# Data mining 781 (COS 781)

| | |
|---|---|
| **Qualification** | Postgraduate |
| **Faculty** | Faculty of Engineering, Built Environment and Information Technology |
| **Module credits** | 15.00 |
| **NQF Level** | 08 |
| **Programmes** | BScHons *Computer Science* |
| **Prerequisites** | No prerequisites. |
| **Contact time** | 2 lectures per week |
| **Language of tuition** | Module is presented in English |
| **Department** | Computer Science |
| **Period of presentation** | Semester 1 or Semester 2 |

**Module content**

Data mining is the analysis of (often large) observational datasets to find unsuspected relationships and summarise the data in novel ways that are both understandable and useful to the data owner(s). The analysis methods fall into two categories: Computational data mining and Statistical data mining. Computational methods originate from Machine Learning, which is a branch of Computer Science (Artificial Intelligence). Statistical methods originate from a branch of Statistics called Statistical pattern recognition. Observational data is data that was collected for some other purpose, e.g. banking data for loan applications and repayments, and is then used for analysis to determine good borrowers and risky borrowers. The objectives of the module are: to introduce the commonly used data mining methods, and to enable the student to acquire practical data mining skills. The module covers Computational and Statistical data mining methods as well as the commonly used process models for data mining projects. The topics covered include: process models (CRISP-DM and SEMMA), exploratory data analysis (univariate and bivariate), dimensionality reduction (feature selection, principal components analysis), descriptive modelling (cluster analysis and association rules), predictive modelling (decision trees, neural networks, K-nearest neighbour, Naive Bayes, ensemble models), statistical modelling (linear and logistic regression) and text mining. It is assumed that students have a basic knowledge of Statistics. It is also highly recommended that students do COS 710 and COS 711, as knowledge of the content of these modules is assumed.

**General Academic Regulations and Student Rules**
The General Academic Regulations (G Regulations) and General Student Rules apply to all faculties and registered students of the University, as well as all prospective students who have accepted an offer of a place at the University of Pretoria. On registering for a programme, the student bears the responsibility of ensuring that they familiarise themselves with the General Academic Regulations applicable to their registration, as well as the

relevant faculty-specific and programme-specific regulations and information as stipulated in the relevant yearbook. Ignorance concerning these regulations will not be accepted as an excuse for any transgression, or basis for an exception to any of the aforementioned regulations. The G Regulations are updated annually and may be amended after the publication of this information.

## Regulations, degree requirements and information

The faculty regulations, information on and requirements for the degrees published here are subject to change and may be amended after the publication of this information.

## University of Pretoria Programme Qualification Mix (PQM) verification project

The higher education sector has undergone an extensive alignment to the Higher Education Qualification Sub-Framework (HEQSF) across all institutions in South Africa. In order to comply with the HEQSF, all institutions are legally required to participate in a national initiative led by regulatory bodies such as the Department of Higher Education and Training (DHET), the Council on Higher Education (CHE), and the South African Qualifications Authority (SAQA). The University of Pretoria is presently engaged in an ongoing effort to align its qualifications and programmes with the HEQSF criteria. Current and prospective students should take note that changes to UP qualification and programme names, may occur as a result of the HEQSF initiative. Students are advised to contact their faculties if they have any questions.