

Do You Know What Your Model is Doing?

How Human Bias Influences Machine Learning

Jim Box— Principal Data Scientist
Elena Snaveley— Senior Data Scientist



10th–13th November
RAI Amsterdam
The Netherlands

EU 2019

The Clinical Data
Science Conference



Machine Learning in Clinical Research

**HEALTH
IT ANALYTICS**
xtelligent HEALTHCARE MEDIA

Home News Features Inter

Population Health Precision Medicine Quality & Governance Tools & Strategies Analytics in Action Focu

FDA: Real-World Data, Machine Learning Critical for Clinical Trials

Real-world data, machine learning, and artificial intelligence will help to support effective clinical trials, says the FDA's Scott Gottlieb.

<https://healthitanalytics.com/news/fda-real-world-data-machine-learning-critical-for-clinical-trials>

Life Science
Leader

Home Magazine Featured Sections Industry Topics Blogs Events Subscribe

Guest Column | July 2, 2019

Machine Learning In Clinical Trials: What Will The Future Hold (And What's Holding Us Back)?

By Kellie Rademacher, Pharm.D., Precision for Value

Former FDA Commissioner Dr. Scott Gottlieb stressed the need for modernizing the clinical trials process in a speech to the Bipartisan Policy Center in January of this year.¹ He is quoted as saying, "digital technologies are one of our most promising tools for making healthcare more efficient." Improving efficiency in clinical trial development is only one potential enhancement that can result from the use of machine learning. Machine learning and artificial intelligence (AI) are often used interchangeably, but that assumption is incorrect. Machine learning is the



<https://www.lifescienceleader.com/doc/machine-learning-in-clinical-trials-what-will-the-future-hold-and-whats-holding-us-back-0001>

Machine Learning Tools Help Predict Clinical Trial Outcomes

Using machine learning tools to forecast clinical trial outcomes can help stakeholders speed the drug and device approval process.



Source: Getty Images

<https://healthitanalytics.com/news/machine-learning-tools-help-predict-clinical-trial-outcomes>

By Jessica Kent

Machine Learning Flow

DATA

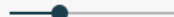
Which dataset do you want to use?



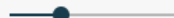
Ratio of training to test data: 50%



Noise: 15



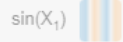
Batch size: 10



REGENERATE

FEATURES

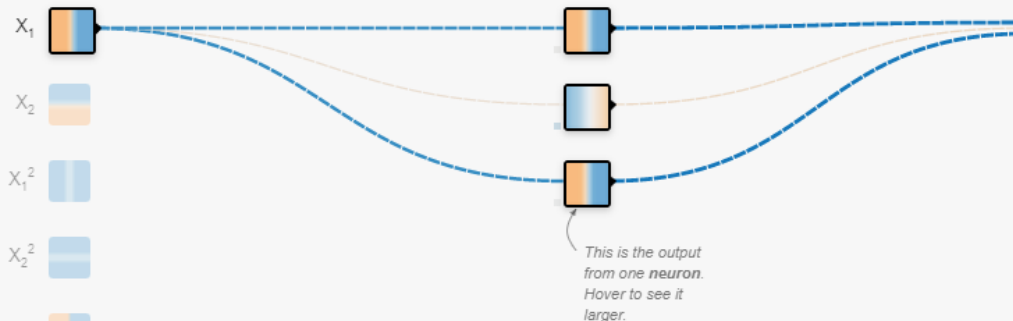
Which properties do you want to feed in?



1 HIDDEN LAYER



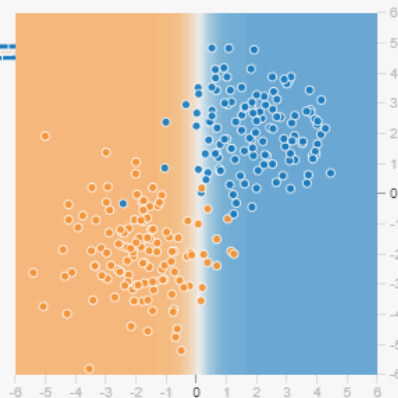
3 neurons



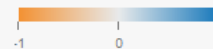
OUTPUT

Test loss 0.085

Training loss 0.106



Colors shows data, neuron and weight values.



