CARTE-Enbridge Bootcamp

Introduction

Welcome!

- My name is Alex Olson
- Senior Research Associate at CARTE
- Bachelor’s in AI from the University of Edinburgh
- Master’s in AI from UofT in collaboration with the School of Cities
- Published papers in collaboration with a wide array of disciplines
- Work closely with students and faculty on all types of AI

Our Teaching Assistants

- Nakul Upadhya
- PhD candidate with Prof. Eldan Cohen
- Researching deep learning and optimization

- Rahul Patel
- PhD candidate with Prof. Elias Khalil
- Researching stochastic and multi-objective optimization for machine learning

Wi-Fi Access

- Eduroam Visitor Access (EVA)
- *Permanent* username and password
- Must be re-authorized each day you visit UofT
- New code each day, same phone number

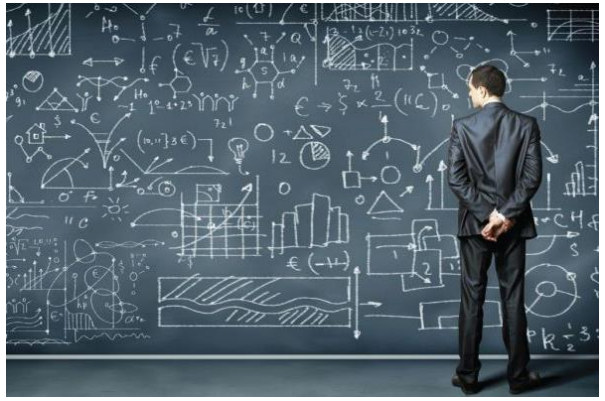
Text `52utoronto`

To `(833) 338-7626`

Introduction

- Analytics play a crucial role in modern decision making across various industries
- Today, we'll explore the basics of analytics and their practical applications

A continuum of analytics

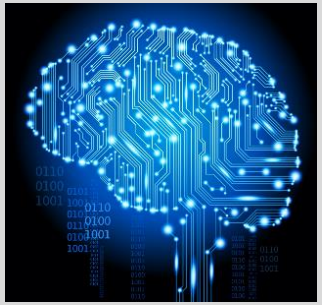
Descriptive	Predictive	Prescriptive
Information →	Insight →	Decisions
		
Use statistics to understand data	Use machine learning to forecast unknown info	Use optimization to make data-driven decisions

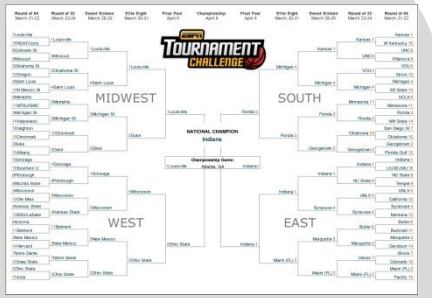
My view on analytics

- Analytics means a lot of different things to different people
  - It is a broad, all-encompassing term
- To me, analytics is about using sophisticated quantitative approaches to understand data, derive insights, and influence decision-making
  - Includes stats, machine learning, optimization, computer science, data visualization, etc.
- Many people consider analytics a rebranding of existing methods

What is analytics?

- Analytics is about using models and data to make better decisions





What is analytics?

- Systematic approach to making informed decisions
- Three key components:
  - Data: Can be quantitative or qualitative
  - Model: Representation of the problem
  - Decision: Identifying course of action using data and model
- Benefits:
  - Improves decision-making accuracy
  - Supports more objective and evidence-based decisions
  - Helps identify trends and patterns in complex situations
- Limitations:
  - Models are simplified versions of reality
  - Models are only as good as the data they're based on
  - Overreliance on models can lead to a lack of critical thinking

Data Collection and Quality

- High-quality data is essential for accurate and reliable analysis
- Common data collection techniques:
  - Surveys and questionnaires
  - Interviews
  - Observations
  - Existing data sources, such as databases or reports
- Data reliability: consistency of measurement
- Data validity: accuracy of the measurements
- To ensure data quality:
  - Use reliable and valid measurement tools
  - Collect data from a representative sample
  - Implement data cleaning procedures to identify and correct errors

Sports Analytics

Forbes / SportsMoney

The Little Black Book of Billionaire Secrets

AUG 18, 2015 @ 03:08 PM 16,089 VIEWS

CHANGING THE GAME: The Rise of Sports Analytics

IBM Watson Teams With Toronto Raptors On Data-Driven Talent Analysis

Posted Feb 10, 2016 by Ron Miller (@ron_miller)

NHL | TRAVIS YOST Feb 9

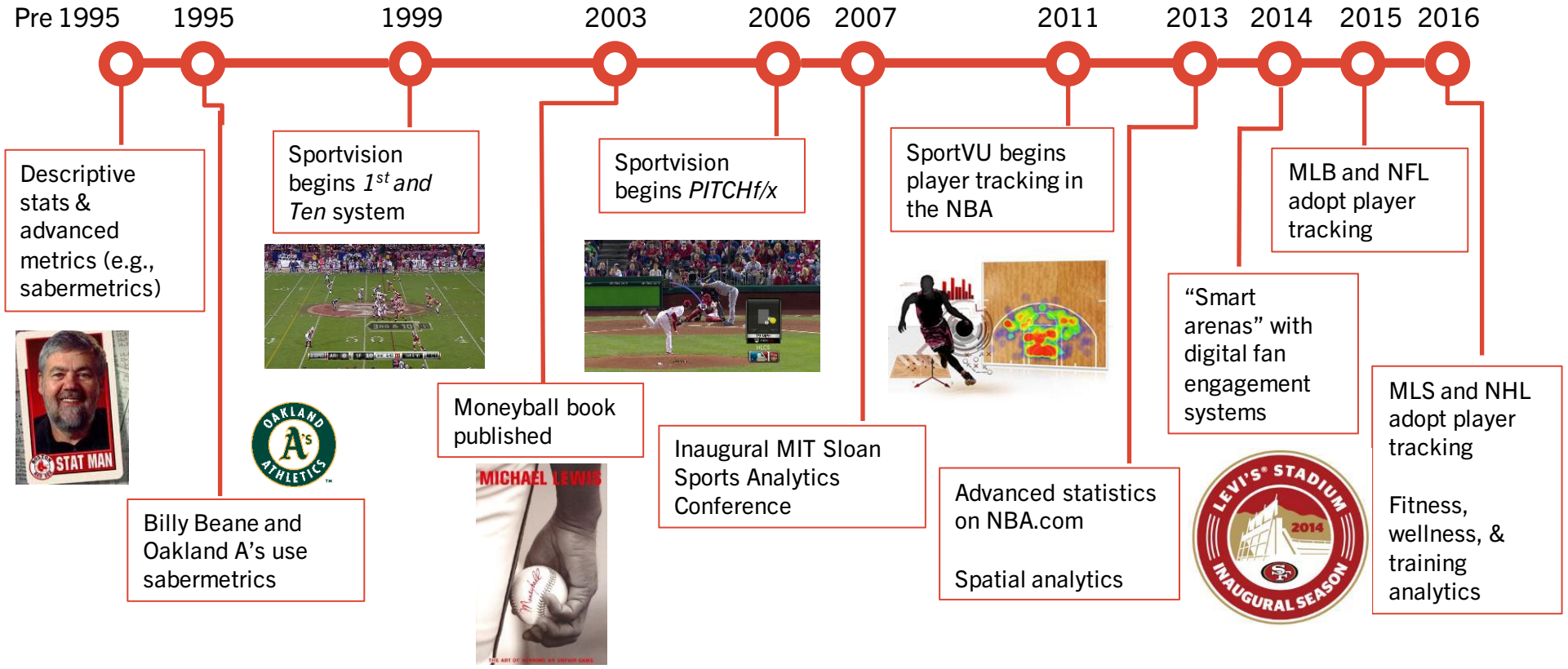
For the teams paying attention, analytics continue to add real value

Maple Leafs creating analytics department: report

New hires to include head of ExtraSkater.com blog

CBC Sports Posted: Aug 19, 2014 2:08 PM ET | Last Updated: Aug 19, 2014 3:56 PM ET

Sports analytics timeline

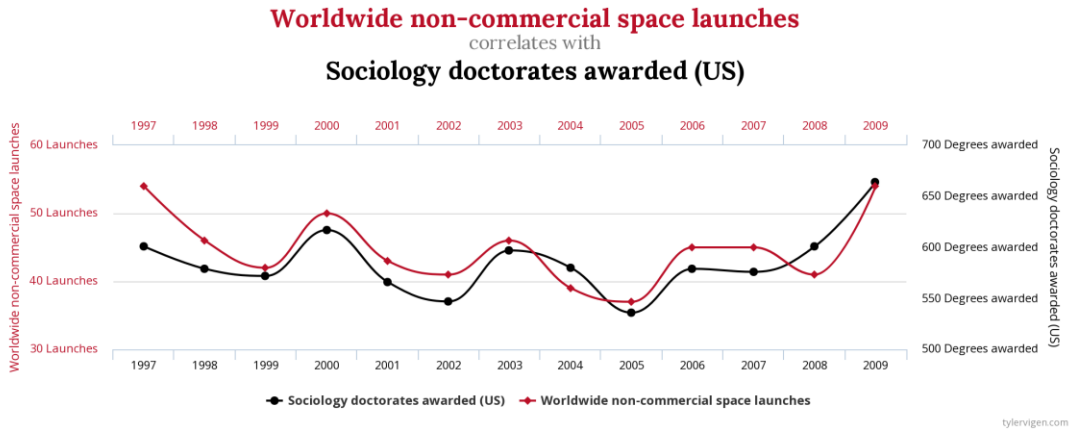


Correlation vs. Causation

- Correlation measures the strength and direction of a relationship between two variables
- Causation refers to a cause-and-effect relationship, where one variable directly influences the other
- It's crucial to remember that a strong correlation doesn't necessarily imply causation

