Data Science II
Course

Course Description:

Data Science is easily one of the fastest growing and most in demand careers in our society. Every industry from farming to finance relies on big data to make decisions. All decisions that we make in our daily life and their consequences are rooted in our ability to analyze and utilize data. Data Science II, is a deeper dive into how statistical analyses can be performed using technical tools and computing skills to solve real world problems. This course goes beyond the fundamental ideas introduced in Data Science I and explores industry level techniques and tools. Real world data is used in a practical manner to extract solutions and insights. The prerequisite knowledge for this course is a strong statistical background, such as that found in AP Statistics, Statistical Reasoning, or the Data Science I course as well as a programming foundation such as that found in Computer Science Principles, AP Computer Science Principles, or AP Computer Science A.

Course Competency 1
DS2-1 Demonstrate how Data Science can be used to impact school, work, and leisure time
1.1: Demonstrates how students currently use data science in their lives, and how data science impacts their lives and possible career choices
1.2: Identify and differentiate between different data governance standards and argue why governance is important.
1.3: Identify ethical issues in data science.
1.4: Identify and compare potential bias issues in data science

Course Competency 2
DS2-2 Formulate Questions to Clarify the problem at hand and formulate 1 or more questions that can be answered with data
2.1: Identify the objectives of data and information management
2.2: Determine whether a problem involves categorical or quantitative data
2.3: Frame a statistical question of interest in terms of measurable data.

Course Competency 3
DS2-3 Design and implement a plan to collect appropriate data to answer the research question
3.1: Describe the factors that must be considered in distributing data effectively and how a simple model can be used to obtain at least a first-cut distribution.

Course Competency 4
DS2-4 Analyze data by selecting appropriate graphical and numerical methods
4.1: Implement advanced spreadsheet functions, automation, and dynamic reporting.
4.2: Utilize various tools (such as the ARIMA model) to analyze time series data.
4.3: Demonstrate the ability to take data and create a dashboard that provides insight to solve real world problems.
4.4: Use graphical and numerical displays to foster further investigation into question of interest

Course Competency 5
DS2-5 Identify the general concepts of databases/data tools and how to utilize design thinking to produce solutions that are clean and thoughtful.
5.1: Identify and distinguish between variations of techniques (Artificial Intelligence, Machine Learning, Deep Learning, etc.)
5.2: Provide definitions of key terms and concepts that describe the database environment
5.3: Describe and build the major components of the database environment and explain how these components interact with each other
5.4: Provide a review of systems development methodologies, particularly the waterfall method and agile programming development and show how database development fits with these methodologies.
5.5: Generate Entity Relationship logical models to represent organization data and plan for database development and infrastructure
5.6: Assess end user data and information requirements and develop a logical model to fit those requirements
5.7: Describe the concept of supertype/subtype relationships and recognize when to use these relationships in data modeling.
5.8: Describe the use of specialization (top-down perspective) and generalization (bottom-upper perspective) as complementary techniques for defining supertype/subtype relationships and understand relationship constraints when modelling.
5.9: Describe the position of logical database design within the overall database development process
5.10: Describe the relational model including the properties of relations, integrity constraints, and well-structured relations.
5.11: Describe the principles and detailed steps involved in mapping Enhanced Entity Relationship diagrams to relations.
5.12: Understand data normalization, functional dependency, and develop a fully normalized Entity Relationship Diagram. Evaluate the normality of a logical data model and correct any anomalies.

Course Competency 6
DS2-6 Build a database based on designed model, identify implementation policies and procedures, and establish plans for testing/debugging a data science solution.
6.1: Describe a Database Management System Language (DMBS) like SQL and summarize its basic operators.
6.2: Illustrate data definition language (DDL) commands for creating tables and views as well as for modifying and dropping tables.
6.3: Formulate single table DMBS (SQL) queries.
6.4: Formulate DMBS (SQL) queries that use functions.
6.5: Show how to establish referential integrity using DMBS (SQL).
6.6: Use of the “group by” and “order by” clauses in DMBS (SQL) queries.
6.7: Demonstrate (DBMS) SQL capabilities such as multiple-table data retrieval (join and other operators such as difference, union, and intersection), explicit and implicit joining, and built-in functions.
6.8: Illustrate the differences between the joining and subquery approaches to manipulating multiple tables in DMBS (SQL)
6.9: Describe triggers and stored procedures and provide examples of how these might be used
6.10: Generate and implement a testing plan for a data management solution implementation.

Course Competency 7
DS2-7 Deploy a data science solution in a production environment, follow implementation procedures, and develop a plan for long term maintenance.
7.1: Describe the differences between the processes of deployment and implementation of solutions.
7.2: Understand the components and key steps to a successful deployment.
7.3: Build and deploy a data management system solution implementation.
7.4: Generate and implement a maintenance plan for a data management solution implementation.

Course Competency 8
DS2-8 Analyze results by interpreting the information provided by the data and how its interpretation supports possible answers to question or problem being investigated.
8.1: Utilize visual reporting and statistic tools to perform, understand, and interpret statistic such as regression analysis, ANOVA, hypothesis testing, and sampling distributions
8.2: Identify and express areas for further study or investigation based on results
8.3: Create a dashboard with appropriate high-level charts, such as heat plots, box and whiskers, etc. to express the data that is being analyzed.