☐ Show test data

☐ Discretize output

<https://playground.tensorflow.org/>



BIASED DATA GIVES BIASED RESULTS

Finding the Right Candidate

Amazon, 2014 - 2017

- **Challenge:** Amazon gets hundreds of applicants for open positions
- **Solution:** Develop some ML algorithms to automate the review, and rank applicants like reviewers rank products. Focus the hiring efforts on the 5-star candidates
- **Data:** Look at all resumes over the past 10 years and score them on if the applicant was hired

Finding the Right Candidate

Amazon, 2014 - 2017

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

8 MIN READ



SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>

Finding the Right Candidate

Amazon, 2014 - 2017

ELENA SNAVELY

100 SAS Campus Drive
Cary, NC 27513

Elena.Snavely@sas.com

PROFESSIONAL OVERVIEW

Dynamic analyst with a proven record of achieving business objectives, committed to maximizing project potential. Experienced in critically evaluating information from disparate sources while proactively communicating results through presentations, reports, and discussions. Skilled at developing and optimizing models, conducting analysis utilizing statistical tools, and realizing business value from their output.

Key accomplishments and abilities:

- Capturing Customer Requirements
- Executing Project Plans
- Data Manipulation and Feature Engineering
- Executive Presentations
- Data Manipulation and Feature Engineering
- Managing and coaching groups
- Communicating effectively
- Design of User Interfaces

EXTRACURRICULAR

WOMEN'S CHESS CLUB CAPTAIN
Smith College, August 2006-May2008

Northampton, Massachusetts

Finding the Right Candidate

Amazon, 2014 - 2017

ELENA SNAVELY

100 SAS Campus Drive
Cary, NC 27513

Elena.Snively@sas.com

PROFESSIONAL OVERVIEW

Dynamic analyst with a proven record of achieving business objectives, committed to maximizing project potential. Experienced in critically evaluating information from disparate sources while proactively communicating results through presentations, reports, and discussions. Skilled at developing and optimizing models, conducting analysis utilizing statistical tools, and realizing business value from their output.

Key accomplishments and abilities:

- Capturing Customer Requirements
- Executing Project Plans
- Data Manipulation and Feature Engineering
- Executive Presentations
- Data Manipulation and Feature Engineering
- Managing and coaching groups
- Communicating effectively
- Design of User Interfaces

EXTRACURRICULAR



WOMEN'S CHESS CLUB CAPTAIN

Smith College, August 2006-May2008

Northampton, Massachusetts

Finding the Right Candidate

Amazon, 2014 - 2017

ELENA SNAVELY

100 SAS Campus Drive
Cary, NC 27513

Elena.Snively@sas.com

PROFESSIONAL OVERVIEW

Dynamic analyst with a proven record of achieving business objectives, committed to maximizing project potential. Experienced in critically evaluating information from disparate sources while proactively communicating results through presentations, reports, and discussions. Skilled at developing and optimizing models, conducting analysis utilizing statistical tools, and realizing business value from their output.

Key accomplishments and abilities:

- Capturing Customer Requirements
- Executing Project Plans
- Data Manipulation and Feature Engineering
- Executive Presentations
- Data Manipulation and Feature Engineering
- Managing and coaching groups
- Communicating effectively
- Design of User Interfaces

EXTRACURRICULAR

WOMEN'S CHESS CLUB CAPTAIN

Northampton, Massachusetts



Smith College August 2006-May2008

Finding the Right Candidate

Amazon, 2014 - 2017

ELENA SNAVELY

100 SAS Campus Drive
Cary, NC 27513

Elena.Snavely@sas.com

PROFESSIONAL OVERVIEW

Dynamic analyst with a proven record of achieving business objectives, committed to maximizing project potential. Experienced in critically evaluating information from disparate sources while proactively communicating results through presentations, reports, and discussions. Skilled at developing and optimizing models, conducting analysis utilizing statistical tools, and realizing business value from their output.

