**ABET Course Syllabus – CS4661**

|  |  |  |  |
| --- | --- | --- | --- |
| **Code** | CS4661 | **Credits** | 3 |
| **Title** | Introduction to Data Science | **Coordinator** | Mohammad Pourhomayoun |

**Course Information**

1. **Catalog Description:** Algorithms, tools, and techniques for extracting knowledge from large-scale data, and use it for future purposes such as prediction, classification, and visualization. Including fundamental Machine Learning and Big Data Manipulation.
2. **Prerequisites**: CS3112.
3. **Contact Hours:** Lecture 3 hours.
4. **Required/Elective:** This course is an elective course in the BS program.

**Textbook**

No textbook is required.

**Course Goals**

The Student Learning Outcomes that are addressed by the course are:

* SLO1. Students will be able to apply concepts and techniques from computing and mathematics to both theoretical and practical problems.
* SLO2. Students will be able to demonstrate fluency in at least one programming language and acquaintance with at least three more.
* SLO3. Students will have a strong foundation in the design, analysis, and application of many types of algorithms.
* SLO5. Students will have the training to analyze problems and identify and define the computing requirements appropriate to their solutions.
* SLO7. Students will be able to communicate effectively orally and in writing.
* SLO9. Students will have the ability to analyze the local and global impact of computing on individuals and society.

Other outcomes of instruction:

At the end of the course, students are able to

* Understand the main concepts, applications, and algorithms of *data analytics* and *machine learning*, including *supervised/unsupervised learning, classification/regression/clustering*, etc.
* Understand and use basic statistical modeling and machine learning algorithms including *nearest neighbor, decision trees, linear regression, logistic regression, gradient descent, k-means*, etc.
* prediction based on statistical tools such as regression, classification, and clustering.
* Understand and resolve the main issues involved in statistical modeling and machine learning including, for example, *overfitting, missing values, outlier samples*, etc.
* Understand and use *result visualization/validation, error identification, cross validation, various metrics for evaluation the prediction accuracy*.
* Understand the basics of *big data manipulation, parallel processing, MapReduce,* and main tools and frameworks.
* Understand and use the main algorithms and techniques for *large-scale data analytics* and *large-scale machine learning*.
* Understand and use tools, software packages, and libraries for data processing, analytics, and machine learning including *python scikit-learn, pandas, numpy*, etc.

**Topics Covered**

*Part 1: Introduction and Applications*

* The concept of Data Science, definitions, integrations of data science, real-life applications and examples, history and context, technology landscape.

*Part 2: Data Analytics and Machine Learning*

* Introduction to machine learning, definitions, real-life applications
* Supervised learning: classification, regression
* Unsupervised learning: Clustering, Dimensionality Reduction
* Model Training/Testing and Error Evaluation
* Data Analytics Challenges

*Part 3: Big Data Manipulation*

* Overview, definition and specifications of big data, sources of big data, challenges
* Introduction to the main approaches for handling big data
* Time/Space complexity, scalability
* MapReduce
* The concept of dimensionality reduction and feature selection for large-scale analytics
* Large-Scale Machine Learning
* MapReduce for machine learning

**ADA statement:** Reasonable accommodation will be provided to any student who is registered with the Office of Students with Disabilities and requests needed accommodation.

**Academic honesty statement:** Students are expected to do their own work and to abide by the University Policy on academic honesty, which is stated in the Schedule of Classes. See http://www.calstatela.edu/academicsenate/handbook/ch5a#cheating.

**Student responsibilities:** Students are responsible for being aware of all announcements that are made in class. Students are responsible for announcements made on days that they are absent. Students must check their CSULA email account regularly for information from the instructor and the Department. Failure to do so may result in missed deadlines or other consequences that might adversely affect students. Note that you can forward this email account to any other account of your choosing.