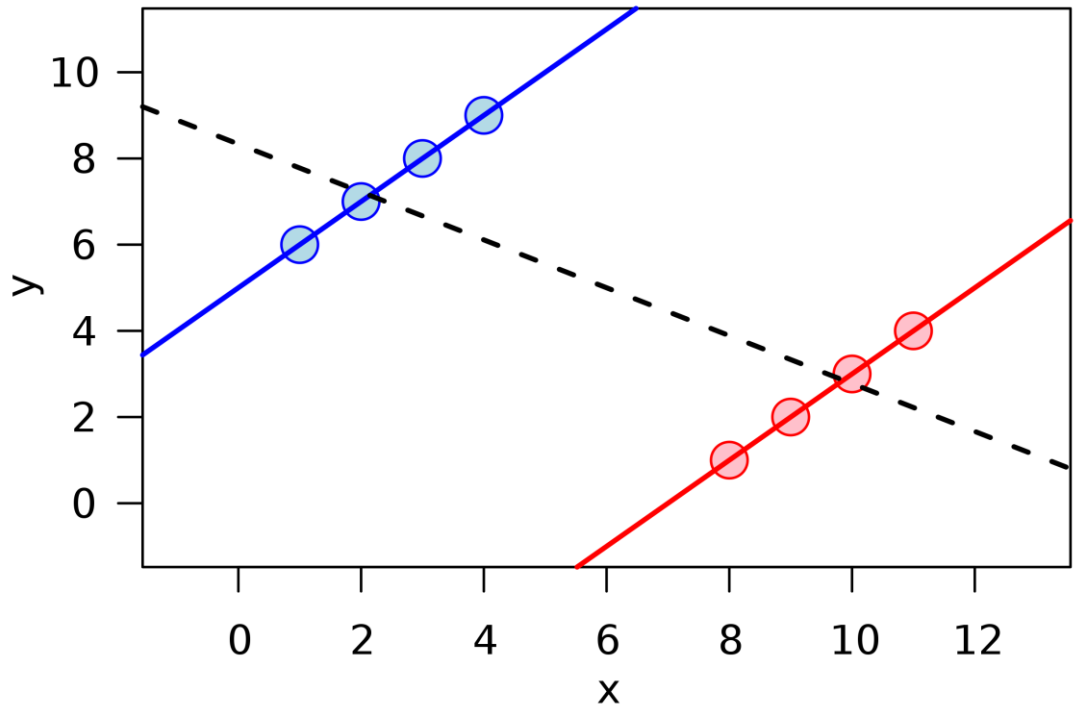
Correlation vs. Causation

- To avoid confusion between correlation and causation:
  - Consider possible confounding variables or third factors
  - Look for evidence of a causal mechanism
  - Test the relationship using controlled experiments or statistical methods



Simpson’s Paradox

- A trend or relationship between two factors seems to exist when you look at separate groups but disappears or even reverses when you combine the groups together.
- To avoid Simpson's Paradox:
  - Investigate data at different levels of aggregation
  - Consider the influence of confounding variables
  - Use caution when combining data from different sources or groups



Simpson's Paradox

- In 1973, UC Berkeley found that men applying were more likely to be admitted than women

	All	Men	Women
Applicants Admitted	41%	44%	35%

Simpson's Paradox

- In 1973, UC Berkeley found that men applying were more likely to be admitted than women
- But when analyzed at a department level, they found only a small subset of departments with a lot of applicants were biased
- Solving the problem required a targeted approach, not a general one

Department	All	Men	Women
A	64%	62%	82%
B	63%	63%	68%
C	35%	37%	34%
D	34%	33%	35%
E	25%	28%	24%
F	6%	6%	7%
Applicants Admitted	**39%**	**45%**	**30%**

Confirmation Bias

- Tendency to seek out, and remember information that supports one's existing beliefs
- Can distort data analysis and lead to inaccurate conclusions
- Impact:
  - Selective data collection: Focusing on evidence that supports a hypothesis while ignoring contradicting evidence
  - Biased interpretation: Favoring analysis results that align with preconceived notions
  - Memory distortion: Recalling information that supports one's beliefs more readily than contradictory information
- Techniques to minimize confirmation bias:
  - Actively seek out disconfirming evidence
  - Conduct blind data analysis, where the researcher is unaware of the group labels or expected outcomes
  - Use objective methods and criteria for data collection and analysis

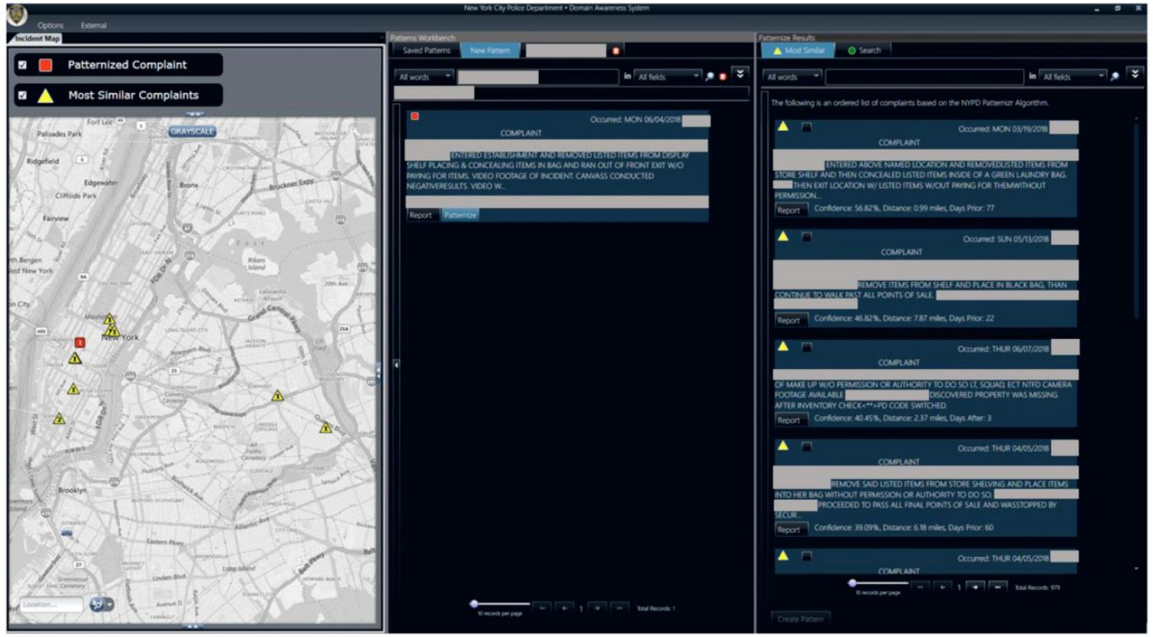
Predictive policing

"The application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crime by making statistical predictions." - Perry et al., 2013

Patternizr

- Decision support tool for identifying crime patterns used by NYPD since 2018
- Uses predictive analytics (machine learning) to identify other crimes that may be part of the same pattern given a “seed” complaint

Figure 6. (Color online) Screenshot of Patternizr in the NYPD Domain Awareness System



Notes. The map of the seed and most similar complaints is in the left panel, the seed complaint is in the middle panel, and the most similar complaints are in the right panel. Grey boxes obscure law enforcement sensitive and personally identifiable information.