Key accomplishments and abilities:

- ✓ • Capturing Customer Requirements
- ✓ • Executing Project Plans
- Data Manipulation and Feature Engineering
- Executive Presentations
- Data Manipulation and Feature Engineering
- Managing and coaching groups
- ✗ • Communicating effectively
- Design of User Interfaces

EXTRACURRICULAR

WOMEN'S CHESS CLUB CAPTAIN
Smith College, August 2006-May2008

Northampton, Massachusetts

Takeaways

What can we learn from this

- Algorithms can only learn from the data they are trained on
- Biases in the training data will be identified and magnified by an algorithm
- Careful consideration of the data available and the goals of the project is vital to success

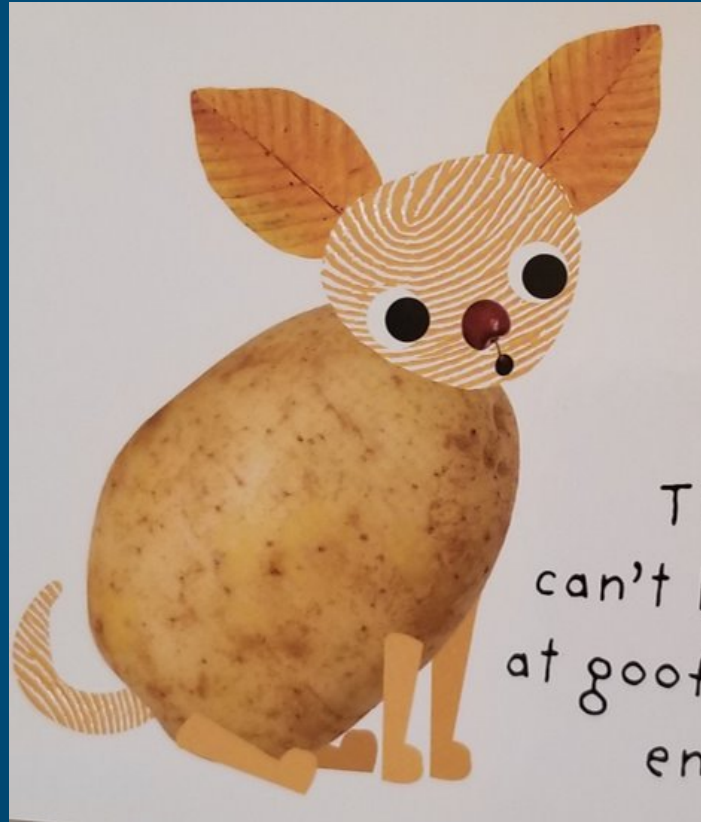
Training Models

Models learn like kids



Training Models

Data Selection Imperative



Happy Dog and Other Furry Friends
written by Robert Newton
illustrated by Ellie Boulwood

Process Bias

Data Must be Representative

Genetics research 'biased towards studying white Europeans'

Ethnic minorities set to miss out on medical benefits of research, scientist warns

People from minority ethnic backgrounds are set to lose out on medical benefits of genetics research due to an overwhelming bias towards studying white European populations, a leading scientist has warned.

In a recent study, published in *Psychiatric Genetics*, Curtis found that a commonly used genetic test to predict schizophrenia risk gives scores that are 10 times higher in people with African ancestry than those with European ancestry. This is not because people with African ancestry actually have a higher risk of schizophrenia, but because the genetic markers used were derived almost entirely from studies of individuals of European ancestry.

