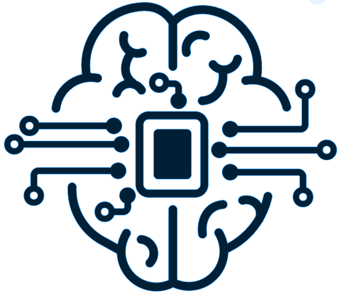


Responsible AI

Josefin Rosén, Ph.D.
Nordic Practice Leader AI & Analytics
SAS Institute

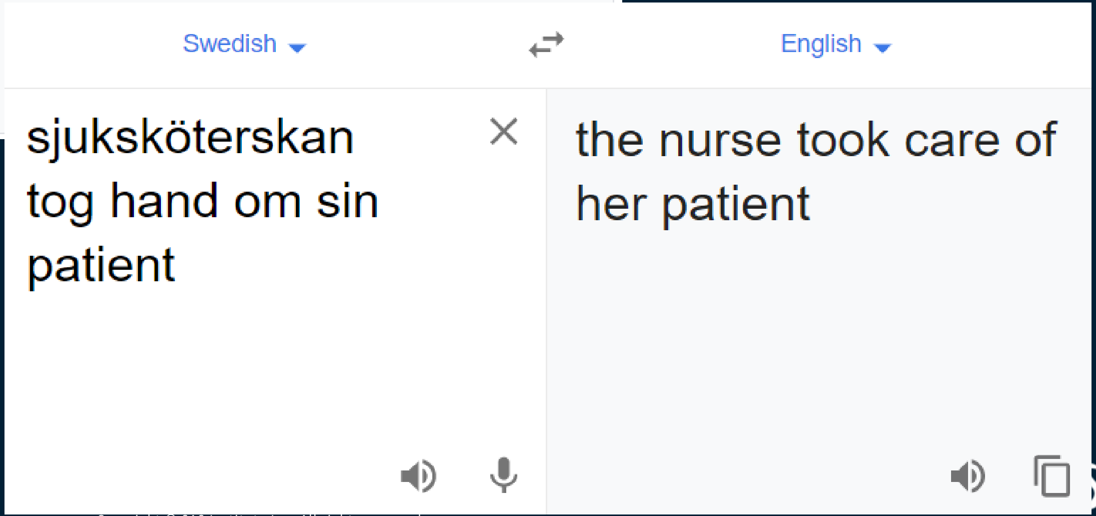
”With great power
comes great
responsibility”
- Uncle Ben, Spider Man


