Data Analytics and Decision Making

DATA ANALYTICS AND DECISION MAKING

ALI ABDULHUSSEIN

University of Windsor
Windsor, ON

CONTENTS

3. Predictive Analytics

4. Prescriptive Analytics

INTRODUCTION

Professor Ali AbdulHussein

Data Analytics is a rapidly-evolving field applicable to many business, engineering, and computer science workplace contexts. Open educational resources are more likely to be better options than costly, traditional commercial textbooks that become outdated rapidly. The digital content in this new online course, *Data Analytics and Decision Making*, introduces learners to fundamental concepts of data analytics and their application in decision making. The emphasis is on the utilization of practical analytic tools in a complex engineering management environment. Topics covered include careers in data analytics; data types, formats, and repositories; big data and cloud computing; stages of data analytics; descriptive analytics; predictive analytics, including statistical learning and machine learning; and prescriptive analytics with decision-making examples.

Download the Course Syllabus

Learning Objectives

At the end of the course, the successful student will know and be able to:

1. describe different career paths in data analysis
2. identify different data types, formats, and repositories
3. identify different ways to select the variables for data analysis
4. explain prediction algorithms and apply them appropriately
5. explain classification algorithms and apply them appropriately implementation
6. evaluate data analytical models using cross validation

Course Need

The course will be offered within the curriculum of graduate degree programs in the Faculty of Engineering at the University of Windsor such as: Master of Engineering, Master of Applied Science, and Master of Engineering Management. The modular design of the course will allow future alignments with other programs.

Labor Market Need

Today's labor market, namely for engineering graduates, demands data analytics skills. Employers are increasingly requiring candidates to be able to deal with large amount of data and be able to interpret and utilize data in decision making. To further verify this market need, we have conducted a labor market analysis (https://www.services.labour.gov.on.ca/) and scanned skills required for jobs with Above Average job outlook for 2017- 2021 such as: computer network technicians, computer and info systems managers, info systems analytics, system consultants and database analysts. Data-related skills were demanded in these jobs.

Need for Online Format

The technical nature of the course lends it the strong potential to be delivered online with high quality. In Summer 2022, this course will be offered for the first time as a required course (GENG-8050) in the Master of Engineering Management program that attracts full-time working professionals. Hence, offering it in a virtual format (fully or partially) adds to programming flexibility.

ACKNOWLEDGEMENTS

This project was developed with funding from the Government of Ontario and eCampusOntario Virtual Learning Strategy.

Data Analytics was developed by Prof. Ali AbdulHussein (Odette School of Business) in collaboration with Dr. Bahman Naderi (Faculty of Engineering) at the University of Windsor. Dr. Bahman contributed draft presentation slides, labs and tutorials, and assignments for the course.

Dr. Nobuko Fujita (Office of Open Learning) provided instructional design support, and Dr. Chris Teplovs recorded the lecture videos.

Dr. Beth Robertson contributed her Equity, Diversity, Decolonization, and Inclusion (EDDI) expertise in reviewing the course materials and providing advice.

Sakshi Arora, our student partner specializing in the Business Analytics stream. Odette Master of Management, contributed to editing data analytics equations for accessibility and multimedia production and post production.

Shreyas Tambe (Office of Open Learning) provided support for video production and post-production.

Patrick Carnevale (OOL Co-op Student; School of Computer Science) provided graphic design support on Dr. Naderi's slides.

Current students in the Master of Management Program, Mayank Ghai, Hasib Imam, Muhammad Shahid, and Spencer Stinson, contributed to the creation of the course by providing feedback on the use of data analytics in their professional engineering contexts and experiences with online learning.

ACCESSIBILITY STATEMENT

While we attempt to make all elements of this resource conform with international accessibility guidelines, we must acknowledge a few accessibility issues:

- Some external resources may not conform with accessibility guidelines.
- Third-party video content may not be captioned, or may be captioned poorly.
- The heading order for some headings do not conform to logical heading order.
- The colour contrast on the Pressbooks Theme Malala header and footer does not meet guidelines and contains a redundant link in the footer.