Impact

- Perpetrator shoplifting power drills and attacked employee with hypodermic needle when confronted
- Analyst ran complaint through Patternizr and discovered similar case in distant precinct
  - Same robbery subtype (shoplifting), several matching words in narrative (drill, needle)
- Combined these two complaints into pattern, along with two other larcenies by the same person
- Passed to detective squad, conducted an investigation, and arrested the perpetrator who later pled guilty to larceny and felony assault

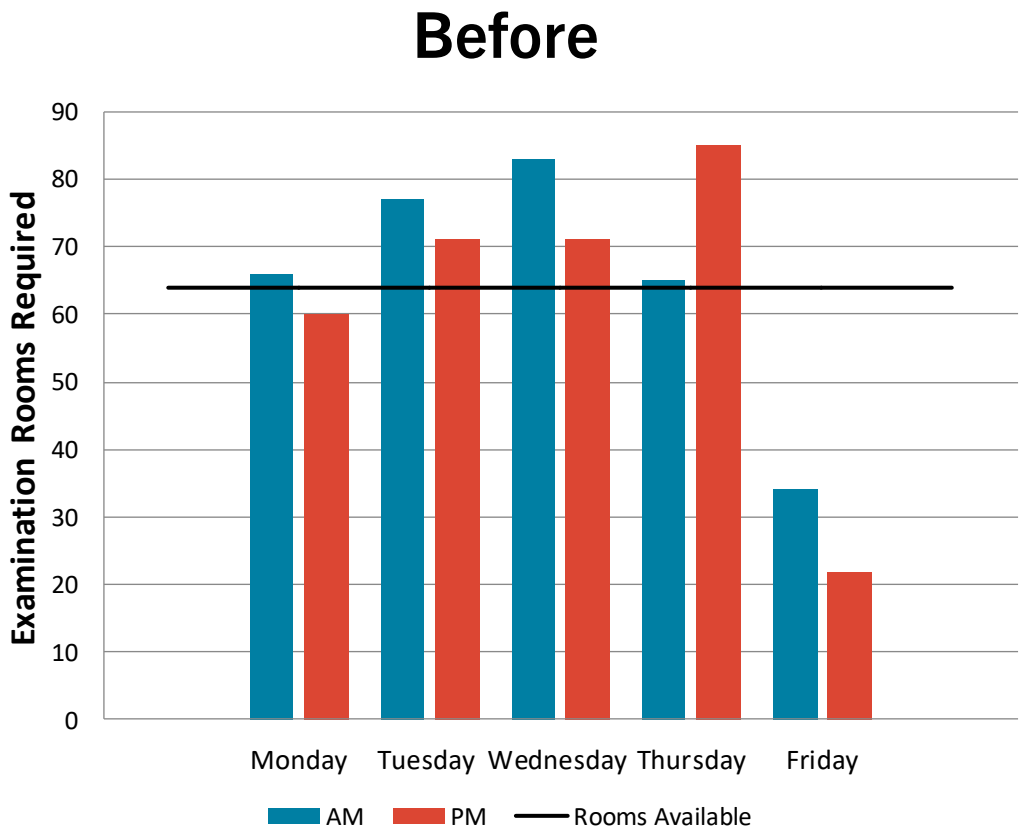
Testimonial

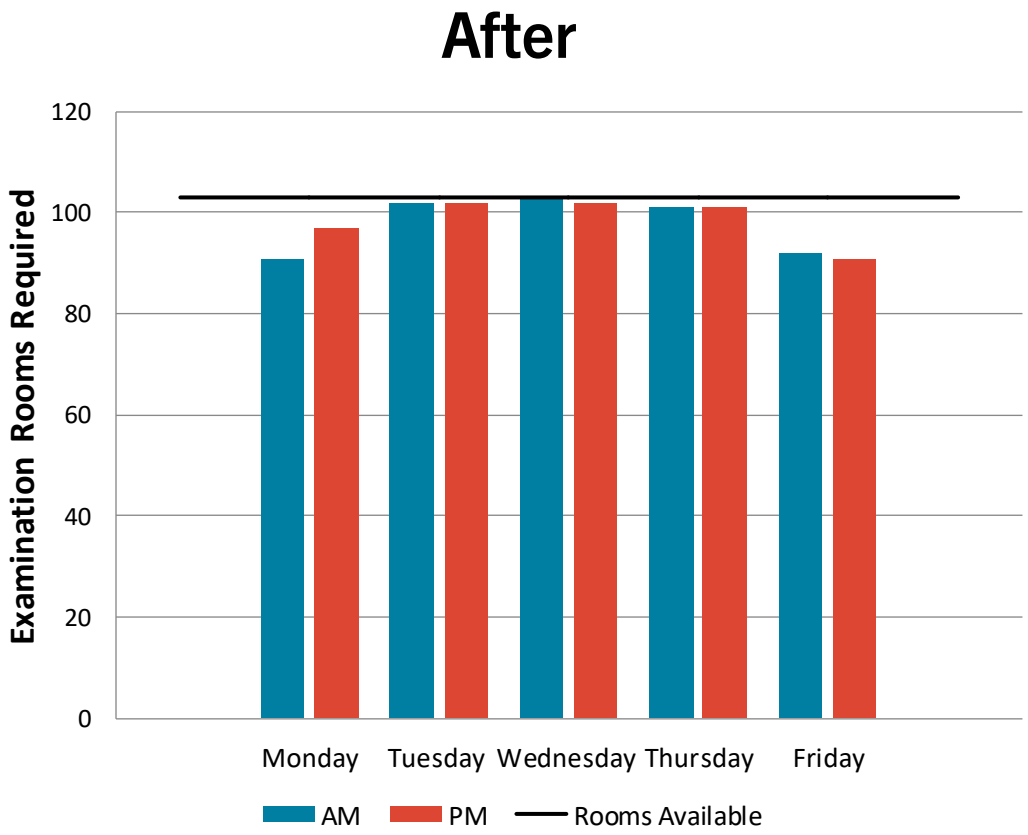
- “Patternizr dramatically improves efficiency compared to traditional methods, and it still allows the analysts that work for me to apply their own thinking and analysis. The science doesn’t overwhelm the art.” — Debra Piehl, NYPD Senior Crime Analyst

Scheduling / HR redeployment

- Women's College Hospital opened new hospital in 2016
- Required moving 300+ weekly clinics from old space into new shared space, necessitating new schedule for thousands of individuals
- Prescriptive analytics (optimization) used to redesign schedule from scratch

Results


Before
Examination Rooms Required
90
80
70
60
50
40
30
20
10
0
Monday
Tuesday
Wednesday
Thursday
Friday
AM
PM
Rooms Available

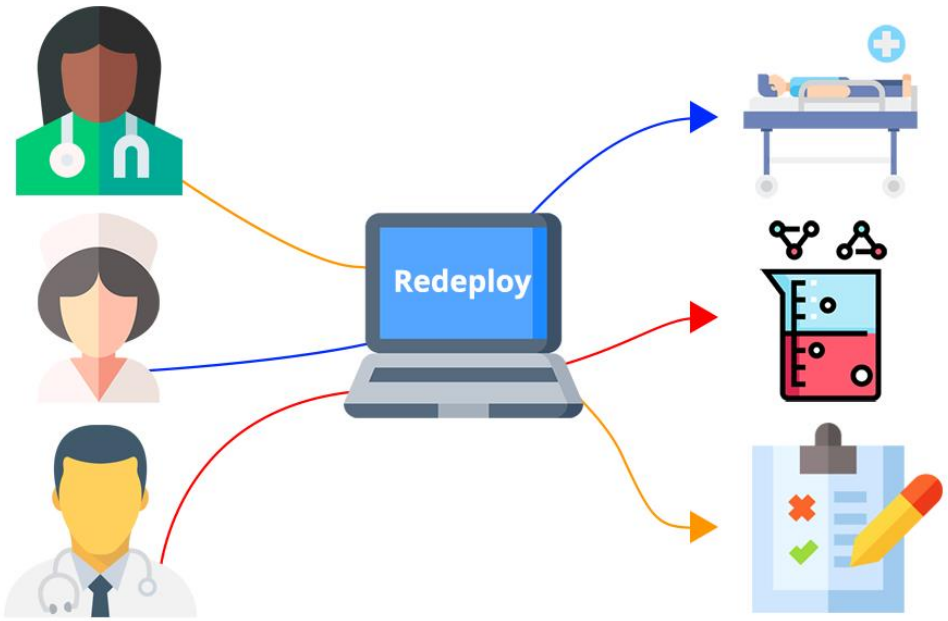

After
Examination Rooms Required
120
100
80
60
40
20
0
Monday
Tuesday
Wednesday
Thursday
Friday
AM
PM
Rooms Available

Impact

- Centralized control of schedule, including optimality, fairness, transparency
- Better resource utilization and space sharing
  - Exam space decreased by over 20% but managed to fit all clinics in new space with 97% utilization
- Proactive planning, new clinics accommodated more easily

- “The model developed by the team provided data-driven decision-making guidance that otherwise would not be possible through a manually designed schedule.” – Heather McPherson, EVP WCH

