Module Specification

Data Mining [TSI]

Version: 2024-25, v2.0, 20 Dec 2024

Contents

Part 1: Information

Module title: Data Mining [TSI]

Module code: UFCEA1-12-M

Level: Level 7

For implementation from: 2024-25

UWE credit rating: 12

ECTS credit rating: 6

College: College of Arts, Technology and Environment

School: CATE School of Computing and Creative Technologies

Partner institutions: Transport and Telecommunication Institute

Field: Computer Science and Creative Technologies

Module type: Module

Pre-requisites: None

Excluded combinations: None

Co-requisites: None

Continuing professional development: No

Professional, statutory or regulatory body requirements: None

Part 2: Description

Overview: This course introduces the core data mining concepts, techniques, algorithms, research issues and practical skills for applying data mining techniques to solve real-world problems. Topics include data understanding and visual data exploration, data preprocessing and transforming, data classification, data clustering etc.

Students will understand principles and concepts in data mining and get insight into data mining techniques and algorithms. Students will study the major data mining

problems as different types of computational tasks (prediction, classification, clustering, etc.) and the algorithms appropriate for addressing these tasks. Will learn how to analyze data through statistical and graphical summarization, supervised and unsupervised learning algorithms. Students learn how systematically evaluate data mining algorithms and understand how to choose algorithms for different analysis tasks. Students are expected to do independent reading of research papers and to do critical review.

Features: Not applicable

Educational aims: To provide students with practical and applied knowledge of how to conduct data mining activities. This includes key concepts in data mining as well as the statistical and modelling techniques necessary to analyse large data sets to generate meaningful intelligence.

Outline syllabus: This module will introduce and cover the following areas:
Introduction. Basic data mining tasks and techniques review. Data Mining Process (CRISP-DM).
Types of data: interval-scaled, binary, nominal, ordinal, mixed types
Data pre-processing: data cleaning, integration and transformation, dimensionality reduction.
Data Exploration and Visualisation using software .
Unsupervised and supervised learning algorithms review. Examples. Measuring Similarity between Objects.
Cluster analysis. Hierarchical algorithms.
 Partitional Methods. Distribution and density-based methods. Cluster analysis validity.
Discriminant analysis as predictive technologies. Canonical Analysis. Classification rule classification functions; Mahalanobis distance-based; Posterior probabilities. Stepwise discriminant analysis.
Computer intensive methods and application for testing classification quality. Main concepts of bootstrap method: the confidence interval estimation.
Essence of dimensionality reduction problem and different methods of its solution. Principal Components Analysis. Factor analysis.
Multiple regression: model building, transformation of variables, interpretation of results. Spatial and Temporal Data. Time Series. ARIMA. Short-term forecast of

transportation volume. ANOVA, MANOVA .

Business Intelligence and Data Mining.

Advanced Methods and Applications. Text Mining. Social Networks. Big Data.

Web-mining: content, usage, structure.

Part 3: Teaching and learning methods

Teaching and learning methods: Learning and teaching will be provided to students in several forms. Lectures will be used to discuss theoretical aspects of the module, like specific methods, measures, limitations, pitfalls etc. The lectures will be supported by presentations, self-reading materials etc. In addition a set of formative assessments will be used to enrich students experience in understanding of data mining methods.

While practical classes are organized as tutoring sessions with teaching staff to develop practical skills on application of data mining methods for specific dataset analysis.

The assessment items, consist from examination, which shall test theoretical aspects of the module, labs are targeted to test gained practical skills, while written assignment would assess research competences on exploring information, critical analysis and discussion. All assessments are subject of individual completion, as module is targeted on development individual skills and competences.

Module Learning outcomes: On successful completion of this module students will achieve the following learning outcomes.

MO1 Apply effectively appropriate data analytics and statistical techniques on available data to discover new relations and deliver insights into research problem or organisational processes and support decision-making.

MO2 Verify data quality and veracity, to recognise value of data, to apply analytics and statistics methods for data preparation, pre-processing.

MO3 Utilise effective visualisation and storytelling methods to create data analytics reports.

Hours to be allocated: 120

Contact hours:

Independent study/self-guided study = 112 hours

Face-to-face learning = 48 hours

Reading list: The reading list for this module can be accessed at readinglists.uwe.ac.uk via the following link https://rl.talis.com/3/uwe/lists/48377AAB-25C5-0850-1028-7D4AE165ACBE.html?lang=en-gb&login=1

Part 4: Assessment

Assessment strategy: The assessment strategy for this module consists of two items, each is subject of individual completion.

Written Examination, which assesses gained theoretical knowledge in the frame of the module.

Portfolio consisting of 4 written reports – students will be provided with a list of topics, they are required to produce a report for each of their selected topic.

For the resit exam, a different version of the exam paper will be used.   For the resit report students will work on a different topic.

Assessment tasks:

Examination (First Sit)
Description: Exam (2 hours).
Weighting: 50 %
Final assessment: Yes
Group work: No
Learning outcomes tested: MO1, MO2

Portfolio (First Sit)

Description: A portfolio of 4 written reports (upper word count of 1200 words)

Weighting: 50 %

Final assessment: No

Group work: No

Learning outcomes tested: MO1, MO2, MO3

Examination (Resit)

Description: Exam (2 hours).

Weighting: 50 %

Final assessment: Yes

Group work: No

Learning outcomes tested: MO1, MO2

Portfolio (Resit)

Description: A portfolio of 4 written reports (upper word count of 1200 words)

Weighting: 50 %

Final assessment: No

Group work: No

Learning outcomes tested: MO1, MO2, MO3

Part 5: Contributes towards

This module contributes towards the following programmes of study:

Computer Science (Data Analytics and Artificial Intelligence) {Double Degree} [TSI] MSc 2024-25