CPIS-490 Syllabus

Catalog Description

CPIS-490 Selected Topics in IS
Credit: 3 (Theory: 3, Lab: 0, Practical: 1)
Prerequisite: None
Classification: Elective

The objective of this course is to explore selected topics about the latest advancements in the field of Information Systems (topics determined by the Council of the Information Systems Department).

Class Schedule

Meet 50 minutes 3 times/week or 80 minutes 2 times/week

Lab/Tutorial 90 minutes 1 times/week

Grade Distribution

Week	Assessment	Grade %
4	Graded Lab Work	15
7	Exam 1	15
8	Homework Assignments 1	5
12	Exam 2	15
12	Homework Assignments 2	10
13	Group Project	10
15	Comprehensive Final Exam	30

Topics Coverage Durations

Topics	Weeks
Introduction to big data analytics and data science (Module 1)	1
Current Practices, the role of a data scientist, application contexts (Module 1 and 2)	1
Analytics Lifecycle (Module 2)	1
Introduction to R (Module 3)	1
Building and Evaluating statistical models (Module 3)	2
Introduction to Analytics Methods: K-Means clustering (Module 4)	1
Apriori and Association Rules (Module 4)	1
Linear Regression (Module 4)	1
Logistic Regression and Analytics evaluation (Module 4)	1
Naïve Bayes Classification and Decision trees (Module 4)	1
Introduction to Time series analysis and Text Analysis (Module 4)	1
Text Analysis using Information Retrieval and the place of Natural Language Processing (NLP) (Module 4)	1
Technology & Tools (Hadoop, Pig, Hive) and the Big	1

Last Articulated

December 24, 2017

Relationship to Student Outcomes

a	b	c	d	e	f	g	h	i	j
x	x	x						x	x

Course Learning Outcomes (CLO)

By completion of the course the students should be able to

1. Identify the reasons for the growth of data and data types, and the need for different tools (like R) and approaches for its address (i)
2. Define the difference between data science and BI, fundamental concepts, architectures and roles pertaining to each in different industry verticals (i)
3. **Describe in detail what happens in each part of the analytics lifecycle in relation to the other including the deliverables of each (j)**
4. Develop R-code that accepts input and gives output, does data manipulation and plotting using R's data types and functions (c)
5. **Demonstrate how to model data using spread, centrality, probability density, visualization and analysis of normal and other distributions (j)**
6. Identify the role of statistical analysis in the analytics lifecycle and the different types of questions (e.g inferential and descriptive) (a)
7. **Demonstrate techniques for turning questions into hypothesis, testing them, evaluating the test, and then answering the original question (b)**
8. **Define fundamental concepts-in and types-of machine learning including classification and clustering techniques with use cases (j)**
9. Examine a k-means clustering problem to explain how to pick k, address assumptions and limitations (i)
10. Describe the association rules and priority properties their use cases (i)
11. **Examine a regression problem to identify the type of regression, implication of coefficients and how to evaluate it using residuals (j)**
12. Interpret coefficients in logistic regression and explain how it can be optimised by the ROC curve (a)
13. Describe bayes and decision-tree classification, and time-series analysis, with their use cases and limitations, and demonstrate them in R (j)
14. Demonstrate how questions can be addressed about text using retrieval, classification and linguistic analysis after data cleaning (j)
15. Describe use cases for hadoop/map-reduce, and demonstrate their use in the context of pig, hive, R and the

CPIS-490 Syllabus

Coordinator(s)

Dr. Sachi Arafat, Assistant Professor