<https://www.theguardian.com/science/2018/oct/08/genetics-research-biased-towards-studying-white-europeans>

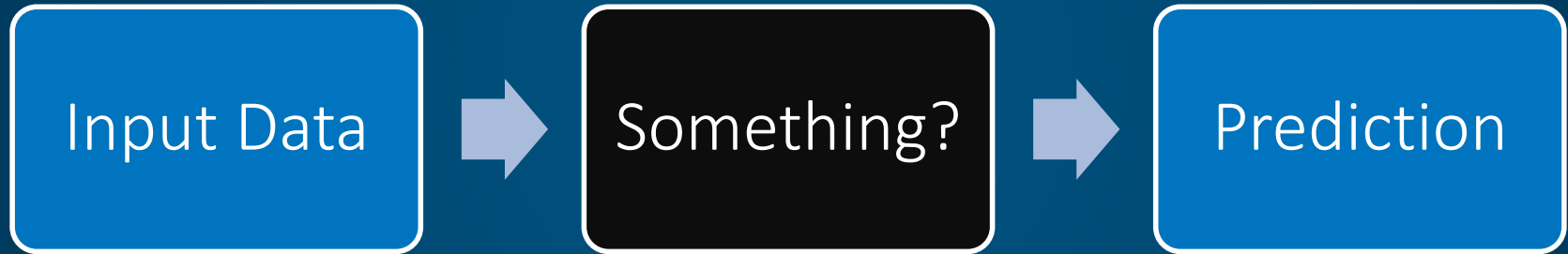


MODELS CAN FIND UNEXPECTED CORRELATIONS

Beware the Black Box Approach

Models can be very complex

Automation Enables Magic



Understand Why the Model Predicts

Image Analysis



Predicted: Wolf
True: Wolf



Predicted: husky
True: husky



Predicted: Wolf
True: Wolf



Predicted: Wolf
True: Wolf



Predicted: husky
True: husky

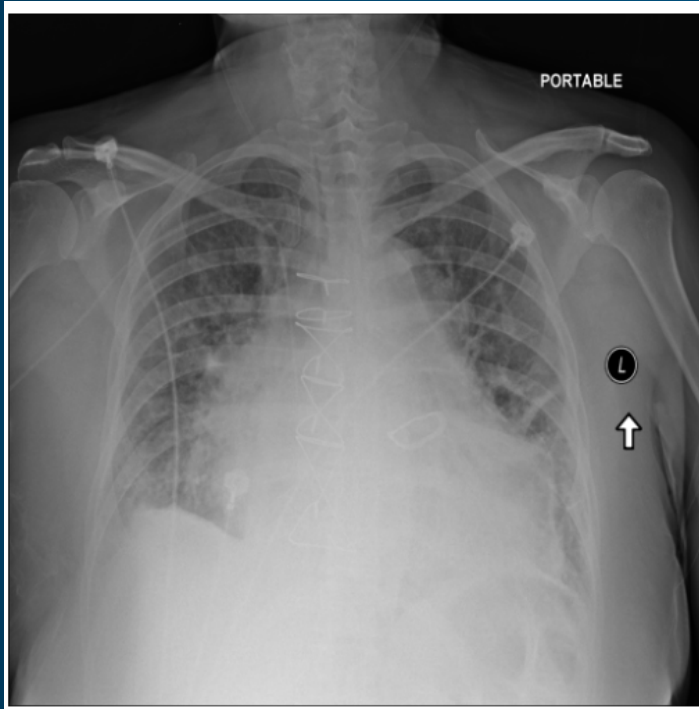


Predicted: Wolf
True: Husky

<https://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf>

Diagnosing Cardiomegaly (Enlarged Heart)

Convolutional Neural Network (CNN)

