Learning Guide Unit 8

Site: University of the People
Course: CS 3440-01 Big Data - AY2025-T1
Book: Learning Guide Unit 8

Printed by: Mejbaul Mubin
Date: Thursday, 5 September 2024, 2:30 PM

Description

Learning Guide Unit 8

Table of contents

Overview

UNIT 8: Future of Big Data

Topics

- Predictive Data Analytics
- Big data integration
- Advancements in big data

Learning Objectives

By the end of this Unit, you will be able to:

1. Discuss the use of predictive data analytics with big data.
2. Explain the developments of big data integration.

Tasks

- Peer assess Unit 7 Written Assignment
- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Make entries to the Learning Journal
- Take and submit the Self-Quiz
- Read the Unit 9 Learning Guide carefully for instructions on the Final Exam
- Take the Review Quiz

Introduction

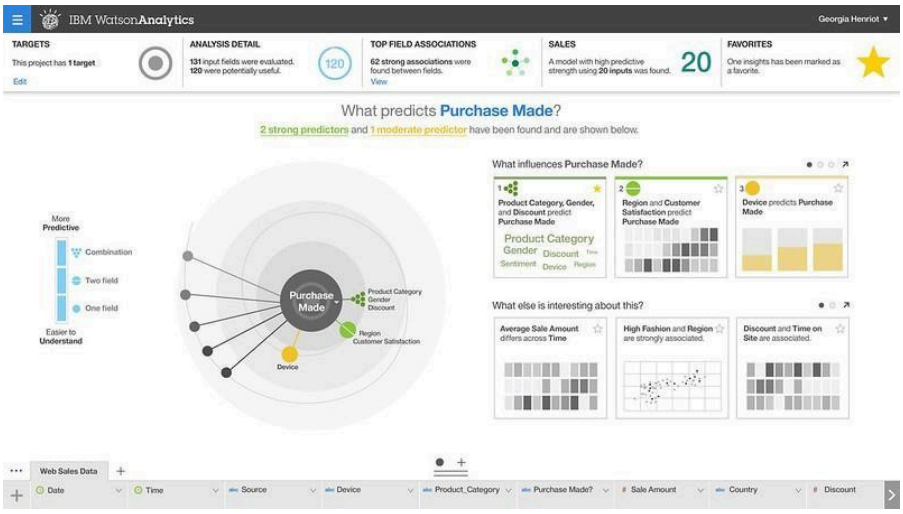

Image: Display of Predictive Analytics Dashboard

Using statistical algorithms, machine learning, predictive analytics, and data determine the likelihood of future outcomes based on historical data. The goal is to provide the best prediction of what will happen in the future rather than simply knowing what has happened. Predictive analytics is being used by businesses to find new opportunities and address challenging issues. Typical uses comprise:

- **Uncovering fraud**: Combining several analytics techniques helps increase pattern recognition and deter illicit activity. High-performance behavioral analytics monitors all network activity in real-time to look for anomalies that could point to fraud, zero-day vulnerabilities, or advanced persistent attacks as cybersecurity concerns escalate.
- **Optimizing advertising campaigns**: Predictive analytics is employed to forecast customer behavior or purchases and to encourage cross-selling opportunities. Predictive models assist firms in luring, keeping, and expanding their most lucrative clients.
- **Improving performance:** Businesses often use predictive models to forecast inventory and manage resources. Predictive analytics is used by airlines to determine ticket prices (SAS.com, 2022).

Data Extraction: What is it?

Data extraction is the process of gathering or extracting various forms of data from many sources, many of which may be erratically organized or entirely unstructured. Data extraction enables the consolidation, processing, and refinement of data to be kept in a single location and later altered. These areas could be on-site, in the cloud, or in combination. ETL (extract, transform, load) and ELT (extract, load, transform) processes begin with data extraction. ETL and ELT are components of a full data integration strategy (Talend.com, 2022). This process is inherently important to successfully create any data warehouse holding big data and should be carefully analyzed during development.