Accessibility Tips

- Links to other pages of the site will always open in the current window.
- Use your screen reader's list headings feature to navigate through the headings within the content of a page.
- Use the Previous and Next links found at the bottom of each page to navigate through the sequence of pages in the site. To access these links most easily, use your screen reader's landmarks list to navigate to the content info region, then press Shift+Tab to back up to the Next links.
- Depending on the operating system and browser being used, font size can be adjusted by pressing a key combination including the plus (+) and minus (-) keys. On Windows systems the key combination is typically "Ctrl+" and on Mac it is "Command+".

ADOPTING OR ADAPTING THE BOOK

Citation & Attribution

The suggested citation for this book in APA format is:

AbdulHussein, A. (2022). *Data Analytics and Decision Making.* University of Windsor. http://ecampusontario.pressbooks.pub/dataanalytics/

The suggested attribution for this textbook is

Data Analytics and Decision Making by Ali AbdulHussein is licensed under a Creative Commons Attribution License 4.0

Share

If you adopt this book, as a core or supplemental source, please report your adoption in order for us to celebrate your support of students' savings! Report your commitment at https://openlibrary.ecampusontario.ca/report-an-adoption/

We invite you to adapt this book further to meet your and your students' needs. Please let us know if you do! If you would like to use Pressbooks, the platform used to make this book, contact eCampusOntario for an account using open@ecampusontario.ca.

If this text does not meet your needs, please check out the full library at www.openlibrary.ecampusontario.ca. If you still cannot find what you are looking for, connect with colleagues and eCampusOntario to explore creating your own open education resource (OER).

eCampusOntario

eCampusOntario is a not-for-profit corporation funded by the Government of Ontario. It serves as a centre of excellence in online and technology-enabled learning for all publicly funded colleges and universities in Ontario and has embarked on a bold mission to widen access to post-secondary education and training in Ontario. This textbook is part of eCampusOntario's open textbook library, which provides free learning resources in a wide range of subject areas. These open textbooks can be assigned by instructors for their classes and can be downloaded by learners to electronic devices. These free and open educational resources are customizable to meet a wide range of learning needs, and we invite instructors to review and adopt the resources for use in their courses.

1. INTRODUCTION TO DATA ANALYTICS AND DECISION MAKING

Photo by Samson on Unsplash

1.1 INTRODUCTION TO DATA ANALYTICS AND DECISION MAKING

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=25

1.3 CAREERS IN DATA ANALYTICS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=27

1.4 DATA TYPES, FORMATS AND REPOSITORIES

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=29

1.5 DATA TECHNOLOGIES: BIG DATA AND CLOUD COMPUTING

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=31

1.6 STAGES OF DATA ANALYTICS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=33

1.7 PREDICTIVE ANALYTICS: STATISTICAL LEARNING & MACHINE LEARNING

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=35

1.8 PRESCRIPTIVE ANALYTICS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=37

Exercises

What decision was being made?
What data (descriptive and predictive) might one need to make the best decision?
What other costs or constraints might you have to consider in routing?
Which other situations might be appropriate for applications of such models?

2. DESCRIPTIVE ANALYTICS

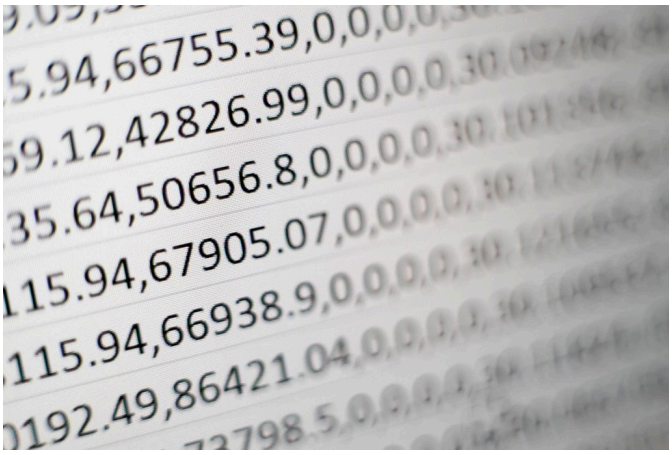


Photo by Mika Baumeisteron Unsplash

2.1 DESCRIPTIVE ANALYTICS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=41

Recommended Readings for the Descriptive Analytics Chapters

Chapter 1. Descriptive Statistics and Frequency Distributions

Chapter 2. The Normal and t-Distributions (only the normal distribution)

This is a chapter from a free, open textbook that has been adapted to the Canadian context. When read online, it allows readers to learn the basic and most commonly-applied statistical techniques in business in an interactive way using Excel spreadsheets.