”With great power
comes great
responsibility”
- Uncle Ben, Spider Man


ON/OFF
NOR./INV.
ON/OFF
ON/OFF
EQUALIZATION
PRE-EQ
PATCH
MASTER
SHIFT
HIGH
PRESENCE

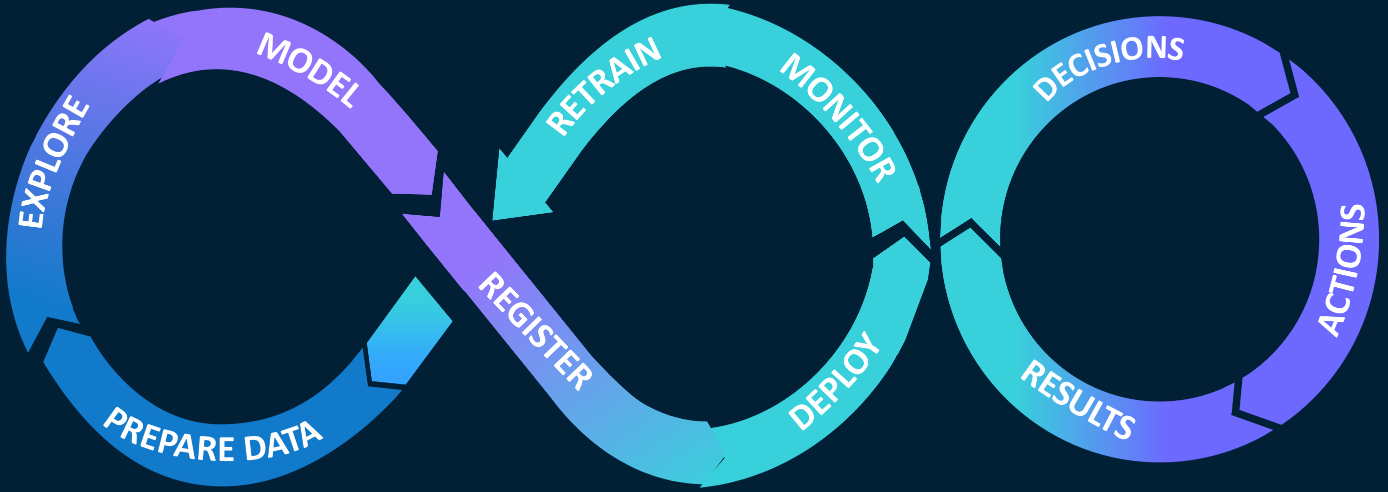
Google translate

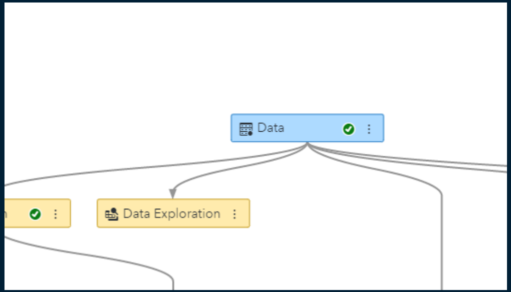

English
Swedish
the nurse took care of his patient
sjuksköterskan tog hand om sin patient

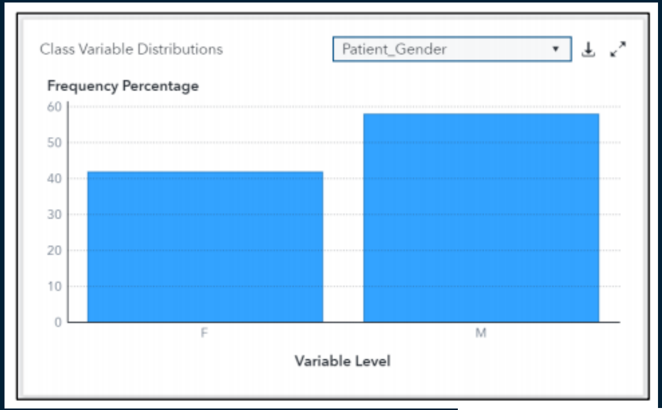

Swedish
English
sjuksköterskan tog hand om sin patient
the nurse took care of her patient

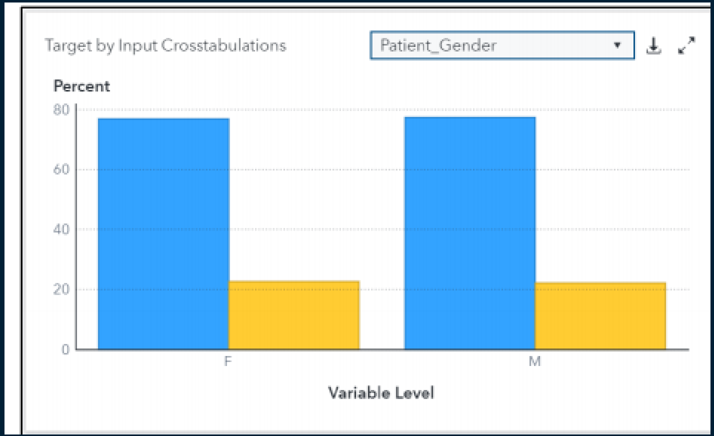

MEN WANTED
FIRST CLASS MECHANICS & GOOD HELPERS OF ALL KINDS ARE BEING HIRED NOW, CONTACT CHARLES SCHMIDT FOR AN AP-PLICATION FOR WORK HERE.
STURGEON BAY SHIPBLDG. & D.D. CO.
STURGEON BAY, WIS.
Lincoln School of Engineering

Responsibility throughout the whole analytics life cycle




Data
Data Exploration


Class Variable Distributions
Patient_Gender
Frequency Percentage
60
50
40
30
20
10
0
F
M
Variable Level

