Subject Description Form

Subject Code	COMP5121
Subject Title	Data Mining and Data Warehousing Applications
Credit Value	3
Level	5
Pre-requisite/ Exclusion	Nil
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) acquire deep understanding on the need for data warehouse in real applications; b) master the typical data warehouse architectures, and be able to analyze and identify the components of data warehouse architecture; c) be able to apply and design a data warehouse in support of business problem solving in real applications; d) develop sophisticated insights on typical knowledge discovery process and the different algorithms available by popular commercial data mining software; and e) obtain hands-on experience with some popular data mining software.
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.

<table>
<tr><td></td><td>

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.

</td></tr>
<tr><td>Teaching/Learning Methodology</td><td>Class activities including - lecture, tutorial, lab, workshop seminar where applicable
A mix of lectures, discussions and case study analysis.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td rowspan="2">Specific Assessment Methods/Tasks</td><td rowspan="2">% weighting</td><td colspan="5">Intended subject learning outcomes to be assessed</td></tr>
<tr><td>a</td><td>b</td><td>c</td><td>d</td><td>e</td></tr>
<tr><td>Assignments, Tests & Projects</td><td>55</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Final Examination</td><td>45</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>Total</td><td>100</td><td colspan="5"></td></tr>
</table>
</td></tr>
<tr><td rowspan="4">Student study effort expected</td><td colspan="2">Class Contact:</td></tr>
<tr><td>Class activities (lecture, tutorial, lab)</td><td>39 hours</td></tr>
<tr><td colspan="2">Other student study effort:</td></tr>
<tr><td>Assignments, Quizzes, Projects, Exams</td><td>66 hours</td></tr>
<tr><td>Reading list and references</td><td colspan="2">
(1) Han, J., and Kamber, M., 2011, Data Mining: Concepts and Techniques, 3rd Ed., Morgan Kaufmann, San Francisco, CA.
(2) Tan, P.N., Steinbach, M., Kumar V., 2014, Introduction to Data Mining, 2nd Ed, Addison Wesley.
(3) Liu, B., 2011, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2nd Ed, Springer.
(4) Golfarelli, M., Rizzi, S., 2009, Data Warehouse Design: Modern Principles and Methodologies, 1st Ed, McGraw-Hill.
(5) Kovalerchuk, B., 2013, Data Mining in Finance: Advances in Relational and Hybrid Methods, Springer.
</td></tr>
</table>