

Data 301

What's next?

Nov. 29, 2022

Remaining Term

- Last lab this week (Microsoft Excel)
- You should be finalizing your Project Milestone 5!
- **Today:** Data Science Retrospective + Final Exam Details
- This **Thursday:** Test 4
- Next **Tuesday:** Review Session (unstructured, bring your questions)
- Next **Thursday:** Bonus Test 4

Retrospective

Course Schedule

Course Schedule

This is the tentative plan for COSC 301 this term. The exact details are subject to change, so this is roughly the plan we will try to follow.

#	Week	Topics	Lab	Logs	Project Milestones	Tests (Thursdays, in person)	Concepts Tested
1	Sept 6-11	Introduction to Data Analytics Course Setup	Lab 1	LL1	-	-	-
2	Sept 12-18	Command-line and Jupyter Notebook	Lab 2	LL2	-	Test 0	Course policies
3	Sept 19-25	Introduction to Version Control with Git	Lab 3	LL3	-	-	-
4	Sept 26 - Oct 2	Introduction to Python	Lab 4	LL4	PM 1	Test 1	Command Line and Git
5	Oct 3 - 9	Working with Data in Python	Lab 5	LL5	-	(Bonus Test 1)	-
6	Oct 10 - 16	Rest and Catchup (no new material)	-	-	PM 2	-	-
7	Oct 17 - 23	Programming in Python	Lab 6	LL6	-	-	-

8	Oct 24 - 30	Data Visualization	Lab 7	LL7	-	Test 2	Python Programming
9	Oct 31 - Nov 6	Exploratory Data Analysis	Lab 8	LL8	PM 3	(Bonus Test 2)	-
10	Nov 7 - 13	Reading week (no new material)	-	-	-	-	-
11	Nov 14 - 20	Tableau	Lab 9	LL9	PM 4	Test 3	Python and Pandas
12	Nov 21 - 27	Microsoft Excel	Lab 10	LL10	-	(Bonus Test 3)	-
13	Nov 28 - Dec 4	Project Wrap-up and Review (no new material)	-	LL11	PM 5	Test 4	Data Visualization and Excel
14	Dec 5 - Dec 8	Final Exam Details (no new material)	-	-	-	(Bonus Test 4)	-

Evaluation

The grading scheme for this course is:

Item	Weight	Due date(s)
Learning Logs	10% (10 x 1%)	Sundays at 6 PM
Labs	30% (10 x 3%)	Saturdays at 6 PM
Project	25% (2% / 5% / 5% / 5% / 8%)	Variable
Tests	20% (4 x 5%)	Thursdays during class
Final Exam	15%	TBD (during the exam period)

There will be 4 tests in this course in an (approximately) bi-weekly schedule. For each of the tests, there will be a bonus test one week later and the better score of the Test-Bonus Test pair will be taken.

Passing requirements

- All students must satisfy ALL conditions to pass the course:
 1. Obtain an average grade of at least 50% on the Labs,
 2. Obtain an average grade of at least 50% on the Test and Exam components together,
 3. Obtain a grade of at least 50% on the course Project,
 4. Obtain an average grade of at least 40% on the Final Exam,
 5. Obtain a grade of at least 50% on the whole course.

If students do not satisfy the appropriate requirements, the student will be assigned the **lower** of their earned course grade or, a maximum overall grade of 45 in the course.

Final Exam Details

Final Exam Info

COSC 301 Take-Home Final Exam Information #209



Firas Moosvi STAFF

Last month in Instructor Announcements



217

INPIN

STAR

WATCHING

VIEWS



Hello everyone, thanks for your patience I will be announcing the details about the final exam for COSC 301 in this post so read carefully!

14

Please like this post so I know you've seen it and read it!

Here are some details about the final exam.

