Analysis Report

Global dataset report

This report is the output of the Amazon SageMaker Clarify analysis. The report is split into following parts:

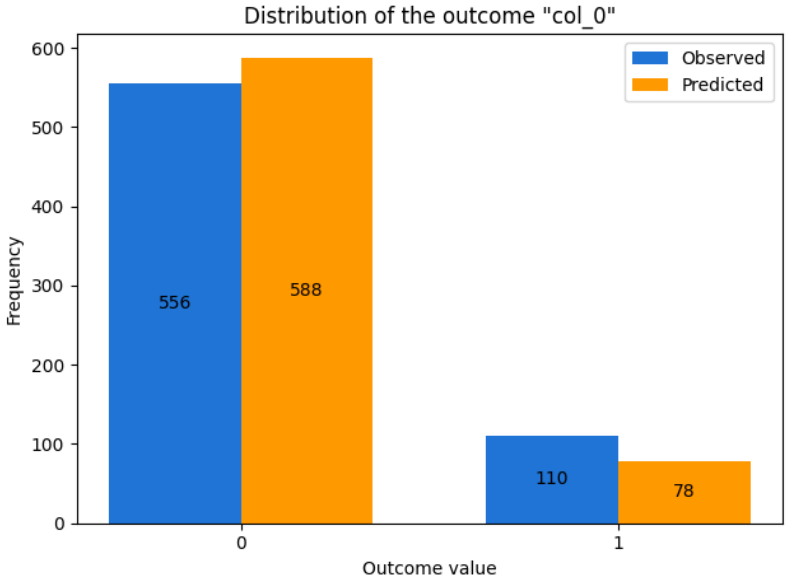
1. Analysis configuration
2. High level model performance
3. Pretraining bias metrics
4. Posttraining bias metrics

Analysis Configuration

Bias analysis requires you to configure the outcome label column, the facet and optionally a group variable. Generating explanations requires you to configure the outcome label. You configured the analysis with the following variables. The complete analysis configuration is appended at the end.

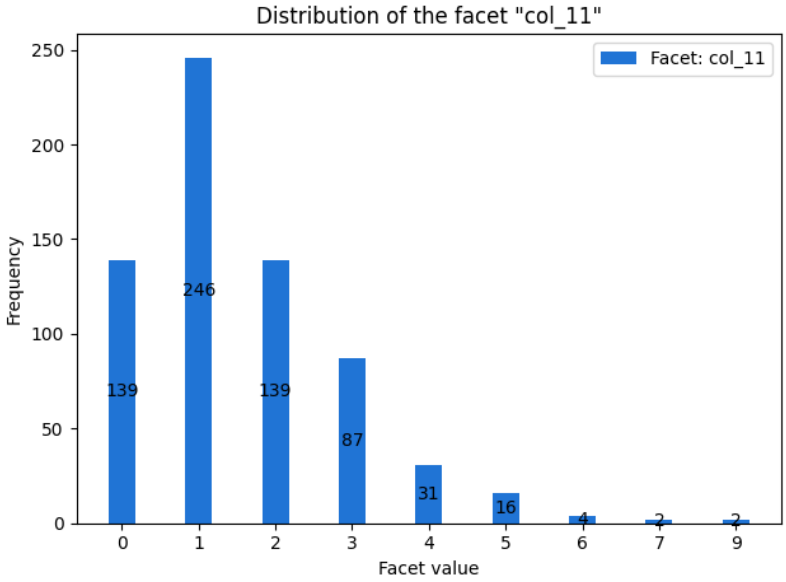
Outcome label: You chose the column `col_0` in the input data as the outcome label. Bias metric computation requires designating the positive outcome. You chose `col_0 = 1` as the positive outcome. `col_0` consisted of values `[0, 1]`.

The figure below shows the distribution of values of `col_0`.



Facet: You chose the column `col_11` in the input data as the facet. `col_11` consisted of values `[0, 1, 2, 3, 4, 5, 6, 7, 9]`. Bias metrics were computed by comparing the inputs `col_11 = 1` with all other inputs, then by comparing inputs `col_11 = 0` with all other inputs, then by comparing inputs `col_11 = 4` with all other inputs, then by comparing inputs `col_11 = 3` with all other inputs, then by comparing inputs `col_11 = 5` with all other inputs, then by comparing inputs `col_11 = 2` with all other inputs, then by comparing inputs `col_11 = 6` with all other inputs, then by comparing inputs `col_11 = 9` with all other inputs, then by comparing inputs `col_11 = 7` with all other inputs.

