

Hvordan kan du oppdaga og reducere skjevheter i SAS Viya

FANS netteverksmøte Data Science/Analytics 2024-03-13

Mathias Lanner | SAS Institute



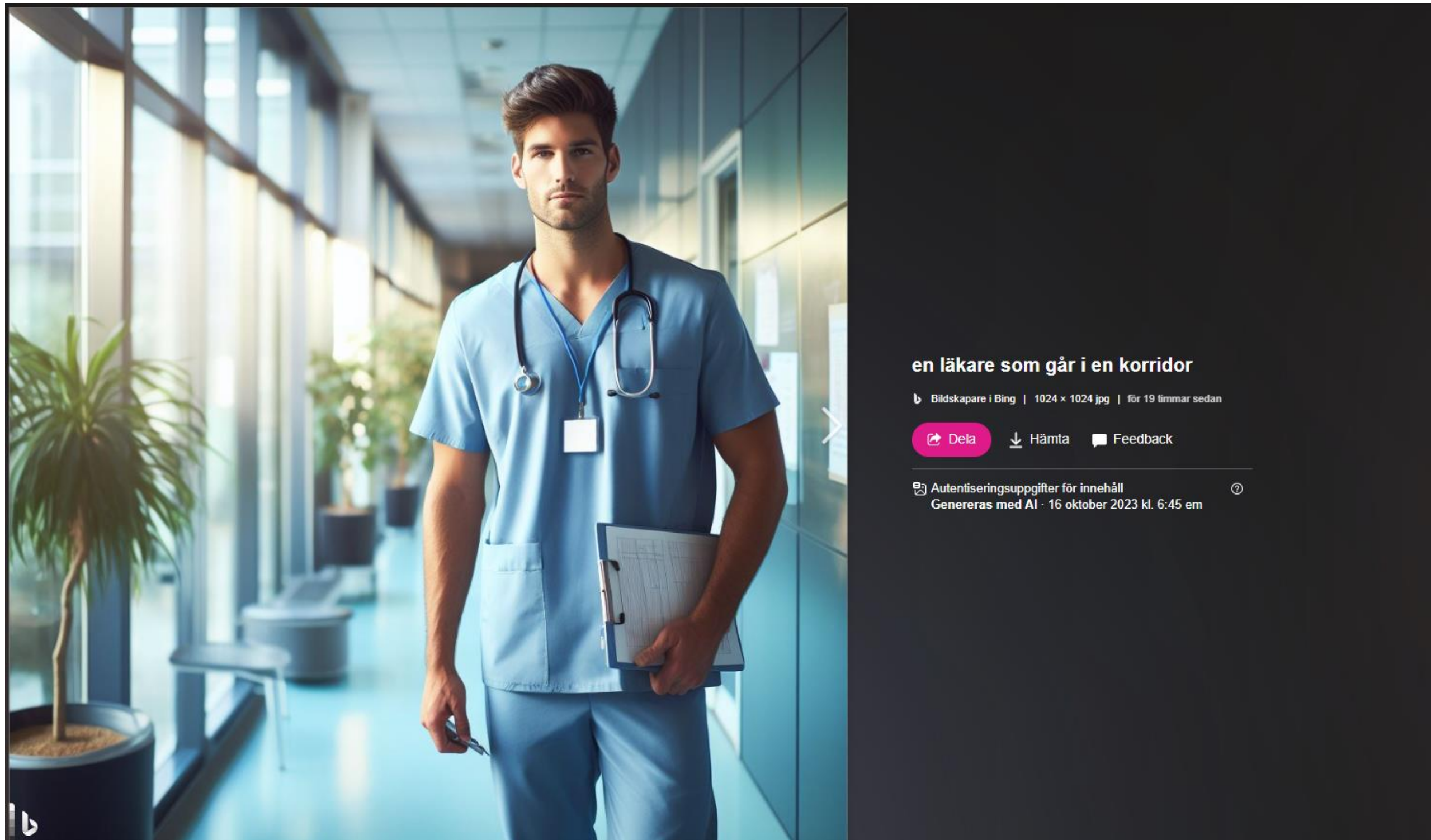
Agenda

- 1  BIAS Introduction
