Machine Learning 101

Lunch & Learn

Sacha Davis | April 17th, 2024

Introduction

- University of Alberta:
  - 2020: B.Sc. Biological Sciences with minor in Computing Science
  - 2023: M.Sc. Computing Science (Thesis-Based)
- Co-Director of the Artificial Intelligence in Medical Systems Society (aimss.ca)
- Research, project management, ML engineering, community engagement, teaching, etc.



Sacha Davis
Co-op work experience Student
Sacha.Davis@albertahealthservices.ca

Presentation Goal #1:
ML Intuition

Presentation Goal #2:
ML Language

Presentation Goal #3:
ML Examples

Presentation Not Goal #1: Technical Details

Presentation Not Goal #2:
Overwhelming you with information

Resources

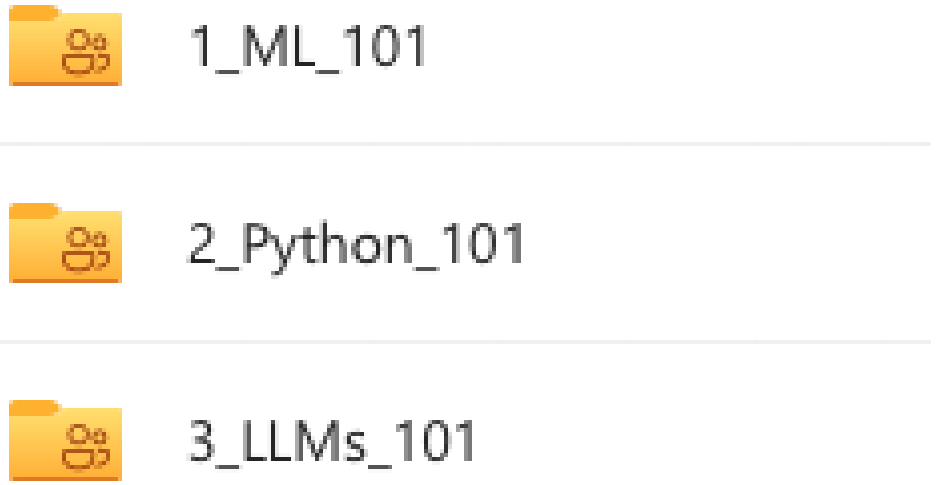
- Slides are available, should have been sent out before over email
- "Machine Learning 101" document that defines, with examples, many of the key terms that are used in the field



ML4PH - Machine Learning 101.docx

Next Lunch & Learn Sessions

- April 17: Machine Learning 101
- May 15: Python 101
- June 19: Introduction to Generative AI (ChatGPT and beyond)



Introduction to Machine Learning

Definitions Rapid Fire: Round #1

What fields are at play?

Machine Learning (ML):

A computer's way of learning from examples or experiences, so it can do tasks without being directly programmed for each one.

Example: Predicting, from administrative data, whether an individual is likely to be “vaccine hesitant” or “vaccine acceptant”.

Definitions Rapid Fire: Round #1

What fields are at play?

Machine Learning (ML):

A computer's way of learning from examples or experiences, so it can do tasks without being directly programmed for each one.

Example: Predicting, from administrative data, whether an individual is likely to be “vaccine hesitant” or “vaccine acceptant”.

Artificial Intelligence (AI):

A broad area of computer science focused on creating machines capable of performing tasks that typically require human intelligence. Machine learning is a part of AI.

“Expert systems” are an example of AI that is not ML.

Definitions Rapid Fire: Round #1

What fields are at play?

Machine Learning (ML):

A computer's way of learning from examples or experiences, so it can do tasks without being directly programmed for each one.

Example: Predicting, from administrative data, whether an individual is likely to be “vaccine hesitant” or “vaccine acceptant”.

Artificial Intelligence (AI):

