Data science thinking: making an impact Part II

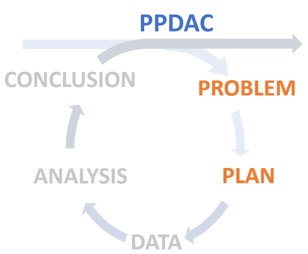
Kostas Sechidis & Mark Baillie
AMLD, Lausanne
March 24th, 2024

NOVARTIS | Reimagining Medicine

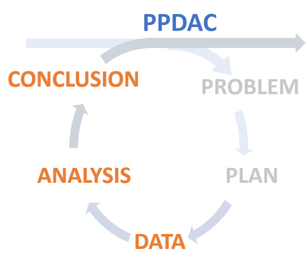
1

Data science thinking: making an impact

Part I



Part II

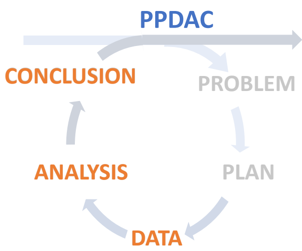


2

Data science thinking: making an impact

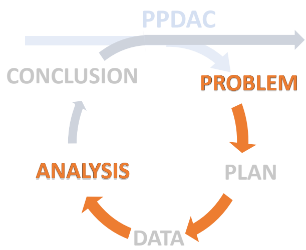
Part II

Part II agenda:
- From question to analysis
- From data to model
- Building machine learning models
- Google Thai project (interactive part)
- One model, many things to consider
- Explaining and communicating analysis outcomes



3

Part II agenda:
- ✓ From question to analysis
- From data to model
- Building machine learning models
- Google Thai project (interactive part)
- One model, many things to consider
- Explaining and communicating analysis outcomes



From question to analysis

Warming up examples from the area of Parkinson’s ...

4

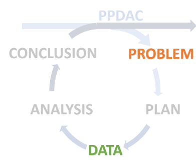
Context setting for this activity

- Parkinson's disease is a neurodegenerative disorder that causes unintended or uncontrollable movements and difficulty with balance and coordination.
- Most prominent cause is when certain nerve cells (neurons) in the brain gradually break down or die.
- Early treatment is very critical since it can slow down the disease progression.
- Testing for Parkinson's can be a lengthy and complex process and many people around the world may not have access to clinical experts to perform this diagnosis.
- Having a DS/AI tool to diagnose Parkinson's in an easy and effective way is critical.

[What is Parkinson's Disease - Parkinson's Nebraska (parkinsonsnebraska.org)]

5

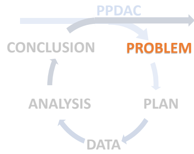
Can we solve this? Question 1:

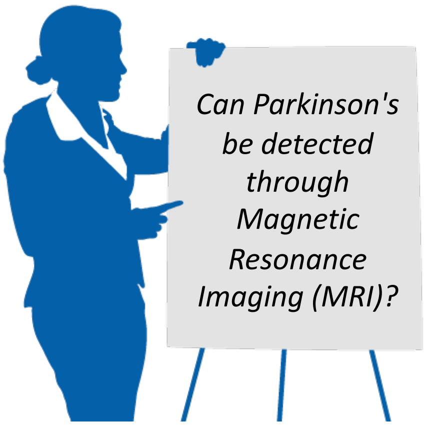


[Max Little: A test for Parkinson's with a phone call]

6

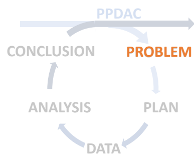
Can we solve this?
Question 2:





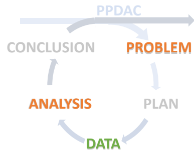
7

Can we solve this?
Question 3:



8

Teaser on analytic solutions



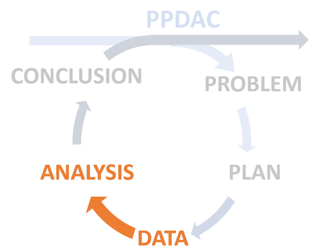
Question/Problem	Analysis via
Can we detect Parkinson’s from voice recordings?	Supervised learning on **audio data**
Can Parkinson's be detected trough Magnetic Resonance Imaging (MRI)?	Supervised learning on **images**
Can we identify genetic factors that lead to higher chances of developing Parkinson’s?	Feature selection on **omics**

In this course, will not learn about the terminologies in details. We will understand how data scientists map questions with various solutions. This can help you work with data scientists effectively.

9

Part II agenda:

- ✓ From question to analysis
- ✓ From data to model
- Building machine learning models
- Google Thai project (interactive part)
- One model, many things to consider
- Explaining and communicating analysis outcomes



From data to model

10

Data is everywhere

- Each person has many characteristics (**variables/features**), e.g., Height, Weight, Blood pressure, Sex, and which together create one **observation**.
- Often there is a variable of interest (**target variable**), e.g., health status.
- Many observations create a **data-set**.