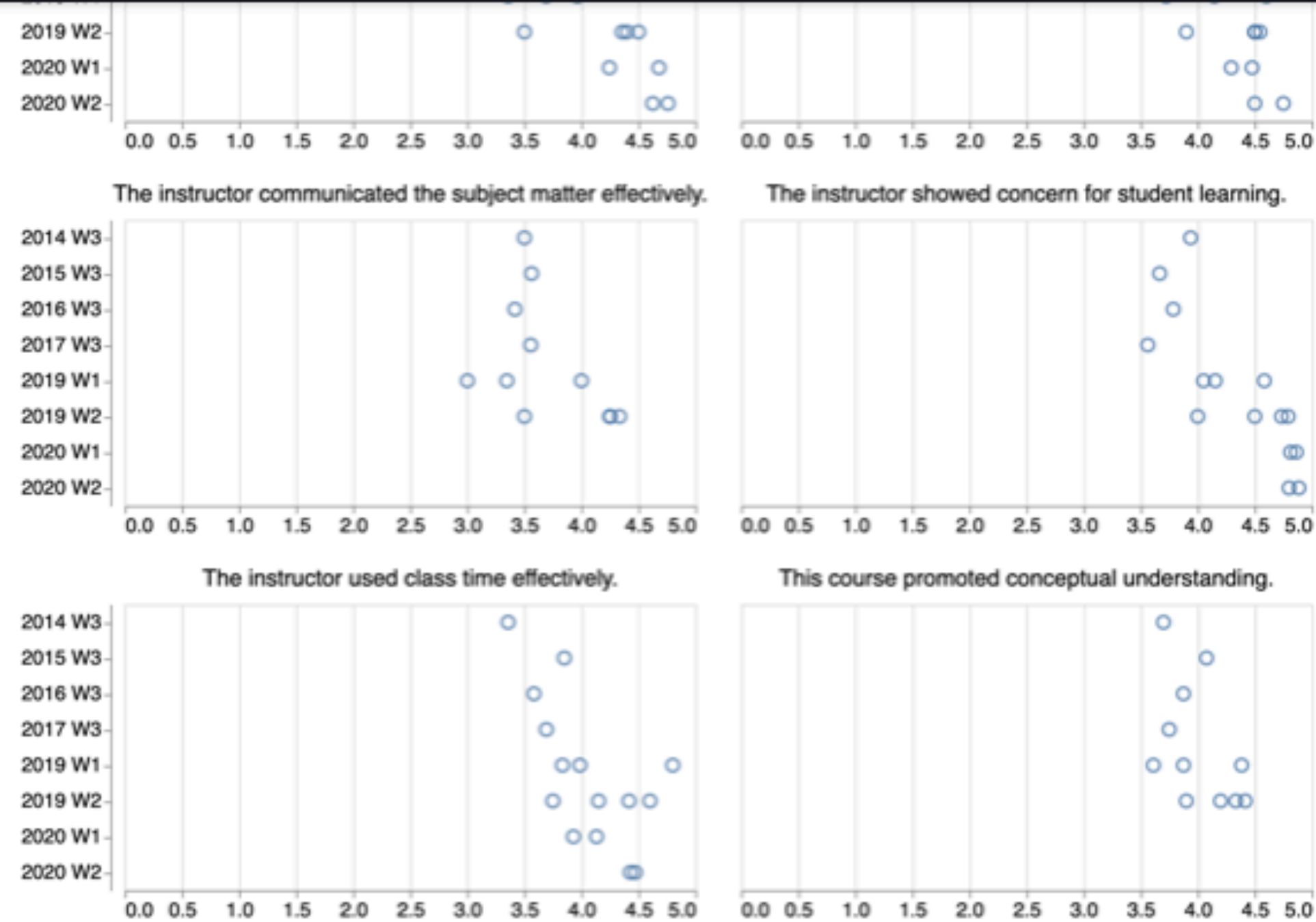
- The final exam will be a take home final exam, delivered as a GitHub Classroom assignment (similar to labs in this course)
- There will be a 72 hour window during which you can start the exam at any time, and you must end it within the 72-hour window.
- The final exam window will **OPEN Sunday December 11th at 14:00 (2 PM)**
- The final exam window will **CLOSE, Dec 14th at 14:00 (2 PM)**.
- The exam **must** be submitted on Gradescope before the close of the final exam window.
- Format: you will get to work with a dataset you have not seen before and be asked to do a comprehensive data analysis, some of it guided, and some of it unguided.
 - You will need to do all the steps of a Data Analysis.
 - You will be given some research questions to answer, based on the data.
 - You will also need to do come up with your own research questions and answer them.
- The exam will not be proctored or invigilated, but the same rules as the Test apply: the exam must be done individually, on your own and will be open-book, open-notes, open-web, no communicating with other humans, no cheating websites like Chegg, CourseHero, Slader, etc...
- If you have questions about the Exam during the window, you can post them on Ed Discussion as a Private question and I will respond to it. For content-related questions, I will usually not be able to answer them.
- Remember that you will need to accept a GH Classroom link (just like with the labs) and then submit your repository link once you are finished with the exam.
- The GH Classroom link will be available on Canvas, inside a "Quiz" called "Final Exam".
- You will also need to commit to the repo and push to GitHub at various points during the exam (I will have instructions in the exam for when you should be committing, and pushing).
- The exam is designed to be completed in about 2.5 hours, but you will have the full 72-hour window to spend on the exam. I highly recommend that you block out a chunk of time and finish the exam in one sitting. You have other exams to deal with as well, and just because I give you 72 hours, does not mean you have to use all of it!! You have other exams as well, so make sure you budget your time and energies accordingly.
- Of course, you're welcome to take breaks (food, sleep, bathroom breaks) etc as needed, you do not need to do the whole exam in one sitting.
- You will need to make sure your JupyterLab installation is functioning, Git and Python is working, and all the packages used in the course are installed. This will be a required aspect of the final exam!
- You will need to frequently commit your work using the Terminal (i.e. NOT GitHub desktop, or the web uploader!) so please make sure you know how to do that. It will be just like the labs and milestones, if you've been keeping up, I don't expect you'll have a problem.
- The exam will contain everything in the course EXCEPT Tableau and Excel. With git, you will be expected to demonstrate proficiency of the basic commands while you are doing the Exam, but there are no specific questions about git.