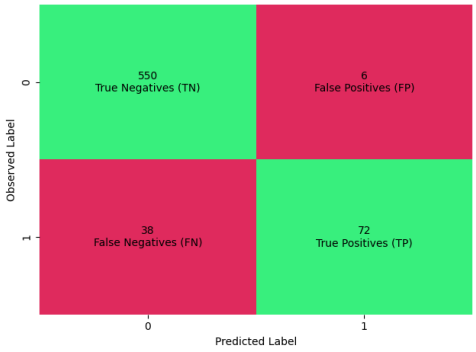
The figure below shows the distribution of values of `col_11`.



High level model performance

Input data points can be divided into different categories based on their observed and predicted label. For instance, a False Negative (FN) is an input with a positive observed label (`col_0 = 1`) but negative predicted label (`col_0 != 1`). A True Negative (TN) is an input whose observed and predicted labels are both negative.True Positives (TP) and False Positives (FP) are defined similarly.

Based on the model predictions, the inputs can be divided into different categories as:

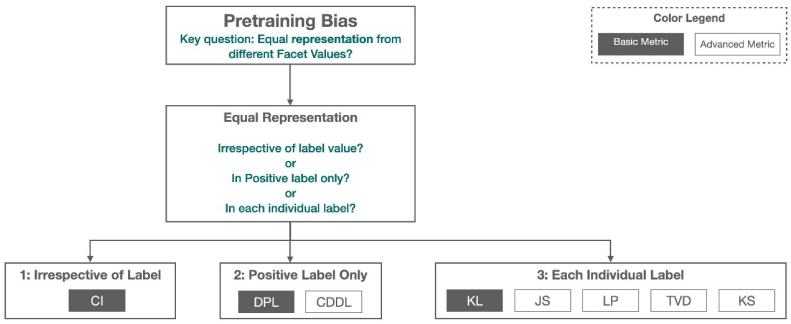


Here are metrics showing the model performance.

Metric	Description	Value
Accuracy	Proportion of inputs assigned the correct predicted label by the model.	0.934
Proportion of Positive Predictions in Labels	Proportion of input assigned in positive predicted label.	0.117
Proportion of Negative Predictions in Labels	Proportion of input assigned the negative predicted label.	0.883
True Positive Rate / Recall	Proportion of inputs with positive observed label correctly assigned the positive predicted label.	0.655
True Negative Rate / Specificity	Proportion of inputs with negative observed label correctly assigned the negative predicted label.	0.989
Acceptance Rate / Precision	Proportion of inputs with positive predicted label that actually have a positive observed label.	0.923
Rejection Rate	Proportion of inputs with negative predicted label that actually have a negative observed label.	0.935
Conditional Acceptance	Ratio between the positive observed labels and positive predicted labels.	1.410
Conditional Rejection	Ratio between the negative observed labels and negative predicted labels.	0.946
F1 Score	Harmonic mean of precision and recall.	0.766

Pre-training Bias Metrics

Pretraining bias metrics measure imbalances in facet value representation in the training data. Imbalances can be measured across different dimensions. For instance, you could focus imbalances within the inputs with positive observed label only. The figure below shows how different pretraining bias metrics focus on different dimensions. For a detailed description of these dimensions, see Learn How Amazon SageMaker Clarify Helps Detect Bias.

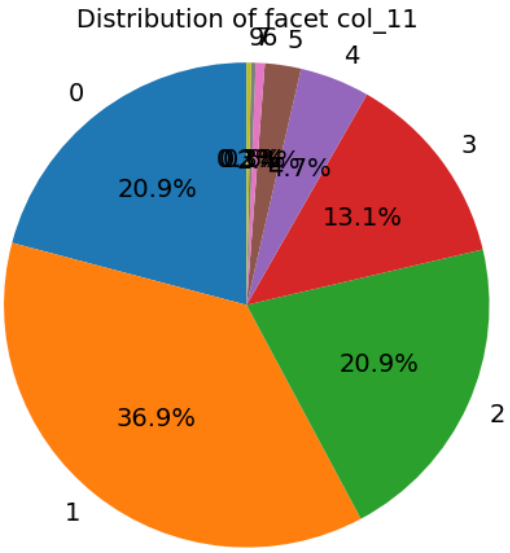


The metric values along with an informal description of what they mean are shown below. For mathematical formulas and examples, see the [Measure Pretraining Bias](https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-measure-data-bias.html) section of the AWS documentation.