11

Creating a dataset

Data set

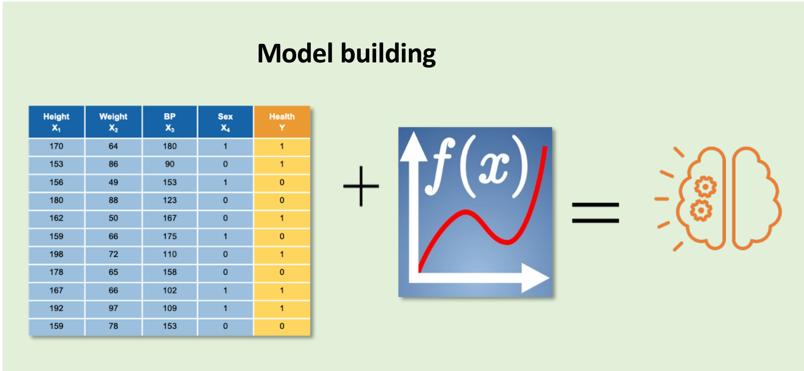
Height X_1	Weight X_2	BP X_3	Sex X_4	Health Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

12

Building a model from a data set: a critical step

Data set

Height X_1	Weight X_2	BP X_3	Sex X_4	Health Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

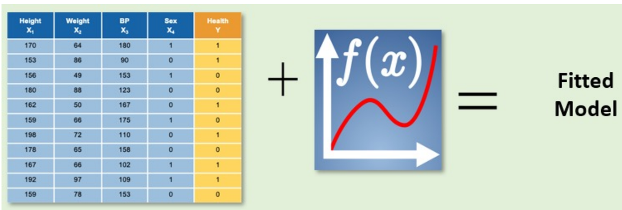


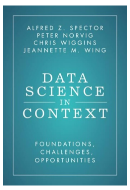
13

But what exactly is a model?

Model is a mathematical representation of the dataset to

- identify relationships between variables
- enable statistical inference
- make predictions about future sets of data
- help to visualize data so that non-analysts and stakeholders can get useful information

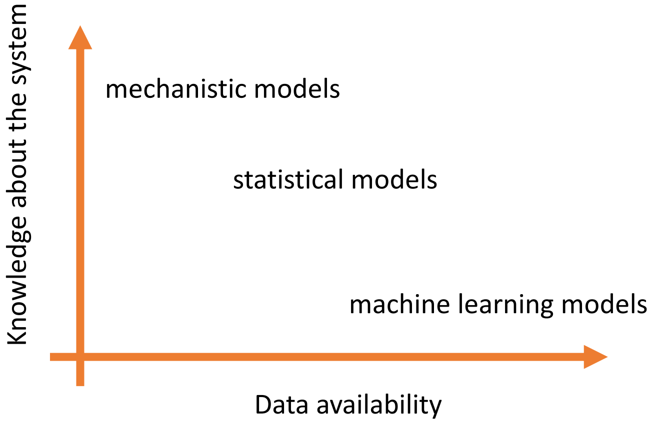




Model is a representation of a subject system – an **abstraction** that emphasizes key ideas about the system and ignores extraneous details.

14

Three modelling approaches used in Drug Development



STATISTICS IN MEDICINE
Statist. Med. **17**, 2501–2508 (1998)

A COMPARISON OF STATISTICAL LEARNING METHODS ON THE GUSTO DATABASE

MARGUERITE ENNIS[1], GEOFFREY HINTON[2], DAVID NAYLOR[3], MIKE REVOW[2] AND ROBERT TIBSHIRANI[4*]

"… we apply a battery of modern, adaptive non-linear learning methods to a large real database of cardiac patient data… we find that none of the methods could outperform a relatively simple logistic regression model previously developed for this problem."

Figured modified from *How to make ecological models useful for environmental management - ScienceDirect*

15

Three modelling approaches used in Drug Development

Quiz

Which type of model (Statistical or Machine learning or Mechanistic model) is used for:

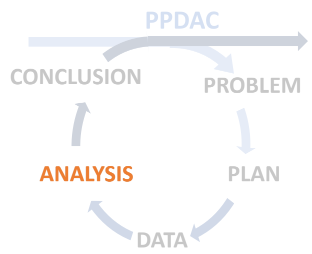
a) **Drug approval**

b) **Chat GPT**

c) **Analyzing how drugs spread in the body**

16

Part II agenda:
- ✓ From question to analysis
- ✓ From data to model
- ✓ Building machine learning models
- Google Thai project (interactive part)
- One model, many things to consider
- Explaining and communicating analysis outcomes



Building machine learning models

Let's unpick what machine learning is, how it works, and what it's used for.

17

Predicting biopsy values from MRIs

Nonalcoholic fatty liver disease (NASH) serious form of fatty liver disease

Liver biopsy: **gold standard** method for determining extent of liver fibrosis and treatment efficacy.
Drawbacks: an **invasive** procedure with all the undesirable side effects and potential complications of a surgical procedure, which can only be performed every 6-12 months.

http://dx.doi.org/10.7869/tg.2012.43

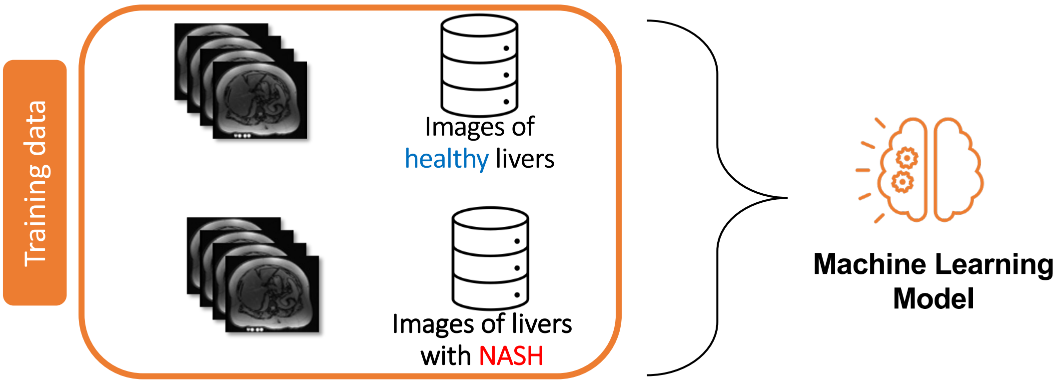
Important question:
Can we replace it with **non-invasive** and possibly more accurate procedures?