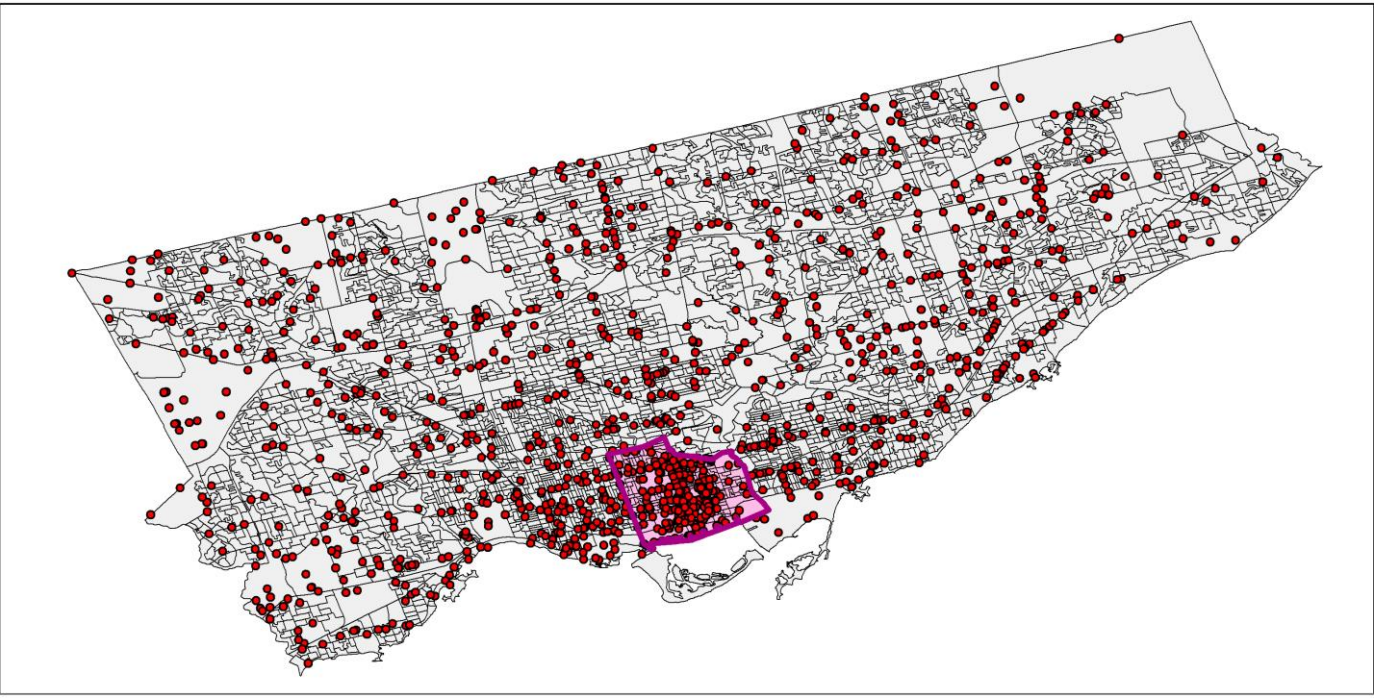
Scheduling / HR redeployment



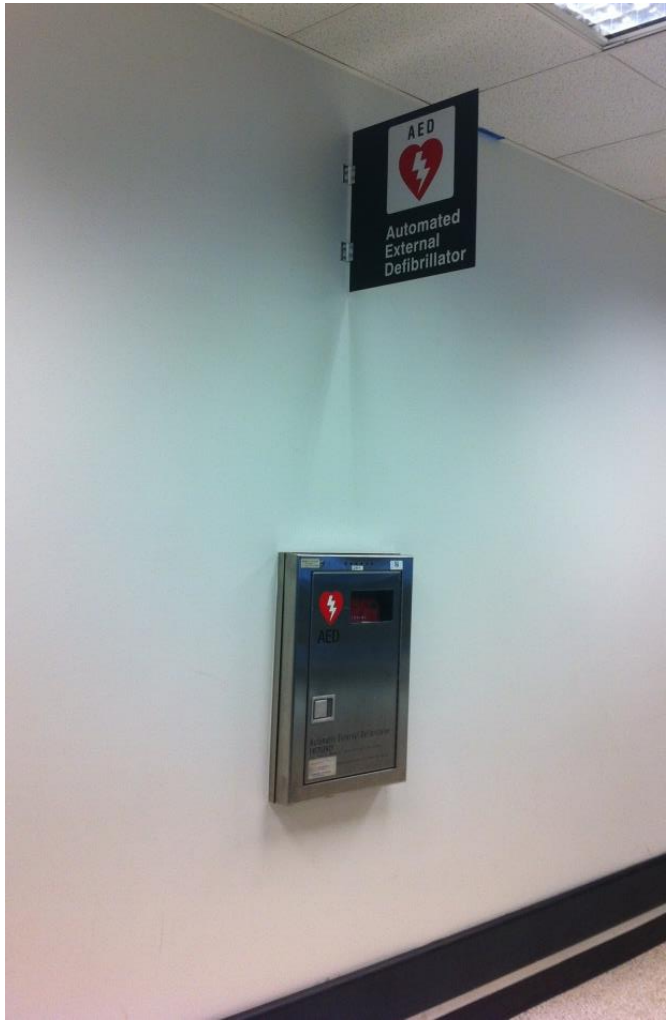
- Hospitals were redeploying hundreds of hospital staff during first wave of COVID-19
- Prescriptive analytics (optimization) tool developed to help automate this process
- https://redeploy.ca

Emergency response and logistics

- Out-of-hospital cardiac arrest kills 400K in North America annually
- 5-10% survival; survival drops 10% per minute delay
- CPR and defibrillation can increase survival dramatically



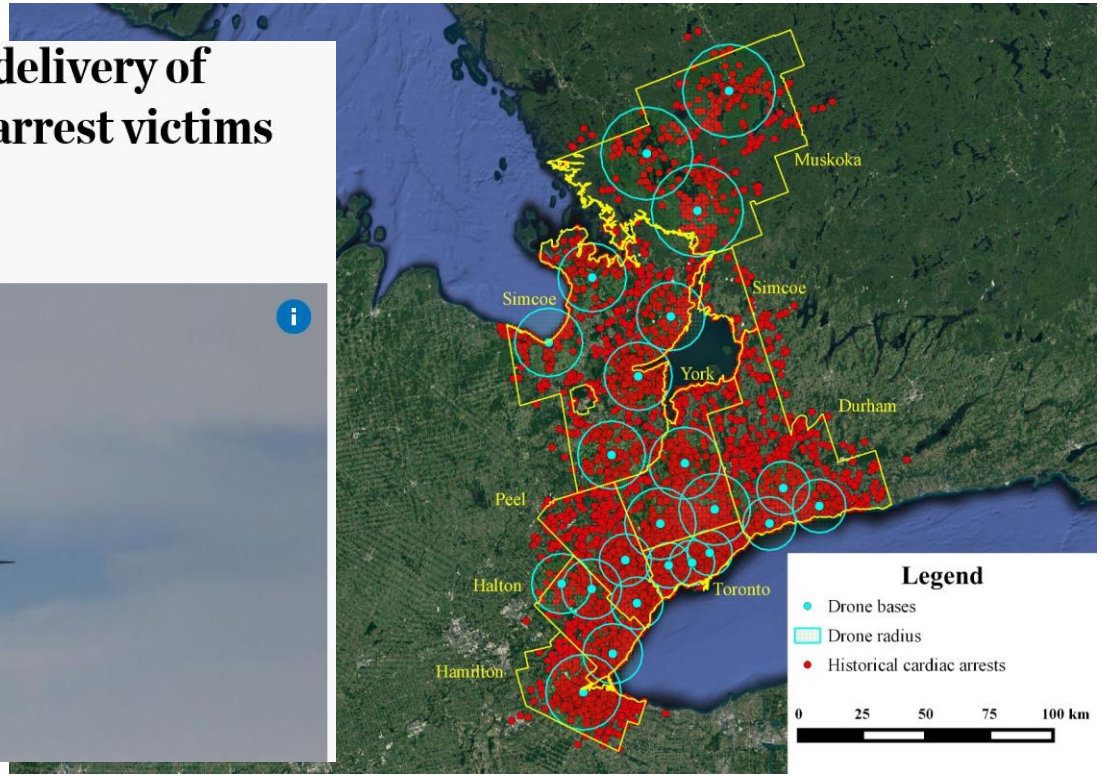
Public defibrillators



Defibrillator, delivered

Peel pilot project tests drone delivery of defibrillators to help cardiac arrest victims

By **Marta Marychuk** Mississauga News
Tues., April 2, 2019 | ⏱2 min. read



Manufacturing

Baosteel, world's 3rd largest steel company, manually planned operations leading to inefficiencies.

