# Subject Description Form

<table>
<thead>
<tr>
<th>Subject Code</th>
<th>COMP5121</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Title</td>
<td>Data Mining and Data Warehousing Applications</td>
</tr>
<tr>
<td>Credit Value</td>
<td>3</td>
</tr>
<tr>
<td>Level</td>
<td>5</td>
</tr>
<tr>
<td>Pre-requisite/ Exclusion</td>
<td>Nil</td>
</tr>
</tbody>
</table>

## Objectives

The objectives of this subject are to enable students to:

1. make more effective use of data stored in databases;
2. create a clean, consistent repository of data within a data warehouse;
3. utilize various levels and types of summarization of data to support management decision making;
4. discover patterns and knowledge that is embedded in the data using different data mining techniques.

## Intended Learning Outcomes

Upon completion of the subject, students will be able to:

a) understand the need for data warehouse;
b) identify components in typical data warehouse architecture;c) design a data warehouse in support of business problem solving;d) understand typical knowledge discovery process and the different algorithms available by popular commercial data mining software; and

## Subject Synopsis/Indicative Syllabus

- Introduction to data warehousing and data mining; possible application areas in business and finance; definitions and terminologies; types of data mining problems.
- Data warehouse and data warehousing; data warehouse and the industry; definitions; operational databases vs. data warehouses.
- Data warehouse architecture and design; two-tier and three-tier architecture; star schema and snowflake schema; data characteristics; static and dynamic data; meta-data; data marts.
- Data replication, data capturing and indexing, data transformation and cleansing; replicated data and derived data; Online Analytical Processing (OLAP); multidimensional databases; data cube.
- Data Mining and knowledge discovery, the data mining lifecycle; pre-processing; data transformation; types of problems and applications.
- Mining of Association Rules; the Apriori algorithm; binary, quantitative and generalized association rules; interestingness measures.
- Classification; decision tree based algorithms; Bayesian approach; statistical approaches, nearest neighbor approach; neural network based approach; Genetic Algorithms based
technique; evaluation of classification model.

- Clustering; k-means algorithm; Hierarchical algorithm; Condorset; neural network and Genetic Algorithms based approach; evaluation of effectiveness.
- Sequential data mining; time dependent data and temporal data; time series analysis; sub-sequence matching; classification and clustering of temporal data; prediction.
- Computational intelligence techniques; fuzzy logic, genetic algorithms and neural networks for data mining.

Teaching/Learning Methodology

Class activities including - lecture, tutorial, lab, workshop seminar where applicable
A mix of lectures, discussions and case study analysis.

Assessment Methods in Alignment with Intended Learning Outcomes

<table>
<thead>
<tr>
<th>Specific Assessment Methods/Tasks</th>
<th>% weighting</th>
<th>Intended subject learning outcomes to be assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a  b  c  d  e</td>
</tr>
<tr>
<td>Assignments, Tests &amp; Projects</td>
<td>55</td>
<td>✓     ✓     ✓     ✓</td>
</tr>
<tr>
<td>Final Examination</td>
<td>45</td>
<td>✓     ✓     ✓     ✓</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>✓     ✓     ✓     ✓</td>
</tr>
</tbody>
</table>

Student study effort expected

Class Contact:
Class activities (lecture, tutorial, lab) 39 hours

Other student study effort:
Assignments, Quizzes, Projects, Exams 65 hours

Total student study effort 104 hours

Reading list and references

5. Kovalerchuk, B., 2013, Data Mining in Finance: Advances in Relational and Hybrid Methods, Springer.