Data 301

What’s next?

Nov. 29, 2021

Remaining Term

- Last lab this week
- **Today**: Data Science Retrospective
- **Wednesday**: Final Exam Details
- **Friday**: Review Session (unstructured, bring your questions)
- **Monday**: Cancelled
- **WEDNESDAY**: Cancelled

Retrospective

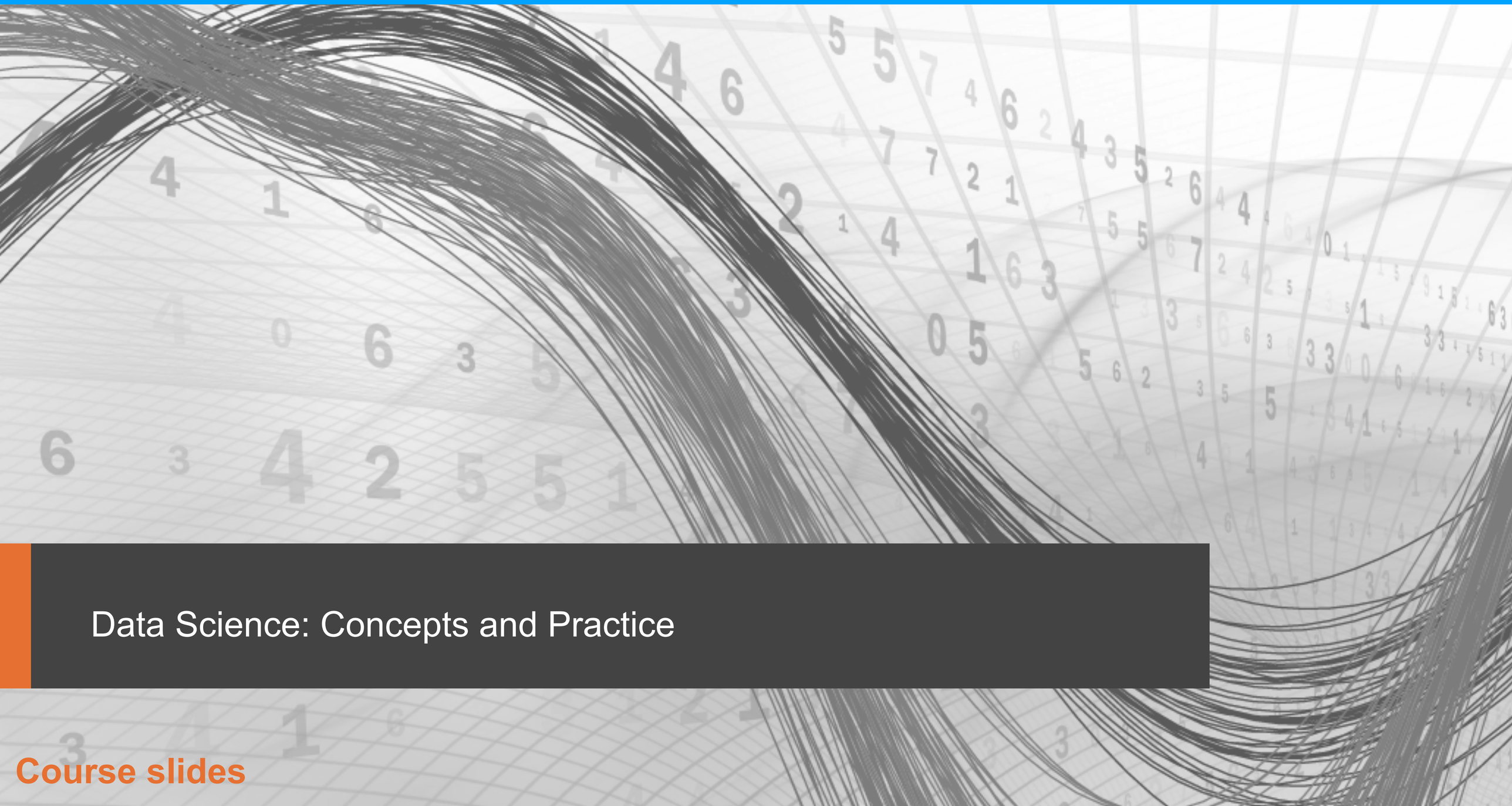
Course Schedule

This is the tentative plan for DATA 301 or 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	Concepts Tested
1	Sept 6-12	Introduction to Data Analytics Course Setup	-		-	Test 0	Course policies
2	Sept 13-19	Command-line and Jupyter Notebook	Lab 1	LL1	-	(Bonus Test 0)	-
3	Sept 20-26	Introduction to Version Control with Git	Lab 2	LL2	PM 1	-	-
4	Sept 27 - Oct 3	Introduction to Python	Lab 3	LL3	-	Test 1	Command Line and Git
5	Oct 4 - 10	Exploratory Data analyis	Lab 4	LL4	PM 2	(Bonus Test 1)	-
6	Oct 11 - 17	Rest and Catchup (no new material)	-	-	-	-	-
7	Oct 18 - 24	Programming in Python	Lab 5	LL5	-	Test 2	Python Programming
8	Oct 25 - Oct 31	Data Visualization	Lab 6	LL6	PM 3	(Bonus Test 2)	
9	Nov 1 - 7	Choosing an appropriate visualization	Lab 7	LL7	-	Test 3	Python and Pandas
10	Nov 8 - 14	Reading week (no new material)	-	-	-	(Bonus Test 3)	-
11	Nov 15 - 21	Tableau and Microsoft Excel	Lab 8	LL8	PM 4	-	-
12	Nov 22 - 28	Microsoft Excel & Project Wrap-up	Lab 9	LL9	PM 5	Test 4	Data Visualization and Excel
13	Nov 29 - Dec 8	Project Wrap-up and Review (no new material)	Lab 10	LL10	-	(Bonus Test 4)	-

Where we are now...