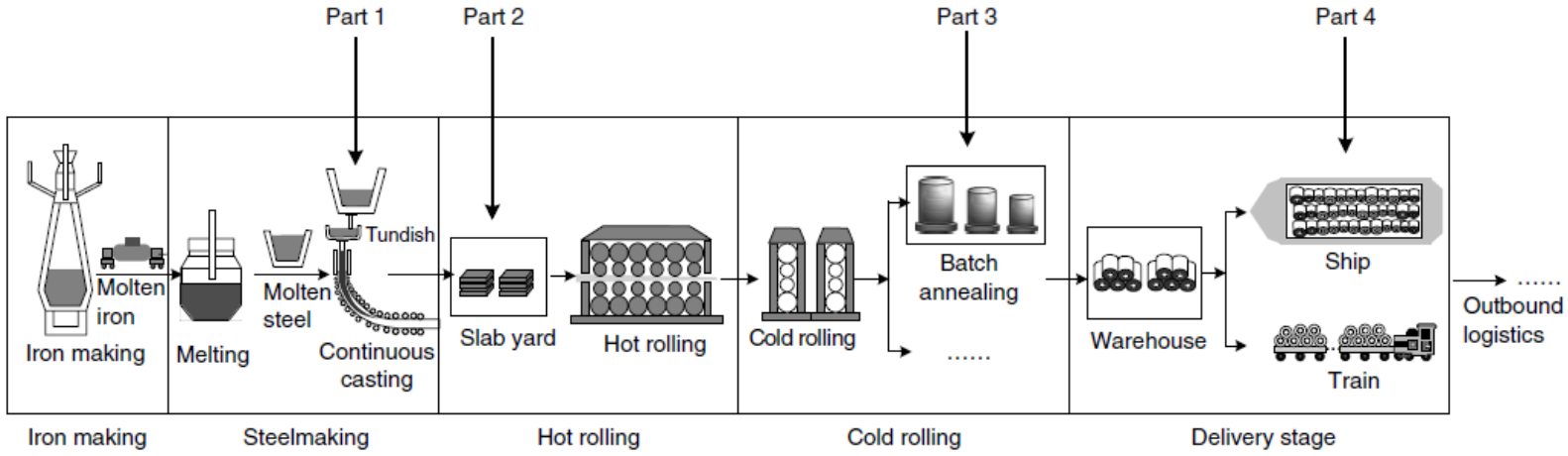
Four bottlenecks to tackle with DSS:

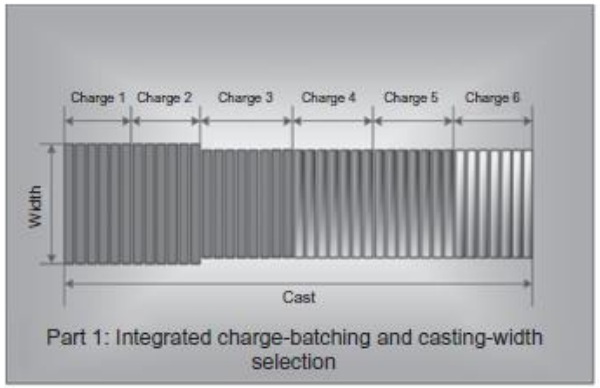
Part 1. Integrated charge-batching and casting-width selection decisions

Part 2. Open-order slab-allocation and slab-reallocation decisions

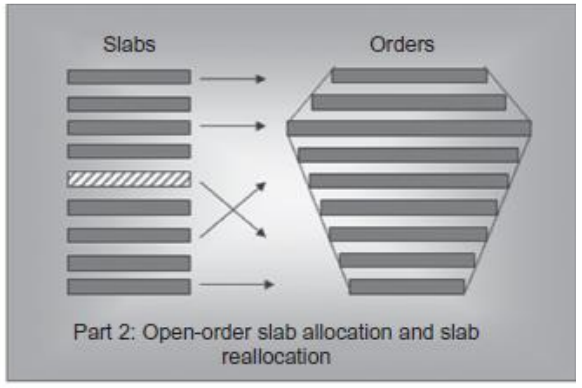
Part 3. Coil batching decisions

Part 4. Ship-consolidation and ship-stowage planning

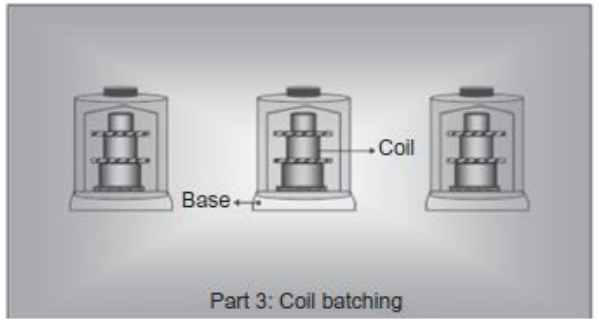




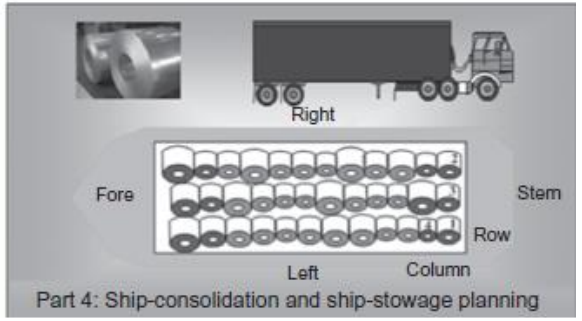
Part 1: Integrated charge-batching and casting-width selection



Part 2: Open-order slab allocation and slab reallocation



Part 3: Coil batching



Part 4: Ship-consolidation and ship-stowage planning

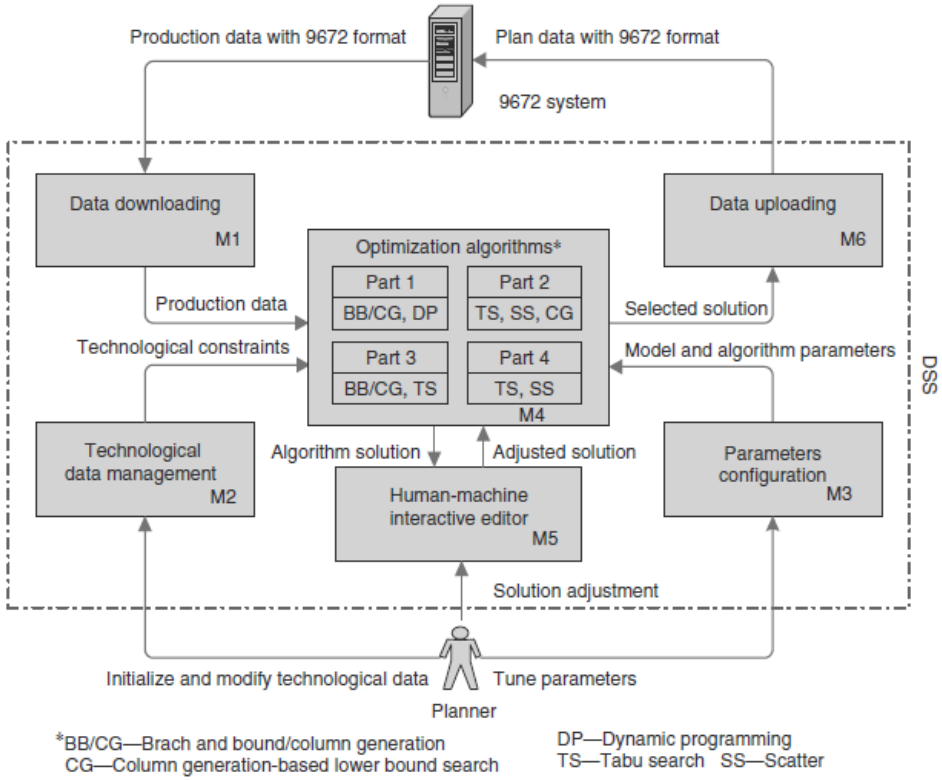
Transforming operations through analytics

Developed and implemented four DSSs, each with six similar modules but calling on different prescriptive analytics (optimization) algorithms.

Cumulative benefits of ~**$77M** over initial 5 years, with ongoing annual benefits of **$20M** estimated.

DSSs	Production line	Total direct economic benefits (million $)	Total indirect economic benefits (million $)	Total economic benefits (million $)
DSS for part 1	Steelmaking shops	14.73	1.92	**16.65**
DSS for part 2	Hot-rolling lines	43.41	0.75	**44.16**
DSS for part 3	Batch annealing line	11.00	0.80	**11.80**
DSS for part 4	Final-product warehouse	4.20		**4.20**
Total		73.34	3.47	**76.81**

Direct: utilization, production rate, reduced scrap
Indirect = quality, reduced inventory space



Retail

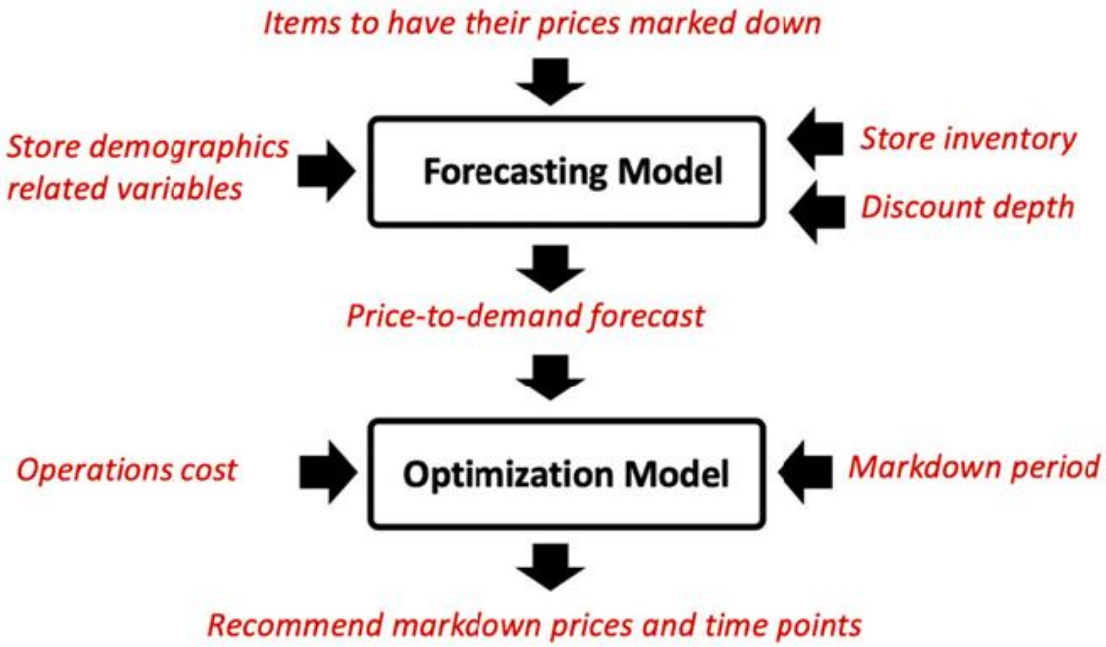
Walmart developed a multi- objective markdown system and deployed it across many merchandising units. The objectives:

1. clear the stores' excess inventory by a specified date
2. improve revenue by minimizing the discounts needed to clear shelves
3. reduce the substantial cost to relabel merchandise in the stores.

