CSCI 6886 V1 | Big Data Analytics
2025/Fall | Wednesdays 6:30 pm – 9:30 pm
Vancouver Campus, Cambie Street 138/139

Instructor: Dr. Ali Safari
Email: a.safarimamaghani@fdu.edu

FAIRLEIGH DICKINSON UNIVERSITY
Vancouver Campus

Acknowledgment of traditional territory
FDU – Vancouver Campus acknowledges that the land on which we gather is the unceded territory of the Coast Salish peoples, including the territories of the xʷməθkwəy̓əm (Musqueam), Sḵwx̱wú7mesh (Squamish), Stó:lō and Səlílwətaʔ/Selilwitulh (Tsleil-Waututh) Nations.

COURSE CATALOGUE DESCRIPTION

The rapid growth of Big Data presents significant challenges for both academia and industry, requiring new technologies for data storage, retrieval, and knowledge discovery in large, often unstructured datasets. This course introduces students to the fundamentals of Big Data analytics, covering data mining, ETL pipelines, and machine learning techniques such as regression and classification. Students will gain hands-on experience with modern Big Data frameworks, including Hadoop (MapReduce) and Apache Spark, for developing parallel algorithms that can efficiently process large-scale data. The course also emphasizes data visualization and the use of free, publicly available platforms such as Google Colab, Databricks Community Edition, and Google BigQuery to ensure practical, accessible learning.

Course credits: 3 **Prerequisite: C**SCI 6623 Database Systems

COURSE OBJECTIVES

a. Course Outcomes

- Understand Big Data concepts and the need for Big Data Analytics
- Build and apply ETL pipelines and preprocessing techniques
- Analyze data using Spark, regression, classification, and advanced analytics
- Learn how to analyze Big Data with MapReduce in big data frameworks like Hadoop, Spark
- Visualize and interpret Big Data insights
- Understand ethical and privacy issues in analytics

b. Student Outcomes Addressed

- Apply CS theory to analyze problems and develop creative solutions
- Work effectively in teams and communicate analytics results
- Innovate with deployment of computing technology

TEXTBOOK AND RESOURCES

Big Data

Required:

- Mining of Massive Datasets, 3rd Edition – Jure Leskovec, Anand Rajaraman, Jeff Ullman, Cambridge University Press, 2020 [ISBN: 9781108476348]
- Lecture Notes and materials on WebCampus

Recommended:

- Databricks Tutorials: Free Spark tutorials, RDDs, DataFrames, MLlib – https://docs.databricks.com/getting-started/index.html
- Apache Spark Official Documentation: https://spark.apache.org/docs/latest/
- Google Colab PySpark Tutorials: https://colab.research.google.com/github/apache/spark-notebook
- Python for Data Analysis – Wes McKinney
- Learning Spark: Lightning-Fast Big Data Analysis, Matei Zaharia, Holden Karau, Andy Konwinski, and Patrick Wendell, O'Reilly Media, 2015 [ISBN13: 9781449358624]
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow – Aurélien Géron
- Storytelling with Data – Cole Nussbaumer Knaflic
- Fairness and Machine Learning – Barocas et al.
- The Ethics of Big Data – Kord Davis

Week 0 - Python References

- Sweigart, A. (2019). Automate the Boring Stuff with Python (2nd ed.). No Starch Press. https://automatetheboringstuff.com/
- Matthes, E. (2023). Python Crash Course (3rd ed.). No Starch Press. https://nostarch.com/python-crash-course-3rd-edition

ATTENDANCE POLICY

Regular attendance and active participation are essential for keeping pace with the course material and engaging in classroom discussions and activities. Students are responsible for all material covered in class and must submit assignments on time, even if a class is missed.

Key Guidelines:

1. Students are expected to attend all classes. Absences should be communicated to the instructor **in advance.**
2. Missed classes without prior communication may result in a **loss of participation points**. Excessive absences may affect your grade or result in being dropped from the course.
3. Makeup quizzes or exams will only be allowed for valid reasons (e.g., illness, emergency) and **require proper documentation**. Whenever possible, inform the instructor before the class is missed.
4. Students are expected to actively engage in class activities and discussions, even if they occasionally miss a session.

Recommendation: Treat this policy as a commitment to stay on track and fully participate, as classroom interaction is vital for mastering the material and completing hands-on projects successfully.