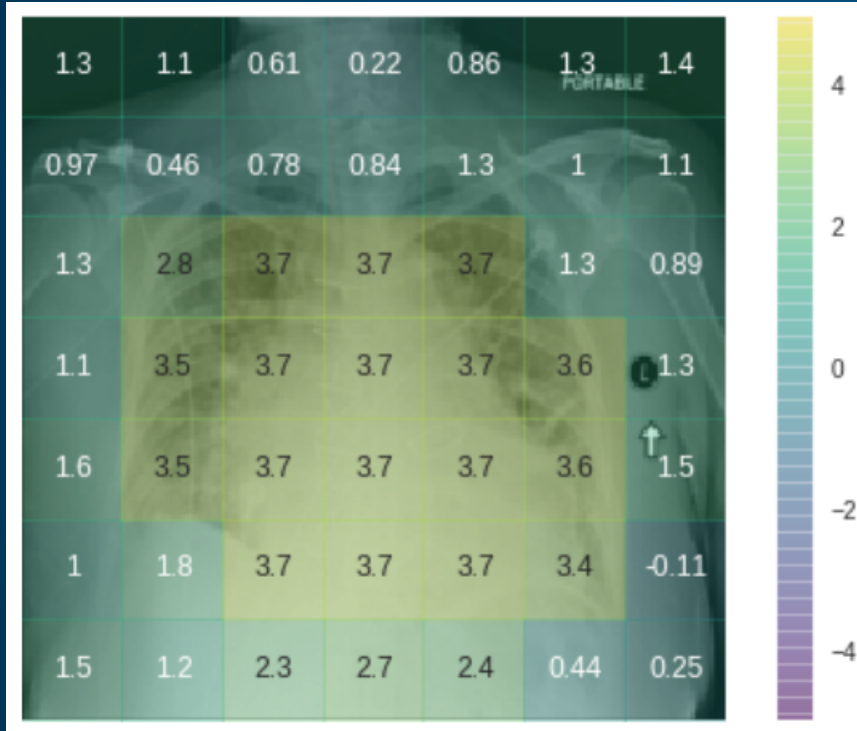


- Model trained on X-rays that had been scored as either positive or negative by radiologists
- Examine some x-rays of people with cardiomegaly to see how the model scored
- This one: $p = .752$, so model seems to be doing the right thing

<https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98>

Diagnosing Cardiomegaly (Enlarged Heart)

Convolutional Neural Network (CNN)