- 2  Fair Tool Action Set
- 3  DEMO
- 4  Bias Mitigation



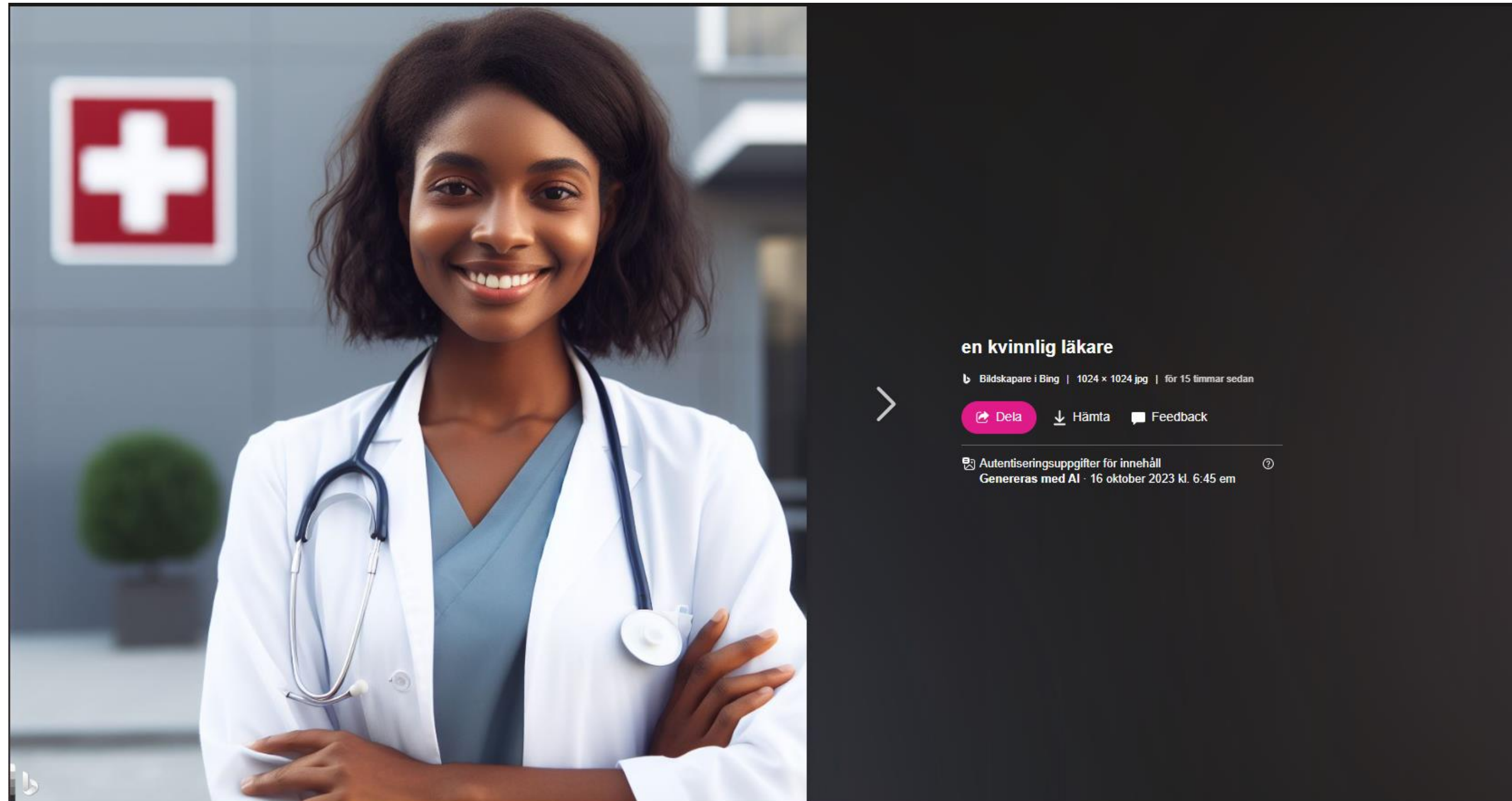
Ett litet experiment

Tänk på en läkare som går i en korridor.....



Ett till

Tänk på en kvinnlig läkare.....



Vet ni vad detta är?