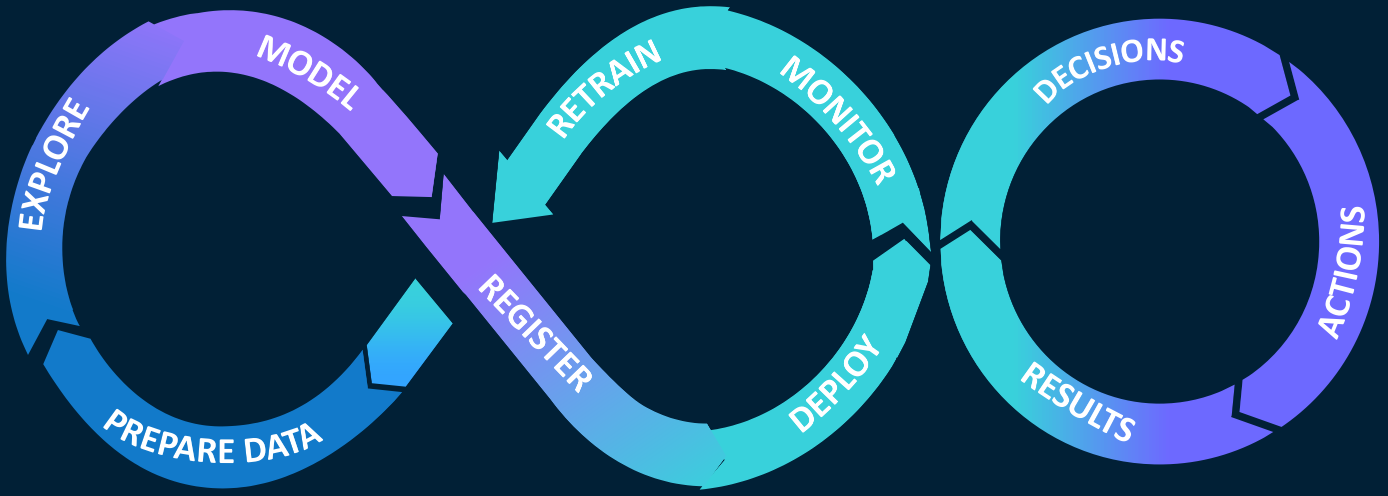

Target by Input Crosstabulations
Patient_Gender
Percent
80
60
40
20
0
F
M
Variable Level