2.2 DATA VISUALIZATION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=43

Exercises

Lab 1

The instructor will go over the following file in class. Load the following file to practice creating data visualizations such as a boxplot and pivot table.

DataVisualization_Template.xls

Lab 2

Load the following file to practice creating data visualizations.

Amazon_Template.xlsx

Data Source:
Amazon Top 50 Bestselling books 2009-2019,
https://www.kaggle.com/sootersaalu/amazon-top-50-bestselling-books-2009-2019

2.3 DATA SUMMARIZATION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=45

In-Class Optional Exercises

The instructor will go over these examples in class. Load the following files to practice analyzing data summarization.

Bankruptcy.csv

Boston Home Price.csv

Data Sources:

https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction

https://www.kaggle.com/fedesoriano/the-boston-houseprice-data

ASSIGNMENT 1

Analyze an Insurance Charge Dataset

Use the dataset, Case_Insurance.csv to answer the following questions:

- Does sex impact the insurance charges?
- Identify the most impactful factor on insurance charges

Note:

Consider the fact that not everyone identifies within a binary of male/female or man/woman. For the purposes of this assignment, we are using the word "sex" to refer to the physiology of the person. A better word to use may be "**gender**." This is because preconceived notions and biases associated with gender, rather than solely the physiology of the person, has been proven to affect health insurance rates and access to health services more generally.

To learn more about these issues, read

Katherine Hay, M. A., et. al. (2019). "Disrupting Gender Norms in Health Systems: Making the Case for Change," *Lancet, 393* (10190), pp. 2535-2549. https://doi.org/10.1016/S0140-6736(19)30648-8

3. PREDICTIVE ANALYTICS

Photo by PAUL SMITHon Unsplash

3.1 PREDICTIVE ANALYTICS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=51

Recommended Reading for Predictive Analytics

Chapter 4. Hypothesis Testing

This is a chapter from a free, open textbook that has been adapted to the Canadian context. When read online, it allows readers to learn the basic and most commonly-applied statistical techniques in business in an interactive way using Excel spreadsheets.

3.2 REGRESSION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=53

Exercises

Lab 3

Load the following file to practice regression.

Case_Advertising_Template.xlsx

Recommended Reading

Chapter 8. Regression Basics

This is a chapter from a free, open textbook that has been adapted to the Canadian context. When read online, it allows readers to learn the basic and most commonly-applied statistical techniques in business in an interactive way using Excel spreadsheets.

3.3 MULTIPLE LINEAR REGRESSION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=55

Exercises

Lab 4

Load the following file to practice multiple linear regression.

Case_Credit_Template.xlsx

3.4 NON-LINEAR RELATIONSHIPS/ POLYNOMIAL REGRESSION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=57

3.5 LOGISTIC REGRESSION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=59

Exercises

Lab 5

Load the following file to practice multiple linear regression.

Logistic_Regression_Template.xlsx

ASSIGNMENT 2

Using Predictive Analytics for Defaulting on Credit Card Payments

Use the dataset, Logistic_Regression_Case_Template, to analyze whether an individual will default on their credit card payment based on their annual income monthly credit card balance, and student status for a subset of 10,000 individuals.

Coefficients of Logistic regression

β_0	β_1	β_2	β_3

Dataset (Y: response variable, X: predictors)

Default	Student?	Balance	Income
Y	X_1	X_2	X_3
0	0	729.5264952	44.36162507
0	1	817.1804066	12.1061347
0	0	1073.549164	31.76713895
0	0	529.2506047	35.70449394
0	0	785.6558829	38.46349588
0	1	919.5885305	7.491558572