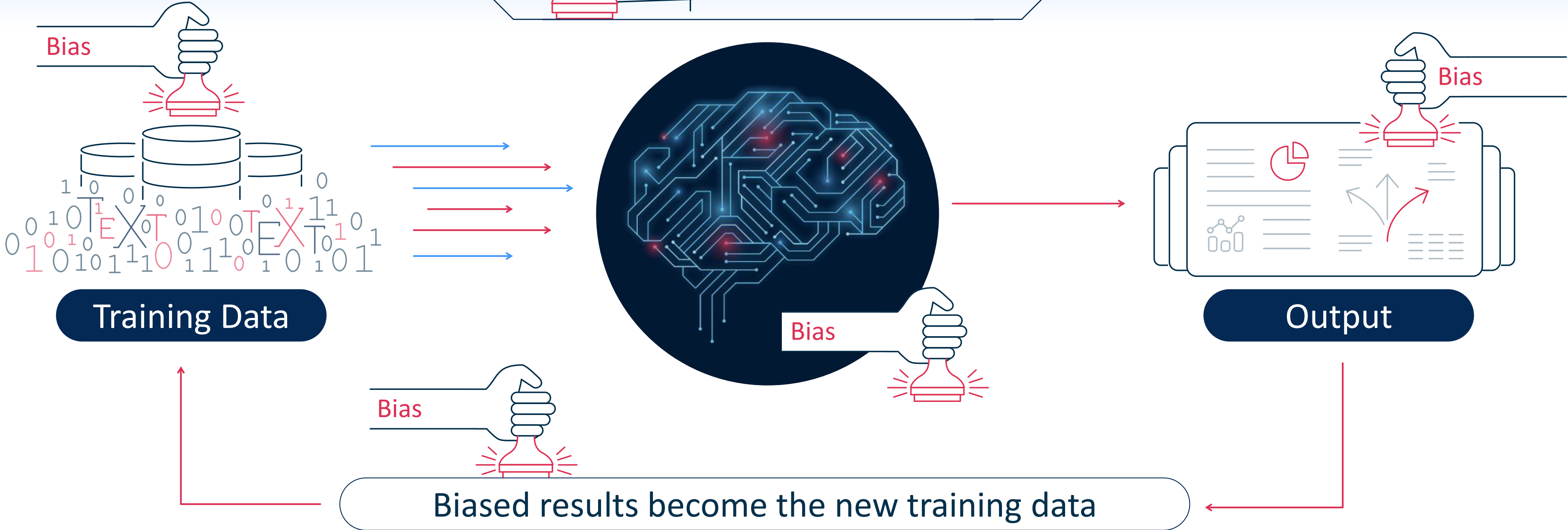
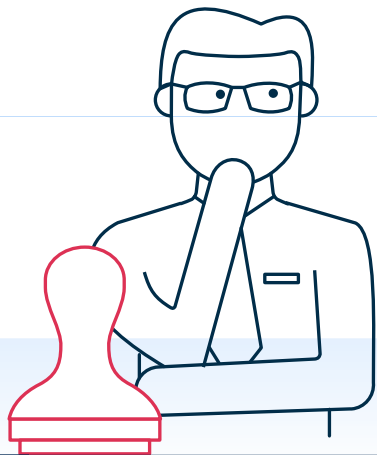


En 54-årig Advisor från ett It-bolag som presenterar BIAS för ett antal SAS-användare

Vad är detta exempel på?