Important: If you believe that the COSC 301 take-home final will conflict with your other scheduled exams, please contact me ASAP so we can work something out. A conflict will occur when you have 3 or more exams (including COSC 301) within the 72-hour window.

Student Experience of Instruction (SEI)

Search...

Past Courses



I am a strong believer in transparency and openness so on this page, I will also be sharing my full teaching evaluations. This is definitely a moment of vulnerability as these are often considered private. However, I think that students should know exactly what they're walking into when they take my course. I am not the perfect instructor, and my teaching is constantly evolving. I also read and reflect upon every comment that I receive so also keep in mind these comments are from the past. I have definitely made mistakes in the past, and am always seeking to improve and better my teaching, as well as your learning.

Without further ado, my teaching evaluations from the past few terms (I will continue adding to this list as time permits):

- [Physics 111, Winter 2020 Term 1 \(Online\)](#)
- [Data 301, Winter 2020 Term 1 \(Online\)](#)

**You should have received an email that the
“Student Experience of Instruction” (SEI)
is now open for this course.**

**Research shows that SEI are flawed because
they are influenced by unconscious and
unintentional biases.**

Despite their flaws, SEIs are used by departments to:

- **Make decisions on Tenure and Promotion**
- **Decide which courses instructors teach**
- **Rate/rank grant applications and awards**

More important to me however, is how you felt about the course content, the structure, and me as an instructor.

I want to hear from you!

My goal is to get at least a 70% response rate on the SEI, the more the merrier!



Account



Dashboard



Courses



Calendar



Inbox



History



Help

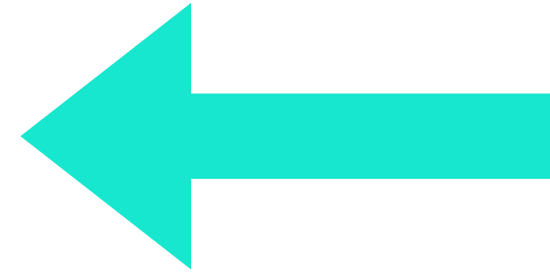


COSC 301 101 2022W1 > Course Evaluation

2022W1

Home

Course Evaluation



Course Content

Discussion

Gradescope

People

PrairieLearn (Tests)