The Role of Statistical Thinking in Biopharmaceutical Research
Frank Bretz & Joel B. Greenhouse
Pages 458-467 | Received 19 Dec 2021, Accepted 19 May 2023, Published online: 24 Jul 2023

18

Predicting biopsy values from MRIs

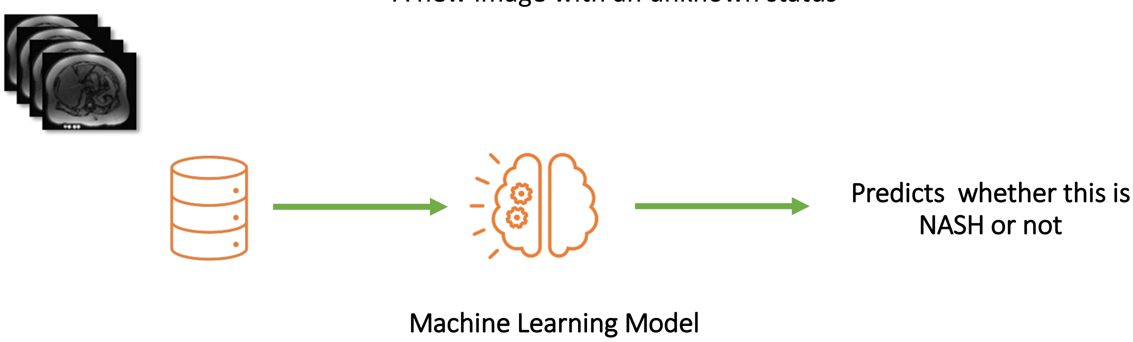
To do this we need MRI images that we know the disease status for each one (healthy or not-healthy)



19

Predicting biopsy values from MRIs

A new image with an unknown status



20

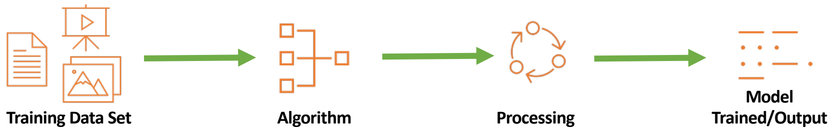
Supervised learning

Model that predicts **target variable (Y)** using values of other **variables (X)** in new data set.

Steps to **build** a model:

- Give a **training dataset** which contains many corresponding values of **Y** and **X**
- Model finds mathematical relationship that maps **X → Y**

Afterwards, the model can be used to **Y** in new data set



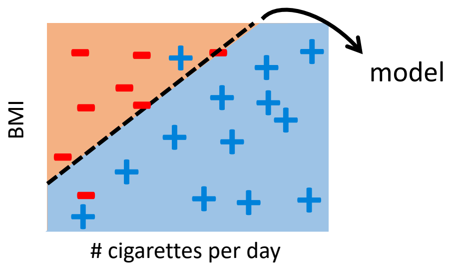
21

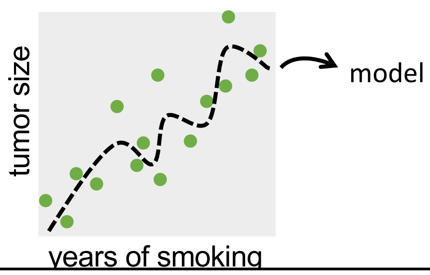
Classification vs Regression learning

In **classification tasks** we are predicting a **categorical target variable**
- Predicting whether a person will develop a cancer
- Predicting whether a person has NASH

VS

In **regression tasks** we are predicting a **continuous target variable**
- Predicting the size of a tumor
- Predicting the actual severity of NASH





22

Supervised vs Unsupervised learning

Supervised learning models are using **input** variables (features) to predict an **outcome** (label) variable of interest.

VS

Unsupervised learning models are given **unlabeled** data and allowed to discover patterns and insights without any explicit guidance or instruction.

Height X_1	Weight X_2	BP X_3	Sex X_4	Health Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

Height X_1	Weight X_2	BP X_3	Sex X_4
170	64	180	1
153	86	90	0
156	49	153	1
180	88	123	0
162	50	167	0
159	66	175	1
198	72	110	0
178	65	158	0
167	66	102	1
192	97	109	1
159	78	153	0

23

Learning from different data types: Structured vs Unstructured data

Structured data: Data that can be stored in a table, and every instance in the table has the same structure (i.e., set of attributes).

VS

Unstructured data: Data (typical large collections of files) that aren't stored in a structured database format (eg row-column format), eg images, text files, audio and video files etc

Height X_1	Weight X_2	BP X_3	Sex X_4	Health Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

24

Areas of ML that focus on unstructured data

Computer vision: learning from **images** lead to the development of **autonomous vehicles**, eg Tesla.

The Amazing Ways Tesla Is Using AI (forbes.com)

Natural Language Processing: learning the from databases of **text** lead to the development of **AI chat boxes**, eg ChatGPT.

What To Know About OpenAI's ChatGPT (forbes.com)

Automatic speech recognition: learning from **audio** lead to the development of **voice assistant devices**, eg Alexa.