BIAS

The ultimate source of bias





Fair Tool Action set

Fair AI Tools Action Set

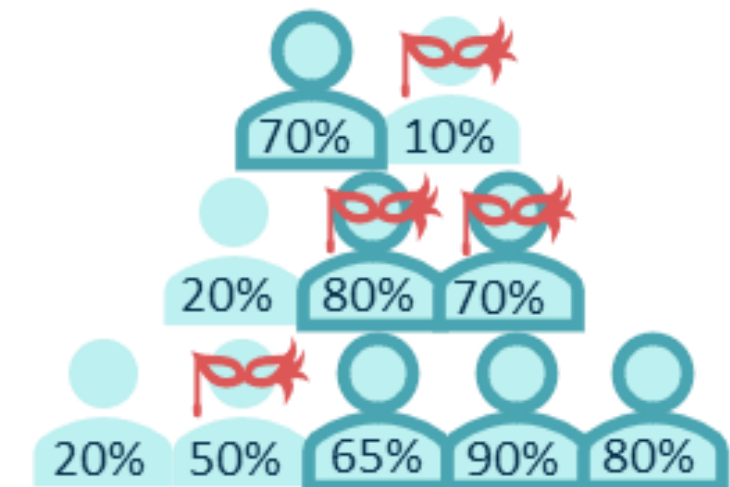
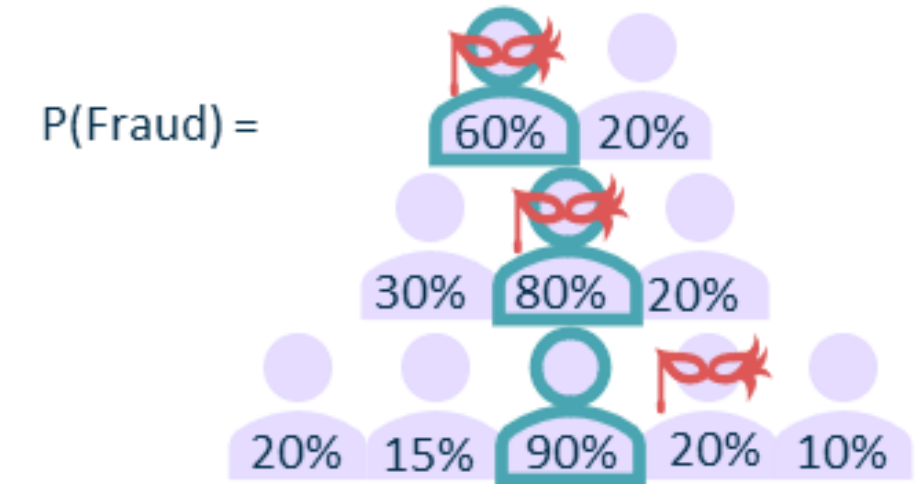
Bias Metrics



- Nominal Bias Statistics
 - DemographicParity – maximum pairwise difference in INTO_EVENT
 - PredictiveParity – maximum pairwise difference in the predicted variable corresponding to the event level
 - EqualAccuracy – maximum pairwise difference in ACC
 - EqualizedOdds – larger of the maximum pairwise differences in TPR and FPR
 - EqualOpportunity – maximum pairwise difference in TPR
- Interval Bias Statistics
 - PredictiveParity – maximum pairwise difference in the predicted variable



Fairness Statistics Explained

Fairness Statistic	Definition	When to use?	Example
Demographic parity (or Statistical parity)	Maximum measured difference in the selection rate of each category of a sensitive variable.	<ul style="list-style-type: none"> - Balance out historical biases that impact the data - Force to select more of minority groups 	<p>Selection Rate = 30%</p> <p>Selection Rate = 60%</p> <p>Demographic Parity = 30%</p>
Predictive parity (or Prediction bias parity)	Maximum measured difference in the average probability to receive a positive outcome between each category of a sensitive variable.	<ul style="list-style-type: none"> - Guarantee that each group within the population has on average the same chance to be selected. 	<p>Avg Probability = 36,5%</p> <p>Avg Probability = 55,5%</p> <p>Predictive Parity = 19%</p>
Equal Accuracy	Maximum measured difference in accuracy between each category of a sensitive variable.	<ul style="list-style-type: none"> - Assess whether the model equally works for each category of sensitive variable. 	<p>Accuracy = 80%</p> <p>Accuracy = 40%</p> <p>Equal Accuracy = 40%</p>
Equalized Opportunity	Maximum measured difference in true positive rate between each category of a sensitive variable.	<ul style="list-style-type: none"> - It is very important to find all the positive examples regardless of the category of sensitive variable (e.g. in fraud, health care...) 	<p>True Positive Rate = 66%</p> <p>True Positive Rate = 50%</p> <p>Equalized Opportunity = 16%</p>
Equalized Odds	Maximum measured difference in true positive rate or false positive rate between each category of a sensitive variable.	<ul style="list-style-type: none"> - It is very important to find all the positive examples regardless of the category of sensitive variable, and simultaneously false positives are extremely costly. 	<p>TPR = 66% // FPR = 14%</p> <p>TPR = 50% // FPR = 66%</p> <p>Equalized Odds = 52%</p>