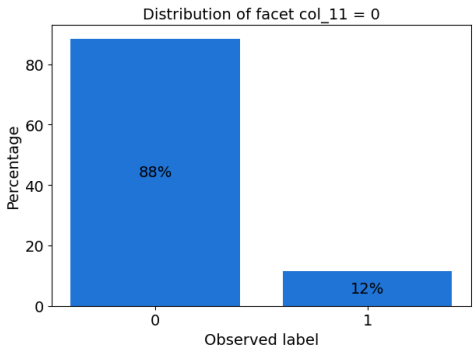
We computed the bias metrics for the label `col_0` using label value(s)/threshold `col_0 = 1` for the following facets:

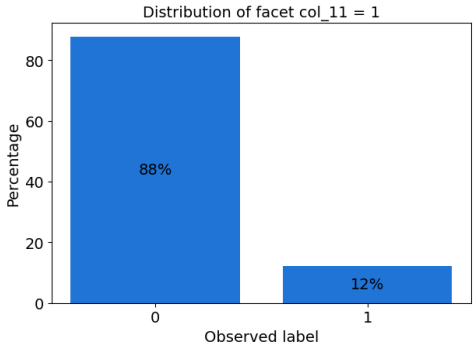
- Facet column: **col_11**

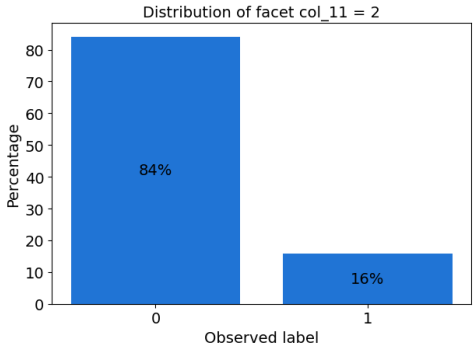
The pie chart shows the distribution of facet column `col_11` in your data.

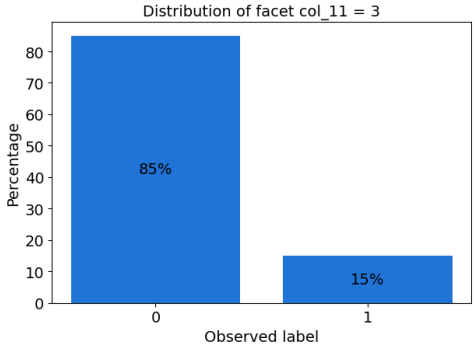


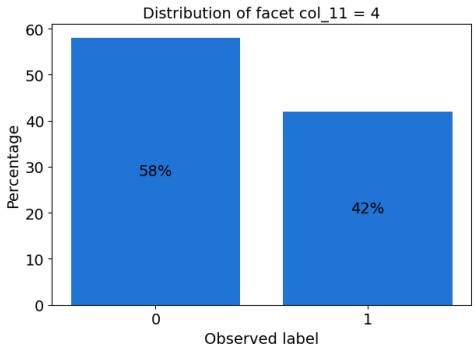
The bar plot(s) below show the distribution of facet column `col_11` in your data.

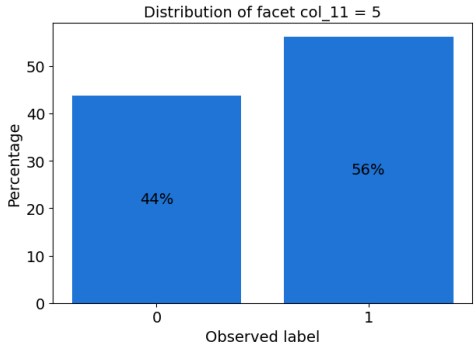


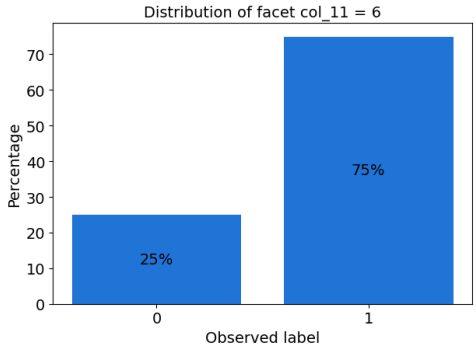

Distribution of facet col_11 = 1
Percentage
Observed label
88%
12%
0
1
0
20
40
60
80

