**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
Senior Research Analyst – Quantitative Research, Watson Financial Services, IBM Canada
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 IBM Watson Analytics 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, MY 150 (first lecture is on September 17)

**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 IBM Watson Analytics

**Data mining and machine learning**

1. Classification (decision trees)
2. Clustering (K-means, Fuzzy C-means, Hierarchical Clustering, DBSCAN)
3. Association rules
4. Advanced supervised machine learning algorithms (Naive Bayes, k-NN, SVM)
5. Intro to ensemble learning algorithms (Random Forest, Gradient Boosting)
6. Data mining case studies in IPython

**Simulation modeling**

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

**Cognitive computing and artificial intelligence**

1. Intro to neural networks and deep learning
2. Text analytics and natural language processing
3. Reinforcement learning
4. Spatio-temporal analytics
5. Cognitive computing case studies in IPython

**Storytelling based on analytics, analytical decision making**

1. Validating analytics
2. Storytelling based on analytics
3. Decision-making based on analytics

## Assignments, Exams and Grading

***Assignment #1 (15%), Assignment #2 (15%)***

***Course Project* (30%)**

***In-Class Group Presentation* (15%)**

***Final Exam* (25%)**

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 14 pts (Assg 1) + 13 pts (Assg 2) + 28 pts (Course Project) + 14 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 25 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 IPython** by W. McKinney, 2012 https://www.amazon.ca/Python-Data-Analysis-Wrangling-IPython/dp/1449319793/
- **Computational Business Analytics** by S. Das, 2013 https://www.amazon.ca/Computational-Business-Analytics-Subrata-Das/dp/1439890706/
- **Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More** by M. Russell, 2013 https://www.amazon.ca/Mining-Social-Web-Facebook-LinkedIn/dp/1449367615/