CEM1002 2018 Syllabus
University of Toronto
5 September 2018

INSTRUCTOR: Prof. Mark S. Fox
Mechanical & Industrial Engineering Department
40 St. George St., Room 8114
Telephone: 416-978-6823; email: msf@eil.utoronto.ca
Office Hours: by appointment

LOCATION: Mondays, 1pm to 3pm.
SF1009

TEXTBOOKS: Provost, F. and Fawcett, T. (2013) *Data Science for Business* (Sebastopol, CA: O'Reilly Media Inc.)

SOFTWARE: Rapidminer (https://rapidminer.com/)

TITLE: Data Analytics and Cities

Course Description
This course provides an introduction to data analytics, including machine learning, illustrating by their application to cities. Topics include Data Discovery and Wrangling, Classification techniques, Similarity Analysis, Text Mining, and City Performance Measurement. The analysis techniques will be demonstrated using the Rapidminer data analytics software tool. Students will be required to present a case study of an existing application of data analytics to a city problem, and will work as part of a team to solve a city problem using the aforementioned techniques.

Course Materiel: Lecture slides and readings, except for the textbook, will be available via the course web site prior to each week's lecture.

Individual Evaluation
1. (10%) Each student will submit a written summary (using the provided HTML template) of a data science application in a city. The summary will be presented on October 1st. The summary will contain the following:
    a. Title, city/state/country, project start and end date
    b. Bibliographic reference for the application
    c. Summary of the application: problem, data science method, datasets used
    d. Issues, outcome and status

2. (35%) Final Exam – downloaded from the course website and to be submitted by via the course website within 2 hours.

Group Evaluation
Each group will choose a city and problem that they will solve using data science methods implemented in Rapidminer. Each group will be composed of 2-3 students to be selected in the second week

Assignment 1	(Due on October 8)	10%
Assignment 2	(Due on October 29)	15%
Assignment 3	(Due on November 26)	20%
Presentation	(December 3)	10%

Warnings:

1. If you don't understand anything, ask questions in class – I am happy to answer them. You can also email me or schedule an appointment.
2. I tell jokes in class, but no one knows if I am joking as I always have a serious look on my face. My kids tell me I am not funny. Humor me by laughing!

Date	Description
10 Sep 18	**1. Introduction to Data Science and Cities, Rapidminer** • Provost and Fawcett, Chapter 1. • Cukier, K. and Mayer-Schoenberger, V. 2013. "The Rise of Big Data," *Foreign Affairs*, May/June, pp. 28-40. • Flowers, M., (2013), "Beyond Open Data: The Data Driven City", In *Beyond Transparency: Open Data and the Future of Civic Innovation*, B. Goldstein and L. Dyson (Eds.), pp. 185-198. • Thompson, E.D., Frolich, E., Bellows, J.C., Bassford, B.E., Skiko, E.J., and Fox, M.S., (2015), "Process Diagnosis System (PDS) – A 30 Year History", *Proceedings of the 27th Conference on Innovative Applications of Artificial Intelligence*, pp. 3928-3933. • Shahrokni, H., Van der Heijde, B., Lazarevic, D., & Brandt, N., (2014), "Big data GIS analytics towards efficient waste management in Stockholm", *Proceedings of the 2014 Conference ICT for Sustainability*.
17 Sep 18	**2. Data Discovery and Wrangling, and the Data Mining Process** • Provost and Fawcett, Chapter 2. • Optional: Kosko, B., (1986), "Fuzzy Cognitive Maps", *International Journal of Man-Machine Studies*, Vol. 24, pp. 65-75. • Barbosa, L., Pham, K., Silva, C., Vieira, M.R., and Freire, J., (2014), "Structured Open Urban Data", *Big Data*, pp. 144-154. • Nalchigar, S., and Fox, M.S., (2017), "Achieving interoperability of smart city data: An analysis of 311 data", *Journal of Smart Cities*, Vol. 3, No. 1.
24 Sep 18	**3. Predictive Modelling – Model Fitting** • Provost and Fawcett, Chapters 3 and 4 • Heaton, J., (201?), "A Non-Mathematical Introduction to Neural Nets", http://www.heatonresearch.com/content/non-mathematical-introduction-using-neural-networks
1 Oct 18	**4. Student Presentations**

	• City Data Analytics Applications
8 Oct 18	**5. Thanksgiving – no class**
15 Oct 18	**6. Overfitting + Examples** • Provost and Fawcett, Chapter 5.
22 Oct 18	**7. Similarity Analysis** • Provost and Fawcett, Chapter 6.
29 Oct 18	**8. Good Models,Visualization, Evidence and Probabilities** • Provost and Fawcett, Chapters 7, 8 and 9.
5 Nov 18	**9 Text Mining** • Provost and Fawcett, Chapter 10.
12 Nov 18	**10. Advanced Topic: Data Interoperability** • Fox, M., (2013), "City Data: Big, Open and Linked", Municipal Interfaces, November.
19 Nov 18	**11. Advanced Topic: Measuring City Performance** • Joss, S. (ed), (2012), "Tomorrow's City Today: Eco-City Indicators, Standards & Frameworks. Bellagio Conference Report", London: University of Westminster. pp. 5-10. • Pires, S.M., Fidelis, T., and Ramos, T.B., (2014), "Measuring and comparing local sustainable development through common indicators: Constraints and achievements", Cities, Vol. 39, pp. 1-9. • Fox, M.S., (2017), "The PolisGnosis Project: Enabling the Computational Analysis of City Performance", *Proceedings of the 2017 Industrial and Systems Engineering Conference*, K. Coperich, E. Cudney, H. Nembhard, eds., Institute for Industrial and Systems Engineering.Fox, M.S., (2013), "The role of ontologies in the publishing and analysis of city indicators", *Computers, Environment and Urban Systems*, Vol. 54 pp. 266-279.
26 Nov 18	**12. Smart Cities** • TBD
3 Dec 18	**13. Team Presentations**
10 Dec 18	**14. Final Exam**