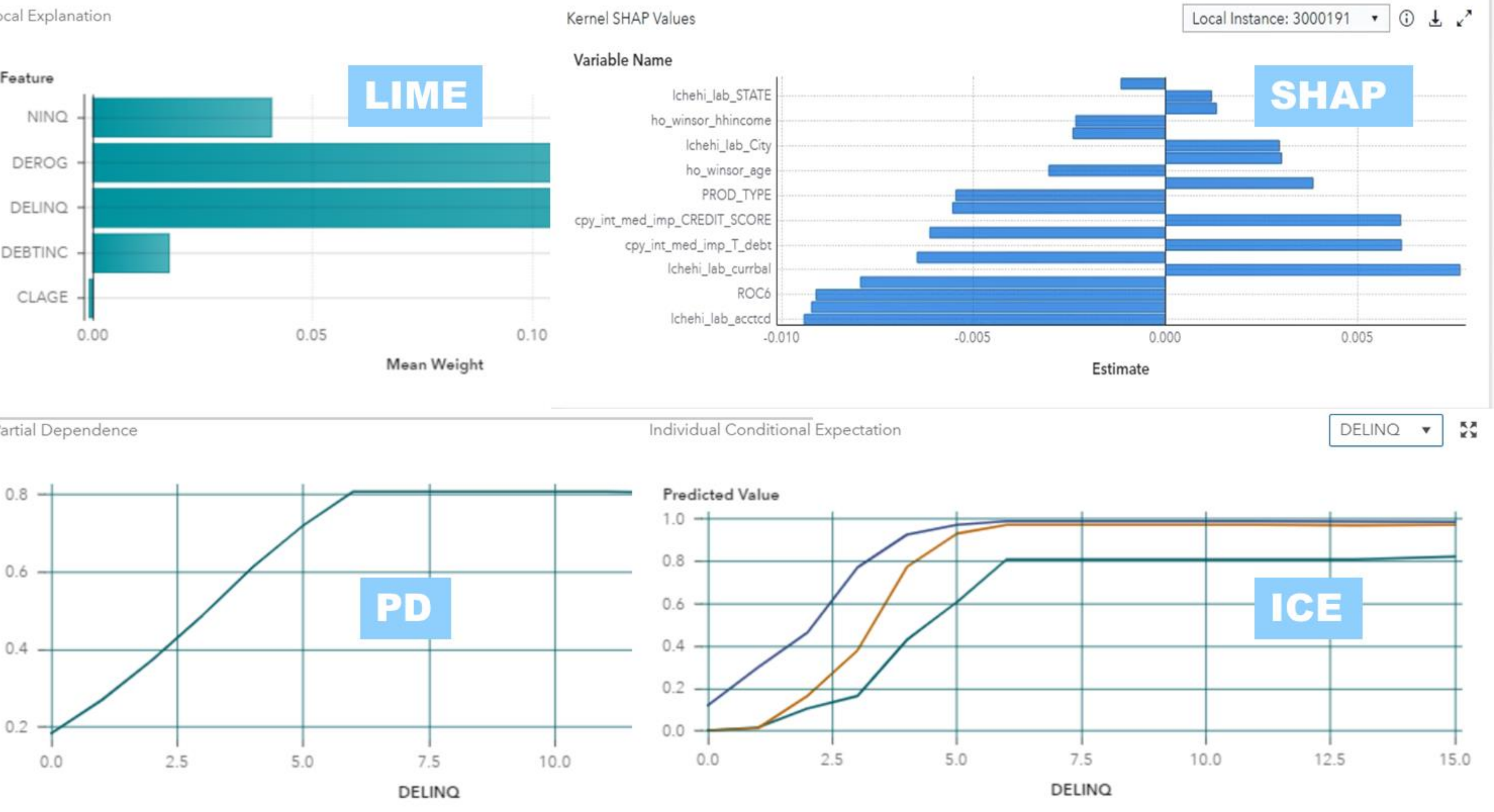

Important Inputs
Relative Importance
Variable Name

“We need to shine a light on the black box. Right now, it is way too dark.”