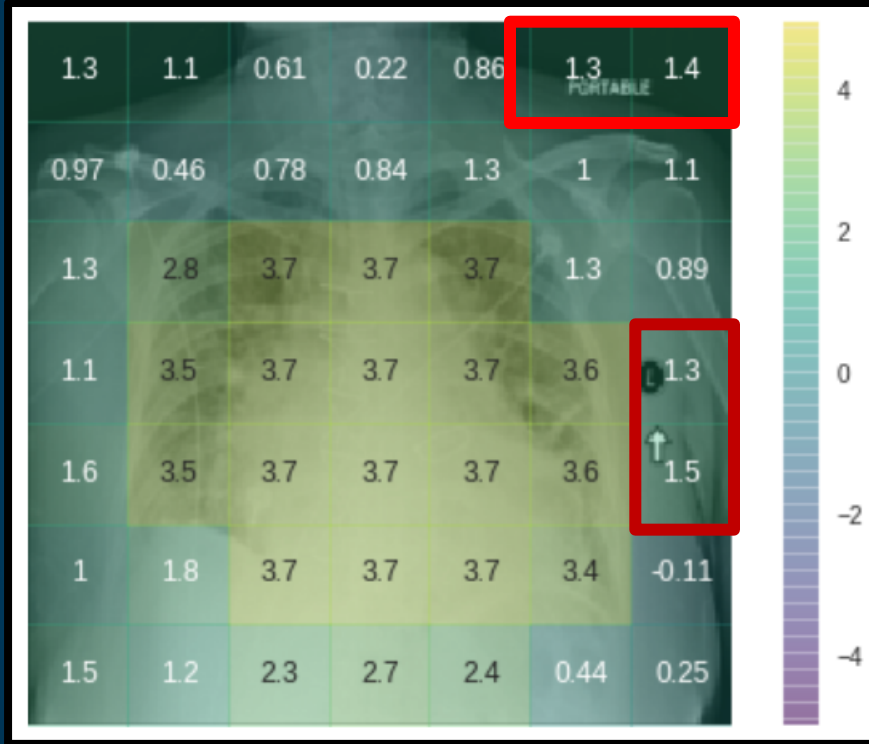


- Closer look at how the model is working
- Heatmap shows how different parts of the image contributed to the prediction (yellow positive, purple negative)
- Seems Focused on the heart area, as expected

<https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98>

Diagnosing Cardiomegaly (Enlarged Heart)

Convolutional Neural Network (CNN)

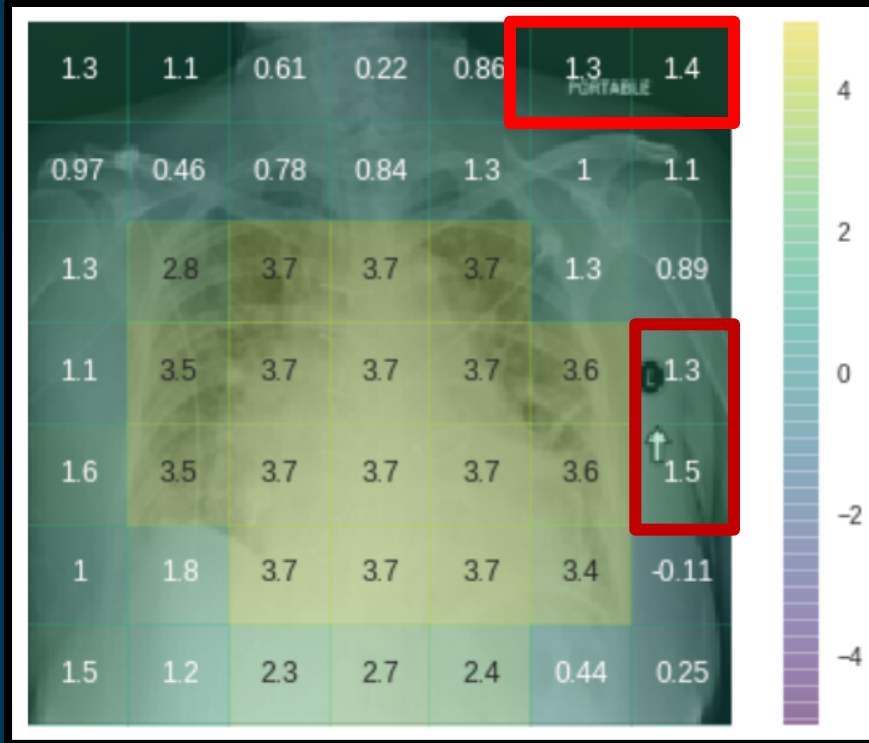


- Closer look at how the model is working
- Heatmap shows how different parts of the image contributed to the prediction (yellow positive, purple negative)
- Seems Focused on the heart area, as expected
- Some surprises, though

<https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98>

Diagnosing Cardiomegaly (Enlarged Heart)

Convolutional Neural Network (CNN)