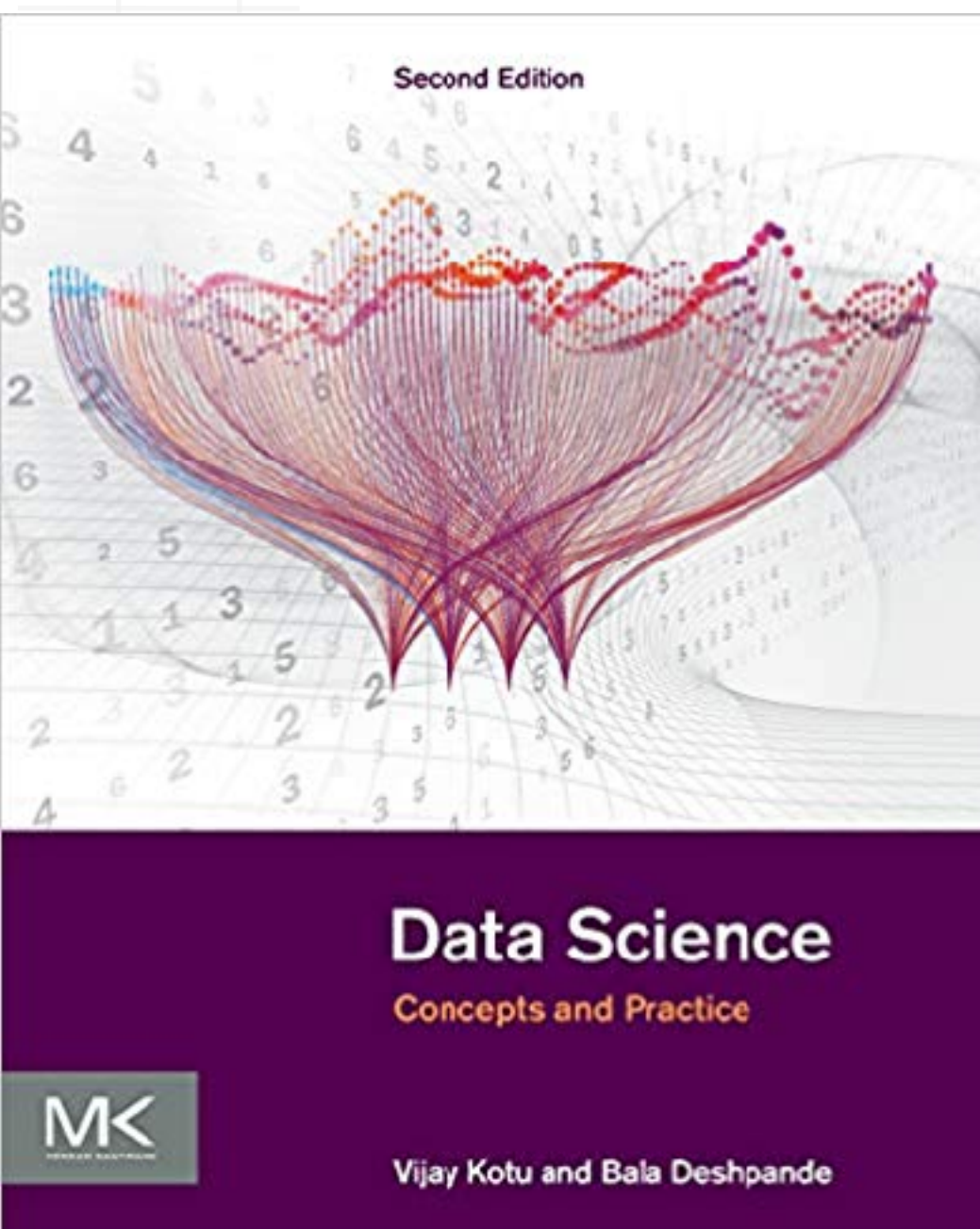
Zoom

Suggestions for how to study for the final exam

Where we are now...

