Analysis Report

Global dataset report

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

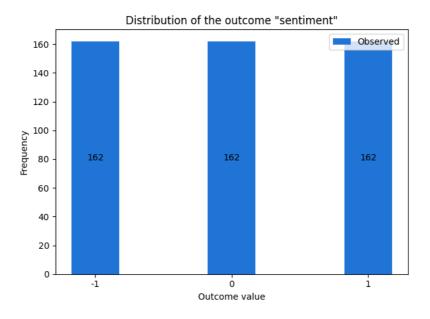
- 1. Analysis configuration
- 2. Pretraining bias metrics

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.

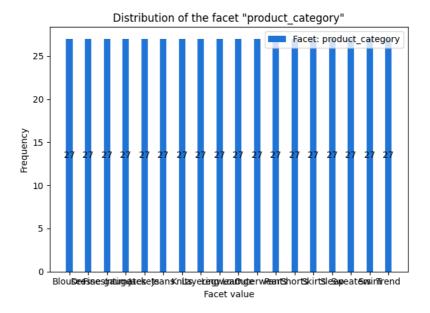
Outcome label: You chose the column sentiment in the input data as the outcome label. Bias metric computation requires designating the positive outcome. You chose sentiment = 1 as the positive outcome. sentiment consisted of values [-1, 0, 1].

The figure below shows the distribution of values of sentiment .



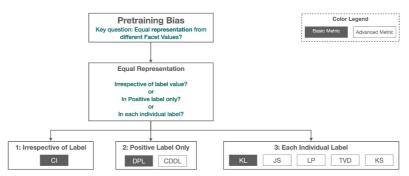
Facet: You chose the column product_category in the input data as the facet. product_category consisted of values ['Blouses', 'Dresses', 'Fine gauge', 'Intimates', 'Jackets', 'Jeans', 'Knits', 'Layering', 'Legwear', 'Lounge', 'Outerwear', 'Pants', 'Shorts', 'Skirts', 'Sleep', 'Sweaters', 'Swim', 'Trend'] . Bias metrics were computed by comparing the inputs product_category = Blouses with all other inputs, then by comparing inputs product_category = Dresses with all other inputs, then by comparing inputs product_category = Fine gauge with all other inputs, then by comparing inputs product category = Intimates with all other inputs, then by comparing inputs product_category = Jackets with all other inputs, then by comparing inputs product_category = Jeans with all other inputs, then by comparing inputs product category = Knits with all other inputs, then by comparing inputs product_category = Layering with all other inputs, then by comparing inputs product_category = Legwear with all other inputs, then by comparing inputs product category = Lounge with all other inputs, then by comparing inputs product_category = Outerwear with all other inputs, then by comparing inputs product_category = Pants with all other inputs, then by comparing inputs product_category = Shorts with all other inputs, then by comparing inputs product_category = Skirts with all other inputs, then by comparing inputs product_category = Sleep with all other inputs, then by comparing inputs product_category = Sweaters with all other inputs, then by comparing inputs product_category = Swim with all other inputs, then by comparing inputs product_category = Trend with all other inputs.

The figure below shows the distribution of values of product_category .



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 <u>Learn How Amazon SageMaker Clarify Helps Detect Bias</u>.

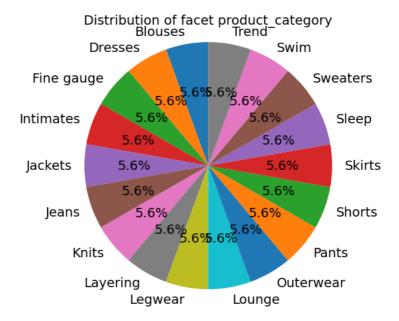


The metric values along with an informal description of what they mean are shown below. For mathematical formulas and examples, see the [Measure Pretraining Bias](https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-measure-data-bias.html) section of the AWS documentation.

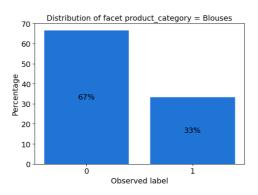
We computed the bias metrics for the label sentiment using label value(s)/threshold sentiment = 1 for the following facets:

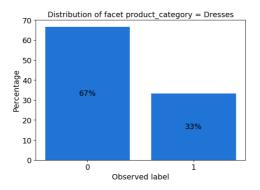
• Facet column: product_category

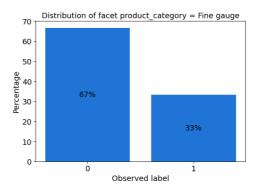
The pie chart shows the distribution of facet column product_category in your data.

