

---

# University of Pretoria Yearbook 2022

---

## Data mining 781 (COS 781)

<b>Qualification</b>	Postgraduate
<b>Faculty</b>	<a href="#">Faculty of Engineering, Built Environment and Information Technology</a>
<b>Module credits</b>	15.00
<b>NQF Level</b>	08
<b>Programmes</b>	<a href="#">BSchHons Computer Science</a>
<b>Prerequisites</b>	No prerequisites.
<b>Contact time</b>	2 lectures per week
<b>Language of tuition</b>	Module is presented in English
<b>Department</b>	Computer Science
<b>Period of presentation</b>	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.

---

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

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.