How Does Amazon's Alexa Really Work? (forbes.com)

25

Part II agenda:
- ✓ From question to analysis
- ✓ From data to model
- ✓ Building machine learning models
- ✓ Google Thai project (interactive part)
- One model, many things to consider
- Explaining and communicating analysis outcomes

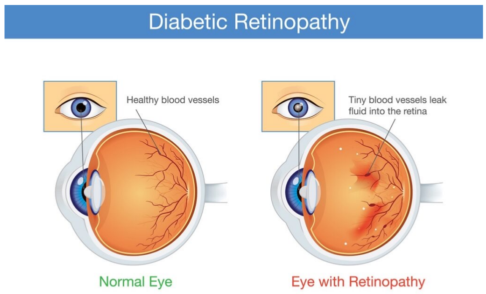
Google Thai project (interactive part)

Screening for diabetic **eye disease**

26

Context setting for this activity

- **Diabetes** is a growing problem around the world, and with diabetes are coming complications.
- One important is **diabetic retinopathy**, which affects blood vessels in the retina and is an eye condition that can cause vision loss and bindless.
- As a result, if you have diabetes, it is important to get a **comprehensive eye exam** at least once a year.



[Understanding the Stages of Diabetic Retinopathy | Elman Retina Group]

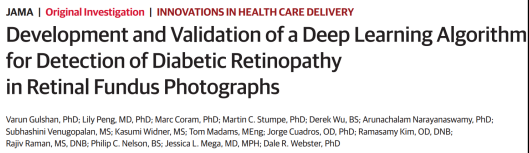
27

Context setting for this activity

In **2018** Google created an AI program to have an **accuracy rate of 95 percent** compared with 74 percent from opticians or eye doctors.

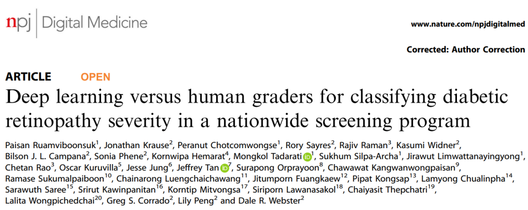
In **Thailand** there is **shortage of clinical specialists** that are qualified to perform this examination. Because of this, nurses conduct screenings by taking photos of the retina and sending them to an ophthalmologist, and this whole **process can take up to 10 weeks** for the patients to hear back, which, as you imagine, **puts the vision of these people in high danger.**

In collaboration with Thailand government the AI system was deployed to **screen for diabetic eye disease**


JAMA | Original Investigation | INNOVATIONS IN HEALTH CARE DELIVERY

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD; Lily Peng, MD, PhD; Marc Coram, PhD; Martin C. Stumpe, PhD; Derek Wu, BS; Arunachalam Narayanaswamy, PhD; Subhashini Venugopalan, MS; Kasumi Widner, MS; Tom Madams, MEng; Jorge Cuadros, OD, PhD; Ramasamy Kim, OD, DNB; Rajiv Raman, MS, DNB; Philip C. Nelson, BS; Jessica L. Mega, MD, MPH; Dale R. Webster, PhD


npj | Digital Medicine

www.nature.com/npjdigitalmed

Corrected: Author Correction

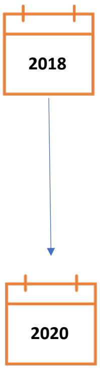
ARTICLE OPEN

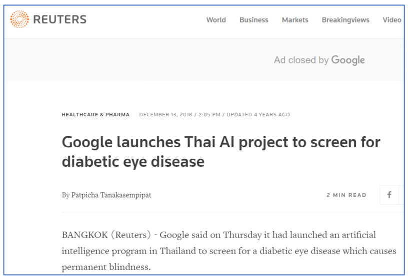
Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program

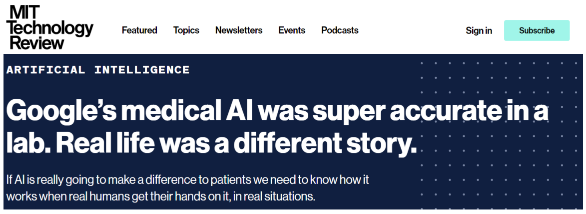
Paisan Ruamviboonsuk[1], Jonathan Krause[2], Peranut Chotcomwongse[1], Rory Sayres[2], Rajiv Raman[3], Kasumi Widner[2], Bilson J. L. Campana[2], Sonia Phene[2], Kornwipa Hemarat[4], Mongkol Tadarati[1], Sukhum Silpa-Archa[1], Jirawut Limwattanayingyong[1], Chetan Rao[3], Oscar Kuruvilla[5], Jesse Jung[6], Jeffrey Tan[7], Surapong Orprayoon[8], Chawawat Kangwanwongpaisan[9], Ramase Sukumalpaiboon[10], Chainarong Luengchaichawang[11], Jitumporn Fuangkaew[12], Pipat Kongsap[13], Lamyong Chualinpha[14], Sarawuth Saree[15], Srirut Kawinpanitan[16], Korntip Mitvongsa[17], Siriporn Lawanasakol[18], Chaiyasit Thepchatri[19], Lalita Wongpichedchai[20], Greg S. Corrado[2], Lily Peng[2] and Dale R. Webster[2]

28

Context setting for this activity



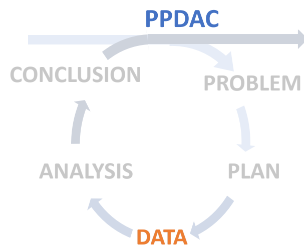

REUTERS
World Business Markets Breakingviews Video
Ad closed by Google
HEALTHCARE & PHARMA DECEMBER 13, 2018 / 2:05 PM / UPDATED 4 YEARS AGO
Google launches Thai AI project to screen for diabetic eye disease
By Patpicha Tanakasempipat
2 MIN READ
BANGKOK (Reuters) - Google said on Thursday it had launched an artificial intelligence program in Thailand to screen for a diabetic eye disease which causes permanent blindness.