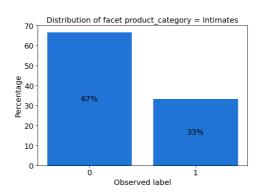


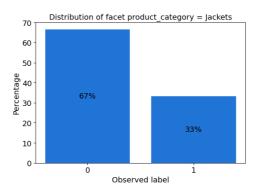
The bar plot(s) below show the distribution of facet column product_category in your data.

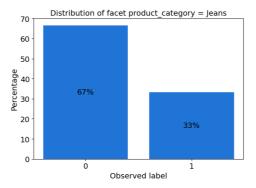


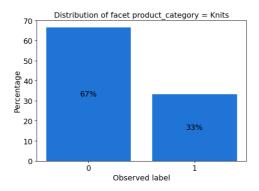


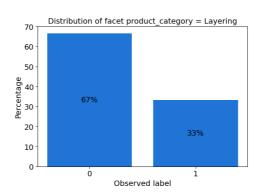


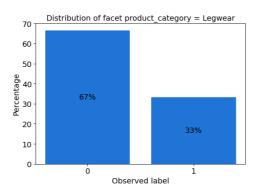


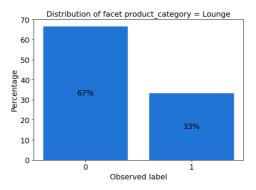


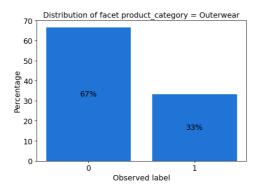


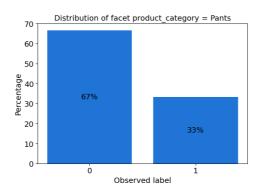


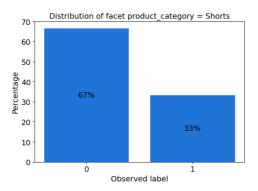


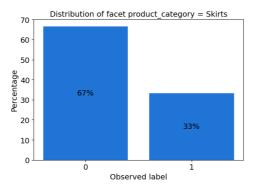


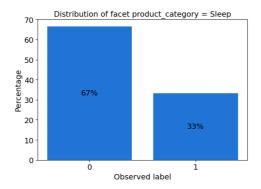


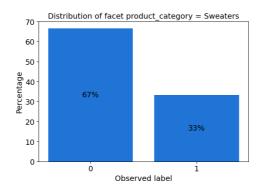


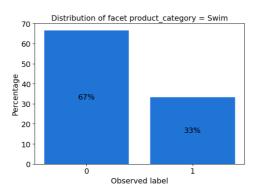


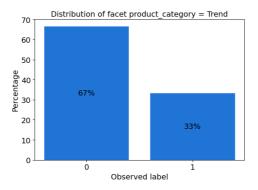












Facet Value(s)/Threshold: product_category = Blouses

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Blouses and rest of the inputs.	0.889
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Blouses and rest of the inputs.	0.000
<u>Jensen-Shannon</u> <u>Divergence (JS)</u>	Measures how much the observed label distributions of facet values <code>product_category = Blouses</code> and rest of the inputs diverge from each other entropically.	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values <code>product_category = Blouses</code> and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values $product_category = Blouses \ and rest of the inputs in the dataset.$	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $product_category = Blouses \ rest of the inputs in the dataset.$	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Blouses and rest of the inputs in the dataset.	0.000

Facet Value(s)/Threshold: product_category = Dresses

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Dresses and rest of the inputs.	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (<u>DPL</u>)	Measures the imbalance of positive observed labels between facet values product_category = Dresses and rest of the inputs.	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values <code>product_category = Dresses</code> and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values <code>product_category = Dresses and rest of the inputs diverge from each other entropically.</code>	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values $product_category = Dresses \ and rest of the inputs in the dataset.$	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $ product_category = Dresses \ rest of the inputs in the dataset. $	0.000
<u>Total Variation</u> <u>Distance (TVD)</u>	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Dresses and rest of the inputs in the dataset.	0.000
Facet Value(s)/Thres	hold: product_category = Fine gauge	
Metric	Description	Value
Metric		value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values <code>product_category = Fine gauge and rest of the inputs.</code>	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Fine gauge and rest of the inputs.	0.000
<u>Jensen-Shannon</u> Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Fine gauge and rest of the inputs diverge from each other entropically.	0.00

Measures how much the observed label distributions of facet values product_category = Fine gauge

Measures a p-norm difference between the observed label distributions associated with facet values

 $\label{thm:local_model} \mbox{Measures half of the L1-norm difference between the observed label distributions associated with}$

Measures maximum divergence between the observed label distributions for facet values

facet values product_category = Fine gauge and rest of the inputs in the dataset.

and rest of the inputs diverge from each other entropically.

product_category = Fine gauge and rest of the inputs in the dataset.

product_category = Fine gauge rest of the inputs in the dataset.