CLASSROOM AND EMAIL ETIQUETTE

Classroom:

- Arrive on time and be prepared.
- Silence phones and personal devices.
- Participate actively and show respect to everyone.
- Eating and drinking in class is discouraged.

Email:

- Use clear subject lines and be concise.
- Maintain a professional tone.
- Respond in a timely manner when seeking assistance.

EVALUATION/ASSESSMENT

Type of evaluation/assessment tool	**% of Overall Grade**
Participation and quizzes	10%
Weekly assignments and labs	15%
Midterm exam	30%
Final project and presentation	15%
Final Exam	30%
TOTAL	**100%**

Grades and Grade Points

Grade	Percentage	GPA
A	90–100	4.00
A-	85-89.9	3.67
B+	80–84.9	3.33
B	75-79.9	3.00
B-	70-74.9	2.67
C+	65-69.9	2.33
C	60-64.9	2.00
F	0-59.9	0

COURSE SCHEDULE

Week	Date	Topic	Suggest reading
0	Sep 3	Python Primer (Self-study recommended)	Sweigart, A. (2019). Automate the Boring Stuff with Python, Ch.1-10
1	Sep 10	Introduction to Big Data	Mining of Massive Datasets, Ch.1
2	Sep 17	Spark Fundamentals	Learning Spark; Apache Spark Docs
3	Sep 24	Data Ingestion & ETL	Databricks Tutorials
4	Oct 1	Preprocessing & Cleaning	McKinney; Hands-On ML (Géron)
5	Oct 8	Regression (1)	Mining of Massive Datasets, Ch.12–13
6	Oct 15	Regression (2)	Mining of Massive Datasets, Ch.12–13
7	**Oct 22**	**Midterm exam**	**All references**
8	Oct 29	Classification (1)	Mining of Massive Datasets Ch.7; Hands-On ML (Géron)
9	Nov 5	Classification (2)	Mining of Massive Datasets Ch.7; Hands-On ML (Géron)
10	Nov 12	Data Visualization	Storytelling with Data (Knaflic)
11	Nov 19	Big Data in Cloud (1)	Databricks Tutorials; BigQuery Docs
12	Nov 26	Big Data in Cloud (2)	Databricks Tutorials; BigQuery Docs
13	Dec 3	Ethics & Privacy	Fairness and Machine Learning (Barocas et al.); Ethics of Big Data (Kord Davis)
14	Dec 10	Final Project Presentations	-
15	**Dec17**	**Final Exam (check exam timetable)**	**All references**

- Please note that this syllabus is a guide to our learning this semester. At any time, the instructor may deem it necessary to adjust the schedule of topics and assignments and will communicate with students accordingly.
- Late assignment submissions will be penalized (20% points are taken off for each day late).
- **Plagiarism will not be tolerated.**

ACADEMIC INTEGRITY

Students are expected to maintain the highest standards of academic honesty. For FDU's complete Academic Integrity Policy, please see Academic Regulations at https://www.fdu.edu/wp-content/uploads/2020/01/academic-regulations.pdf or https://www.fdu.edu/academic-integrity/.

STUDENTS WITH DISABILITIES

Students with documented medical, psychological or learning disabilities, who feel they may need in-class academic adjustments, reasonable modifications, and/or auxiliary aids and services while taking this course, should first contact the Disability Support Services (DSS) to discuss their specific needs. At the Florham Campus, including the School of Pharmacy & Health Sciences and study abroad programs, contact the Director of Disability Support Services at 973-443-8079. At the Metropolitan Campus, online and off-campus programs, contact the Director of Disability Support Services at 201-692-2076. At the **Vancouver Campus**, contact the Deputy Campus Executive at 604-648-4463. Once the academic adjustments,

modifications, or auxiliary aids and services are approved by DSS, make an appointment to see the professor.

RESOURCES FOR STUDENT SUCCESS

Students in this course are invited to reach out to one or more of our student support offices for support in this course, other courses, or any aspect of their FDU experience. In addition, any student with a concern who does not know where to turn may reach out to the Chief Student Experience Officer on the relevant campus. Here at the Vancouver Campus, our Chief Student Experience Officer is 'Jobin Mojtabavi, jobin_mojtabavi@fdu.edu, 604-648-446'