**Arab Academy for Science and Technology and Maritime Transport**
**Information Systems Curriculum**
**Course Syllabus**

| **Course Code:** IS467 | **Course Title:** Big Data Analytics | **Classification:** E | **Coordinator's Name:** Prof. Dr. Mohamed Kholief<br>**Lecturer:** Prof. Dr. Mohamed Kholief | **Credit:** 3 |
|---|---|---|---|---|
| **Pre-requisites:**<br>**BA203**<br>**CS366** | **Co-requisites:** None | **Schedule:**<br>Lecture 2 hrs.<br>Tutorial 2 hrs. | | |

**Office Hours: (Room 209)**
**Tuesday 10:30 a.m. -12:30 p.m.**

**Course Description:**
This course is an introduction to Big Data. The course explains key Big Data concepts, theory and terminology, as well as fundamental technologies and techniques. The course explains how Big Data can propel an organization forward by solving a spectrum of previously intractable business problems. It also introduces key analysis techniques and technologies and show how a Big Data solution environment can be built and integrated to offer competitive advantages.

**Textbook:**
Thomas Erl, Wajid Khattak, Paul Buhler, *Big Data Fundamentals: Concepts, Drivers & Techniques,* Pearson

**References:**
Tom White, *Hadoop The Definitive Guide, O'Reilly*

| **Course Objectives:** | **Contribution to Program Student Outcomes:** |
|---|---|

| | |
|---|---|
| 1. Discovering Big Data's fundamental concepts and what makes it different from previous forms of data analysis and data science | SO 1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions. |
| 2. Understanding the business motivations and drivers behind Big Data adoption, from operational improvements through innovation | SO 1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions. |
| 3. Planning strategic, business-driven Big Data initiatives | SO 2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. |
| 4. Addressing considerations such as data management, governance, and security | SO 2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. |
| 5. Recognizing the 5 "V" characteristics of datasets in Big Data environments: volume, velocity, variety, veracity, and value | SO 1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions. |
| 6. Clarifying Big Data's relationships with OLTP, OLAP, ETL, data warehouses, and data marts | SO 1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions. |
| 7. Working with Big Data in structured, unstructured, semi-structured, and metadata formats | SO 2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. |
| 8. Increasing value by integrating Big Data resources with corporate performance monitoring | SO 2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. |
| 9. Understanding how Big Data leverages distributed and parallel processing | SO 1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions. |

| | |
|---|---|
| 10. Using NoSQL and other technologies to meet Big Data's distinct data processing requirements | SO 2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. |
| 11. Leveraging statistical approaches of quantitative and qualitative analysis<br>Applying computational analysis methods, including machine learning | SO 2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. |

**Course Outline:**

1. Introduction and Big Data Overview
2. Big Data state of practice, data scientist characteristics, case studies
3. Big Data Analytics Lifecycle
4. Data Preparation and Model Planning, Data Analysis and Exploration
5. Statistics for Model Planning
6. 7th week exam
7. Data Visualization for Big Data
8. Big Data BI: Clustering, Association Rules
9. Big Data BI: Linear Regression, Logistics Regression
10. Big Data BI: Classification, Naive Bayes, Decision Trees
11. Big Data BI: Text Analysis, Time-Series Analysis
12. 12th week exam
13. Hadoop, Map Reduce, Hadoop Eco System
14. In Database Analytics
15. Putting it all together: big data projects' deliverables
16. Final Exam

**Grade Distribution:**

**7th Week Assessment (30%):**
Exam (20%) + Homework Assignments and/or quizzes 10%

**12th Week Assessment (20%):**
Exam (15%) + Assignments 5%

**Year Work (10%):**
Course Project (10%)

**Final Exam (40%)**

**Policies:**

**Attendance:**
AASTMT Education and Study Regulations (available at aast.edu)

**Academic Honesty:**
AASTMT Education and Study Regulations (available at aast.edu)

**Late Submission:**
*Late submissions are graded out of 75% (1 week late), 50% (2 weeks late), 25% (3 weeks late), 0% (more than 3 weeks late)*