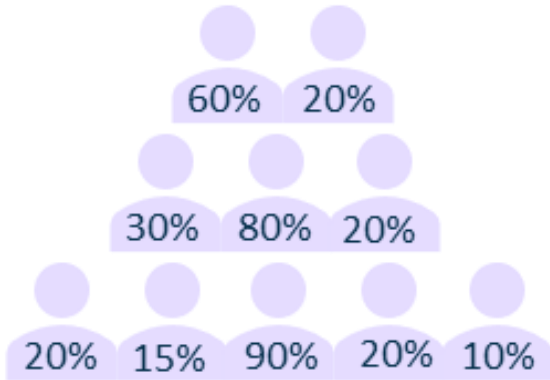
 predicted fraudster
 true fraudster

Fairness Statistics Explained

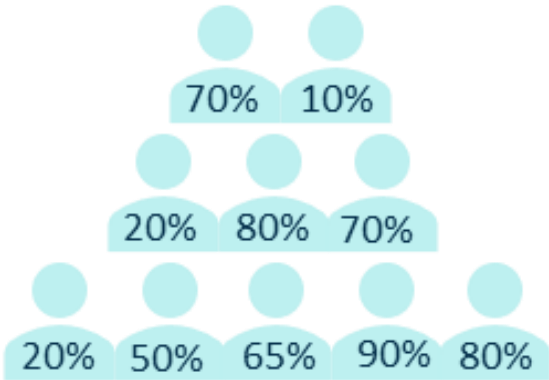
Predictive parity (or Prediction bias parity) – Maximum measured difference in the average probability to receive a positive outcome between each category of a sensitive variable.

Example: How much does the average probability to be granted a loan differ between people with origin “Norway” versus “Sweden” (versus other nationalities)?

When to use? Guarantee that each group within the population has on average the same chance to be selected.



Average Probability:
36,5%



Average Probability
55,5%

Predictive Parity = 19%

(|55,5%-36,5%|)

Fair AI Tools Action Set(Fairness)

Required Specifications

- An input table. Typically, this the data table that you use to test the model.
- A scoring function or model output variables
- Predicted target variable/variables (P_Bad1/P_Bad0)
- A sensitive variable to assess for bias. The model's performance and average predictions are calculated for each level of the sensitive variable and compared in order to produce bias metrics.
- A response variable to use for model assessment.
- A list of response levels. For models that predict a nominal response variable, you must specify the level of the response variable that corresponds to each predicted target variable.

How to Assess Model Fairness with SAS Studio?

fairAITools.assessBias Action

```

Code   Node   Notes
-----
1  ⊖  proc cas;
2      fairAITools.assessBias /
3          modelTableType = "NONE",
4          predictedVariables = {"P_high_low_flag0", "P_high_low_flag1"}
5          response = "high_low_flag",
6          responseLevels = {"0", "1"}
7          event = "1",
8          sensitiveVariable = "sex"
9          table = {name="adultScored", caslib="casuser"};
10     run;
11     quit;

```

		Bias Metrics		
Bias Statistic	Bias Statistic Label	Bias Statistic Value	Base Level	Compare Level
DemographicParity	Demographic Parity (Statistical Parity)	0.1783	Male	Female
PredictiveParity	Predictive Parity	0.1905	Male	Female
EqualAccuracy	Equal Accuracy	0.1007	Female	Male
EqualizedOdds	Equalized Odds	0.0786	Male	Female
EqualOpportunity	Equal Opportunity	0.0786	Male	Female



How to Assess Model Fairness with Model Studio?

SAS - Model Studio

Model Studio - Build Models

MitigateBias

Data Pipelines Pipeline Comparison Insights

Variable Name	Format	Role	Level	Label	Max
<input type="checkbox"/> _PartInd_		Partition	Binary	Partition Indicator	1.00
<input type="checkbox"/> age		Input	Interval		90.00
<input type="checkbox"/> capital_gain		Input	Interval		99,999
<input type="checkbox"/> capital_loss		Input	Interval		4,356.00
<input type="checkbox"/> education		Input	Nominal		
<input type="checkbox"/> education_num		Input	Nominal		16.00
<input type="checkbox"/> fnlwgt		Input	Interval		1,484.00
<input type="checkbox"/> high_low_flag		Target	Binary		1.00
<input type="checkbox"/> hours_per_week		Input	Interval		99.00
<input type="checkbox"/> marital_status		Input	Nominal		
<input type="checkbox"/> native_country		Rejected	Nominal		
<input type="checkbox"/> occupation		Input	Nominal		
<input type="checkbox"/> race		Input	Nominal		
<input type="checkbox"/> relationship		Input	Nominal		
<input checked="" type="checkbox"/> sex		Input	Binary		
<input type="checkbox"/> workclass		Input	Nominal		