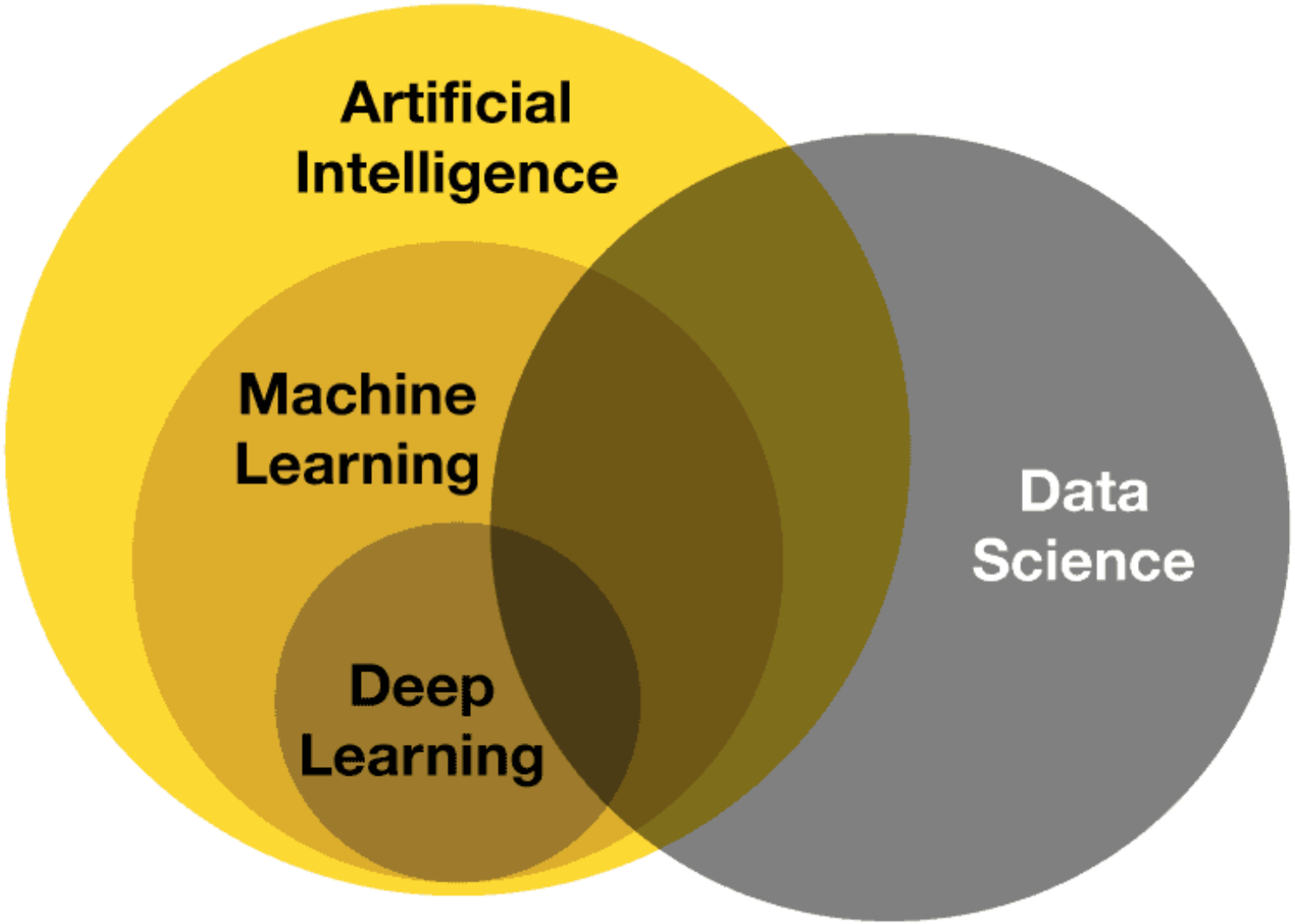
A broad area of computer science focused on creating machines capable of performing tasks that typically require human intelligence. Machine learning is a part of AI.

“Expert systems” are an example of AI that is not ML.

Data Science:

The practice of using statistics, computer science, visualization, and knowledge about a specific area (like health) to understand and make decisions from large amounts of data.

Creating a Tableau dashboard is an example of Data Science that is not AI or ML.


Artificial Intelligence
Machine Learning
Deep Learning
Data Science

[Source]

Where is ML happening?