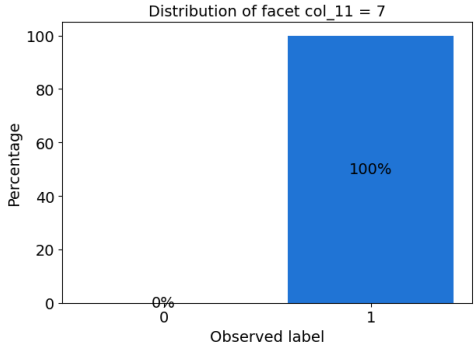

Distribution of facet col_11 = 2
Percentage
Observed label
84%
16%
0
1
0
10
20
30
40
50
60
70
80

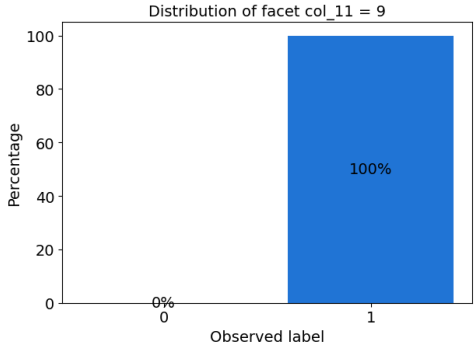

Distribution of facet col_11 = 3
Percentage
Observed label
85%
15%
0
1
0
10
20
30
40
50
60
70
80


Distribution of facet col_11 = 4
Percentage
Observed label
58%
42%
0
1
0
10
20
30
40
50
60









Facet Value(s)/Threshold: `col_11 = 1`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 1` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 1` and rest of the inputs.	0.261	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 1` and rest of the inputs.	0.069	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 1` and rest of the inputs diverge from each other entropically.	0.004	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 1` and rest of the inputs diverge from each other entropically.	0.019	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 1` and rest of the inputs in the dataset.	0.069	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 1` rest of the inputs in the dataset.	0.097	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 1` and rest of the inputs in the dataset.	0.069	None

Facet Value(s)/Threshold: `col_11 = 0`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 0` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 0` and rest of the inputs.	0.583	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 0` and rest of the inputs.	0.063	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 0` and rest of the inputs diverge from each other entropically.	0.004	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 0` and rest of the inputs diverge from each other entropically.	0.017	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 0` and rest of the inputs in the dataset.	0.063	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 0` rest of the inputs in the dataset.	0.089	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 0` and rest of the inputs in the dataset.	0.063	None

Facet Value(s)/Threshold: `col_11 = 4`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 4` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 4` and rest of the inputs.	0.907	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 4` and rest of the inputs.	-0.267	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 4` and rest of the inputs diverge from each other entropically.	0.045	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 4` and rest of the inputs diverge from each other entropically.	0.166	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 4` and rest of the inputs in the dataset.	0.267	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 4` rest of the inputs in the dataset.	0.377	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 4` and rest of the inputs in the dataset.	0.267	None

Facet Value(s)/Threshold: `col_11 = 3`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 3` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 3` and rest of the inputs.	0.739	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 3` and rest of the inputs.	0.018	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 3` and rest of the inputs diverge from each other entropically.	0.000	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 3` and rest of the inputs diverge from each other entropically.	0.001	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 3` and rest of the inputs in the dataset.	0.018	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 3` rest of the inputs in the dataset.	0.026	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 3` and rest of the inputs in the dataset.	0.018	None

Facet Value(s)/Threshold: `col_11 = 5`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 5` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 5` and rest of the inputs.	0.952	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 5` and rest of the inputs.	-0.407	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 5` and rest of the inputs diverge from each other entropically.	0.094	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 5` and rest of the inputs diverge from each other entropically.	0.356	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 5` and rest of the inputs in the dataset.	0.407	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 5` rest of the inputs in the dataset.	0.576	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 5` and rest of the inputs in the dataset.	0.407	None

Facet Value(s)/Threshold: `col_11 = 2`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 2` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 2` and rest of the inputs.	0.583	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 2` and rest of the inputs.	0.009	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 2` and rest of the inputs diverge from each other entropically.	0.000	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 2` and rest of the inputs diverge from each other entropically.	0.000	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 2` and rest of the inputs in the dataset.	0.009	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 2` rest of the inputs in the dataset.	0.012	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 2` and rest of the inputs in the dataset.	0.009	None