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

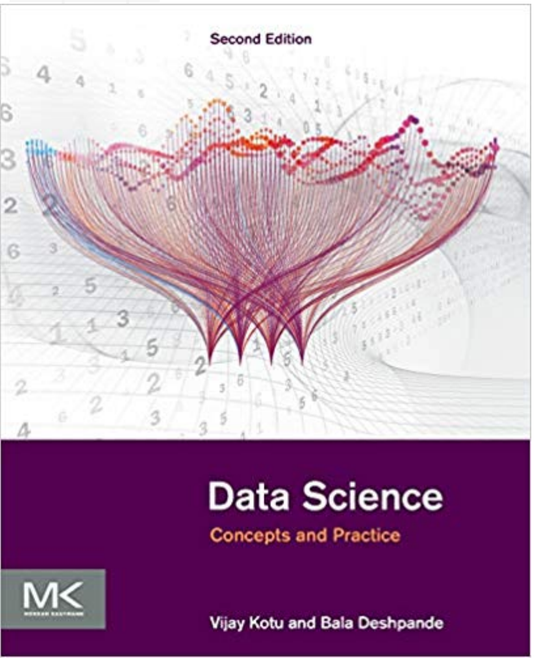


Data Science: Concepts and Practice

Course slides

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

Course Book



Data Science: Concepts and Practice

Authors : Vijay Kotu & Bala Deshpande
Publisher : Morgan Kaufmann