Data Science: Concepts and Practice

Course Book



Data Science: Concepts and Practice

Authors : Vijay Kotu & Bala Deshpande
Publisher : Morgan Kaufmann

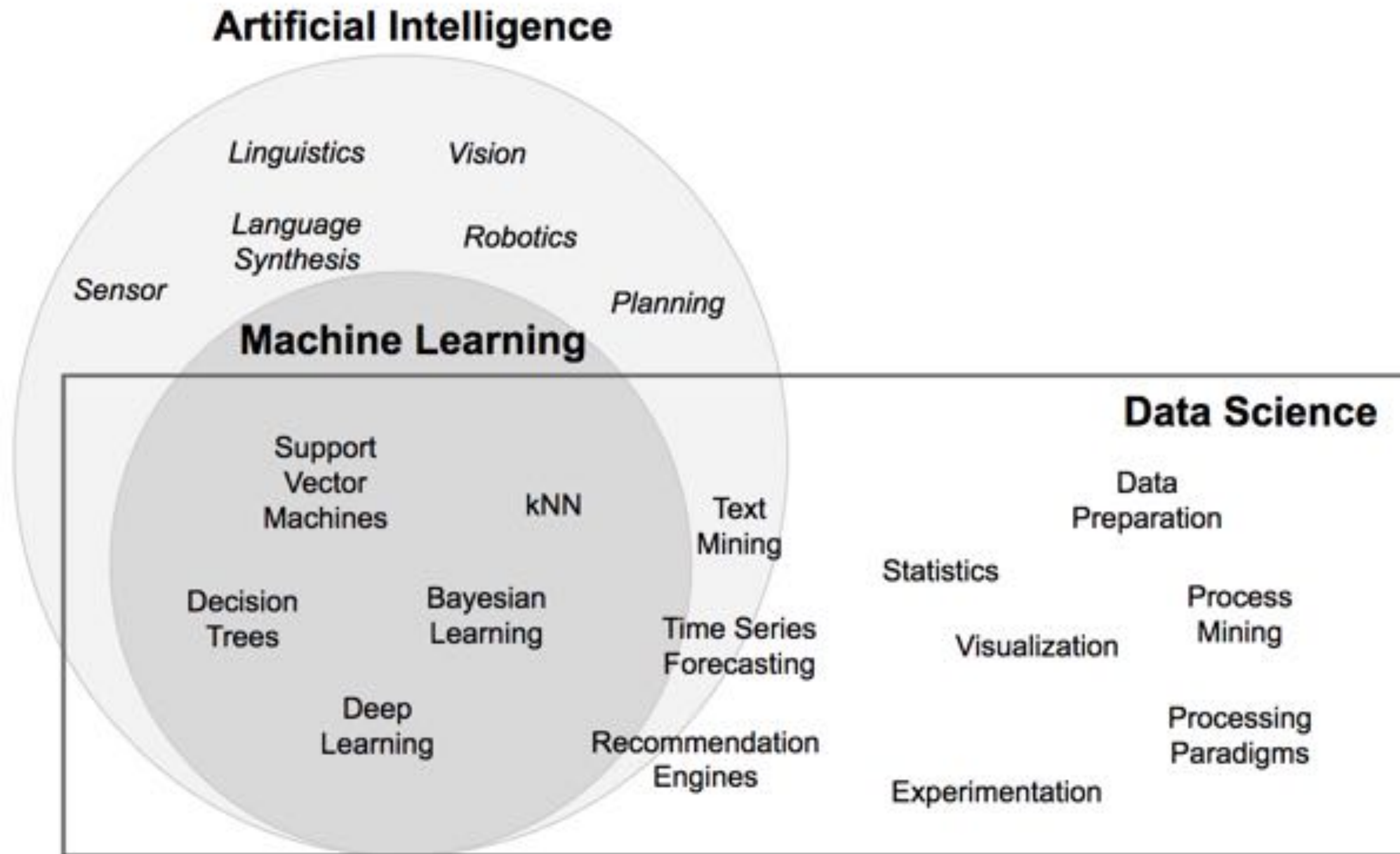
Course Software



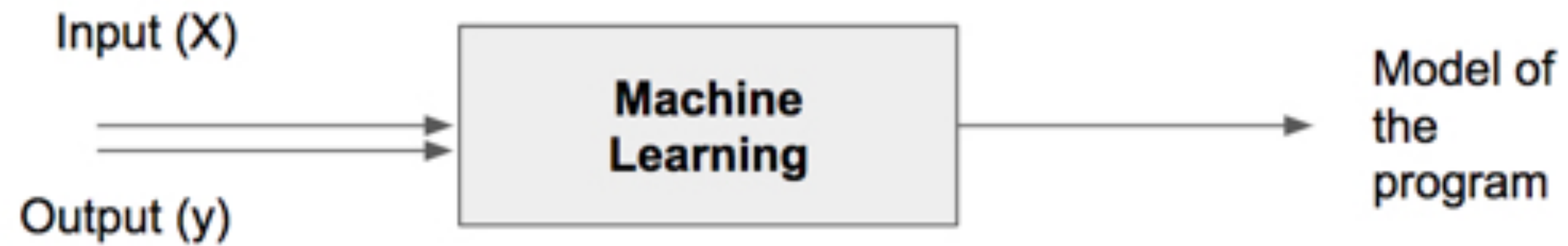
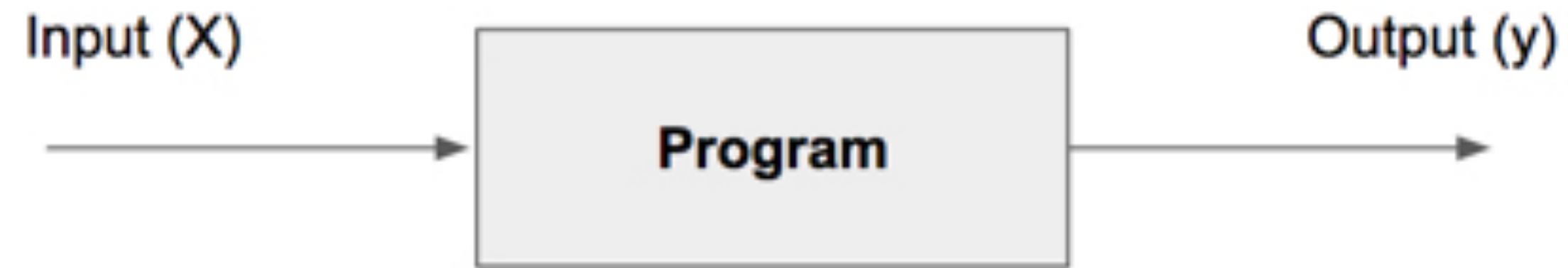
www.rapidminer.com

