## Knight Foundation School of Computing and Information Sciences

**Course Title:** Introduction to Data Mining **Date:** 11/11/19

**Course Number:** CAP 4770

**Number of Credits:** 3

| | |
|---|---|
| **Subject Area:** Computer Systems | **Subject Area Coordinator:** Leonardo Bobadilla<br>**email:** bobadilla@cs.fiu.edu |
| **Catalog Description:** Data mining applications, data preparation, data reduction and various data mining techniques such as association, clustering, classification, anomaly detection. | |
| **Textbook:** Data Mining: Concepts and Techniques, 2nd Edition.<br>Jiawei Han and Micheline Kamber,<br>Morgan Kaufmann Publishers (ISBN: 1-55860-901-6) | |
| **References:** Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, 2nd Edition.<br>Ian H. Witten and Eibe Frank<br>Morgan Kaufmann Publishers (ISBN: 0-12-088407-0) | |
| **Prerequisites Courses:** COP 3530 | |
| **Corequisites Courses:** COP 4710 | |

Type: Elective for CS (Applications group).

Prerequisites Topics:

- Fundamental data structures
- Basic database concepts
- Discrete math
- Programming languages

Course Outcomes:

1. Be familiar with the data mining process
2. Be familiar with various data mining applications
3. Be familiar with major data mining tasks and algorithms
4. Exposure to the principles and theoretical foundations behind major data mining approaches
5. Be able to undertake the systematic and comparative evaluation of data mining procedures

6. Demonstrate a basic understanding of selecting and applying data mining techniques to the solutions of real world problems

## Relationship between Course Outcomes and Program Outcomes

| BS in CS: Program Outcomes | Course Outcomes |
|---|---|
| a) Demonstrate proficiency in the foundation areas of Computer Science including mathematics, discrete structures, logic and the theory of algorithms | 3,4 |
| b) Demonstrate proficiency in various areas of Computer Science including data structures and algorithms, concepts of programming languages and computer systems. | 1, 3, 4, 5,6 |
| c) Demonstrate proficiency in problem solving and application of software engineering techniques | 2,5,6 |
| d) Demonstrate mastery of at least one modern programming language and proficiency in at least one other. | |
| e) Demonstrate understanding of the social and ethical concerns of the practicing computer scientist. | 2,6 |
| f) Demonstrate the ability to work cooperatively in teams. | |
| g) Demonstrate effective communication skills. | |

## Assessment Plan for the Course & how Data in the Course are used to assess Program Outcomes

Student and Instructor Course Outcome Surveys are administered at the conclusion of each offering, and are evaluated as described in the School's Assessment Plan: https://abet.cs.fiu.edu/csassessment/

Knight Foundation School of Computing and Information Sciences
CAP 4770
Introduction to Data Mining

## Outline

| Topic | Number of Lecture Hours | Outcome |
|---|---|---|
| • Data Mining Introduction<br>o What is data mining<br>o Data mining functionalities<br>o Data mining applications<br>o Data mining systems | 2 | 1,2,3,6 |
| • Data Preparation<br>o Descriptive data summarization<br>o Data cleaning<br>o Data integration and transformation<br>o Data reduction<br>o Data discretization and concept hierarchy generation | 5 | 1,6 |
| • Data Warehouse and OLAP<br>o Data warehouse introduction<br>o Multidimensional data model<br>o Data cube and OLAP | 3 | 1,2,4 |
| • Frequent Pattern Mining<br>o Basic concepts<br>o Efficient and scalable frequent itemset mining methods<br>o Correlation analysis<br>o Sequential pattern mining<br>o Graph and tree mining | 8 | 3,4,5,6 |
| • Classification and Prediction<br>o Decision tree induction<br>o Bayesian classification<br>o Support vector machines<br>o Nearest neighbor methods<br>o Other classification methods<br>o Ensemble methods<br>o Performance evaluation | 10 | 3,4,5,6 |
| • Clustering Analysis<br>o Partitioning methods<br>o Hierarchical methods<br>o Density-based methods<br>o Outlier analysis | 6 | 3,4,5,6 |
| • Advanced topics and applications | 3 | 2,3,4,6 |

Knight Foundation School of Computing and Information Sciences
CAP 4770
Introduction to Data Mining

## Course Outcomes Emphasized in Laboratory Projects / Assignments

| | Outcome | Number of Weeks |
|---|---|---|
| 1 | Association rule mining<br><br>Outcome:1,2,3,5,6 | 3 |
| 2 | Classification<br><br>Outcomes: 1,2,3,5,6 | 3 |
| 3 | Clustering<br><br>Outcomes: 1,2,3,5,6 | 3 |

## Oral and Written Communication

No significant coverage

| Written Reports | | Oral Presentations | |
|---|---|---|---|
| Number Required | Approx. Number of pages | Number Required | Approx. Time for each |
| 3 | 5 | 0 | 0 |

## Social and Ethical Implications of Computing Topics

No significant coverage

| Topic | Class time | student performance measures |
|---|---|---|
| | | |

## Approximate number of credit hours devoted to fundamental CS topics

| Fundamental CS Area | Core Hours | Advanced Hours |
|---|---|---|
| **Algorithms:** | | **3.0** |
| **Software Design:** | | |
| **Computer Organization and Architecture:** | | |
| **Data Structures:** | | |
| **Concepts of Programming Languages** | | |

Knight Foundation School of Computing and Information Sciences
CAP 4770
Introduction to Data Mining

## Theoretical Contents

| Topic | Class time |
|---|---|
| Statistics, Information theory, and Liner algebra | 2.0 |

## Problem Analysis Experiences

1. Data mining application: Identify a problem that would be amenable to data mining

## Solution Design Experiences

1. Association rule mining
2. Classification
3. Clustering analysis

## The Coverage of Knowledge Units within Computer Science Body of Knowledge[1]

| Knowledge Unit | Topic | Lecture Hours |
|---|---|---|
| IM2 | Database systems | 1 |
| IM3 | Data modeling | 2 |
| DS6 | Discrete probability | 1 |
| IM10 | Data Mining | 35 |

[1]See https://www.acm.org/binaries/content/assets/education/cs2013_web_final.pdf for a description of Computer Science Knowledge units