Academic Research*

Tech Industry*

Health and Wellness

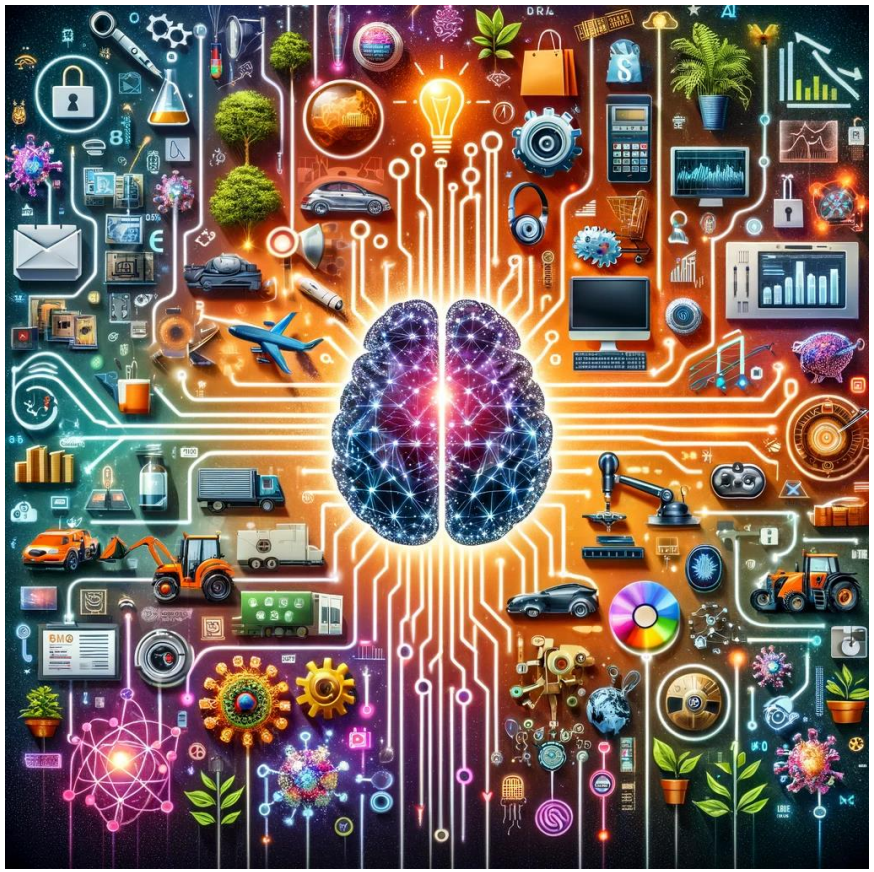
Retail

Finance

Manufacturing

Transportation

Source: DALL-E



Agriculture

Entertainment

Customer Service

Smart Homes

Security

Education

Source: DALL-E



Small Data
Simple algorithms
Low Compute
Statistical Methods

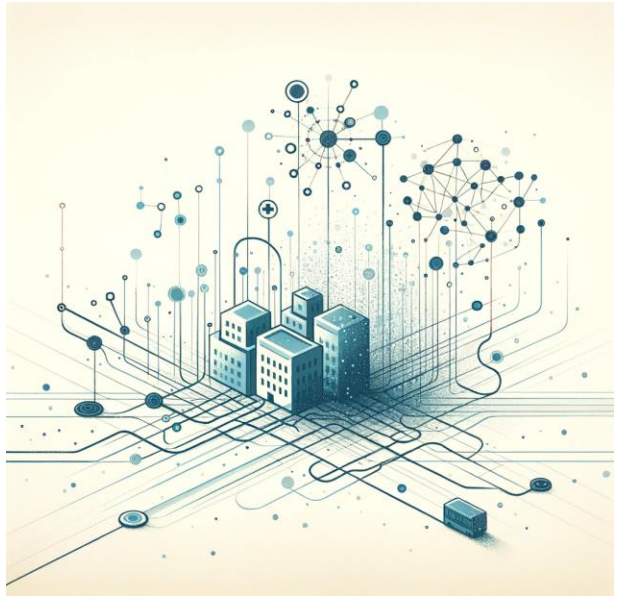
time

Big Data
Sophisticated Algorithms
High Compute
Machine Learning Methods

Promise for ML in health



Precision Health



Hospital Dynamics



Disease Dynamics

... and more!

Source: DALL-E

Promise for ML in health

- AI can help us become proactive rather than reactive
  - Anticipate and prevent fires rather than putting them out

[Source]

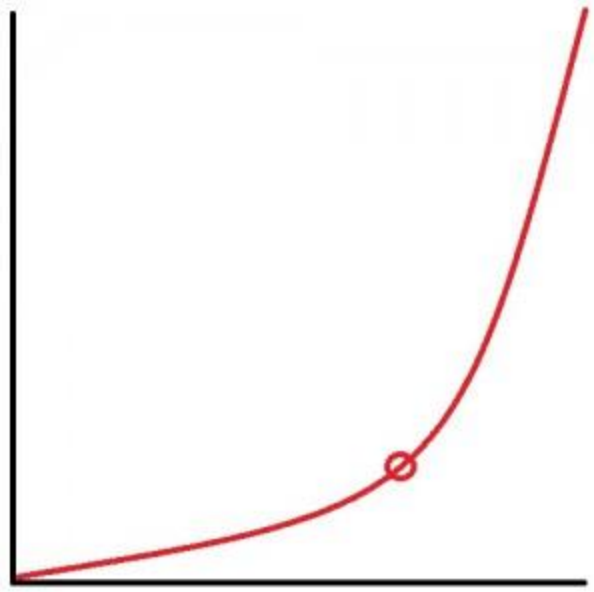


Promise for ML in health

- AI can help us become proactive rather than reactive
  - Anticipate and prevent fires rather than putting them out

