IT-2070: INTRODUCTION TO DATA SCIENCE AND ANALYTICS

Cuyahoga Community College

Viewing: IT-2070 : Introduction to Data Science and Analytics

Board of Trustees:
March 2023

Academic Term:
Fall 2023

Subject Code
IT - Information Technology

Course Number:
2070

Title:
Introduction to Data Science and Analytics

Catalog Description:
Broad coverage of topics key to data scientists to convert information to knowledge. Focus is on current data analytics methods to address business problems.

Credit Hour(s):
3

Lecture Hour(s):
2

Lab Hour(s):
2

Requisites

Prerequisite and Corequisite
IT-1025 Information Technology Concepts for Programmers and IT-1050 Programming Logic.

Outcomes

Course Outcome(s):
Identify and apply key Data Science and Analytics principles.

Objective(s):
1. Understand the evolution of Computerized Decision Support to Data Science and Analytics
2. Differentiate and apply big data processing concepts.
3. Identify characteristics of data and data storage.
4. Describe organizational impacts of analytics applications including project scope.
5. Describe ethical and legal issues related to analytics implementation.
6. Explain drivers for big data adoption including media types and cloud computing.
7. Discuss data procurement, privacy, security, governance, and provenance.
8. Understand the taxonomy of data, statistical modeling, and visualization.
9. Recognize the big data analytics lifecycle.
10. Identify and demonstrate data analysis techniques.

Course Outcome(s):
Define and explain descriptive, predictive, and prescriptive analytics.

Essential Learning Outcome Mapping:
Critical/Creative Thinking: Analyze, evaluate, and synthesize information in order to consider problems/ideas and transform them in innovative or imaginative ways.

Objective(s):

1. Explain the basic definitions and concepts of data warehousing.
2. Define data mining objectives, benefits, and applications of data mining.
3. Apply methods and algorithms to the data mining process.
4. Apply text analytics and rationale for text mining.
5. Analyze descriptive, predictive, and prescriptive analytics.
6. Define the taxonomy of data, statistical modeling, and visualization.

Course Outcome(s):

Explain and demonstrate modern methods and techniques used in data analytics.

Essential Learning Outcome Mapping:

Critical/Creative Thinking: Analyze, evaluate, and synthesize information in order to consider problems/ideas and transform them in innovative or imaginative ways.

Objective(s):

1. Analyze modern big data analytic approaches including Hadoop, MapReduce, and NoSQL. (not only SQL)
2. Describe big data platforms and services.
3. Discuss big data storage concepts and technologies.
4. Identify and apply enterprise technologies for big data intelligence.

Methods of Evaluation:

a. Participation and discussion
b. Online research
c. Written reports
d. Demonstrations
e. Projects
f. Tests
g. Quizzes

Course Content Outline:

a. Evolving needs for decision support and analytics
    i. Group communication and collaboration
    ii. Data management
    iii. Data availability
    iv. Data and societal issues
    v. Cloud computing
    vi. Internet of things
b. Data characteristics
    i. Volume
    ii. Velocity
    iii. Variety
    iv. Veracity
    v. Value
c. Types of data
    i. Structured
    ii. Unstructured
    iii. Semi-structured
    iv. Metadata
d. Big data solutions
    i. Data procurement
    ii. Privacy
    iii. Ethics
    iv. Security
    v. Provenance
    vi. Governance

e. Data analytics lifecycle
   i. Problem evaluation
   ii. Data identification
   iii. Data acquisition and filtering
   iv. Data extraction
   v. Data validation and cleansing
   vi. Data aggregation and representation
   vii. Data analysis
   viii. Data visualization
   ix. Utilization of analysis results
f. Data warehouses
   i. Data warehousing history, process and architecture
   ii. ETL (data extraction, transformation and load)
   iii. OLAP (online analytic processing) vs. OLTP (online transaction processing) and OLAP operations
   iv. Scalability and security issues
   v. Data marts
g. Big data storage
   i. Clusters
   ii. File systems and distributed file systems
   iii. NoSQL
   iv. Sharding
   v. replication
   vi. CAP theorem
   vii. ACID
   viii. BASE
   ix. Stream analytics
h. Big data processing
   i. Parallel and distributed
   ii. Hadoop
   iii. Batch and real-time modes
i. Big data analysis techniques
   i. Quantitative and qualitative
   ii. Data mining
   iii. Statistical, machine, semantic and visual analysis
j. Descriptive analytics
   i. Taxonomy of data
   ii. Statistical modeling concepts: descriptive statistics for descriptive analytics (measures of centrality tendency, mean, median, mode, range, variance, standard deviation, quartiles and interquartile range, box-and-whiskers plot)
   iii. Regression modeling for inferential statistic
   iv. Visual analytics - basic charts and graphs / information dashboards
k. Predictive analytics
   i. Data mining
   ii. Cluster analysis for data mining and association rule mining
   iii. Data mining software tools
   iv. Text analytics and text mining
   v. Sentiment analysis
   vi. Web mining and search engines
l. Prescriptive analytics
   i. Model-based decision making
   ii. Components of Decision Support Mathematical models
   iii. Mathematical programming optimization
   iv. Multiple goals, sensitivity analysis, what-if analysis and goal-seeking
   v. Decision tables and trees
   vi. Simulation
m. Cloud and location based analytics
   i. DaaS (Data as a Service), SaaS (Software as a Service), PaaS (Platform as a Service) and, IaaS (infracture as a Service)
   ii. Cloud deployment models
   iii. Location-based analytics (geospatial, real-time)

n. Legal, Privacy and Ethics
   i. Mobile user privacy
   ii. Ethics in decision making and support
   iii. Homeland security and Individual privacy
   iv. Collecting information about individuals
   v. Current technology issues in privacy and analytics
   vi. Legal issues
o. Careers in Data Analytics
   i. Use BLS to research outlook
   ii. Education and skillsets
p. Utilize industry-standard software to perform data analysis techniques listed in points 10-12 above. This software may include but is not limited to:
   i. SAS
   ii. R
   iii. Excel
   iv. ArcGis
   v. Python
   vi. Jupyter Notebooks

Resources

Mount, George. *Advancing into Analytics*. Sebastopol, CA: O'Reilly Media, Inc., 2021.

Shah, Chirag. *A Hands-On Introduction to Data Science*. 1st. Cambridge University Press, 2020. April 2, 2020.

McCoy, Scott. *Murach's Python for Data Analysis*. Fresno: Mike Murach & Associates, Inc., 2021. 2021.

Shmueli, Galit. *Data Mining for Business Analytics: Concepts, Techniques and Applications in Python*. Hoboken, NJ: John Wiley & Sons, Inc., 2020.

Wade, Ryan. *Advanced Analytics in Power BI with R and Python: Ingesting, Transforming, Visualizing*. New York, NY: Springer Science +Business Media, 2020.

Sharda, R., Delen, D., & Turban, E. (2018) *Business Intelligence, Analytics, and Data Science a Managerial Perspective*, Boston: Pearson.

Evans, J. R. (2017) *Business analytics: methods, models, and decisions,*Harlow, Essex, England: Pearson Education Limited.

Resources Other

a. SAS basics: http://video.sas.com/#category/videos/how-to-tutorials
b. SAS OnDemand for Academics: https://www.sas.com/en_us/software/on-demand-for-academics.html
c. Jupyter Notebooks https://jupyter.org/
d. Google Colaboratory https://colab.research.google.com/
e. Teaching and Learning with Jupyter https://jupyter4edu.github.io/jupyter-edu-book/

Top of page
Key: 5104