Facet Value(s)/Threshold: `col_11 = 6`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 6` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 6` and rest of the inputs.	0.988	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 6` and rest of the inputs.	-0.588	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 6` and rest of the inputs diverge from each other entropically.	0.187	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 6` and rest of the inputs diverge from each other entropically.	0.766	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 6` and rest of the inputs in the dataset.	0.588	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 6` rest of the inputs in the dataset.	0.832	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 6` and rest of the inputs in the dataset.	0.588	None

Facet Value(s)/Threshold: `col_11 = 9`

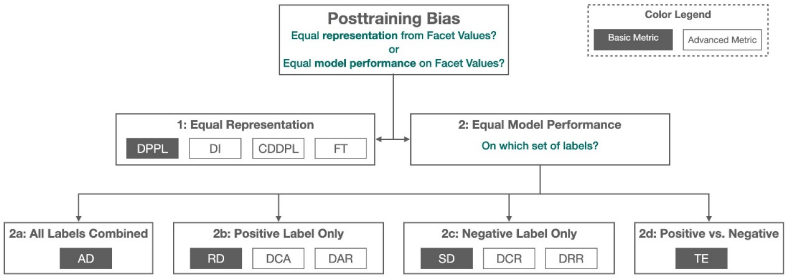
Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 9` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 9` and rest of the inputs.	0.994	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 9` and rest of the inputs.	-0.837	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 9` and rest of the inputs diverge from each other entropically.	0.168	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 9` and rest of the inputs diverge from each other entropically.	-0.295	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 9` and rest of the inputs in the dataset.	0.837	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 9` rest of the inputs in the dataset.	0.837	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 9` and rest of the inputs in the dataset.	0.419	None

Facet Value(s)/Threshold: `col_11 = 7`

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 7` and rest of the inputs in the dataset.	None	Error: see Clarify job output
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `col_11 = 7` and rest of the inputs.	0.994	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `col_11 = 7` and rest of the inputs.	-0.837	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `col_11 = 7` and rest of the inputs diverge from each other entropically.	0.168	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `col_11 = 7` and rest of the inputs diverge from each other entropically.	-0.295	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `col_11 = 7` and rest of the inputs in the dataset.	0.837	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `col_11 = 7` rest of the inputs in the dataset.	0.837	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `col_11 = 7` and rest of the inputs in the dataset.	0.419	None

Post-training Bias Metrics

Posttraining bias metrics measure imbalances in model predictions across different inputs. The figure below shows how different posttraining metrics target different types of imbalances over inputs. For a detailed description of these types, see Learn How Amazon SageMaker Clarify Helps Detect Bias.

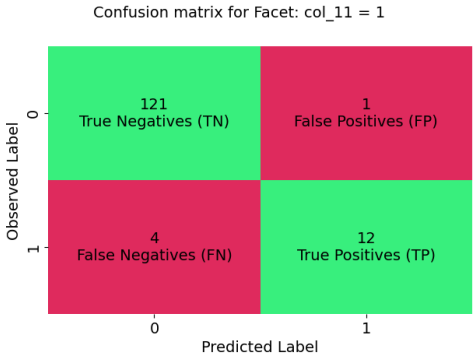


Bias can also result form imbalances in the model outcomes even when the facet value is not considered. The metric computing these imbalances is GE. The metric values along with an informal description of what they mean are shown below. For mathematical formulas and examples, see the [Measure Posttraining Data and Model Bias] (https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-measure-post-training-bias.html) section of the AWS documentation.

We computed the bias metrics for the label `col_0` using label value(s)/threshold `col_0 = 1` for the following facets:

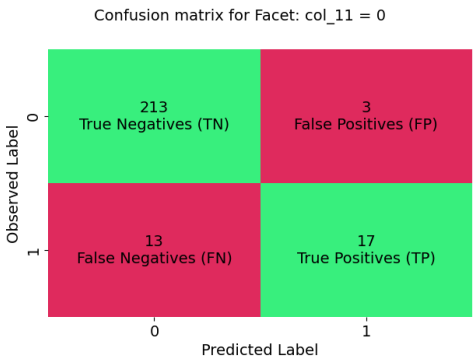
- Facet column: **col_11**

Facet Value(s)/Threshold: `col_11 = 1`



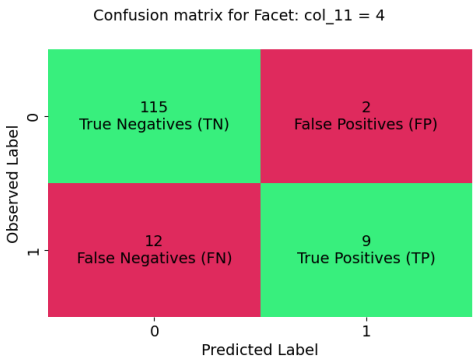
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 1` and rest of the inputs.	-0.002	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 1` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 1` and rest of the inputs.	0.098	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 1` and rest of the inputs for predicted positive outcomes (acceptances).	-0.121	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 1` and rest of the inputs for negative outcomes (rejections).	0.017	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 1` and rest of the inputs.	0.589	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 1` and rest of the inputs.	0.057	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 1` and rest of the inputs.	0.012	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 1` and assesses whether similar members from rest of the inputs have different model predictions.	0.033	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 1` and rest of the inputs.	0.121	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 1` and rest of the inputs.	-0.005	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 1` and rest of the inputs.	-4.000	None

Facet Value(s)/Threshold: `col_11 = 0`



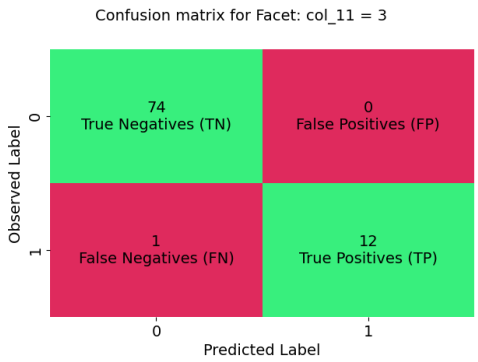
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 0` and rest of the inputs.	-0.038	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 0` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 0` and rest of the inputs.	0.000	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 0` and rest of the inputs for predicted positive outcomes (acceptances).	0.215	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 0` and rest of the inputs for negative outcomes (rejections).	0.039	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 0` and rest of the inputs.	0.758	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 0` and rest of the inputs.	0.030	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 0` and rest of the inputs.	0.042	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 0` and assesses whether similar members from rest of the inputs have different model predictions.	0.043	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 0` and rest of the inputs.	-0.112	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 0` and rest of the inputs.	0.003	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 0` and rest of the inputs.	-2.800	None

Facet Value(s)/Threshold: `col_11 = 4`



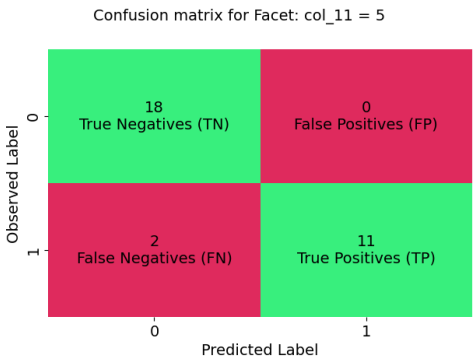
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 4` and rest of the inputs.	-0.002	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 4` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 4` and rest of the inputs.	-0.090	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 4` and rest of the inputs for predicted positive outcomes (acceptances).	0.266	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 4` and rest of the inputs for negative outcomes (rejections).	-0.047	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 4` and rest of the inputs.	3.363	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 4` and rest of the inputs.	-0.249	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 4` and rest of the inputs.	-0.037	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 4` and assesses whether similar members from rest of the inputs have different model predictions.	0.258	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 4` and rest of the inputs.	-0.217	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 4` and rest of the inputs.	0.011	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 4` and rest of the inputs.	Infinity	None

Facet Value(s)/Threshold: `col_11 = 3`



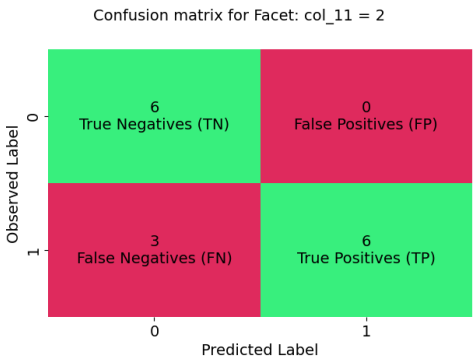
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 3` and rest of the inputs.	-0.063	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 3` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 3` and rest of the inputs.	-0.091	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 3` and rest of the inputs for predicted positive outcomes (acceptances).	0.386	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 3` and rest of the inputs for negative outcomes (rejections).	0.047	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 3` and rest of the inputs.	1.210	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 3` and rest of the inputs.	-0.024	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 3` and rest of the inputs.	0.059	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 3` and assesses whether similar members from rest of the inputs have different model predictions.	0.126	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 3` and rest of the inputs.	-0.305	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 3` and rest of the inputs.	0.012	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 3` and rest of the inputs.	Infinity	None

