Appendix A: Course Syllabus

Credits	3
Number	CIS501
Title	Data Mining: Finding the Data and Models that Create Value
Pre-requisites	Students are expected to have background knowledge in mathematics and computer science/information technology. In particular, students will need to have a very good grasp of the following areas or topics: 1. Probability and statistics 2. Software development in at least one general purpose programming language (and preferably more) 3. Algorithms
Co-requisites	None
Catalogue Text	Introduction to a class of methods known as data mining or machine learning that assist managers in recognizing patterns and making intelligent use of massive amounts of electronic data collected via the internet, e-commerce, electronic banking, point-of-sale devices, bar-code readers, and intelligent machines. Topics selected from logistic regression; association rules; tree-structured classification and regression; cluster analysis; discriminant analysis; and neural network methods. Examples of successful applications in areas such as credit ratings, fraud detection, database marketing, customer relationship management, investments, and logistics are covered. Introduction to data-mining software.

Learning Outcomes

Academic knowledge: Students will be able to demonstrate knowledge and understanding of:

1. The strengths and limitations of popular data mining techniques

Intellectual skills: Students will be able to:

2. Identify business applications of data mining techniques
3. Identify areas of future research in data mining systems

Subject practical skills: Upon completion, students will be able to do the following:

4. Perform powerful data analysis using existing software packages

Transferable skills: Upon completion, students will be able to:

5. Communicate effectively about technical issues, in particular data mining related topics.

Relationship of course objectives to program outcomes	
Program	Use and apply current technical concepts and practices in core computing

Outcome 1	and information technologies.
Program Outcome 2	Analyze a problem, and identify and define the computing requirements appropriate to its solution.
Program Outcome 4	Communicate effectively with a range of audiences.

Syllabus

Approximate breakdown by week:

1. Data Mining Overview,
2. Prediction and Classification with k-Nearest Neighbors
3. Classification and Bayes Rule, Naïve Bayes
4. Decision Trees
5. Decision Trees II
6. Discriminant Analysis
7. Logistic Regression Case
8. **Mid semester break**
9. Neural Networks
10. Neural Networks II
11. Regularization and training algorithms.
12. k-Means clustering, hierarchical clustering
13. Visualization, Principal Components
14. Association rules, recommendation systems: collaborative filtering
15. Final project presentations
16. **Final examination**

Assessment

1. Assignments and class participation (20%)
2. Exams (midterm and final) (40%)
3. Project work (40%)

Course Texts

- Bramer, Max. *Principles of Data Mining*. Springer, 2007.
- Hand, Mannila, and Smyth. *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.
- Krzysztof J. Cios , Witold Pedrycz , Roman W. Swiniarski , Lukasz A. Kurgan . *Data Mining: A Knowledge Discovery Approach*. Springer, 2007.

Additional Reading

- Liu, Bing. Web Data Mining (2nd Ed). Springer, 2011. (Mainly “Part 1: Data Mining Foundations”)
- Shmueli, Patel, and Bruce. *Data Mining for Business Intelligence.* Wiley-Interscience, 2006.
- Berry and Linoff. *Mastering Data Mining*. New York, NY: Wiley, 2000.
- Daniel Larose. *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley-Interscience, 2005.
- Thomas Runkler, *Data Analytics: Models and Algorithms for Intelligent Data Analysis,* Springer Vieweg, 2012.