Oliver Schabenberger, COO & CTO at SAS

Responsibility throughout the whole analytics life cycle



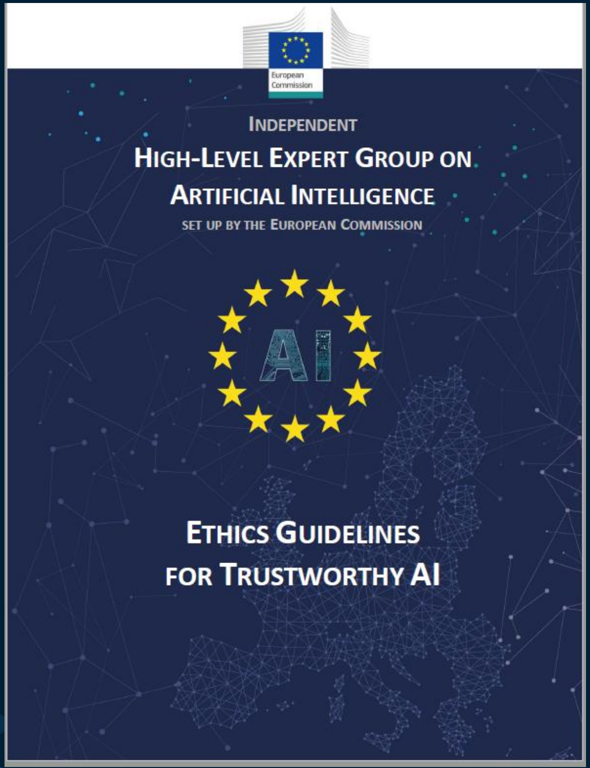

ocal Explanation
LIME
Feature
NINQ
DEROG
DELINQ
DEBTINC
CLAGE
0.00
0.05
0.10
Mean Weight
Kernel SHAP Values
Local Instance: 3000191
SHAP
Variable Name
lchehi_lab_STATE
ho_winsor_hhincome
lchehi_lab_City
ho_winsor_age
PROD_TYPE
cpy_int_med_imp_CREDIT_SCORE
cpy_int_med_imp_T_debt
lchehi_lab_currbal
ROC6
lchehi_lab_acctcd
-0.010
-0.005
0.000
0.005
Estimate
artial Dependence
PD
0.8
0.6
0.4
0.2
0.0
2.5
5.0
7.5
10.0
DELINQ
Individual Conditional Expectation
DELINQ
ICE
Predicted Value
1.0
0.8
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
DELINQ

Keep humans in the loop!


When we invest in AI, we also have a responsibility to ensure that it is ethical and sustainable

EU’s Ethics Guidelines for Trustworthy AI

Seven key requirements that AI systems should meet:



- Human agency & oversight
- Technical robustness & safety
- Privacy and Data governance
- **Transparency**
- **Diversity, non-discrimination and fairness**
- Societal and **environmental well-being**
- **Accountability**

Click here to see this page in other languages: Chinese German Japanese Korean Russian

Artificial intelligence has already provided beneficial tools that are used every day by people around the world. Its continued development, guided by the following principles, will offer amazing opportunities to help and empower people in the decades and centuries ahead.

Research Issues

1) **Research Goal:** The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.

2) **Research Funding:** Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:

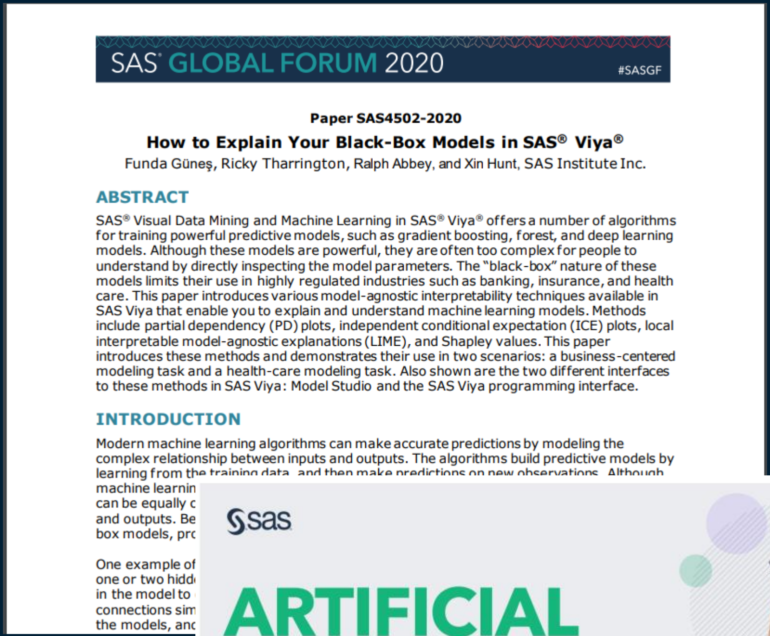
- How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?
- How can we grow our prosperity through automation while maintaining people's resources and purpose?
- How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI?
- What set of values should AI be aligned with, and what legal and ethical status should it have?

3) **Science-Policy Link:** There should be constructive and healthy exchange between AI researchers and policy-makers.

4) **Research Culture:** A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.

5) **Race Avoidance:** Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.

Learn more


SAS GLOBAL FORUM 2020 #SASGF

Paper SAS4502-2020

How to Explain Your Black-Box Models in SAS® Viya®

Funda Güneş, Ricky Tharrington, Ralph Abbey, and Xin Hunt, SAS Institute Inc.

