

IT-1080: INTRODUCTION TO DATA ANALYTICS

Cuyahoga Community College

Viewing: IT-1080 : Introduction to Data Analytics

Board of Trustees:

May 2022

Academic Term:

Fall 2022

Subject Code

IT - Information Technology

Course Number:

1080

Title:

Introduction to Data 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):

4

Lecture Hour(s):

3

Lab Hour(s):

2

Requisites

Prerequisite and Corequisite

IT-1025 Information Technology Concepts for Programmers or concurrent enrollment.

Outcomes

Course Outcome(s):

Identify and discuss key Data Analytics principles.

Objective(s):

1. Explain the evolution of Computerized Decision Support to Data Analytics/Science.
2. Explain big data processing concepts.
3. Describe the job opportunities in Data Analytics and the required skill sets.
4. Define privacy issues related to data mining.
5. Identify characteristics of data and data storage including types and two and three dimensional representations.
6. Describe organizational impacts of analytics applications.
7. Describe ethical and legal issues related to analytics implementation.
8. Explain drivers for big data adoption including media types and cloud computing.
9. Discuss data procurement, privacy, security, governance and provenance.
10. Define the taxonomy of data, statistical modeling, and visualization.
11. Recognize the big data analytics lifecycle.
12. 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. Describe methods and algorithms of data mining.
4. Describe text analytics and rationale for text mining.
5. Differentiate text analytics, text mining, and data mining.
6. Explain prescriptive analytics.
7. 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. Explain modern data mining software tools.
2. Discuss modern big data analytic approaches including Hadoop, MapReduce, and NoSQL. (not only SQL)
3. Describe big data platforms and services.
4. Explain stream analytics.
5. Describe and demonstrate Geospatial analytics.
6. Discuss big data storage concepts and technologies.
7. Identify enterprise technologies for big data intelligence.

Methods of Evaluation:

1. Participation and discussion
2. Online research
3. Written reports
4. Demonstrations
5. Projects
6. Tests
7. Quizzes

Course Content Outline:

1. Evolving needs for decision support and analytics
 - a. Group communication and collaboration
 - b. Data management
 - c. Data availability
 - d. Data and societal issues
 - e. Cloud computing
 - f. Internet of things
2. Data characteristics
 - a. Volume
 - b. Velocity
 - c. Variety
 - d. Veracity
 - e. Value
3. Types of data
 - a. Structured
 - b. Unstructured
 - c. Semi-structured
 - d. Metadata
4. Big data solutions
 - a. Data procurement
 - b. Privacy
 - c. Ethics
 - d. Security

- e. Provenance
- f. Governance
- 5. Data analytics lifecycle
 - a. Problem evaluation
 - b. Data identification
 - c. Data acquisition and filtering
 - d. Data extraction
 - e. Data validation and cleansing
 - f. Data aggregation and representation
 - g. Data analysis
 - h. Data visualization
 - i. Utilization of analysis results
- 6. Data warehouses
 - a. Data warehousing history, process and architecture
 - b. ETL (data extraction, transformation and load)
 - c. OLAP (online analytic processing) vs. OLTP (online transaction processing) and OLAP operations
 - d. Scalability and security issues
 - e. Data marts
- 7. Big data storage
 - a. Clusters
 - b. File systems and distributed file systems
 - c. NoSQL
 - d. Sharding
 - e. replication
 - f. CAP theorem
 - g. ACID
 - h. BASE
 - i. Stream analytics
- 8. Big data processing
 - a. Parallel and distributed
 - b. Hadoop
 - c. Batch and real-time modes
- 9. Big data analysis techniques
 - a. Quantitative and qualitative
 - b. Data mining
 - c. Statistical, machine, semantic and visual analysis
- 10. Descriptive analytics
 - a. Taxonomy of data
 - b. 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)
 - c. Regression modeling for inferential statistic
 - d. Visual analytics - basic charts and graphs / information dashboards
- 11. Predictive analytics
 - a. Data mining vs. Statistics
 - b. Data mining process
 - c. Cluster analysis for data mining and association rule mining
 - d. Data mining software tools
 - e. Text analytics and text mining
 - f. NLP (Natural Language Processing)
 - g. Text mining applications and process
 - h. Sentiment analysis
 - i. Web mining and search engines
- 12. Prescriptive analytics
 - a. Model-based decision making
 - b. Components of Decision Support Mathematical models
 - c. Mathematical programming optimization
 - d. Multiple goals, sensitivity analysis, what-if analysis and goal-seeking

- e. Decision tables and trees
- f. Simulation

13. Cloud and location based analytics

- a. DaaS (Data as a Service), SaaS (Software as a Service), PaaS (Platform as a Service) and, IaaS (infrastructure as a Service)
- b. Cloud deployment models
- c. Location-based analytics (geospatial, real-time)

14. Legal, Privacy and Ethics

- a. Mobile user privacy
- b. Ethics in decision making and support
- c. Homeland security and Individual privacy
- d. Collecting information about individuals
- e. Current technology issues in privacy and analytics
- f. Legal issues

15. Careers in Data Analytics

- a. Use BLS to research outlook
- b. Education and skillsets

16. Utilize industry-standard software to perform data analysis techniques listed in points 10-12 above. This software may include but is not limited to:

- a. SAS
- b. R
- c. Excel
- d. ArcGis

Resources

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

Cody, R. (2015) *An Introduction to SAS® University Edition*, An introduction to SAS® university edition.

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

Erl, T., Khattak, W., Buhler, P. (2016) *Big data fundamentals: concepts, drivers techniques*,Boston: Prentice Hall, ServiceTech Press.

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

Resources Other

1. Get Started with ArcGis Online: <http://learn.arcgis.com/en/projects/get-started-with-arcgis-online/>
2. Excel visualization basics: <https://www.gcflearnfree.org/excel2016/>
3. SAS basics: <http://video.sas.com/#category/videos/how-to-tutorials>
4. SAS OnDemand for Academics: https://www.sas.com/en_us/software/on-demand-for-academics.html

Top of page

Key: 2480