Facet Value(s)/Threshold: `col_11 = 5`



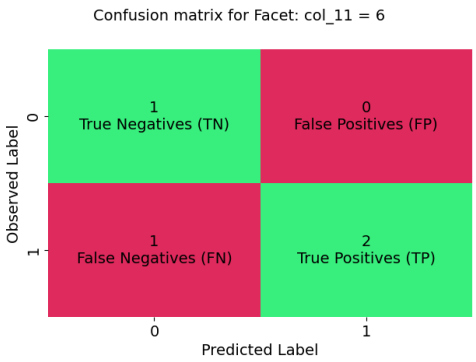
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 5` and rest of the inputs.	0.124	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 5` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 5` and rest of the inputs.	-0.083	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 5` and rest of the inputs for predicted positive outcomes (acceptances).	-0.097	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 5` and rest of the inputs for negative outcomes (rejections).	-0.250	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 5` and rest of the inputs.	3.385	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 5` and rest of the inputs.	-0.264	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 5` and rest of the inputs.	-0.239	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 5` and assesses whether similar members from rest of the inputs have different model predictions.	0.312	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 5` and rest of the inputs.	-0.013	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 5` and rest of the inputs.	0.011	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 5` and rest of the inputs.	Infinity	None

Facet Value(s)/Threshold: `col_11 = 2`



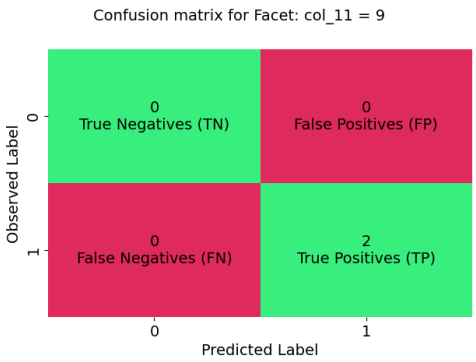
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 2` and rest of the inputs.	0.044	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 2` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 2` and rest of the inputs.	0.106	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 2` and rest of the inputs for predicted positive outcomes (acceptances).	-0.500	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 2` and rest of the inputs for negative outcomes (rejections).	-0.031	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 2` and rest of the inputs.	0.689	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 2` and rest of the inputs.	0.039	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 2` and rest of the inputs.	-0.038	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 2` and assesses whether similar members from rest of the inputs have different model predictions.	0.065	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 2` and rest of the inputs.	0.250	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 2` and rest of the inputs.	-0.008	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 2` and rest of the inputs.	-0.500	None

Facet Value(s)/Threshold: `col_11 = 6`



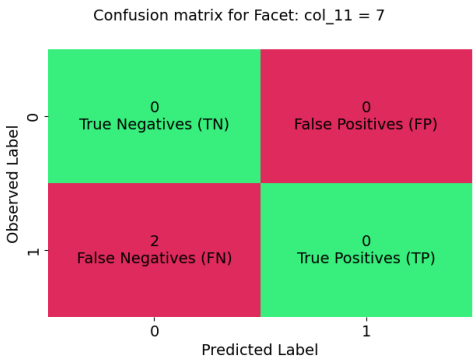
Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 6` and rest of the inputs.	0.185	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 6` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 6` and rest of the inputs.	-0.079	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 6` and rest of the inputs for predicted positive outcomes (acceptances).	-0.092	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 6` and rest of the inputs for negative outcomes (rejections).	-0.447	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 6` and rest of the inputs.	4.355	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 6` and rest of the inputs.	-0.385	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 6` and rest of the inputs.	-0.437	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 6` and assesses whether similar members from rest of the inputs have different model predictions.	0.500	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 6` and rest of the inputs.	-0.012	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 6` and rest of the inputs.	0.011	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 6` and rest of the inputs.	Infinity	None

Facet Value(s)/Threshold: `col_11 = 9`



Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 9` and rest of the inputs.	0.937	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 9` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 9` and rest of the inputs.	0.923	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 9` and rest of the inputs for predicted positive outcomes (acceptances).	-Infinity	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 9` and rest of the inputs for negative outcomes (rejections).	-0.949	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 9` and rest of the inputs.	0.000	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 9` and rest of the inputs.	0.117	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 9` and rest of the inputs.	-0.939	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 9` and assesses whether similar members from rest of the inputs have different model predictions.	0.000	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 9` and rest of the inputs.	0.667	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 9` and rest of the inputs.	-0.989	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 9` and rest of the inputs.	Infinity	None