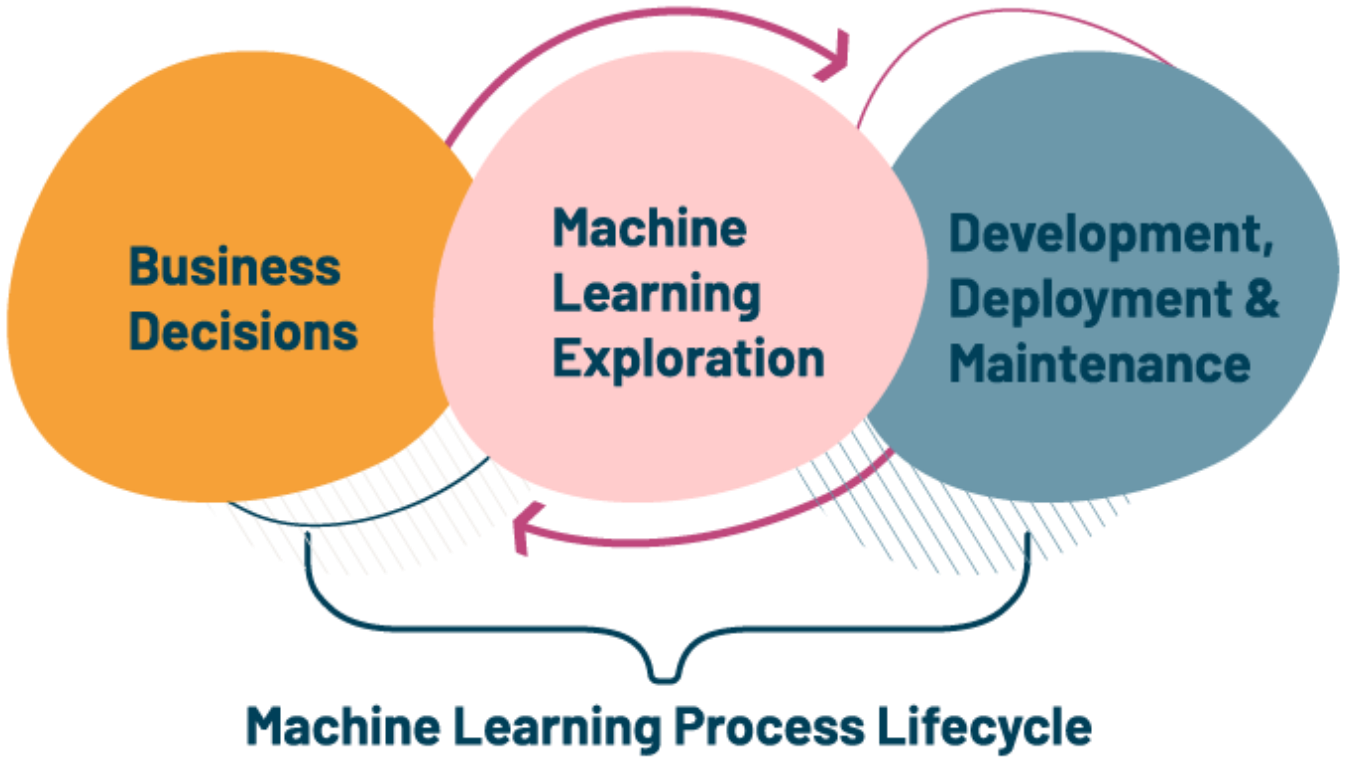
- "Start with scraps" – small wins can have a compounding effect over time

[Source]



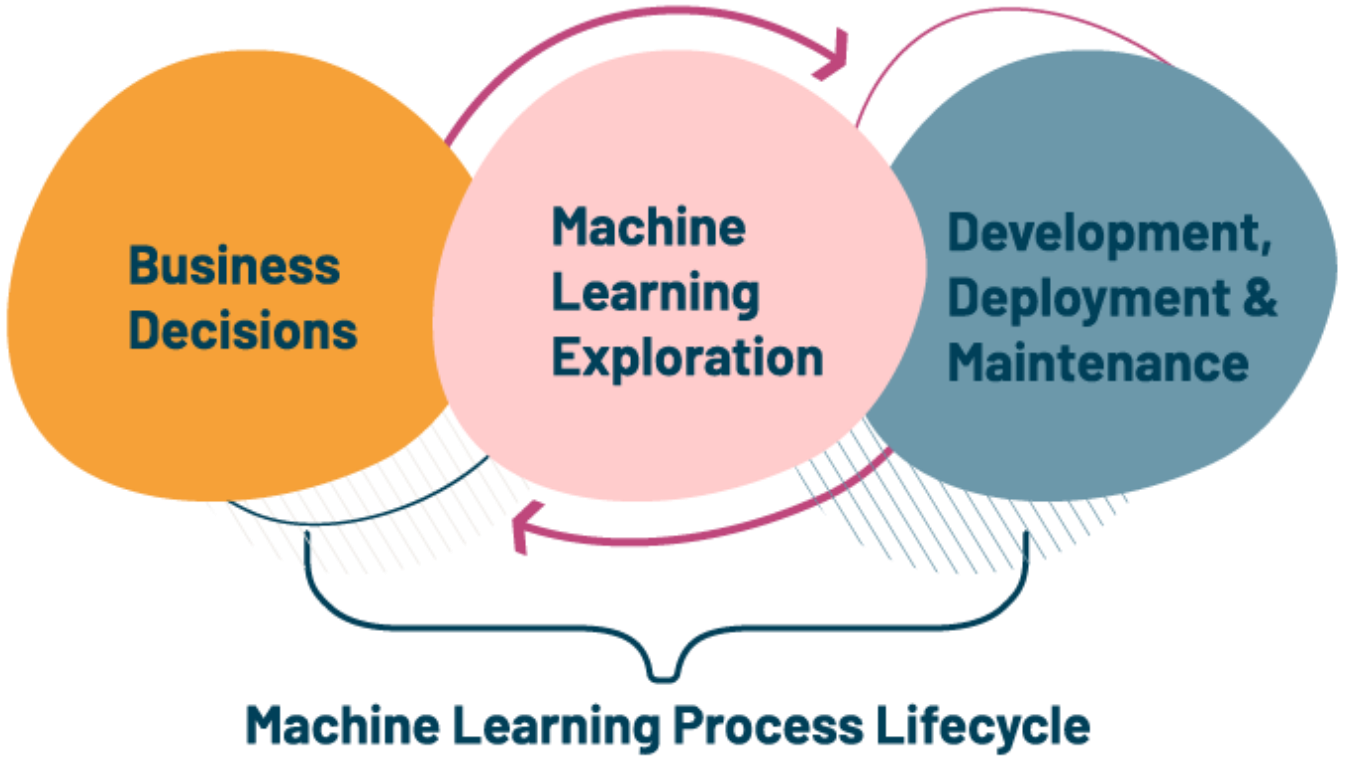
Stages of a Machine Learning Project

ML Project In Practice



[Source - amii.ca]