The Algorithmic Approach was a shift from the traditional rule-based, tiered discounting process (e.g., 25%–50%–75%) for mark downs

The Markdown Process Consists of:

- Forecasting to Determine the Price-to-Demand Relationship
- Optimization to Determine the Most Favorable Markdown Prices and Markdown Period



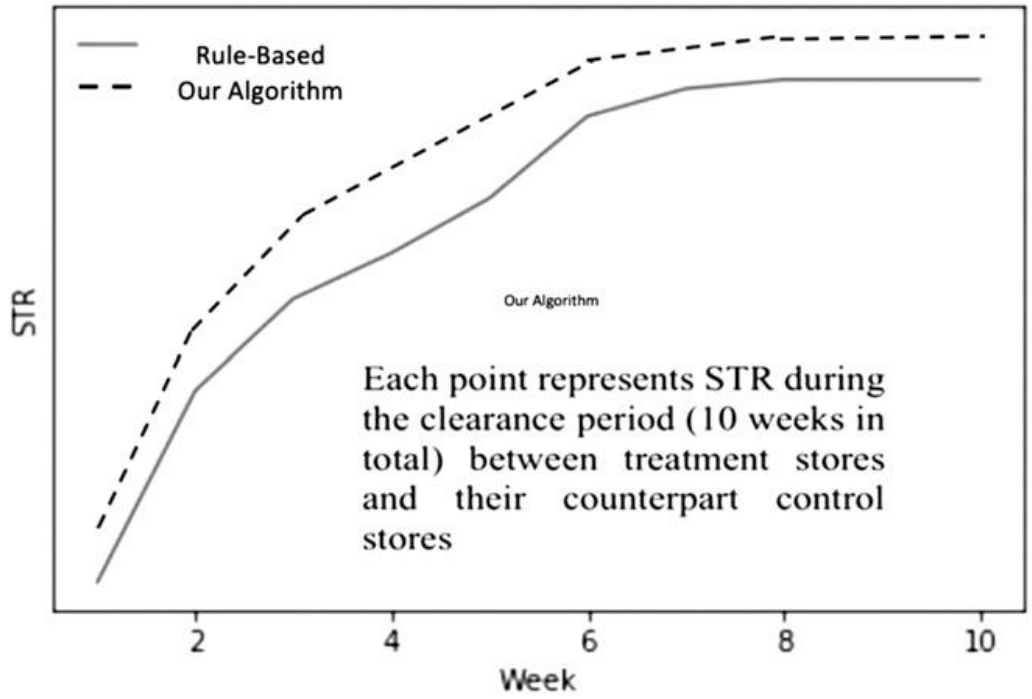
Benefits

- Starting in 2019, after 6 months of testing, the approach was implemented across all USA stores.

- The result was a high performance model with a price-adjustment policy tailored to each store.
  - Increased its STR* by 21%
  - Reduced its costs by 7%

* Sell Through Rate**:** # of units sold during the markdown period divided by its inventory at the beginning of the markdown

The Evolution of STR as a Function of Time for Furniture Products

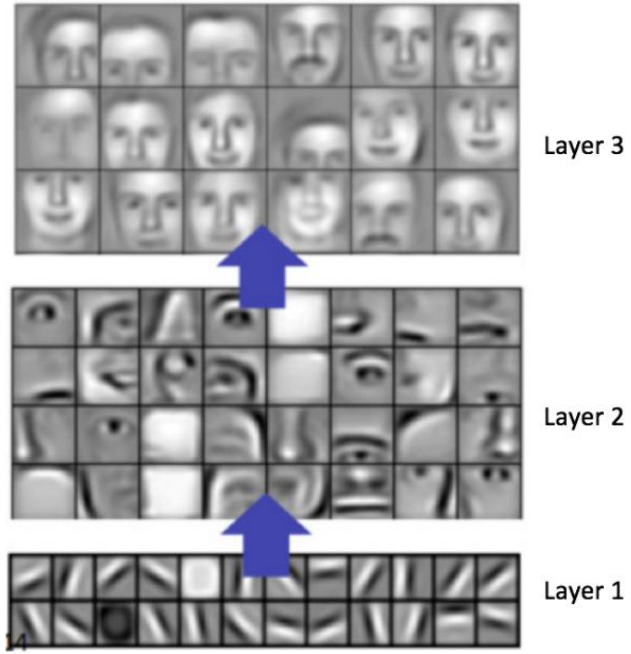


Case studies → Generalizable ideas

1. Predictive policing → finding similarity in patterns
2. Scheduling / HR redeployment → optimizing task assignment
3. Emergency response and logistics → facility location
4. Manufacturing → production, supply chain
5. Retail → revenue & inventory mgmt

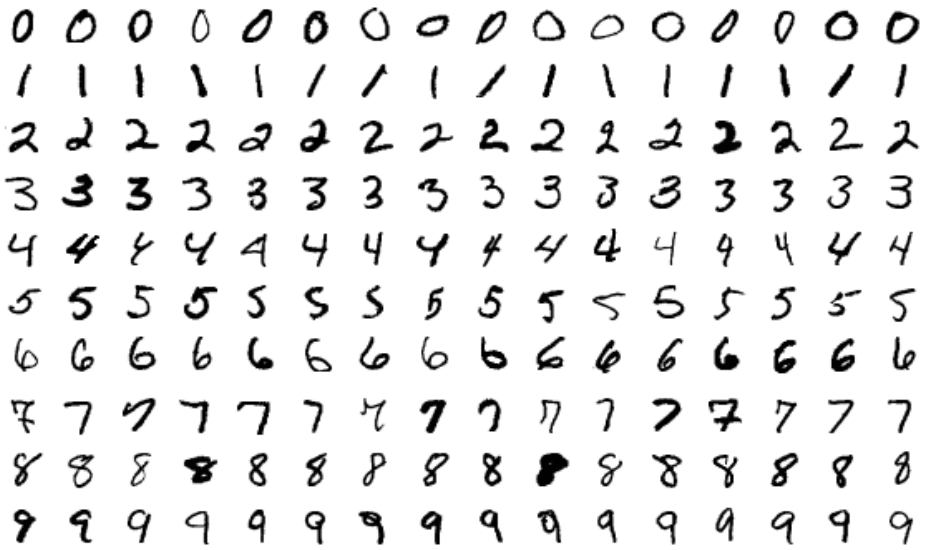
Exciting ML advance #1: image data

- One of the earliest success stories in deep learning
- Faster computers + better machine learning techniques (pioneered at U of T) allowed for the success of Convolutional Neural Networks (CNNs)
- Each layer recognizes ‘features’ of an image that are successively more complex



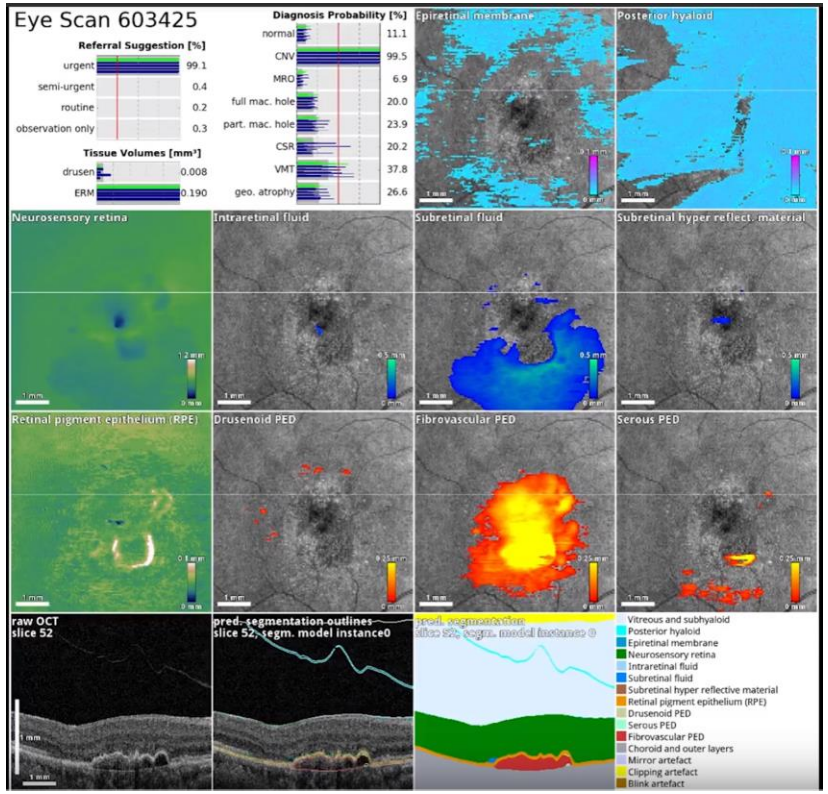
Exciting ML advance #1: image data

- An early application of this technology: automatically reading handwritten addresses and bank cheques
- Pre-CNN this required hand designing features to be recognized
- Today, handwriting can be read with an accuracy of 99.83% - using almost the same core technology as in 1998



