by going back to the same Turnitin link where you originally submitted your work and you will be able to view or download the feedback provided.

.
You will also be able to access your results via Moodle.

How can I give feedback on the module?
You are welcome to discuss your views with your lecturer on the module at any time. Views may also be expressed through your Programme Leader or via Programme Committee Meetings (PCM). During the course of the module, you will be encouraged to fill in the anonymous evaluation questionnaires to assist the University in its monitoring and planning. Such questionnaires are important for the benefit of your fellow and future students. We would be grateful for your full and prompt co-operation in completing them in a constructive and objective way. You will be guided on how to access the Online Student Survey System at: https://apiit.atlassian.net/l/c/6SKKYy1V.

Students who do not complete their student survey by the stipulated date indicated on their examination schedule (posted on the webspace) could find their results withheld until they complete their end semester survey.

9.0 Academic Integrity

This module requires that you demonstrate what you have learnt and that you have achieved the learning outcomes of the module. The University requires you to comply with the regulations on academic conduct. Academic misconduct includes but is not restricted to cheating in examinations, making - up data and plagiarism.

Plagiarism is the use of someone else's work (words, images, tables or ideas etc) without acknowledging the source. This includes materials from the internet as well as library books and the work of another person. Plagiarism is an assessment offence and any individual (who is suspected of plagiarism) will be referred to the University Academic Dishonesty Board. Please refer to Academic Dishonesty Policy at https://lms2.apiit.edu.my/course/view.php?id=750 for further information.

10.0 Module Descriptor

The module descriptor for this module is available on Moodle