ML Project In Practice

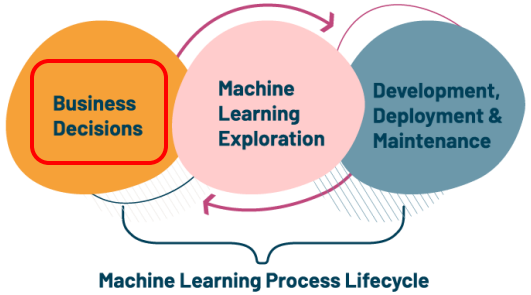


[Source - amii.ca]

Why is it so iterative?

- Data limitations
- Poor model performance
- Unanticipated findings in data exploration
- Changes in business context
- Proxy is insufficient
- Poor evaluation metrics
- Technical feasibility issues
- Ethical considerations
- Regulatory compliance concerns
- Failure point in human-computer interaction

Formulating a Good ML Question





Too broad or too vague

e.g., "Can we improve healthcare?"

Lacks actionability

e.g., "What are the patterns in COVID outbreaks?"



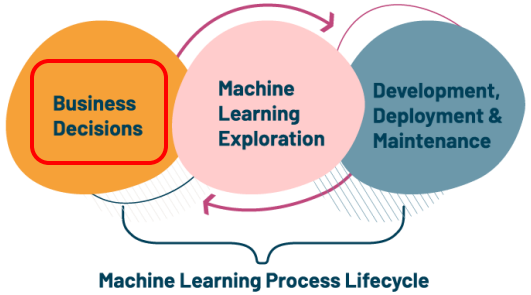
Specific and Measurable

e.g., "Can we predict which patients are at high risk of readmission within 30 days after discharge from hospital records to reduce readmissions by 10%?"

Aligned with Business Objectives

e.g., "How can we use customer transaction data to personalize marketing campaigns and increase conversion rates by 15%?"

Formulating a Good ML Question



A well-formulated business question is SMART:

S: Specific
M: Measurable
A: Actionable
R: Relevant
T: Time-bound



Specific and Measurable

e.g., "Can we predict which patients are at high risk of readmission within 30 days after discharge from hospital records to reduce readmissions by 10%?"

Aligned with Business Objectives

e.g., "How can we use customer transaction data to personalize marketing campaigns and increase conversion rates by 15%?"

Definitions Rapid Fire: Round #2

Important Data Concepts

Sample:

A singular entity for which a **model** can make a **prediction**. Synonymous with “example”, “instance”, “individual”, or a “row” of a dataset. May contain many **attributes**.

Example: A patient, for which we have captured many sociodemographic and medical details.

Definitions Rapid Fire: Round #2

Important Data Concepts

Sample:

A singular entity for which a **model** can make a **prediction**. Synonymous with “example”, “instance”, “individual”, or a “row” of a dataset. May contain many **attributes**.

Example: A patient, for which we have captured many sociodemographic and medical details.

Attribute:

A descriptive element containing information about the **samples** in a **dataset**. Synonymous with a “column”, “variable”, and “feature”. When the attribute is used for **prediction**, synonymous with “predictor”, “covariate”, or “regressor”.

Example: Patient age.

Definitions Rapid Fire: Round #2

Important Data Concepts

Sample:

A singular entity for which a **model** can make a **prediction**. Synonymous with “example”, “instance”, “individual”, or a “row” of a dataset. May contain many **attributes**.

Example: A patient, for which we have captured many sociodemographic and medical details.

Attribute:

A descriptive element containing information about the **samples** in a **dataset**. Synonymous with a “column”, “variable”, and “feature”. When the attribute is used for **prediction**, synonymous with “predictor”, “covariate”, or “regressor”.

Example: Patient age.

Target:

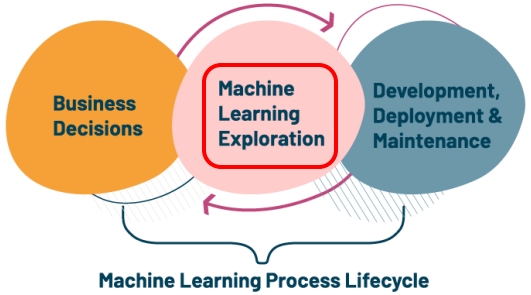
The **attribute** we wish to use **ML** to predict or detect. The **predicted** target is synonymous with the **model’s** “output”.

Example: A “1” if an individual is vaccine-hesitant, and a “0” otherwise.