0.000

0.000

0.000

0.000

Facet Value(s)/Threshold: product_category = Intimates

Facet Value(s)/Threshold: product_category = Jackets

Kullback-Leibler

Divergence (KL)

Lp-norm (LP)

Total Variation

Distance (TVD)

(KS)

Kolmogorov-Smirnov

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Intimates and rest of the inputs.	0.889
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Intimates and rest of the inputs.	0.000
<u>Jensen-Shannon</u> <u>Divergence (JS)</u>	Measures how much the observed label distributions of facet values <code>product_category = Intimates</code> and rest of the inputs diverge from each other entropically.	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values product_category = Intimates and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Intimates and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $ product_category = Intimates $	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Intimates and rest of the inputs in the dataset.	0.000

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Jackets and rest of the inputs.	0.889
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values <code>product_category = Jackets and rest of the inputs.</code>	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values <code>product_category = Jackets and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values product_category = Jackets and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Jackets and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Jackets rest of the inputs in the dataset.	0.000
<u>Total Variation</u> <u>Distance (TVD)</u>	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Jackets and rest of the inputs in the dataset.	0.000
acet Value(s)/Thresl	nold: product_category = Jeans	
Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Jeans and rest of the inputs.	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Jeans and rest of the inputs.	0.000
<u>Jensen-Shannon</u> <u>Divergence (JS)</u>	Measures how much the observed label distributions of facet values <code>product_category = Jeans and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values <code>product_category = Jeans</code> and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Jeans and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Jeans rest of the inputs in the dataset.	0.000
<u>Total Variation</u> <u>Distance (TVD)</u>	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Jeans and rest of the inputs in the dataset.	0.000
acet Value(s)/Thres	nold: product_category = Knits	
Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values <code>product_category = Knits and rest of the inputs.</code>	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Knits and rest of the inputs.	0.000

value	Description	Metric
0.889	Measures the imbalance in the number of inputs with facet values <code>product_category = Knits and rest of the inputs.</code>	Class Imbalance (CI)
0.000	Measures the imbalance of positive observed labels between facet values product_category = Knits and rest of the inputs.	<u>Difference in</u> <u>Proportions of Labels</u> (DPL)
0.000	Measures how much the observed label distributions of facet values <code>product_category = Knits and rest of the inputs diverge from each other entropically.</code>	<u>Jensen-Shannon</u> <u>Divergence (JS)</u>
0.000	Measures how much the observed label distributions of facet values <code>product_category = Knits and rest of the inputs diverge from each other entropically.</code>	<u>Kullback-Leibler</u> <u>Divergence (KL)</u>
0.000	Measures maximum divergence between the observed label distributions for facet values product_category = Knits and rest of the inputs in the dataset.	Kolmogorov-Smirnov (KS)
0.000	Measures a p-norm difference between the observed label distributions associated with facet values $product_category = Knits rest \ of \ the \ inputs \ in \ the \ dataset.$	<u>Lp-norm (LP)</u>
0.000	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Knits and rest of the inputs in the dataset.	Total Variation Distance (TVD)

Facet Value(s)/Threshold: product_category = Layering

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Layering and rest of the inputs.	0.889
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values <code>product_category = Layering</code> and rest of the inputs.	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Layering and rest of the inputs diverge from each other entropically.	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values product_category = Layering and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Layering and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $product_category = Layering rest of the inputs in the dataset.$	0.000
<u>Total Variation</u> <u>Distance (TVD)</u>	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Layering and rest of the inputs in the dataset.	0.000
acet Value(s)/Thresh	nold: product_category = Legwear	
Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values $ \begin{array}{c} \text{product_category} = \text{Legwear} \text{and} \\ \text{rest of the inputs}. \end{array} $	0.889
Difference in		
Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values <code>product_category = Legwear</code> and rest of the inputs.	0.000
Proportions of Labels		0.000
Proportions of Labels (DPL) Jensen-Shannon	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear	
Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Legwear	0.000
Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler Divergence (KL) Kolmogorov-Smirnov	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures maximum divergence between the observed label distributions for facet values	0.000
Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler Divergence (KL) Kolmogorov-Smirnov (KS)	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures maximum divergence between the observed label distributions for facet values product_category = Legwear and rest of the inputs in the dataset. Measures a p-norm difference between the observed label distributions associated with facet values	0.000 0.000 0.000
Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler Divergence (KL) Kolmogorov-Smirnov (KS) Lp-norm (LP) Total Variation Distance (TVD)	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures maximum divergence between the observed label distributions for facet values product_category = Legwear and rest of the inputs in the dataset. Measures a p-norm difference between the observed label distributions associated with facet values product_category = Legwear rest of the inputs in the dataset. Measures half of the L1-norm difference between the observed label distributions associated with	0.000 0.000 0.000 0.000
Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler Divergence (KL) Kolmogorov-Smirnov (KS) Lp-norm (LP) Total Variation Distance (TVD)	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures maximum divergence between the observed label distributions for facet values product_category = Legwear and rest of the inputs in the dataset. Measures a p-norm difference between the observed label distributions associated with facet values product_category = Legwear rest of the inputs in the dataset. Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Legwear and rest of the inputs in the dataset.	0.000 0.000 0.000 0.000
Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler Divergence (KL) Kolmogorov-Smirnov (KS) Lp-norm (LP) Total Variation Distance (TVD) acet Value(s)/Thresh	Legwear and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically. Measures maximum divergence between the observed label distributions for facet values product_category = Legwear and rest of the inputs in the dataset. Measures a p-norm difference between the observed label distributions associated with facet values product_category = Legwear rest of the inputs in the dataset. Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Legwear and rest of the inputs in the dataset. mold: product_category = Lounge	0.000 0.000 0.000 0.000

