About this Course

Once called “knowledge discovery in databases,” advances in processing power and speed over the last decade have allowed users to move beyond manual, tedious, and time-consuming practices to quick, easy data analysis that harnesses the power of machine learning and high-performance computing. This course will introduce you to the fundamentals of data mining and pattern recognition. You will gain a deeper understanding of data through hands-on experience in the topic areas of big data analysis, classification, clustering, and association rule mining. Advanced topics such as reinforcement learning, deep learning, transfer learning and Deep Mind for Google will also be covered. By the end of the course, you will be able to apply state of the art data mining technology to real world applications, analyze and compare competing techniques, and design optimal solutions for a given set of application driven constraints.

Specific topics covered include:

- Data Mining Fundamentals
- Machine Learning
- Data Collection
- Deep Learning
- Data Visualization
- Reinforcement Learning
- Data Mining Algorithms

Required Prior Knowledge and Skills

- Intermediate understanding of core concepts of data mining
- Basics of statistics
- Programming (language such as Python or MATLAB)
Learning Outcomes

Learners completing this course will be able to:

• Differentiate among major data mining techniques such as classification, cluster analysis, and association rule mining
• Apply common data mining algorithms to discover relationships and patterns in large datasets
• Implement more advanced learning algorithms such as deep learning and reinforcement learning
• Utilize open source tools to build a data mining project to solve a specific problem

Projects

• Project 1: Activity Recognition
• Project 2: User Dependent Analysis

Course Content

Instruction

• Video lectures
• Office hours
• Live sessions with instructional team
• Knowledge checks
• Practice quizzes
• Discussion questions

Assessment

• Graded quizzes (auto-graded)
• Individual project(s) (instructor-graded)
• Midterm exam (auto-graded)
• Final exam (auto-graded)

Estimated Workload/Time Commitment Per Week

Approximately 20 hours per week.

Technology Requirements

Hardware

• Standard with major OS

Software and Other

• Standard – technology integrations will be provided through Coursera
Course Outline

**Unit 1: Core Concepts of Data Mining**

**Learning Objectives**
1.1 Explain the history and purpose of data mining across multiple disciplines
1.2 Differentiate what is and what is not data mining
1.3 Describe different data mining tasks
1.4 Recognize attributes of data needed for data mining
1.5 Review and summarize data exploration techniques for use in initial data analysis
   - Welcome and Start Here
   - Module 1: History and Purpose of Data Mining
   - Module 2: Data Attributes Needed for Data Mining
   - Module 3: Review of Initial Data Exploration Techniques
   - Assignment: Activity Recognition Direction

**Unit 2: Existing Techniques: Classification**

**Learning Objectives**
2.1 Define classification and classification applications
2.2 Compare and contrast common classification techniques
2.3 Apply common algorithms used in data mining
   - Module 1: Introduction to Classification
   - Module 2: Introduction to Classification Tasks
   - Module 3: Classification Issues

**Unit 3: Alternative Classification Techniques**

**Learning Objectives**
3.1 Define Instance Based Classifiers
3.2 Use the basics of probability theory to calculate the Bayes Classifier
3.3 Use the probability estimation to calculate Naive Bayes Classifier
3.4 Recognize the basic structure of Neural Networks
3.5 Identify the Perceptron learning algorithm
3.6 Recall the Artificial Neural Networks learning model
3.7 Explain the underlying concepts behind support vector machines and why they work
Module 1: Alternative Data Mining Techniques
Module 2: Artificial Neural Networks
Module 3: Support Vector Machines

Unit 4: Clustering

Learning Objectives
4.1 Define cluster analysis
4.2 Differentiate what is and what is not cluster analysis
4.3 Categorize different types of clusters
4.4 Use common algorithmic measures to evaluate clusters
4.5 Analyze DB Scan in relation to other clustering methods
   - Module 1: K Means Clustering
   - Module 2: Hierarchical Clustering
   - Module 3: Cluster Validity
   - Module 4: DBSCAN

Unit 5: Association Rule Mining

Learning Objectives
5.1 Apply the Mining Association Rules to discover relationships in large datasets
5.2 Use inferencing techniques to analyze association rule analysis results
5.3 Identify ways to reduce the computational complexity of frequent itemset generation
5.4 Describe how to efficiently generate rules from frequent datasets
   - Module 1: Introduction to Basic Concepts of Association Rule Mining
   - Module 2: Apriori Principle