Global data collecting is expanding at an unheard-of rate. Talented people who can evaluate the data and make conclusions are needed to keep up with this enormous expansion. The International Data Corporation (IDC) estimates that even by 2025, the total amount of data generated worldwide may reach 175 zettabytes. To put it into perspective, it would require 12.5 billion storage devices to download all 175ZB using the biggest disk drive presently in use. The use of data in practically every business depends on how it is stored, processed, and utilized, and the future of big data speaks to continued digital expansion. Computer scientists and experts are paving the road for future innovation by gleaning knowledge from enormous amounts of data and converting it (Marville University, 2022). Big data advancements are being made in many different areas, and as we make our way through this unit you will learn about one of those advancements, blockchain technology.

In this unit, we will learn what predictive data analytics is and how it is used by big data, along with extensive data integration and what the future of big data holds.

References

Maryville University (2022). *The future of big data in business: Using data analytics to provide insight.* from https://online.maryville.edu/blog/future-big-data/

Predictive analytics. What it is and why it matters. (2022). SAS. https://www.sas.com/en_us/insights/analytics/predictive-analytics.html#:~:text=Predictive%20analytics%20is%20the%20use,will%20happen%20in%20the%20future.

What is data extraction? Definition and examples. (2022). Talend. from https://www.talend.com/resources/data-extraction-defined/

Reading Assignment

Read the following to gain an understanding of what predictive analytics is and how it helps big data provide analysis outputs to companies and researchers. You will learn about big data integration and the steps involved to load big data into a data warehouse for data mining and analysis activities. Finally, you will see what some future advancements in big data are and how they can help with the dissemination of information, as well as privacy and security issues.

Ahmed, I., Ahmad, M., Jeon, G., & Piccialli, F. (2021). A framework for pandemic prediction using big data analytics. *Big Data Research,* 25, 100190. https://doi.org/10.1016/j.bdr.2021.100190

- This article provides a framework used in pandemic prediction using big data analytics.

Arputhamary, B., & Arockiam, L. (2015). Data integration in big data environment. *Bonfring International Journal of Data Mining, 5*(1), 01-05. _http://www.journal.bonfring.org/papers/dm/volume5/BIJ-8001.pdf

- This article covers data integration in a big data environment.

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28,* 45-59._https://doi.org/10.1016/j.inffus.2015.08.005

- This article shows advancements in big data, social medial interactions, and new challenges.

Deepa, N., Pham, Q-V., Nguyen, D. C., Bhattacharya, S., Prabadevi, B., Gadekallu, T. R., Maddikunta, P. K. R., Fang, F., & Pathirana, P. N. (2022). A survey on blockchain for big data: Approaches, opportunities, and future directions. *Future Generation Computer Systems, 131*, 209-226. https://doi.org/10.1016/j.future.2022.01.017

- This article provides insights into the future of big data using blockchain technologies.

Mon Oo, M. C., & Thein, T. (2022). An efficient predictive analytics system for high dimensional big data. *Journal of King Saud University - Computer and Information Sciences*, *34*(1), 1521-1532. https://doi.org/10.1016/j.jksuci.2019.09.001 licensed under CC BY-NC-ND 4.0

- This article shows a predictive analytics system used with high-dimensional big data.

Patel, J. (2019). Overcoming data silos through big data integration. *International Journal of Education ((IJIT), 4* (4). https://famebook.com/jounals/IJIT/paper/IJIT003.pdf

- This article shows how to overcome data silos in big data integration and development.

Selvaraj, P., & Marudappa, P. (2018). A survey of predictive analytics using big data with data mining. International Journal of Bioinformatics Research and Applications, *14*(3), 269-282. https://www.researchgate.net/profile/Poornima-Selvaraj/publication/326071541_A_survey_of_predictive_analytics_using_big_data_with_data_mining/links/5c51132f92851c22a39a385b/A-survey-of-predictive-analytics-using-big-data-with-data-mining.pdf

- This article shows how data mining and big data interact and what type of analytics can be used in researching big data repositories.