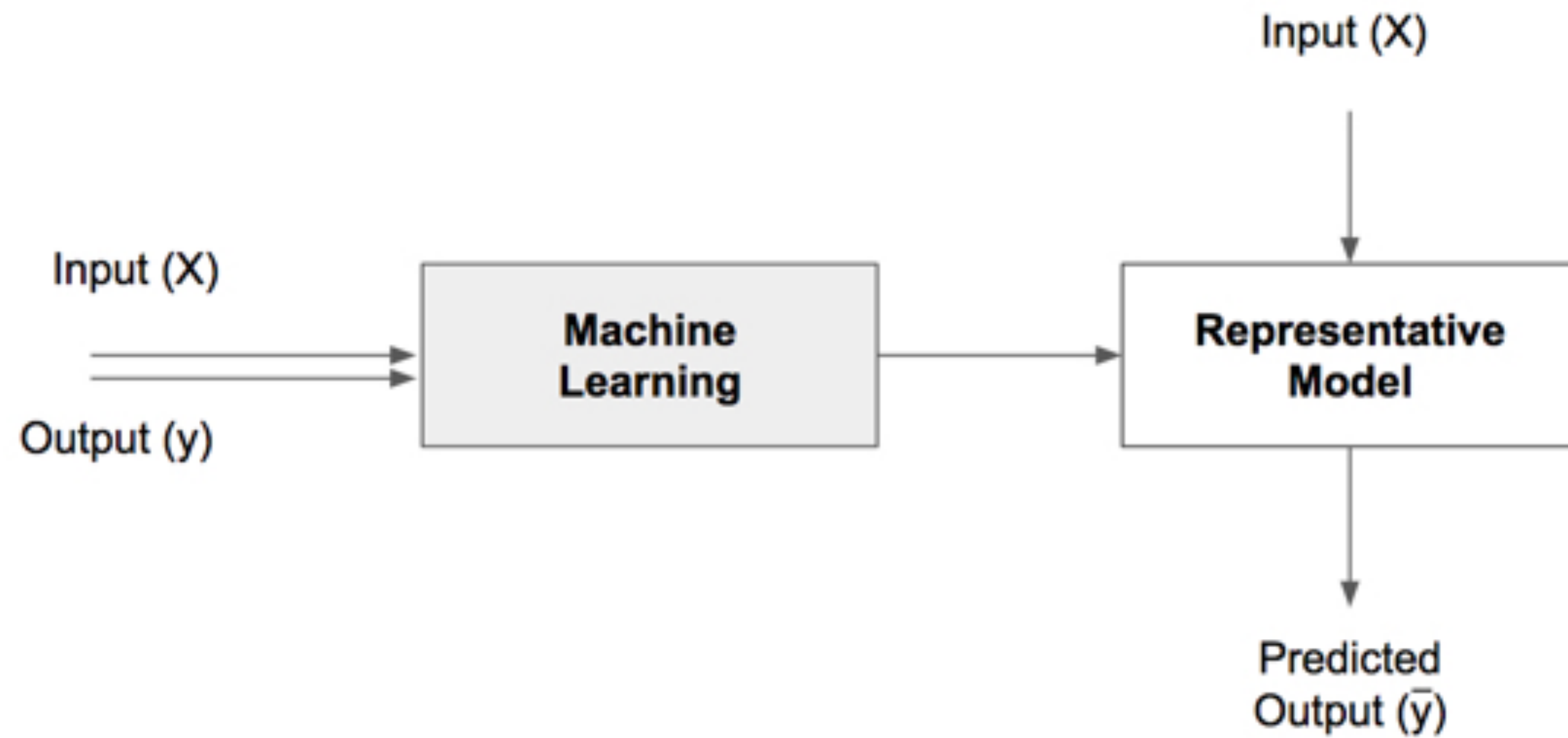
Free Download

What is Data Science

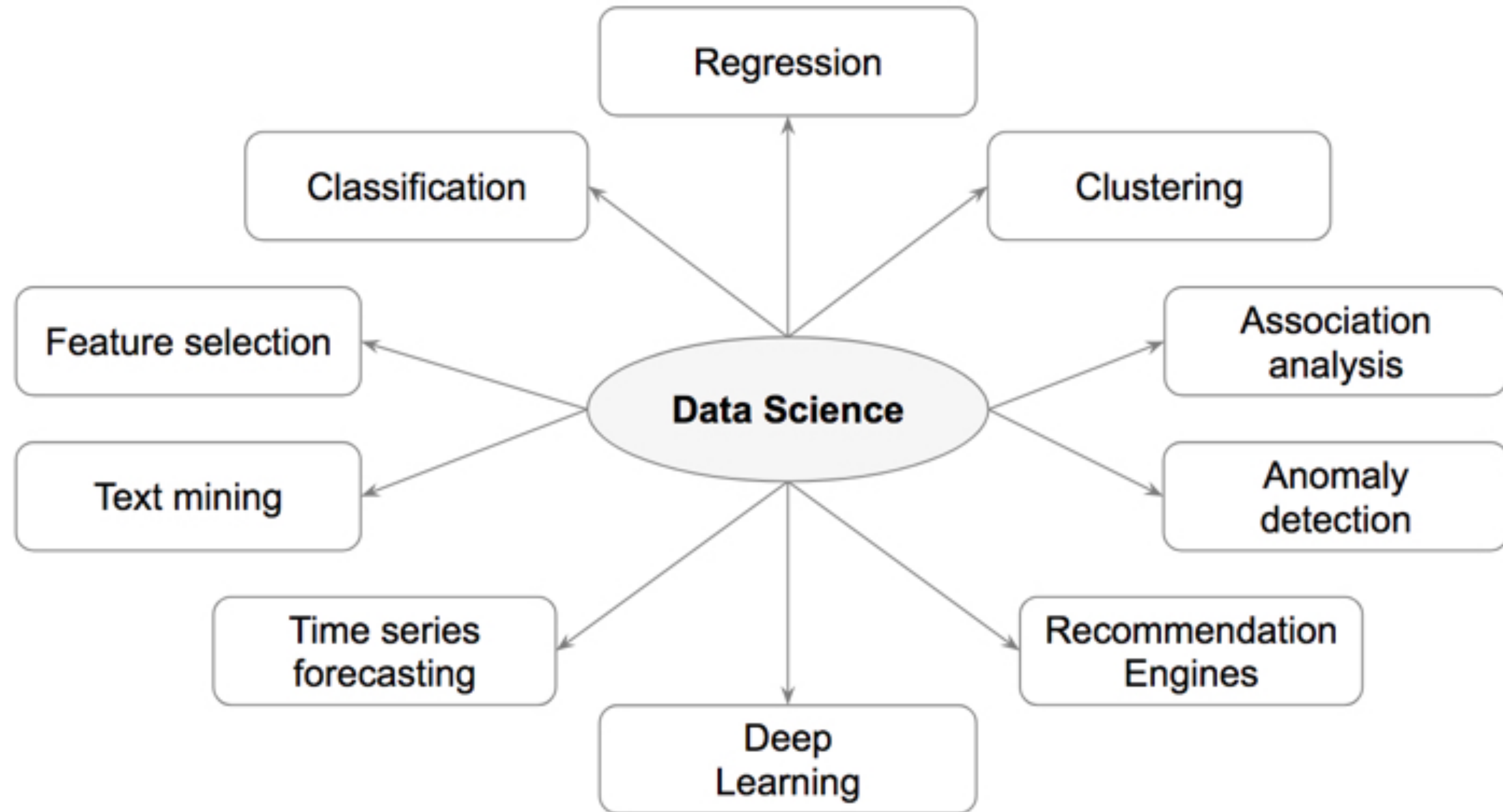


Models





Types of Data Science



Slides courtesy of Vijay Kotu and “Data Science: Concepts and Practice”

Tasks	Description	Algorithms	Examples
Classification	Predict if a data point belongs to one of predefined classes. The prediction will be based on learning from known data set.	Decision Trees, Neural networks, Bayesian models, Induction rules, K nearest neighbors	Assigning voters into known buckets by political parties eg: soccer moms. Bucketing new customers into one of known customer groups.
Regression	Predict the numeric target label of a data point. The prediction will be based on learning from known data set.	Linear regression, Logistic regression	Predicting unemployment rate for next year. Estimating insurance premium.
Anomaly detection	Predict if a data point is an outlier compared to other data points in the data set.	Distance based, Density based, LOF	Fraud transaction detection in credit cards. Network intrusion detection.
Time series	Predict if the value of the target variable for future time frame based on history values.	Exponential smoothing, ARIMA, regression	Sales forecasting, production forecasting, virtually any growth phenomenon that needs to be extrapolated
Clustering	Identify natural clusters within the data set based on inherit properties within the data set.	K means, density based clustering - DBSCAN	Finding customer segments in a company based on transaction, web and customer call data.
Association analysis	Identify relationships within an itemset based on transaction data.	FP Growth, Apriori	Find cross selling opportunities for a retailer based on transaction purchase history.