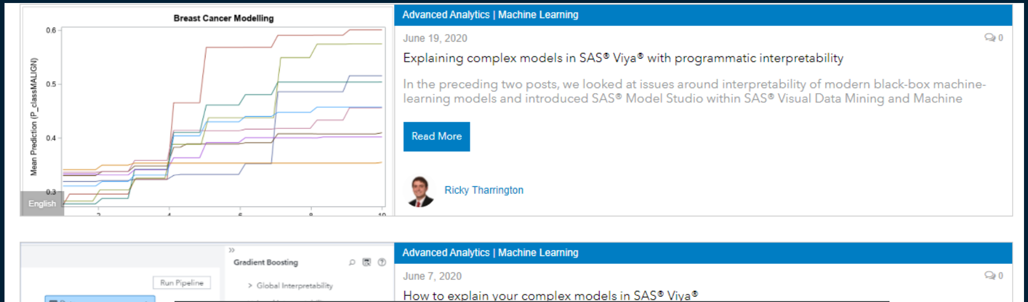
ABSTRACT

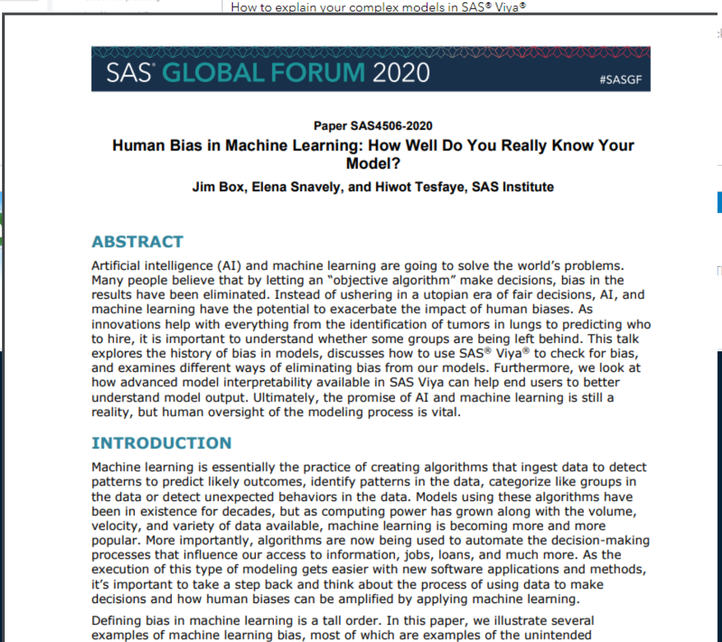
SAS® Visual Data Mining and Machine Learning in SAS® Viya® offers a number of algorithms for training powerful predictive models, such as gradient boosting, forest, and deep learning models. Although these models are powerful, they are often too complex for people to understand by directly inspecting the model parameters. The "black-box" nature of these models limits their use in highly regulated industries such as banking, insurance, and health care. This paper introduces various model-agnostic interpretability techniques available in SAS Viya that enable you to explain and understand machine learning models. Methods include partial dependency (PD) plots, independent conditional expectation (ICE) plots, local interpretable model-agnostic explanations (LIME), and Shapley values. This paper introduces these methods and demonstrates their use in two scenarios: a business-centered modeling task and a health-care modeling task. Also shown are the two different interfaces to these methods in SAS Viya: Model Studio and the SAS Viya programming interface.

INTRODUCTION

Modern machine learning algorithms can make accurate predictions by modeling the complex relationship between inputs and outputs. The algorithms build predictive models by learning from the training data, and then make predictions on new observations. Although machine learnin can be equally c and outputs. Be box models, pro

One example of one or two hidd in the model to connections sim the models, and




SAS GLOBAL FORUM 2020 #SASGF

Paper SAS4506-2020

Human Bias in Machine Learning: How Well Do You Really Know Your Model?

Jim Box, Elena Snavely, and Hiwot Tesfaye, SAS Institute

ABSTRACT

Artificial intelligence (AI) and machine learning are going to solve the world's problems. Many people believe that by letting an "objective algorithm" make decisions, bias in the results have been eliminated. Instead of ushering in a utopian era of fair decisions, AI, and machine learning have the potential to exacerbate the impact of human biases. As innovations help with everything from the identification of tumors in lungs to predicting who to hire, it is important to understand whether some groups are being left behind. This talk explores the history of bias in models, discusses how to use SAS® Viya® to check for bias, and examines different ways of eliminating bias from our models. Furthermore, we look at how advanced model interpretability available in SAS Viya can help end users to better understand model output. Ultimately, the promise of AI and machine learning is still a reality, but human oversight of the modeling process is vital.