MIT Technology Review
Featured Topics Newsletters Events Podcasts Sign in Subscribe
ARTIFICIAL INTELLIGENCE
Google's medical AI was super accurate in a lab. Real life was a different story.
If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.

TECH & SCIENCE
Google's AI Health Screening Tool Claimed 90 Percent Accuracy, but Failed to Deliver in Real World Tests

29

Let's discuss, what do you think went wrong?



- **Gradability**: the ability to read an image and make an assessment.
- After the deployment of the DL system the researchers observed that on average **20% of the images couldn't be read by the model**.
- Several causes:
  - Limited time to align patients
  - Imperfect lighting conditions (eg non-darkened environment)
  - No time to let pupils adjust for each photo (eg clinics weren't using dilation drops)

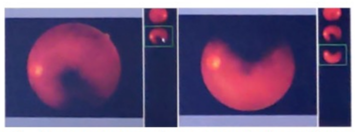
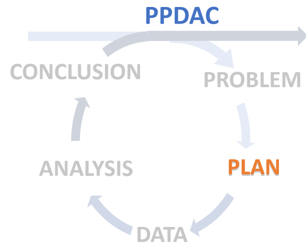


Figure 5. A nurse attempts to form a composite image of one eye by taking two images of the same eye, with varied lighting.

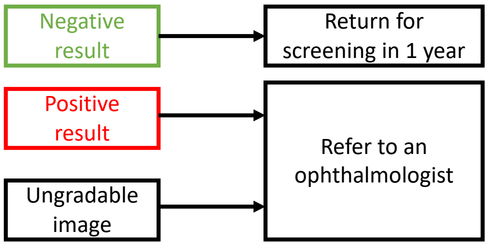
[A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy (acm.org)]

30

Let’s discuss, what do you think went wrong?



- **Challenges with the protocol**
- **Patient informed consent**
  The informed consent process was a big challenge and was made more complicated by the need to explain the AI system.
- **Internet speed and connectivity**



A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy (acm.org)

31

… here is how many issues were fixed

- Fixed cameras
- Added curtain to camera set up
- Amended protocol to wait for an ophthalmologist to review ungradable images before referring patients (a combination of AI and a human clinician was the best)
- Continued model improvements

A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy (acm.org)

32

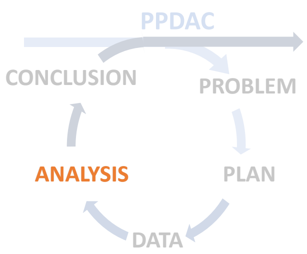
As a summary

- Remember the various **environmental and contextual factors** that need to be considered while building the ML model.
- Bring in your **domain expertise, ask questions, review deliverables** keeping these factors in mind.
- You need to **keep a close eye in the data and the model performance from early enough** to make sure that the quality of the data is in the level you were expecting.

[A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy (acm.org)]

33

Part II agenda:
- ✓ From question to analysis
- ✓ From data to model
- ✓ Building machine learning models
- ✓ Google Thai project (interactive part)
- ✓ One model, many things to consider
- Explaining and communicating analysis outcomes



One model, many things to consider

Is our model good enough? Is our model fair enough?
How our model takes decisions?

34

How "good" is our model depends on the question we want to answer

We are testing a patient for a life-threatening disease (a cost-sensitive decision).

- A **false positive (false alarm)** will lead to further tests which will eventually reveal the misdiagnosis.
- A **false negative** means that the disease is left undetected and thus untreated, with **potentially lethal results.**

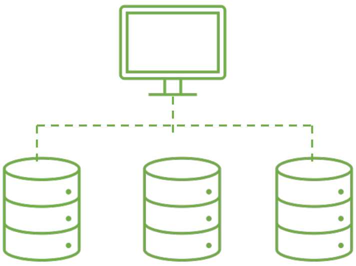
35

Thinks to consider when building a model: We shouldn't bias our findings

How much we trust our model it is equivalent to how much we trust the **answer** it provides to our **question**

We should not tailor our model to a specific dataset.

E.g., when we have some interesting finding in our training data, we want to be confident that this will also hold in different dataset.

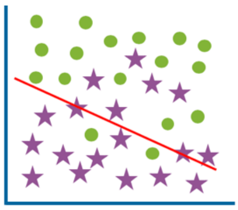


36

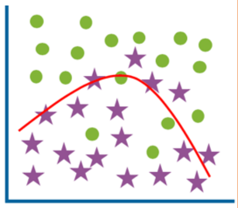
Thinks to consider when building a model: We shouldn’t bias our findings

Can you identify the “best” model to separate circles from stars?

Option A



Option B



Option C

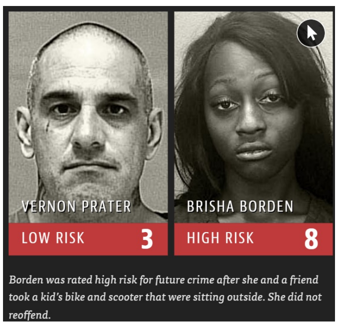
37

Ethical aspects and fairness

Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn’t Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe’s assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from



Taken from [Machine Bias — ProPublica]

38

Examples of (potential) bias from healthcare

AI classified black patients as lower risk vs. equally sick white patients due to lower past healthcare spend

Asthma patients with pneumonia predicted to be low risk due to good outcomes after aggressive treatment

Women with heart attack symptoms do not get this diagnosis suggested by AI tool, men with identical characteristics do

Image datasets used for training skin cancer detection algorithms primarily include patients with light skin tones

Controversy about race adjustment in algorithms like the calculation of eGFR

Test for neurological diseases, only worked for native English speakers with a particular Canadian accent

All 91 COVID-19 models in a systematic review judged to be at high-risk of bias

via Björn Holzhauer

39

Association vs Causation
Does carrying a lighter cause lung cancer?