Exciting ML advance #1: image data

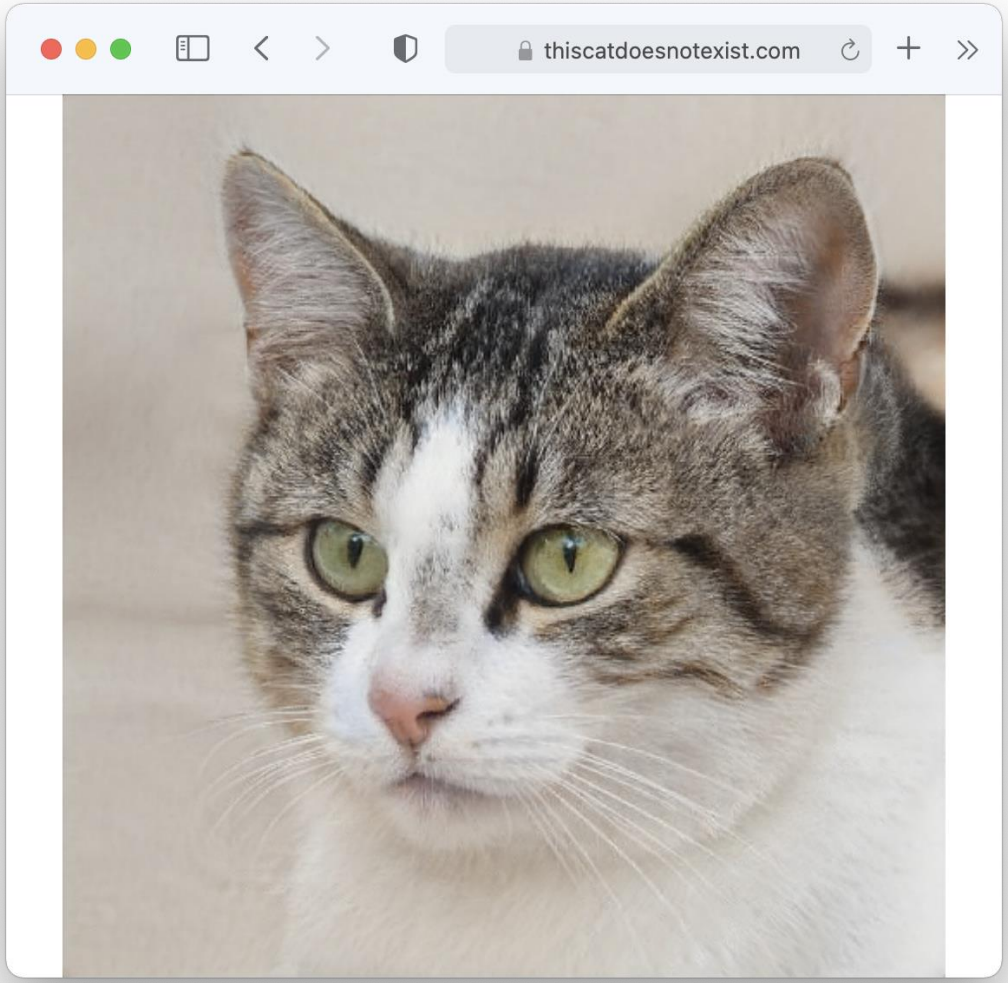
- Hugely beneficial in medical applications
- Retinal disease identification: 95% accuracy (better than humans!), while saving doctors time
- CNNs are also widely used in drug discovery, identifying promising compounds



Clinically applicable deep learning for diagnosis and referral in retinal disease

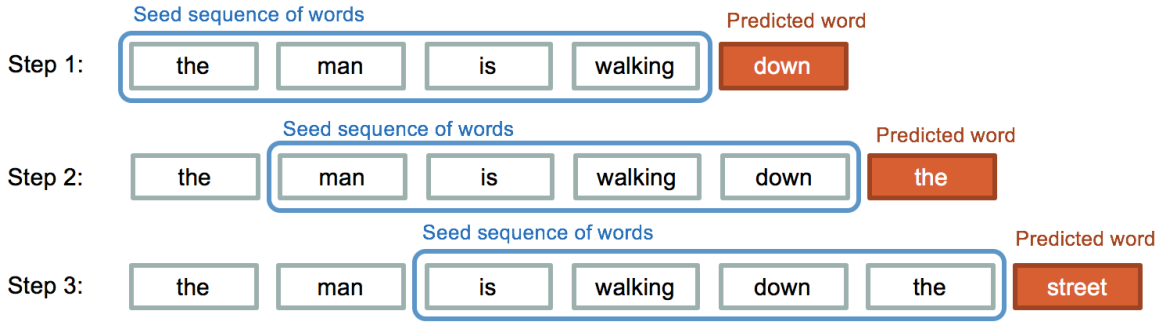
Exciting ML advance #1: image data

- Facebook automatically captions images for people who can’t see them
- Generating realistic images: powerful applications for film, television and video games



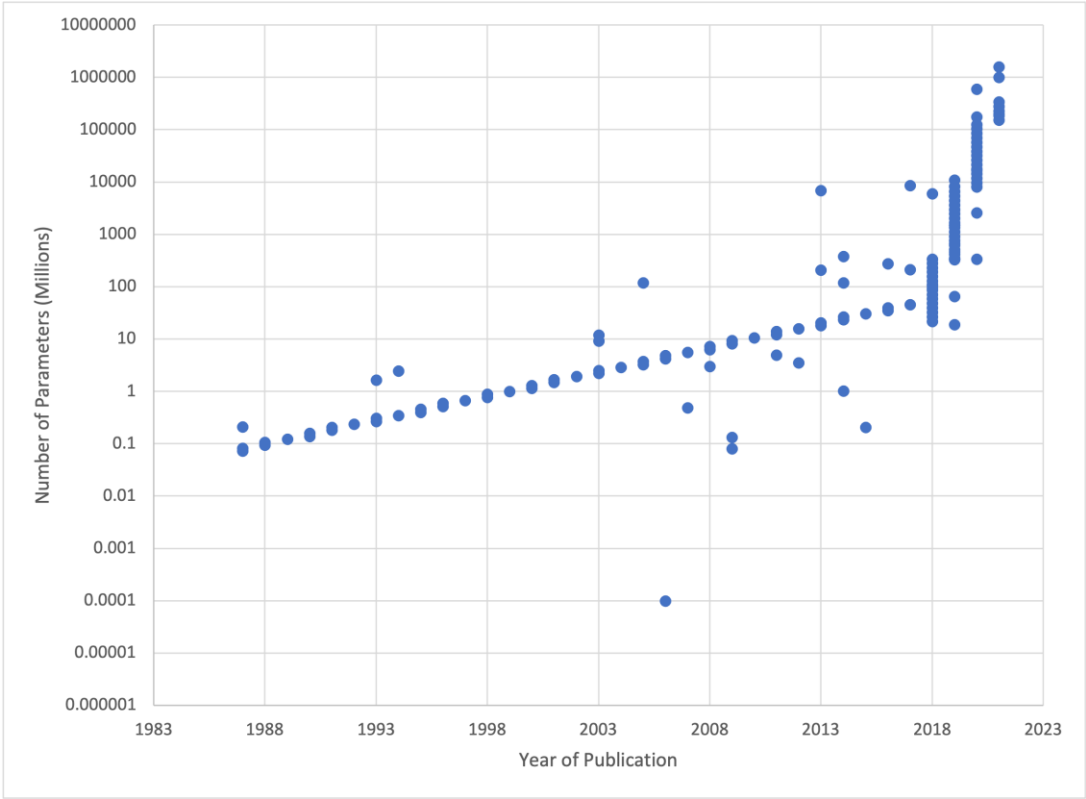
Exciting ML advance #2: LLMs

- From an early advancement, to a recent one
- To generate text, select the word most likely to appear next
- How do we estimate likelihood? By looking at lots of text