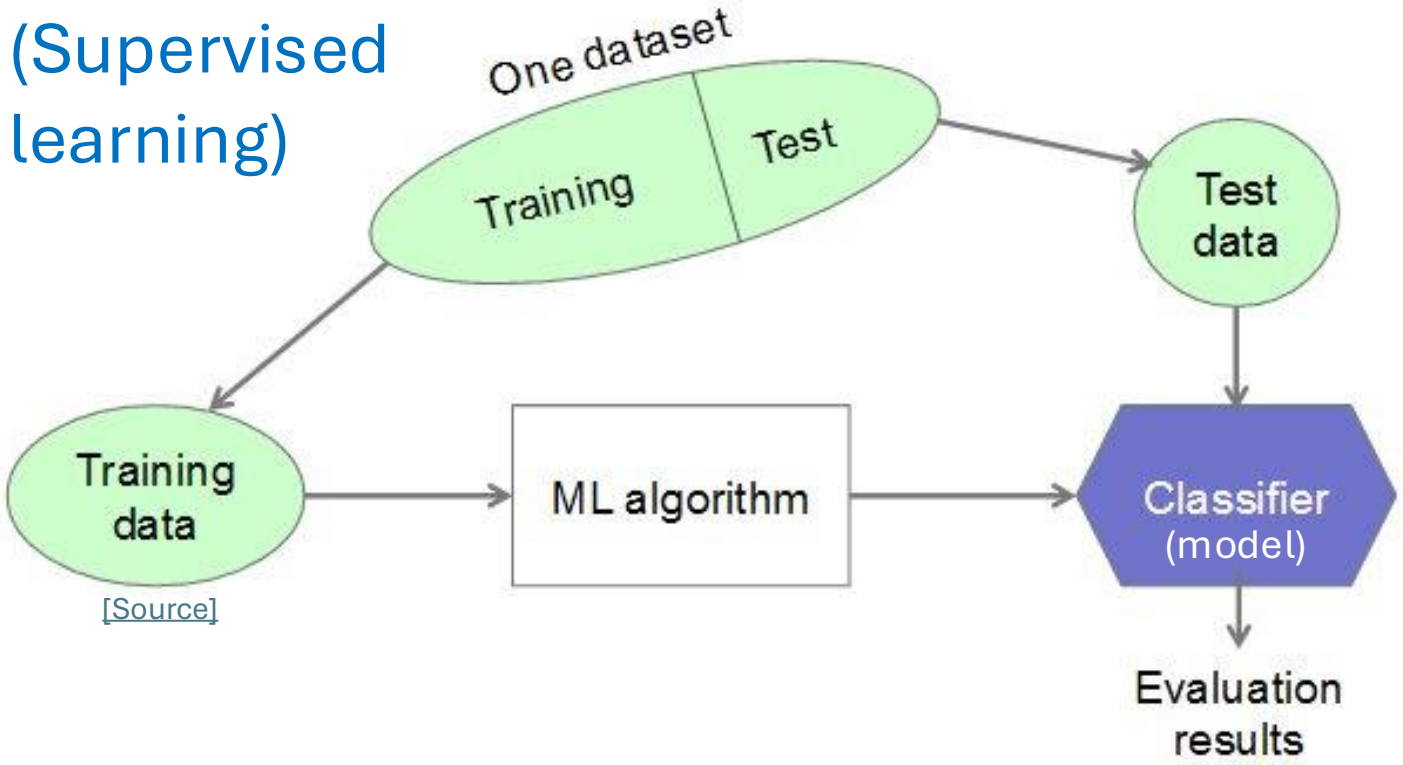
[Source]

	A	B	C	D	E	F	G
1	Date	Average price in USD	Total Sold	Smal Avocados Sold	Large Avocados Sold	Extra Large Avocados Sold	City
2	12/27/2022	1.29	319746	292097	27351	298	Atlanta
3	12/20/2022	1.38	272580	251774	20702	103	Atlanta
4	12/13/2022	1.26	356227	324932	31019	276	Atlanta
5	12/6/2022	1.37	306947	283024	23741	182	Atlanta
6	11/29/2022	1.29	279486	250289	28890	308	Atlanta
7	11/22/2022	1.3	300032	269799	29732	501	Atlanta
8	11/15/2022	1.43	292814	263916	28442	456	Atlanta
9	11/8/2022	1.42	285413	250441	34483	488	Atlanta
10	11/1/2022	1.29	353690	290458	62980	253	Atlanta
11	10/25/2022	1.39	301727	236814	64608	304	Atlanta
12	10/18/2022	1.39	288733	228710	59732	291	Atlanta
13	10/11/2022	1.25	347580	255933	91047	600	Atlanta
14	10/4/2022	1.26	359222	265797	92780	644	Atlanta
15	9/27/2022	1.07	324659	262107	61871	681	Atlanta

Machine Learning Exploration

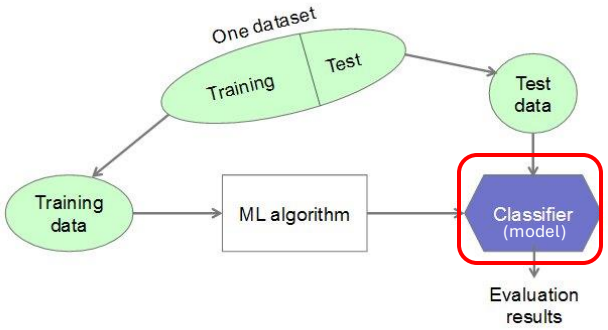


(Supervised learning)



[Source]