Unit 6: Big Data Tools

Learning Objectives
6.1 Describe components that comprise deep learning
6.2 Implement a deep neural network using common tools such as Keras or Theano
6.3 Describe the structure and usage of Restricted Boltzmann Machines
6.4 Design Restricted Boltzman Machine algorithm to create a movie recommendation application
6.5 Describe the structure of deep autoencoders, and describe different application scenarios where they can be used
6.6 Apply deep autoencoders to on a sample data to derive low dimensional representations
6.7 Compare open source tools that allow for fast implementation of data mining tasks

- Module 1: Deep Learning Introduction

**Unit 7: Reinforcement Learning**

**Learning Objectives**

7.1 Describe agents, environments, states, actions and rewards that comprise reinforcement learning
7.2 Describe a Markov decision process and how it is different from Markov chains
7.3 Describe the difference between an MDP learning and reinforcement learning
7.4 Describe usage of reinforcement learning in automatically solving Atari games
7.5 Describe slate MDP and its difference from MDP
7.6 Describe the usage of attention in Reinforcement Learning
7.7 Describe the usage of Reinforcement Learning in a commercially used recommendation system (Deep Mind)
7.8 Describe multiple ways to solve a Reinforcement Learning problem
   - Module 1: Reinforcement Learning
   - Module 2: Markov Decision Process
   - Module 3: Solving Reinforcement Learning Problems

**Unit 8: Course Wrap-Up**

**Learning Objectives**

8.1 Complete the final exam
8.2 Optional: Complete and submit your Portfolio Inclusion Report
**About ASU**

Established in Tempe in 1885, Arizona State University (ASU) has developed a new model for the American Research University, creating an institution that is committed to access, excellence and impact.

As the prototype for a New American University, ASU pursues research that contributes to the public good, and ASU assumes major responsibility for the economic, social and cultural vitality of the communities that surround it. Recognizing the university’s groundbreaking initiatives, partnerships, programs and research, U.S. News and World Report has named ASU as the most innovative university all three years it has had the category.

The innovation ranking is due at least in part to a more than 80 percent improvement in ASU’s graduation rate in the past 15 years, the fact that ASU is the fastest-growing research university in the country and the emphasis on inclusion and student success that has led to more than 50 percent of the school’s in-state freshman coming from minority backgrounds.

**About Ira A. Fulton Schools of Engineering**

Structured around grand challenges and improving the quality of life on a global scale, the Ira A. Fulton Schools of Engineering at Arizona State University integrates traditionally separate disciplines and supports collaborative research in the multidisciplinary areas of biological and health systems; sustainable engineering and the built environment; matter, transport and energy; and computing and decision systems. As the largest engineering program in the United States, students can pursue their educational and career goals through 25 undergraduate degrees or 39 graduate programs and rich experiential education offerings. The Fulton Schools are dedicated to engineering programs that combine a strong core foundation with top faculty and a reputation for graduating students who are aggressively recruited by top companies or become superior candidates for graduate studies in medicine, law, engineering and science.

**About the School of Computing, Informatics, & Decision Systems Engineering**

The School of Computing, Informatics, and Decision Systems Engineering advances developments and innovation in artificial intelligence, big data, cybersecurity and digital forensics, and software engineering. Our faculty are winning prestigious honors in professional societies, resulting in leadership of renowned research centers in homeland security operational efficiency, data engineering, and cybersecurity and digital forensics. The school’s rapid growth of student enrollment isn’t limited to the number of students at ASU’s Tempe and Polytechnic campuses as it continues to lead in online education. In addition to the Online Master of Computer Science, the school also offers an Online Bachelor of Science in Software Engineering, and the first four-year, completely online Bachelor of Science in Engineering program in engineering management.
Dr. Banerjee is an Assistant Research Professor at School of Computing Informatics and Decision Systems Engineering, Arizona State University. His research interests include pervasive computing in healthcare and analysis, safety verification of embedded system software. Dr. Banerjee currently focuses on data driven analysis and modeling in many different domains including diet monitoring, gesture recognition, and biological process modeling. He works closely with government agencies such as Food and Drug Administration and medical agencies such as Mayo Clinic. Dr. Banerjee is also interested in hybrid system-based modeling and safety verification of closed loop control systems which interact with the physical environment, also known as Cyber-Physical Systems. In addition, his work includes developing management algorithms for sustainable data centers using renewable sources of energy.