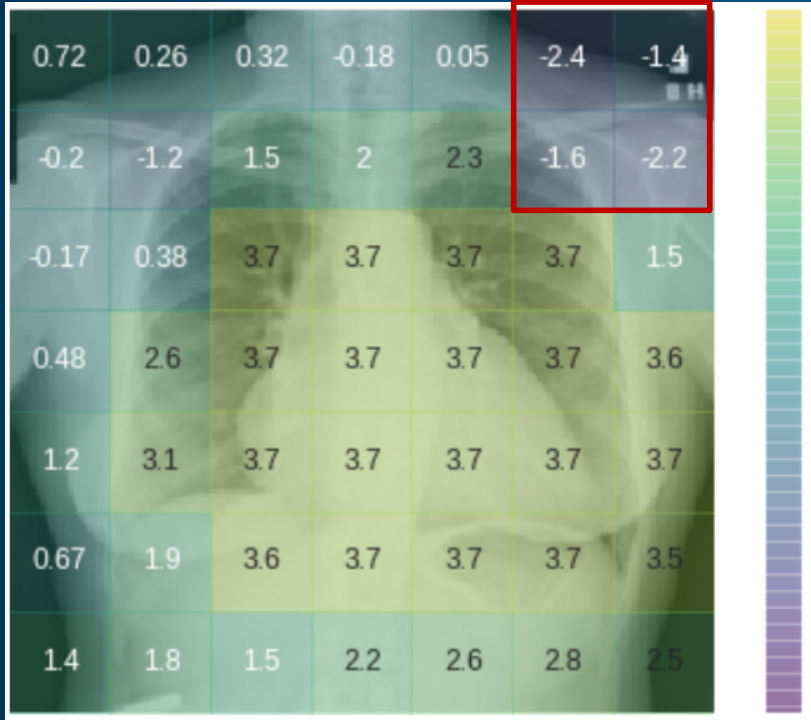


- The model figured out that when a portable x-ray was used, those patients were more likely to have the indication
- Portable x-rays are used more frequently when patients are sick and can't come to the imaging room
- Basically, the model showed that sick people are sick

<https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98>

Diagnosing Cardiomegaly (Enlarged Heart)

Convolutional Neural Network (CNN)



- Different patient with the condition
- Model gave 93.7% prediction of cardiomegaly
- Note the downgrade for missing the “Portable” – this image was taken in an outpatient setting

<https://medium.com/@jrzech/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98>

Takeaways

Understanding is the Key to Reliability

- Results can be surprising, but should not be inexplicable
- Understand why the model is predicting what it is
 - Can be easy with a list of variables
 - Can be challenging with black box models
- Models that can't be explained often are not implemented

Model Implementation

Understanding is the Key to Reliability



- 2009 contest to improve recommendation engine by 10%
- Team hit the target a few days before end of contest
- Netflix awarded \$1 Million prize
- Never implemented the model due to complexity

<https://www.wired.com/2012/04/netflix-prize-costs/>



WHAT CAN YOU DO?

AWARENESS OF THE ISSUE IS VITAL

Human Bias in Machine Learning

Steps in the Right Direction

- Frame the Problem
 - Understand the question being asked
 - Make sure the training data is representative of the population you to which you want apply the model
 - Understand the goals – are you trying to find job applicants that are like the ones you already have, or are you trying to find the “best” candidates

Human Bias in Machine Learning

Steps in the Right Direction

- Define Fairness
 - Think about how you are using the data, is it applicable to your whole population
 - Are there regulatory or ethical rules about the types of inputs you can use?
 - Are you using variables that can serve as proxies of disqualified variables (like race)?
 - Codifying this can be uncomfortable, be sure people know why you are taking this approach

Human Bias in Machine Learning

Steps in the Right Direction

- Check for Bias
 - Study the input data, be sure it is representative and as free from bias as possible
 - Think about possible sources of human bias – awareness here is key
 - Test the results of the model – do they change unexpectedly based on factors such as age, race or gender?

Human Bias in Machine Learning

Steps in the Right Direction

- Examine Model Interpretability
 - Understand why the model is producing the results – don't just rely on a black box
 - This is an iterative process, many cycles might be required as you are examining and explaining results
 - Get buy-in from Subject Matter Experts – should make medical sense
 - Explore statistical methods for explaining results
 - Local Interpretable Model-Agnostic Explanations (LIME)
 - Partial Dependence
 - Individual Conditional Expectation (ICE)



Questions?

Check paper for useful links

Jim Box
Jim.Box@sas.com

Elena Snavelly
Elena.Snavelly@sas.com