MIE1624HF – Introduction to Data Science and Analytics

Department of Mechanical and Industrial Engineering, University of Toronto

Oleksandr Romanko, Ph.D., Adjunct Professor, University of Toronto
Associate Director, Financial Risk Quantitative Research, SS&C Algorithmics
oleksandr.romanko@utoronto.ca

Course title: Introduction to Data Science and Analytics (MIE1624HF)

Course description: The objective of the course is to learn analytical models and overview quantitative algorithms for solving engineering and business problems. Data science or analytics is the process of deriving insights from data in order to make optimal decisions. It allows hundreds of companies and governments to save lives, increase profits and minimize resource usage. Considerable attention in the course is devoted to applications of computational and modeling algorithms to finance, risk management, marketing, health care, smart city projects, crime prevention, predictive maintenance, web and social media analytics, personal analytics, etc. We will show how various data science and analytics techniques such as basic statistics, regressions, uncertainty modeling, simulation and optimization modeling, data mining and machine learning, text analytics, artificial intelligence and visualizations can be implemented and applied using Python. Python and Tableau, Power BI are modeling and visualization software used in this course. Practical aspects of computational models and case studies in Interactive Python are emphasized.

Lectures: Tuesday, 6:00pm-9:00pm, SF 1101 (first lecture is on September 10)

Office Hours: After the lecture or by appointment

Teaching Assistants: TBA

Course Outline

Introduction to data science and analytics

1. Data science concepts
2. Application areas of quantitative modeling

Python programming, data science software

1. Introduction to Python
2. Comparison of Python, R and Matlab usage in data science

Basic statistics

1. Random variables, sampling
2. Distributions and statistical measures
3. Hypothesis testing
4. Statistics case studies in IPython

Overview of linear algebra

1. Linear algebra and matrix computations
2. Functions, derivatives, convexity

Optimization

1. Unconstrained non-linear optimization algorithms
2. Overview of constrained optimization algorithms
3. Optimization case studies in IPython

Modeling techniques, regression

1. Mathematical modeling process
2. Linear regression
3. Logistic regression
4. Regression case studies in IPython

Data visualization and visual analytics

1. Visual analytics
2. Visualizations in Python, and visual analytics in Tableau and Power BI

Advanced machine learning

1. Decision trees
2. Advanced supervised machine learning algorithms (Naive Bayes, k-NN, SVM)
3. Intro to ensemble learning algorithms (Random Forest, Gradient Boosting)
4. Intro to neural networks and deep learning
5. Text analytics and natural language processing
6. Generative AI
7. Clustering (K-means, Fuzzy C-means, Hierarchical Clustering, DBSCAN)
8. Dimensionality reduction
9. Association rules
10. Overview of reinforcement learning
11. Machine learning case studies in IPython

Simulation modeling

1. Random number generation
2. Monte Carlo simulations
3. Simulation case studies in IPython

Introduction to Deep Learning

1. Mathematics of neural networks
2. Introduction to Deep Learning
3. Convolutional Neural Networks (CNN)
4. Recurrent Neural Networks (RNN) and Autoencoders
5. Transformers and Large Language Models

Assignments, Exams and Grading (tentative)

Assignment #1 (12%), Assignment #2 (15%), Assignment #3 (15%) *– individual assignments*

Course Project (20%) *– group project*

In-Class Group Presentation (12%) *– group assignment*

Final Exam (26%) *– individual exam*

If a student gets less than 50% mark at the Final Exam, her/his course mark will be reduced one letter grade down. E.g., a student got 11 pts (Assg 1) + 14 pts (Assg 2) + 14 pts (Assg 3) + 19 pts (Course Project) + 11 pts (In-Class Presentation) + 12 pts (Final Exam) = 81 pts that corresponds to A- course mark, but because a student got 12 pts out of 26 pts at the Final Exam (less than 50%), the course mark will be reduced from A- to B+.

Recommended References and Readings

Course lecture notes are self-contained, IPython case studies would be run and discussed-in class

- **A Cook-book of Mathematics** by V. Vinogradov, 1999 http://www.cerge-ei.cz/pdf/lecture_notes/LN01.pdf
- **Getting Started with Data Science: Making Sense of Data with Analytics** by M. Haider, 2015 https://www.amazon.ca/Getting-Started-Data-Science-Analytics/dp/0133991024/
- **Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter** by W. McKinney, 2022 https://www.amazon.ca/Python-Data-Analysis-Wrangling-Jupyter/dp/109810403X/
- **Artificial Intelligence Programming with Python** by P. Xiao, 2022 https://www.amazon.ca/Artificial-Intelligence-Programming-Python-Zero/dp/1119820863/
- **Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More** by M. Russell and M. Klassen, 2019 https://www.amazon.ca/Mining-Social-Web-Facebook-Instagram/dp/1491985046/