Facet Value(s)/Threshold: `col_11 = 7`



Metric	Description	Value	Error
Accuracy Difference (AD)	Measures the difference between the prediction accuracy for facet values `col_11 = 7` and rest of the inputs.	-0.066	None
Conditional Demographic Disparity in Predicted Labels (CDDPL)	Measures the disparity of predicted labels between facet values `col_11 = 7` and rest of the inputs as a whole, but also by subgroups dictated by Age.	None	Error: see Clarify job output
Difference in Acceptance Rates (DAR)	Measures the difference in the ratios of the observed positive outcomes (TP) to the predicted positives (TP + FP) between facet values `col_11 = 7` and rest of the inputs.	-0.079	None
Difference in Conditional Acceptance (DCAcc)	Compares the observed labels to the labels predicted by the model. Assesses whether this is the same across facet values `col_11 = 7` and rest of the inputs for predicted positive outcomes (acceptances).	0.421	None
Difference in Conditional Rejection (DCR)	Compares the observed labels to the labels predicted by the model and assesses whether this is the same across facet values `col_11 = 7` and rest of the inputs for negative outcomes (rejections).	-0.946	None
Disparate Impact (DI)	Measures the ratio of proportions of the predicted labels for facet values `col_11 = 7` and rest of the inputs.	8.737	None
Difference in Positive Proportions in Predicted Labels (DPPL)	Measures the difference in the proportion of positive predictions between facet values `col_11 = 7` and rest of the inputs.	-0.886	None
Difference in Rejection Rates (DRR)	Measures the difference in the ratios of the observed negative outcomes (TN) to the predicted negatives (TN + FN) between facet values `col_11 = 7` and rest of the inputs.	-0.935	None
Counterfactual Fliptest (FT)	Examines each input with facet value `col_11 = 7` and assesses whether similar members from rest of the inputs have different model predictions.	1.000	None
Generalized entropy (GE)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.035	None
Recall Difference (RD)	Measures the difference between the recall, aka true positive rate, of the model for facet values `col_11 = 7` and rest of the inputs.	-0.352	None
Specificity difference (SD)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `col_11 = 7` and rest of the inputs.	-0.989	None
Treatment Equality (TE)	Measures the difference in the ratio of false positives to false negatives between facet values `col_11 = 7` and rest of the inputs.	-6.333	None

Appendix: Analysis Configuration Parameters


```
{
  "dataset_type": "text/csv",
  "headers": [
    "col_0",
    "col_1",
    "col_2",
    "col_3",
    "col_4",
    "col_5",
    "col_6",
    "col_7",
    "col_8",
    "col_9",
    "col_10",
    "col_11",
    "col_12",
    "col_13",
    "col_14",
    "col_15",
    "col_16",
    "col_17",
    "col_18",
    "col_19",
    "col_20",
    "col_21",
    "col_22",
    "col_23",
    "col_24"
```

```
    "col_24",
    "col_25",
    "col_26",
    "col_27",
    "col_28",
    "col_29",
    "col_30",
    "col_31",
    "col_32",
    "col_33",
    "col_34",
    "col_35",
    "col_36",
    "col_37",
    "col_38",
    "col_39",
    "col_40",
    "col_41",
    "col_42",
    "col_43",
    "col_44",
    "col_45",
    "col_46",
    "col_47",
    "col_48",
    "col_49",
    "col_50",
    "col_51",
    "col_52",
    "col_53",
    "col_54",
    "col_55",
    "col_56",
    "col_57",
    "col_58",
    "col_59",
    "col_60",
    "col_61",
    "col_62",
    "col_63",
    "col_64",
    "col_65",
    "col_66",
    "col_67",
    "col_68",
    "col_69"
],
"label": "col_0",
"label_values_or_threshold": [
    1
],
"facet": [
    {
        "name_or_index": "col_11"
    }
],
"methods": {
    "report": {
        "name": "report",
        "title": "Analysis Report"
    },
    "pre_training_bias": {
        "methods": "all"
    },
    "post_training_bias": {
        "methods": "all"
```

```
        }
      },
      "predictor": {
        "model_name": "churn-bias-model-2026-02-06-214155",
        "instance_type": "ml.m5.xlarge",
        "initial_instance_count": 1,
        "accept_type": "text/csv",
        "content_type": "text/csv",
        "content_template": "$features"
      }
    }
```