3.6 K-NEAREST NEIGHBOURS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=64

Exercises

Lab 6

Load the following file to practice KNN

KNN_template.xlsx

3.7 CROSS VALIDATION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=66

Exercises

Cross Validation Tutorial

Load the following file for a cross validation tutorial

CrossValidation_Tutorial.xlsx

3.8 RESAMPLING

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=68

3.9 FEATURE SELECTION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=70

4. PRESCRIPTIVE ANALYTICS

Photo by NASA on Unsplash

4.1 PRESCRIPTIVE ANALYTICS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=74

4.2 MINIMUM COST NETWORK FLOW PROBLEM (MCNFP)

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=76

Exercises

Lab 8

Load the following files to practice MCNFP

Minimum CostFlow Problem.docx

Minimum CostFlow_Template.xlsx

4.3 ROUTING

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=78

Exercises

Lab 9

Load the following file to practice routing

TSP_Template.xlsx

4.4 SIMULATION

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=80

Exercises

Lab 10 – Part 1 of 4

Load the following file to practice simulation

Pierre's Bakery.xlsx

4.5 INVESTMENT MANAGEMENT

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=82

Exercises

Lab 10 – part 2 of 4

Load the following files to practice investment management

Fisherperson.xlsx

NPV.xlsx

4.6 STOCHASTIC DECISION TREE ANALYSIS

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=84

;

Exercises

Lab 10 – part 3 of 4

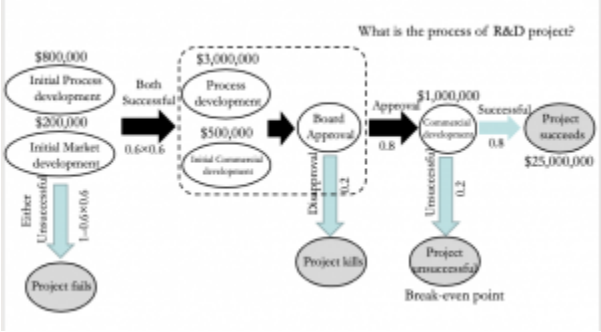
Load the following file to practice stochastic decision tree analysis

Stochastic Decision Tree.xlsx

ASSIGNMENT 3

Case Study – Applying Stochastic Optimization

We recommend using an Ivey Case Study, Research and Development at ICI: Anthraquinone (1999, revised 2010) by Peter C. Bell. This case is available for purchase, per person for a low cost (CAD 9.00) and has an accompanying Microsoft Excel model available.



4.7 REVENUE MANAGEMENT

One or more interactive elements has been excluded from this version of the text. You can view them online here: https://ecampusontario.pressbooks.pub/dataanalyticsvls1/?p=89

Exercises

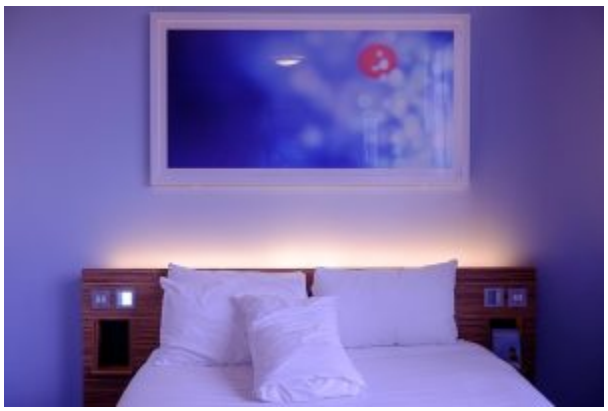
Lab 10 – Part 4 of 4

Load the following file to practice revenue management

Revenue (Airline).xlsx

ASSIGNMENT 4

Revenue Management at a Hotel



A hotel with 100 rooms considers entering into the booking market. There are two demands:

- Early demand: Before one week to the date U(80, 120)
- Late demand: During one week to the date U(10, 50)

Early customers first book and pay $10 which is non-refundable. They can cancel the booking anytime before one week to the date. In the case of finalizing, they pay another $50. Late customers pay $80 which is non-refundable.

The hotel currently plans for booking policy. Past data shows that at least 80% of early customers finalize the booking. If overbooking is realized, the hotel pays a penalty of $150 to each booked customer.

Use the dataset, Assignment Hotel.xlsx, to analyze how the hotel can maximize profit.

Maximize Profit (X) = Booking fee + Early demand profit + Late demand profit – Overbooked cost