How to Talk About ML Models



Learning Style

Task Type

Input Data Type

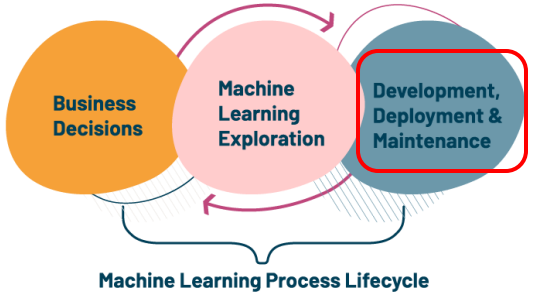
Algorithm Type

Explainability

Computational Load/complexity

Adaptivity

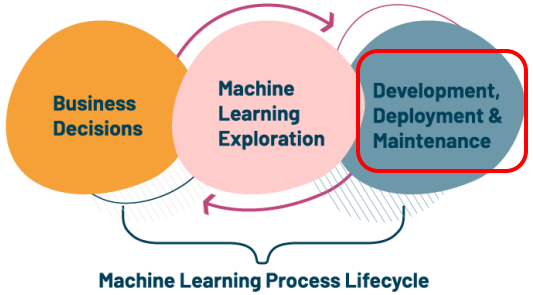
Development, Deployment, and Maintenance

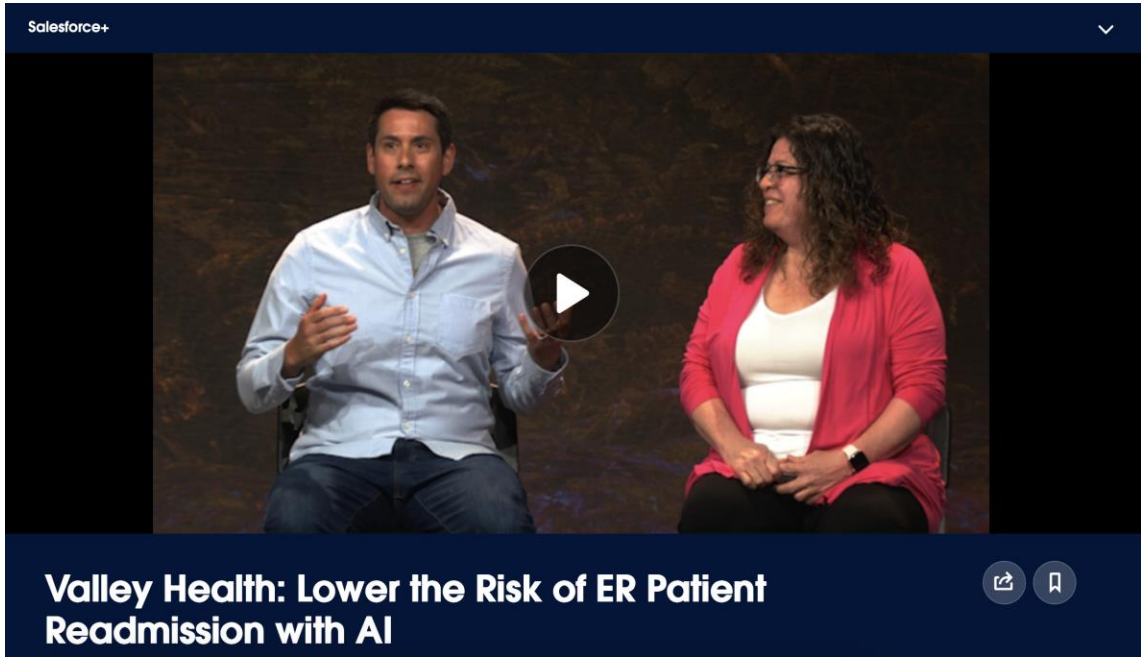


Machine Learning Operations (MLOps) [Source]

	Dedicated Data Scientist	Common Pool: Data Scientist, ML Engineers, Data Engineers, DevOps , Privacy & Security Experts					
Resources							
Phase	Model Development	Machine Learning Model Lifecycle					
	Model Development	**Data Pipelines**	**Model Prep**	**Model Deployment**	**DevOps**	**Catalogs**	**Governance**
Key Activities	• EDA • Model training	• Data pipelines • Data cleanup • Data transformations • Dataset management • Data drift monitoring • Automation scripts	• Experimentation • Model training • Model optimization • Model scoring • Validations • Automation scripts	• Scaled deployment • Deployment monitoring • Real-time VS Batch • Model drift monitoring • Scoring scrips	• CI/CD Pipelines • Testing/ monitoring • Logging/ Notifications • Code version • Model versioning • Data versioning	• Model catalogs • Dataset catalogs • Featureset catalogs • Usage governance	• GDPR/CCPA • Security/ compliance • Access controls • Transparency • Reporting

Development, Deployment, and Maintenance

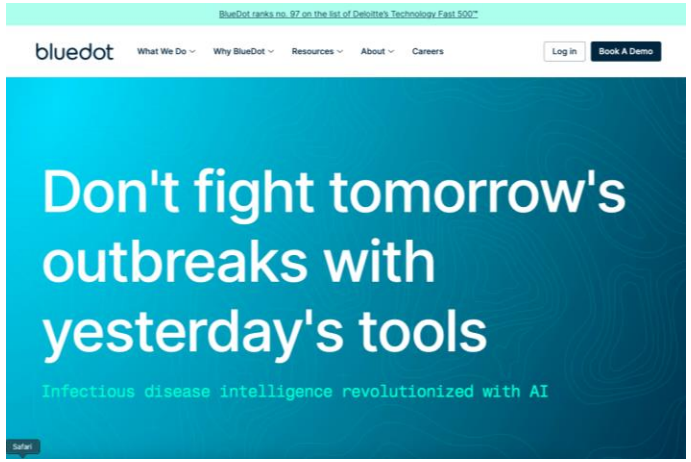




[Video Link]