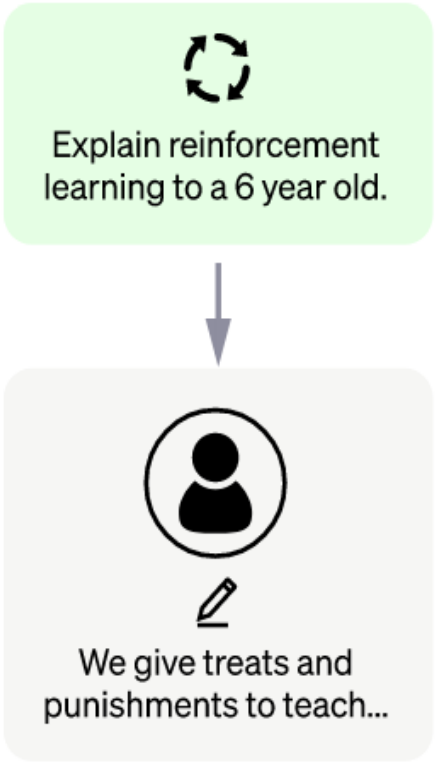
Exciting ML advance #2: LLMs

- Latest models are capable of learning from much more data
- Both thanks to technological improvements, and a willingness to spend more money



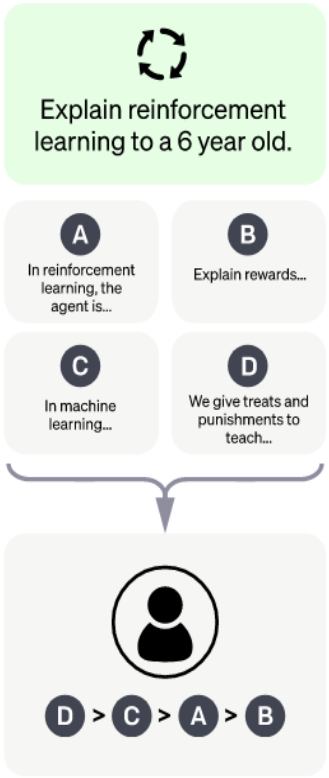
The training innovation of ChatGPT

Human annotators write answers to questions



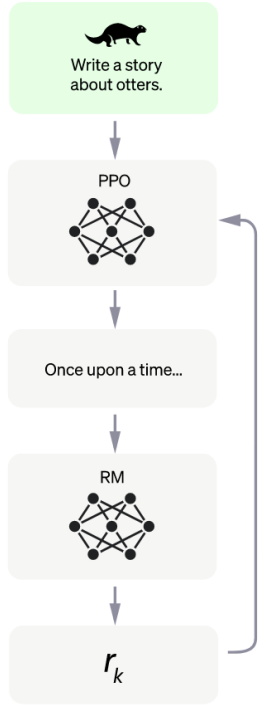
The generalist GPT-3 model is taught from these Q&A pairs

Human annotators write more answers, and someone else ranks them



A separate model learns to rate the quality of an answer

GPT writes answers to sampled questions



The reward model rates each answer, allowing GPT to keep learning

Figures: https://openai.com/blog/chatgpt/

Foundations of Critical Thinking

Foundations of Critical Thinking

- Actively and skillfully analyzing, synthesizing, and evaluating information to reach well-founded conclusions
- Essential for effective data analysis and decision making
- Components of critical thinking:
  - Clarity: Ensuring the meaning of statements and questions is clear
  - Accuracy: Verifying the truth of information
  - Precision: Providing sufficient detail
  - Relevance: Ensuring the information is applicable to the issue at hand
  - Depth: Analyzing the underlying factors of a problem
  - Breadth: Considering alternative viewpoints
  - Logic: Ensuring a coherent and rational flow in reasoning

Enhancing Clarity and Accuracy

- To improve clarity in data analysis:
  - Ask clear and specific questions
  - Define terms and concepts precisely
  - Avoid vague or ambiguous language
- To ensure accuracy:
  - Verify information by cross-checking sources
  - Be cautious of potential misinformation or biased sources
  - Maintain a healthy skepticism and question assumptions

Precision and Relevance

- To enhance precision:
  - Use accurate and specific language
  - Break down complex issues into smaller, more manageable components
  - Clearly outline the scope and boundaries of the analysis
- To maintain relevance:
  - Continually assess how the data and analysis relate to the issue
  - Prioritize information based on its importance and applicability
  - Avoid straying from the main topic or getting lost in unrelated details

Depth, Breadth and Logic

- To enhance depth:
  - Identify and address any assumptions or biases that may impact the analysis
  - Consider potential consequences or implications of the findings
  - Evaluate the strengths and limitations of the analysis methods used
- To ensure breadth:
  - Actively seek out and evaluate differing opinions or interpretations
  - Assess the problem from multiple angles or dimensions
  - Recognize the potential impact of contextual factors on the analysis
- To maintain logic:
  - Ensure a coherent and rational flow in reasoning and arguments
  - Identify any logical inconsistencies or fallacies
  - Evaluate the strength and relevance of evidence used to support conclusions

Centre for Analytics and AI Engineering (CARTE)

CARTE is the hub for analytics and AI translation at the #1 engineering school in Canada

- Combining analytics/AI with sector-specific knowledge
- Enabling successful application of analytics/AI to problems across industry and society

CARTE at a glance

100

Faculty affiliates

400+

Grad students and PDFs

- Chemical Eng
- Civil & Mineral Eng
- Elect. & Comp. Eng
- Inst. Biomedical Eng
- Mech. & Ind. Eng
- Materials Science Eng
- UTIAS
- ISTEP

Steering Committee

Timothy Chan
MIE
(Director)

Margaret Cheng
IBBME/ECE

Stark Draper
ECE

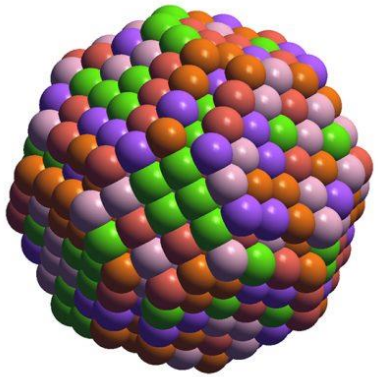
Elias Khalil
MIE

Shoshanna Saxe
CivMin

Somayeh Sadat
Assistant Director

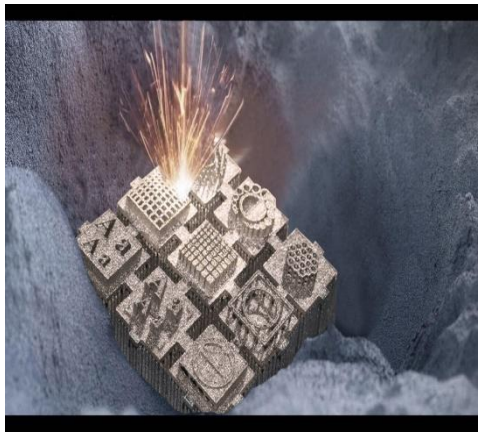
Alex Olson
Research Associate

Focus on translating AI ideas



High throughput simulations, ML, and experimental data to enable design of high entropy alloys for structural applications

Utilizing residential smart thermostats data to provide insights and methods to reduce energy use and improve comfort



Integrating Additive Manufacturing and AI to develop a “self-driving 3D printing system” that senses/predicts defects and corrects processing parameters adaptively

Ambient monitoring of gait in long-term care facilities to automatically identify residents who are at a high risk of falling

Pillars of CARTE

Research

- **CARTE Seed: Funding for high-impact multi-disciplinary research**
- **CARTE Match: Connecting faculty with A/AI projects with students**
- **In-house A/AI research support**
- **AI/ML Research drop-in clinics**

Education

- **ML Bootcamp for Faculty**
- **Conference De-Brief Sessions**
- **Pathway of analytics courses**
- **Information repository on cloud computing resources**
- **Experiential learning: MITACS Accelerate internship announcements**

Partnerships

- **Industry speaker seminars**
- **Applied research projects (with leveraged funds)**
- **A/AI Consulting**
- **Executive & technical training**
- **Support to identify top talent (competitions) and recruit (job board, industry days)**


MAGNA
BOMBARDIER
Johnson & Johnson
VALE
Hitachi High-Tech
TOYOTA
IBM
CLEARPATH ROBOTICS™
ONTARIO CLEAN WATER AGENCY
AGENCE ONTARIENNE DES EAUX
PHILIPS
GenomeCanada
Teck
HATCH
Toronto Water Authority
ONTARIO POWER GENERATION
400+
industry partners
GM
EllisDon
FUJITSU
SUNCOR ENERGY
Syncrude
Ford
GLENCORE
Celestica™
LG Electronics
MDA™
SAMSUNG
SIEMENS
DUPONT
HRG
HIT ROBOT GROUP
McEWEN MINING
Pratt & Whitney
A United Technologies Company
TELUS

Analytics/AI INNOVATION ECOSYSTEM

GitHub & Google Colab

`tinyurl.com/enbbootcamp`

CARTE-Enbridge AI Bootcamp

Workshop Materials

This repository contains the materials for the Enbridge AI Bootcamp.

Day 1

Morning

Lab 1-1a: Introduction to Python

Open in Colab

Lab 1-1b: Introduction to NumPy and Pandas

Open in Colab

Afternoon

Lab 1-2: Introduction to Scikit-Learn

Open in Colab