INTRODUCTION

Machine learning is essentially the practice of creating algorithms that ingest data to detect patterns to predict likely outcomes, identify patterns in the data, categorize like groups in the data or detect unexpected behaviors in the data. Models using these algorithms have been in existence for decades, but as computing power has grown along with the volume, velocity, and variety of data available, machine learning is becoming more and more popular. More importantly, algorithms are now being used to automate the decision-making processes that influence our access to information, jobs, loans, and much more. As the execution of this type of modeling gets easier with new software applications and methods, it's important to take a step back and think about the process of using data to make decisions and how human biases can be amplified by applying machine learning.

Defining bias in machine learning is a tall order. In this paper, we illustrate several examples of machine learning bias, most of which are examples of the unintended

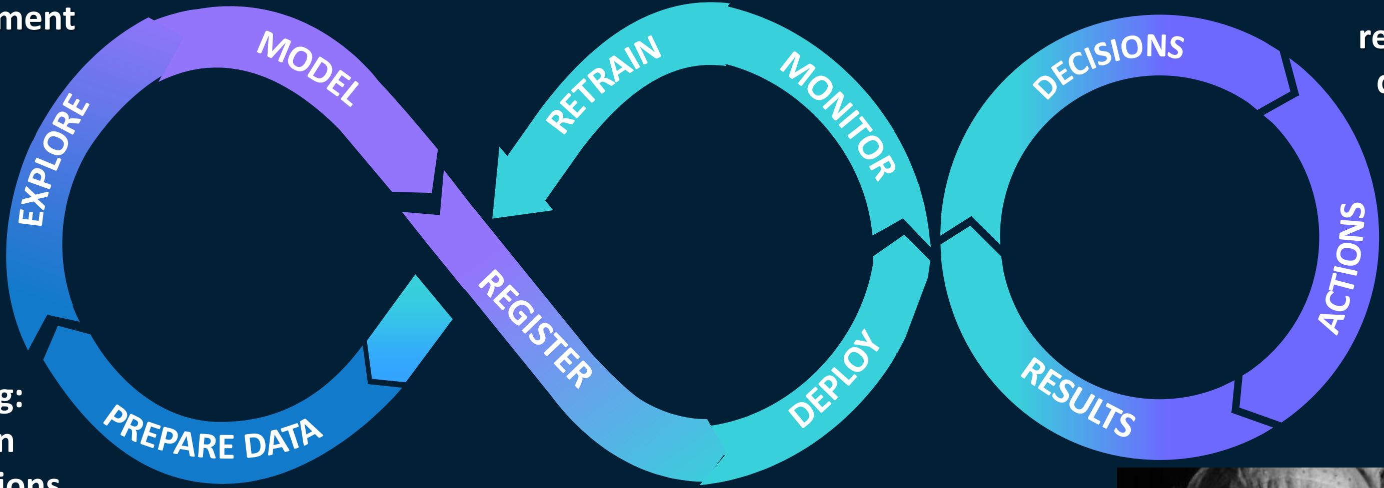
Responsibility throughout the whole analytics life cycle

In-processing:
Expose bias during model development

Post-processing:
Explain predictions
Expose bias in predictions

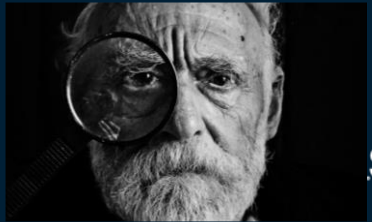
Post-processing:
Monitor and manage performance, track versions and data used

Post-processing:
Manage responsible decisions



Pre-processing:
Data selection
Explore distributions, representations, privacy issues, unfair variables etc.

Keep humans in the loop!



Thank you!

josefin.rosen@sas.com

www.linkedin.com/in/josefinrosen

@rosenjosefin