Height X_1	Weight X_2	BP X_3	Lighter X_4	Lung Cancer Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

There's no question that carrying a lighter in your pocket is associated with having lung cancer

*...and Machine learning models are very powerful to detect these **associations**.*

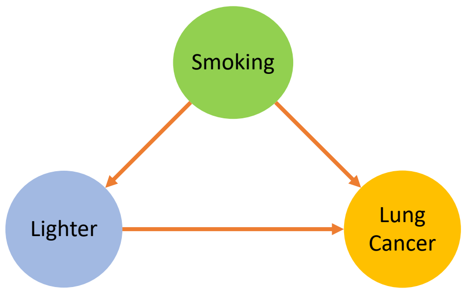
*... but it doesn't mean that carrying a lighter **causes** lung cancer.*

40

Association vs Causation
Does carrying a lighter cause lung cancer?

Height X_1	Weight X_2	BP X_3	Lighter X_4	Lung Cancer Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

Carrying a lighter and lung cancer are both caused by **a common factor**, which is **smoking**, but they don't cause each other

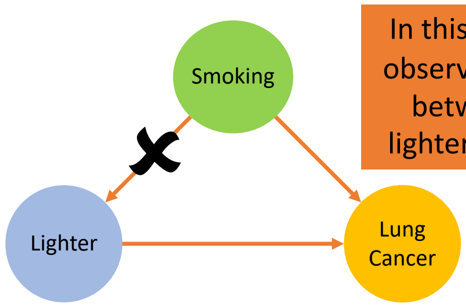


41

Association vs Causation
Does carrying a lighter cause lung cancer?

Height X_1	Weight X_2	BP X_3	Lighter X_4	Lung Cancer Y
170	64	180	1	1
153	86	90	0	1
156	49	153	1	0
180	88	123	0	0
162	50	167	0	1
159	66	175	1	0
198	72	110	0	1
178	65	158	0	0
167	66	102	1	1
192	97	109	1	1
159	78	153	0	0

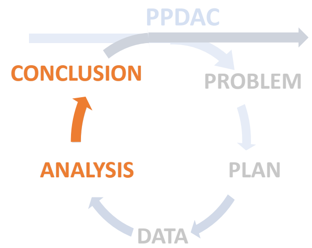
Let's run a controlled experiment, like our clinical trials:
We ask at random half of the people to carry a lighter, and the other half not to.



In this data we will not observe any correlation between carrying a lighter and lung cancer

42

Part II agenda:
- ✓ From question to analysis
- ✓ From data to model
- ✓ Building machine learning models
- ✓ Google Thai project (interactive part)
- ✓ One model, many things to consider
- ✓ Explaining and communicating analysis outcomes



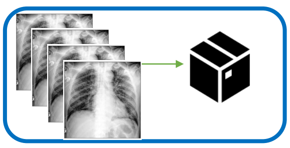
Explaining and communicating

How does our model take decisions?

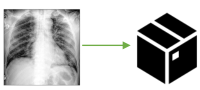
43

Question: Can we diagnose COVID using chest X-ray images?

Train model

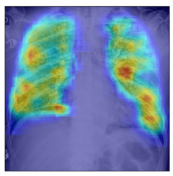


New person



Prediction: Positive to COVID

Explanation: The red region is important to accurately classify this X-ray

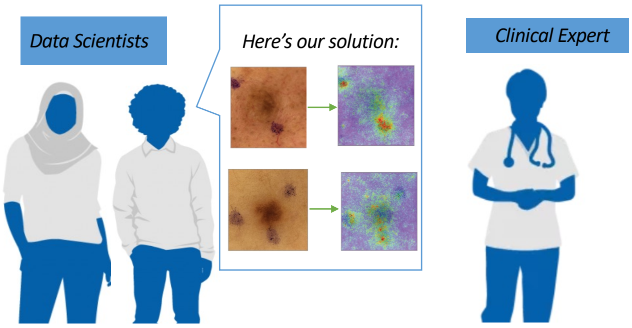


This type **of heat maps** helps us to understand how important are different regions in the decision of the model

Deep Learning COVID-19 Features on CXR using Limited Training Data Sets

44

Question:
Can we diagnose melanoma from photos?