Video Resources

FME Channel. (2020, February 11). *What is data integration, and how does it work?* [Video]. YouTube. https://www.youtube.com/watch?v=zj0ZxjxHOAs

This video explains data integration and how it can be used to incorporate data into your big data warehouse.

Kelly, K. (2020, July 1). *The future of big data.* [Video]. YouTube. https://www.youtube.com/watch?v=JBcplC8AFuE

- This video goes into the future of big data and its future use.

Discussion Assignment

Predictive analytics play an important role in discovering the underlying reasons and findings in any big data analysis project.

- Discuss any two uses of predictive data analytics with big data. How are these two uses important to businesses' operational planning?

Your Discussion should be a minimum of 200 words in length and not more than 300 words. Please include a word count. Following the APA standard, use references and in-text citations for the textbook and other sources.

Use APA citations and references for the textbook and any other sources; you should use at least 1 APA citation and reference, but you can use more if needed. Refer to the UoPeople APA Tutorials in the LRC for help with APA citations. You are required to post an initial response to the question/issue presented in the Forum and then respond to at least 3 of your classmates' initial posts. You should also respond to anyone who has responded to you. Don't forget to rate your classmates' postings according to the Rating Guidelines. Review the Discussion Forum rating guidelines to see how your classmates will rate your post.

After posting an appropriate, meaningful, and helpful response to your three classmates, you must rate their posts on a scale of 0 (unsatisfactory) to 10 (excellent).

10 (A) - Excellent, substantial, relevant, insightful, enriching, and stimulating contribution to the discussion. Also, it uses external resources to support positions where required and/or applicable.
8 - 9 (B) - Good, quite substantial, and insightful, but missing minor details which would have otherwise characterized it as an excellent response.
6 - 7 (C) - Satisfactory insight and relevance, but required some more information and effort to have warranted a better rating.
4 - 5 (D) - Limited insight and relevance of the material; more effort and reflection needed to have warranted a satisfactory grading.
0 - 3 (F) - Unsatisfactory insight/relevance or failure to answer the question, reflecting a poor or limited understanding of the subject matter and/or the guidelines of the question.

The rating scores are anonymous; therefore, do NOT mention in your remarks the separate rating score you will give the peer. The instructor is the only person who knows which score matches the comment given to a peer. Some classmates may worry that some peers will not provide a fair rating or be unable to provide accurate corrections for grammar or other errors. It is the instructor's responsibility to ensure fairness and accuracy.

Learning Journal

For organizations to gain any value from their big data repositories, the data must first be integrated from its source, transformed based on business and data rules then finally loaded into the data warehouse that holds the data. This process is called ETL (Extract, Transform and Load) and is used by data analysts and architecture to load the data warehouse for analysis.

- Explain the process of ETL in data integration for big data repository, and describe each step and its importance in the process.

The rubric detailing how you will be graded for this assignment can be found within the unit's assignment on the main course page.

Self-Quiz

The Self-Quiz gives you an opportunity to self-assess your knowledge of what you have learned so far.

The results of the Self-Quiz do not count towards your final grade. However,the quiz is an important part of the University's learning process and it is expected that you will take it to ensure understanding of the materials presented. Reviewing and analyzing your results will help you perform better on future Graded Quizzes and the Final Exam.

Please access the Self-Quiz on the main course homepage; it is listed inside the Unit.

Review Quiz

The Review Quiz will test your knowledge of all materials learned in this course. The results of this quiz will not count towards your final grade, but will help guide you in more thoroughly reviewing necessary topics and adequately preparing for the Final Exam.

Please access the Review Quiz under Unit 9 on the main course homepage; it will be listed inside the Unit.

Checklist

- Peer assess Unit 7 Written Assignment
- Read the Learning Guide and Reading Assignments
- Participate in the Discussion Assignment (post, comment, and rate in the Discussion Forum)
- Make entries to the Learning Journal
- Take the Self-Quiz
- Complete and submit the anonymous Course Evaluation
- Read the Unit 9 Learning Guide carefully for instructions on the Final Exam
- Take the Review Quiz