Examples of Successful ML Projects

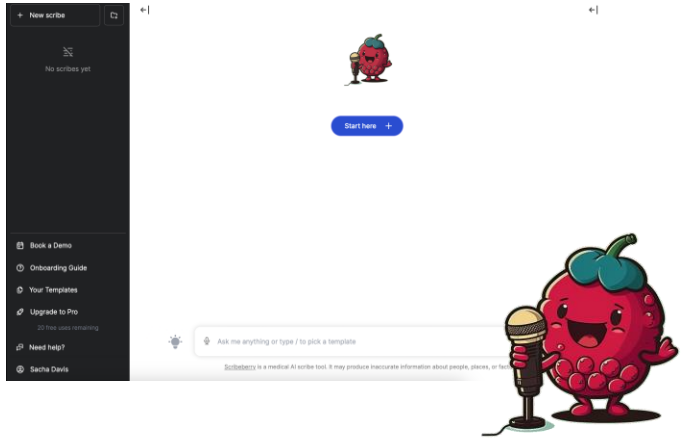
BlueDot



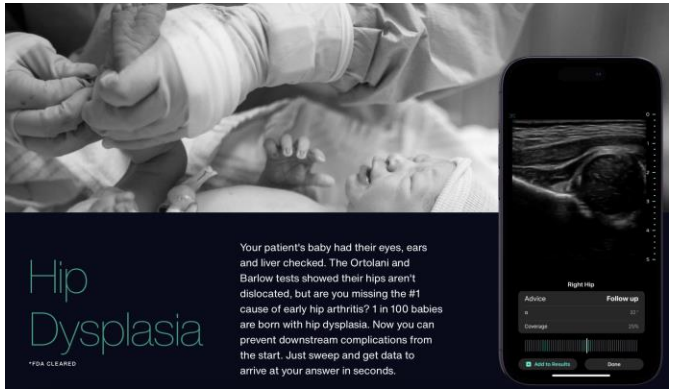
SmokeBeat



Scribeberry



MEDO/Exo



Final Considerations

What is holding us back?



Source: DALL-E

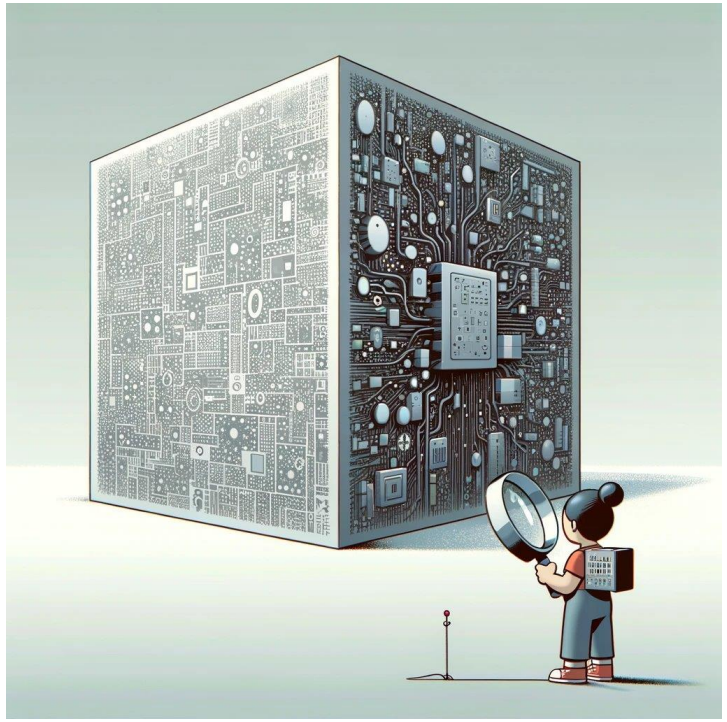
Limitations within the AI model:

- Only knows the data it's trained on
- Unembodied
- Only understand associations, NOT causality*

- This means ML models:
  - Are generally uncreative
  - Have limited understanding of context
  - Don't know their own shortcomings

* In most models

What is holding us back?



Source: DALL-E

Limitations from outside the AI model:

- Data issues
- Mismatches and misunderstandings between fields
- Implementation is hard!
- Human psychology stuff: knowledge, trust, buy-in, hype, etc.

Bonus: How do I keep up?



https://www.nytimes.com/column/hard-fork



NEJM AI

https://ai-podcast.nejm.org



https://www.youtube.com/@aiexplained-official



https://importai.substack.com/



https://www.healthcareitnews.com/

Scan to get the application form to your inbox!



aimss.ca/conference

Thank you!

Questions?

ML4PH - Machine Learning 101.docx



Sacha Davis
Co-op work experience Student
Sacha.Davis@albertahealthservices.ca