sex

Role: Input

Level: Binary

Order: Default

Transform: Default

Impute: Default

Assess this variable for bias

SAS - Model Studio

Model Studio - Build Models

MitigateBias > "GB MARITAL STATUS WEIGHT" Results

Summary Output Data

Node Assessment Fairness and Bias

Performance Bias Parity

Prediction Bias Parity

Performance Bias

Prediction Bias

Demo – Assess Bias

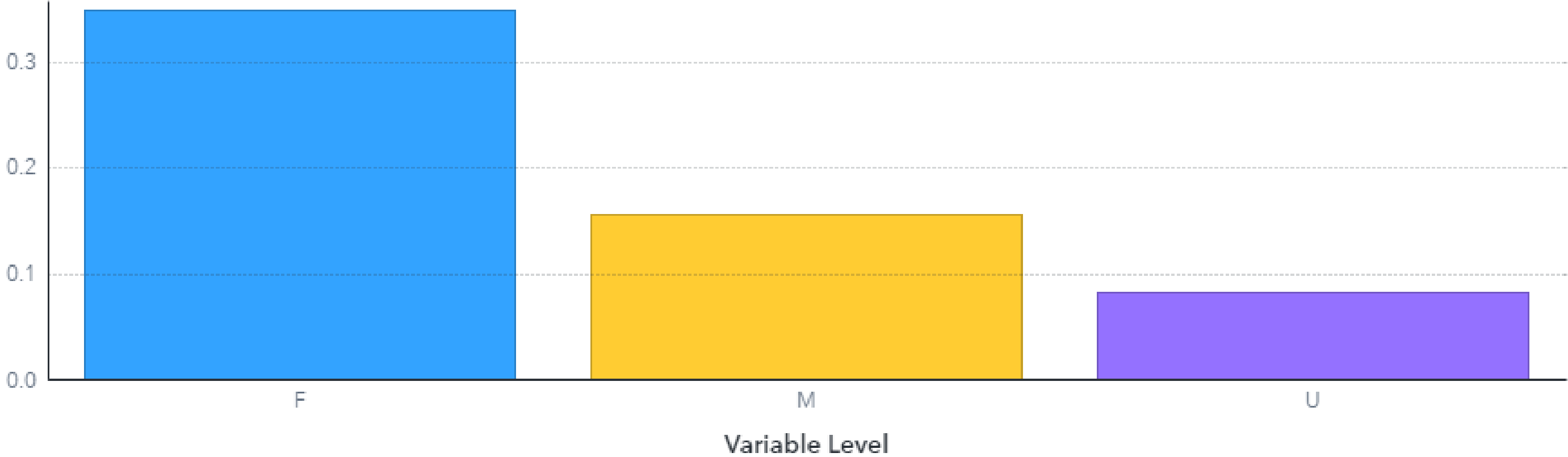
Bias Mitigation

How to Mitigate Bias in your Model?

Prediction Bias

DemGender ⌵ ⓘ ⌵ ↗

Average Prediction for Event



Fair AI Tools Action Set (Mitigate)

```
proc cas;
fairAITools.mitigateBias /
  biasMetric="EQUALOPPORTUNITY",
  event="Bought",
  learningRate="0.01",
  maxIters="15",
  predictedVariables={"P_TargetbuyBought", "P_TargetbuyNo"},
  response="Targetbuy",
  responseLevels={"Bought", "No"},
  sensitiveVariable="DemGender",
  table="ORGANICS_TRAIN",
  tolerance="0.005",
  trainProgram="
    decisionTree.gbtTreeTrain result=train_res /
      table=table,
      weight=weight,
      target=""TargetBuy"",
      inputs= {
        ""DemAffl"", ""DemAge"", ""DemGender"",
        ""DemHomeowner"", ""PromClass"", ""Promtime""},
      nominals={"TargetBuy", "DemGender", "PromClass"},
      nBins=50,
      quantileBin=True,
      maxLevel=5,
      maxBranch=2,
      leafSize=5,
      missing=""USEINSEARCH"",
      minUseInSearch=1,
      binOrder=True,
      varImp=True,
      mergeBin=True,
      encodeName=True,
      nTree=15,
      seed=12345,
      ridge=1,
      savestate={
        name=""org_gb_astore"",
        replace=True
      }
    ;
    astore.score result=score_res /
      table=table,
      casout=casout,
      copyVars=copyVars,
      rstore=""org_gb_astore""
  ;
  ;
  tuneBound="True";
run;
```

You need the following pieces:

- BiasMetrics
- Dataset
- ML-model must Support Weight parameter
- Scoring Function