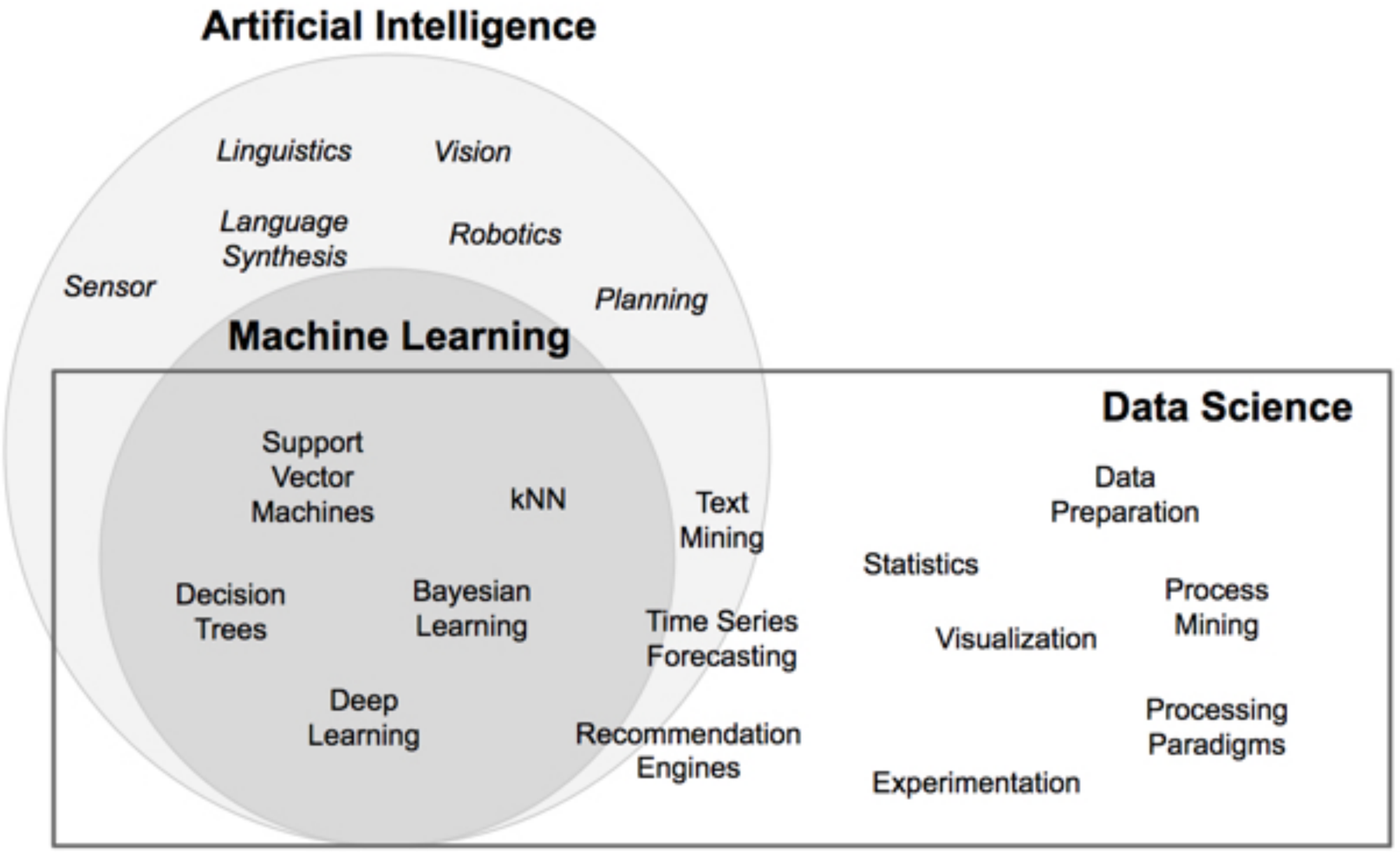
Course Software

rapidminer

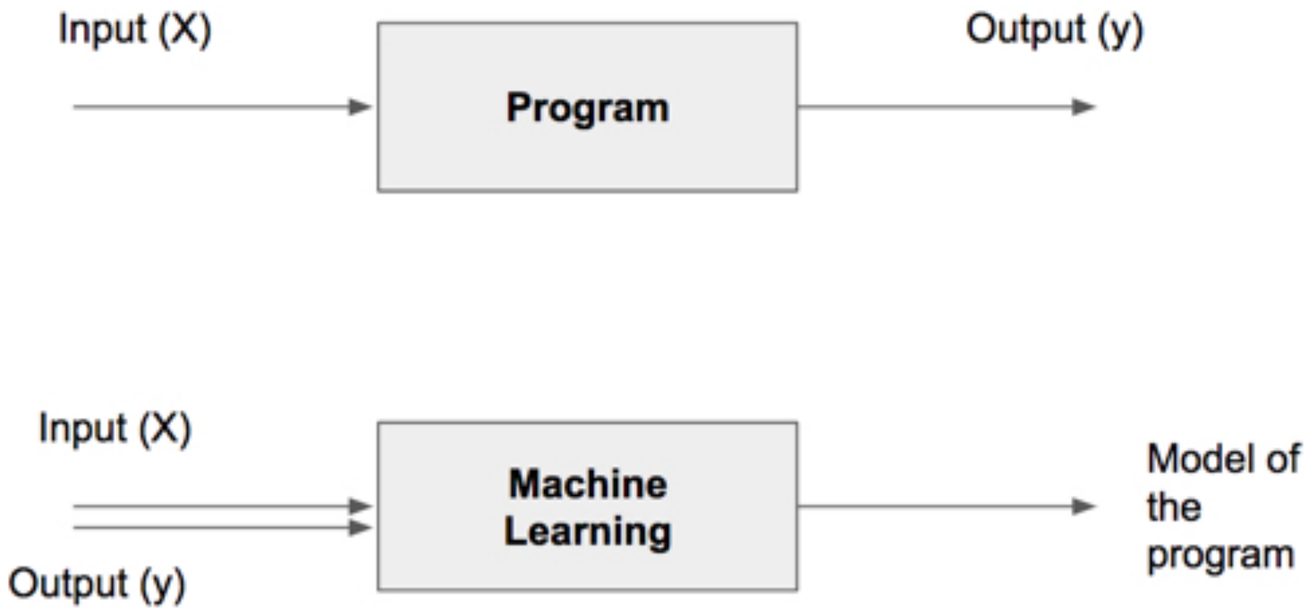
www.rapidminer.com

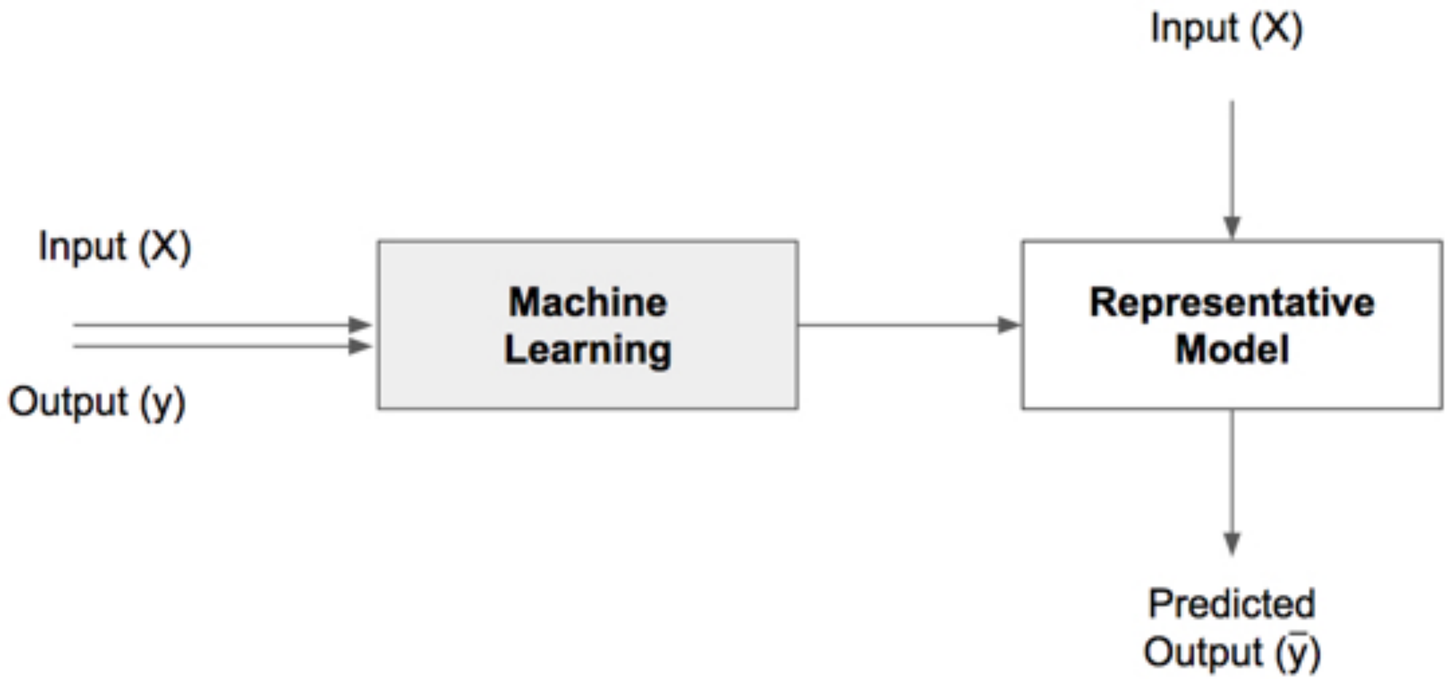
Free Download

What is Data Science

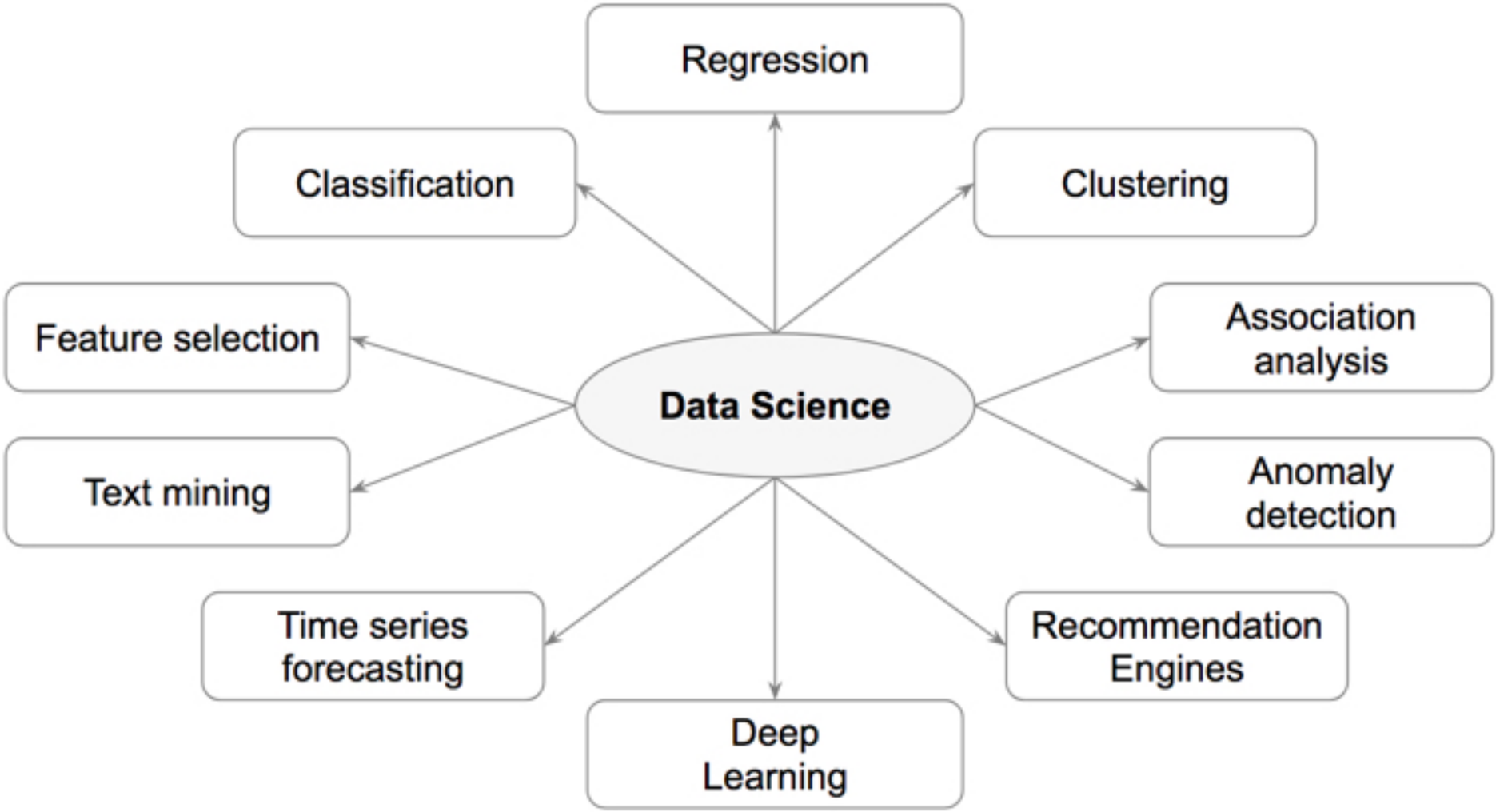


Models


Input (X)
Program
Output (y)
Input (X)
Output (y)
Machine Learning
Model of the program


Input (X)
Output (y)
Machine Learning
Representative Model
Input (X)
Predicted Output (ȳ)

Types of Data Science



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 retailor based on transaction purchase history.

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

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

Data 101

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

Data 311

Course Schedule / Browse Courses / DATA / DATA 311

Campus: UBC Okanagan | Session: 2021 Winter

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)

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