Course outline

Process Basics

**Data Science
Process**

Data Exploration

Model Evaluation

Core Algorithms

Classification

Decision Trees

Rule Induction

k-Nearest Neighbors

Naïve Bayesian

Artificial Neural Networks

Support Vector Machines

Ensemble Learners

Regression

Linear Regression

Logistic Regression

Association Analysis

Apriori

FP-Growth

Clustering

k-Means

DBSCAN

Self-Organizing Maps

Common Applications

Text Mining

Time Series Forecasting

Anomaly Detection

Feature Selection

What to do at UBCO


Save To Worklist

DATA 101 Making Predictions with Data

Introduction to the techniques and software for handling real-world data. Topics include data cleaning, visualization, simulation, basic modelling, and prediction making.

Credits: 3

- Choose one section from all 2 activity types. (e.g. Lecture and Laboratory)

Status	Section	Activity	Term	Mode of Delivery	Interval	Days	Start Time	End Time	Section Comments	Course Requires In-Person Attendance
	DATA 101 101	Lecture	1	Online		Tue Thu	12:30	14:00	 Section Comments	No

Save To Worklist

DATA 311 Machine Learning

Regression, classification, resampling, model selection and validation, fundamental properties of matrices, dimension reduction, tree-based methods, unsupervised learning. Credit will be granted for only one of STAT 311 or DATA 311.

Credits: 3

Pre-reqs: Either (a) [STAT 230](#) or (b) a score more than 75% in one of [APSC 254](#), [BIOL 202](#), [PSYO 373](#); and one of [COSC 111](#), [APSC 177](#).

- Choose one section from all 2 activity types. (e.g. Lecture and Laboratory)

**Course Requires
In-Person Attendance**

Status	Section	Activity	Term	Mode of Delivery	Interval	Days	Start Time	End Time	Section Comments	Course Requires In-Person Attendance
	DATA 311 101	Lecture	1	Online		Mon Wed	11:00	12:30	 Section Comments	No

WHAT YOU NEED TO KNOW

Data Minor

Program Overview ^

Courses ^

Degree Options ^

The following links take you to data science program details in the Okanagan Academic Calendar, a comprehensive guide to all programs, courses, services, and academic policies at the University of British Columbia:

Bachelor of Science (BSc)

- [Major in data science](#)
- [Data science honours](#)
- [Minor in data science](#)

Communications and Rhetoric Certificate

Students can receive a [Certificate in Communications and Rhetoric](#) by completing 15 credits through courses from four thematic interdisciplinary and relational clusters, and a final capstone project. This is an add on to any major; some credits can be double-counted.

Graduate Studies

An accelerated, 10-month professional [Master of Data Science](#) program is available at the Okanagan campus, or students can choose to pursue MA, MSc or PhD degrees in the [Interdisciplinary Graduate Studies](#) program which offers innovative themed options, as well as individualized options.

Link to resource:

<https://cmps.ok.ubc.ca/undergraduate/data-science/>

Master of Data Science

UBC Master of Data Science

Home Why Data Science? Programs Admissions Why UBC? Employers Contact Us

Apply Now



UBC MASTER OF DATA SCIENCE

Data in Action: Cross-Lingual NER for Low-Resource Languages Student Capstone Project

Working with Seattle-based AI start-up, Seasalt.ai, students from UBC's Master of Data Science in Computational Linguistics program created a universal NER (Named Entity Recognition) system that applied transfer learning from high-resource language datasets to low-resource languages. This allowed crucial information to be extracted from previously underrepresented languages, like Indonesian, Javanese, Malay, Vietnamese, Tagalog, Croatian, and Czech, for use across a variety of Natural Language Processing tasks.

Read more

Have questions about the MDS Program? Attend our Admissions Q&A.
[Register Here.](#)

Applications for **September 2022** are now open. Deadline is **January 31, 2022.**
[Review Admission Requirements.](#)

UBC Master of Data Science / Turning data into knowledge

Turning data into knowledge

Data is everywhere. Continuously generated and collected across every domain, it is a vast and largely untapped resource of information with the potential to reveal insights about every aspect of our lives and the world we live in. However, the ability to uncover these insights is a highly specialized skill possessed by far too few.

UBC's Master of Data Science program was designed to address this workforce gap by equipping students with the technical skills, practical experience, and most importantly, the confidence to seize opportunities in an ever-expanding field.

The program is offered at both the UBC Vancouver campus and the UBC Okanagan campus, with an additional Master of Data Science in Computational Linguistics offered at the Vancouver campus.

Explore Data Science

"It's estimated that Canada will see a shortage of up to 19,000 professionals with data and analytical skills."
- Canada's Big Data Consortium

See you on Thursday!