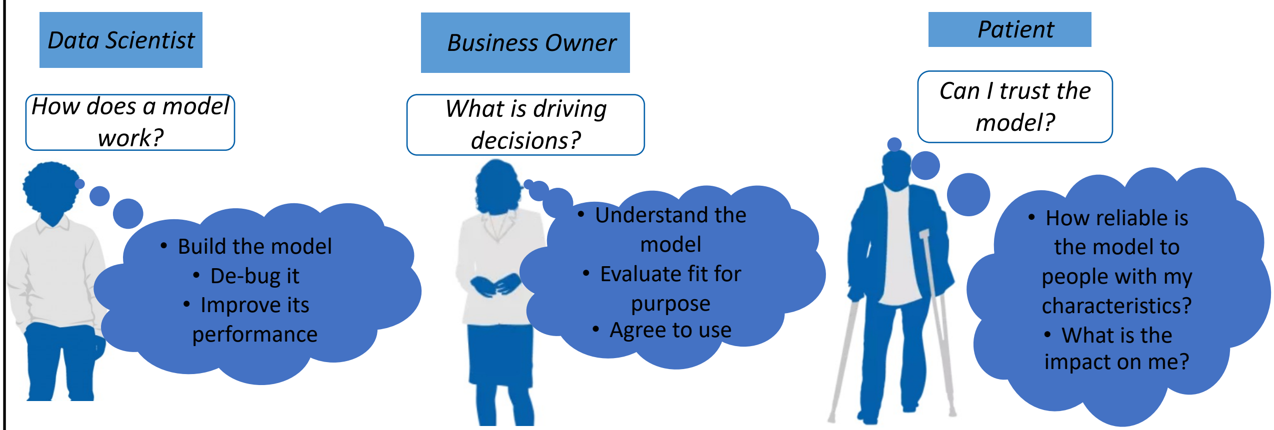
Explanation: Heat maps reveal that **skin markings are of high relevance for neural network's prediction** of malignant melanomas, while the nevus itself is mostly ignored

Conclusion: What the model used to decide was different from the nevus itself, **therefore it is critical for the data scientist to work closely with the domain experts.**

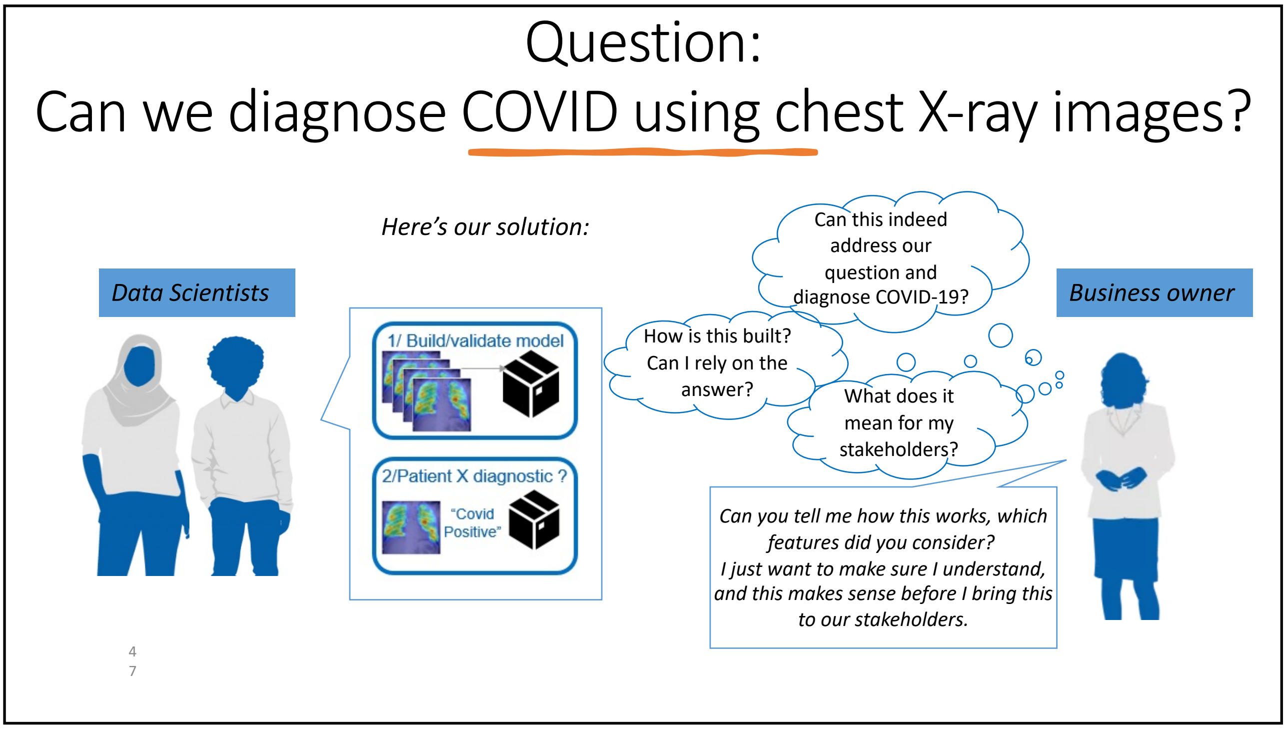
Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition

45

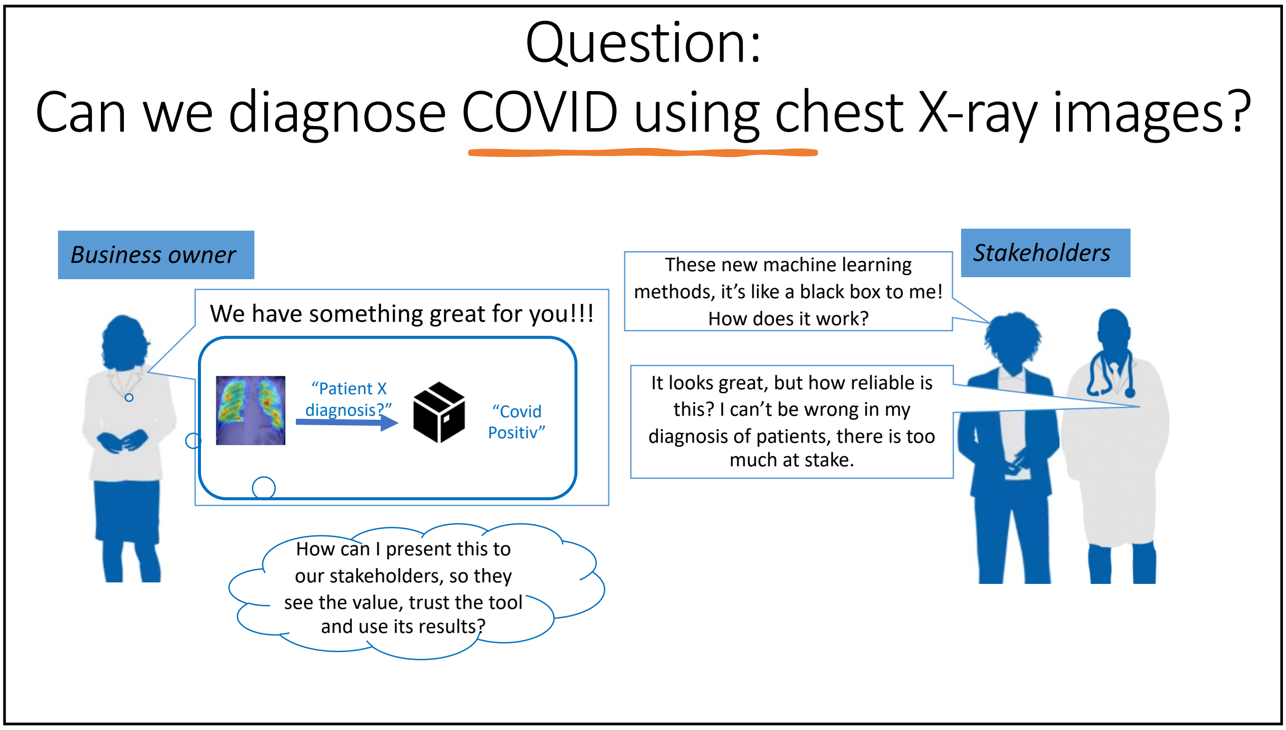
Different stakeholders, different concerns, different ways to communicate



46


Question:
Can we diagnose COVID using chest X-ray images?
Here's our solution:
Data Scientists
1/ Build/validate model
2/Patient X diagnostic ?
"Covid Positive"
Can this indeed address our question and diagnose COVID-19?
How is this built? Can I rely on the answer?
What does it mean for my stakeholders?
Business owner
Can you tell me how this works, which features did you consider? I just want to make sure I understand, and this makes sense before I bring this to our stakeholders.
4
7

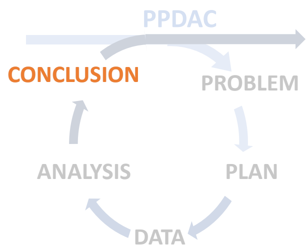
47


Question:
Can we diagnose COVID using chest X-ray images?
Business owner
We have something great for you!!!
"Patient X diagnosis?"
"Covid Positiv"
How can I present this to our stakeholders, so they see the value, trust the tool and use its results?
These new machine learning methods, it's like a black box to me! How does it work?
Stakeholders
It looks great, but how reliable is this? I can't be wrong in my diagnosis of patients, there is too much at stake.

48

From analysis results to interpretation and effective communication

- Results such as outputs from the analysis strategy are not the end product.
- They **need to be interpreted and contextualized** with the domain experts before any form of communication.
- **Stories can make complex concepts more understandable** as people can relate to them. Visuals can be a powerful way to convey stories.



49

Three principles for effective (visual) communication

- ➢Have a clear purpose
- ➢Show the data clearly
- ➢Make the message obvious

https://graphicsprinciples.github.io/

Tutorial | Open Access

Effective Visual Communication for the Quantitative Scientist

Marc Vandemeulebroecke, Mark Baillie, Alison Margolskee, Baldur Magnusson

First published: 22 July 2019 | https://doi.org/10.1002/psp4.12455 | Citations: 9

50

Three principles for effective (visual) communication

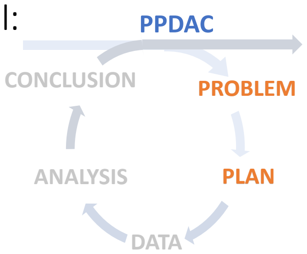
- 1st principle: Have a clear purpose
  - Understand the question you are trying to answer
  - Identify the quantitative evidence to answer that question
  - Know your audience and focus the design to support their needs
- 2nd principle: Show the data clearly
  - Have an appropriate graph type to display your data
  - Be faithful to the data and avoid misrepresentation
  - Maximize data to ink ratio (reduce distraction, less is more)
- 3rd principle: Make the message obvious
  - Use proximity and alignment to aid in comparisons
  - Use colors and annotations to highlight important details
  - Use a meaningful title to bring your message across

51

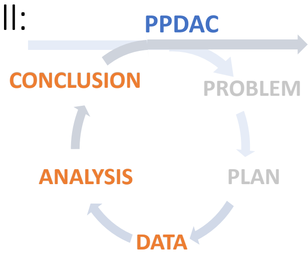
Wrap-up …

- Data science thinking is an integrated set of thinking skills and practices refocused for answering questions with data.
- A good workflow is an established set of habits that help drive you forward towards your goal. They enable complexity to scale in the right areas.
- This workflow demonstrates the steps for abstracting and solving a real problem. An impactful solution requires a clear understanding of how things work.

Part I:



Part II:



52

Material and Acknowledgments

- All the material and more are online: https://datascience-thinking.github.io
- Special thanks to ...
  - Conor Moloney
  - Carsten Philipp Mueller
  - Malika Cremer
  - Peter Krusche
  - Björn Holzhauer
  - Janice Branson
  - David Ohlssen

53

Thank you

for attending this workshop, any

questions?

54