Lounge and rest of the inputs. (DPL) Measures how much the observed label distributions of facet values product_category = Lounge and Jensen-Shannon 0.000 Divergence (JS) rest of the inputs diverge from each other entropically. $\label{thm:measures} \mbox{Measures how much the observed label distributions of facet values} \mbox{ product_category = Lounge } \mbox{ and } \mbox{ and } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ and } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ and } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \mbox{ product_category = Lounge } \mbox{ of facet values } \mbox{ product_category = Lounge } \m$ <u>Kullback-Leibler</u> 0.000 Divergence (KL) rest of the inputs diverge from each other entropically. Measures maximum divergence between the observed label distributions for facet values Kolmogorov-Smirnov 0.000 product_category = Lounge and rest of the inputs in the dataset. Measures a p-norm difference between the observed label distributions associated with facet values Lp-norm (LP) 0.000 product_category = Lounge rest of the inputs in the dataset. $\label{thm:local_model} \mbox{Measures half of the L1-norm difference between the observed label distributions associated with}$ **Total Variation** 0.000 Distance (TVD) facet values product_category = Lounge and rest of the inputs in the dataset. Facet Value(s)/Threshold: product_category = Outerwear

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values <code>product_category = Outerwear</code> and rest of the inputs.	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Outerwear and rest of the inputs.	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values <code>product_category = Outerwear</code> and rest of the inputs diverge from each other entropically.	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values product_category = Outerwear and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Outerwear and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $ product_category = Outerwear \ rest of the inputs in the dataset. $	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values <code>product_category = Outerwear</code> and rest of the inputs in the dataset.	0.000
Facet Value(s)/Thres	hold: product_category = Pants	
Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Pants and rest of the inputs.	0.889
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Pants and rest of the inputs.	0.000
<u>Jensen-Shannon</u> <u>Divergence (JS)</u>	Measures how much the observed label distributions of facet values <code>product_category = Pants and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values product_category = Pants and rest of the inputs diverge from each other entropically.	0.000

Measures maximum divergence between the observed label distributions for facet values

Measures a p-norm difference between the observed label distributions associated with facet values

 $\label{thm:local_model} \mbox{Measures half of the L1-norm difference between the observed label distributions associated with}$

product_category = Pants and rest of the inputs in the dataset.

facet values <code>product_category = Pants and rest of the inputs in the dataset.</code>

product_category = Pants rest of the inputs in the dataset.

0.000

0.000

0.000

Facet Value(s)/Threshold: product_category = Shorts

Kolmogorov-Smirnov

Lp-norm (LP)

Total Variation

Distance (TVD)

(KS)

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Shorts and rest of the inputs.	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values <code>product_category = Shorts and rest of the inputs.</code>	0.000
<u>Jensen-Shannon</u> <u>Divergence (JS)</u>	Measures how much the observed label distributions of facet values <code>product_category = Shorts and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values <code>product_category = Shorts</code> and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values $product_category = Shorts \ \ and rest of the inputs in the dataset.$	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $ product_category = Shorts $	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Shorts and rest of the inputs in the dataset.	0.000
Facet Value(s)/Thres	hold: product_category = Skirts	

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values <code>product_category = Skirts and rest of the inputs.</code>	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Skirts and rest of the inputs.	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values <code>product_category = Skirts and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values product_category = Skirts and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Skirts and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values product category = Skirts rest of the inputs in the dataset.	0.000
	F	
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset.	0.000
Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with	0.000
Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset.	
<u>Distance (TVD)</u> acet Value(s)/Thresl	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset. nold: product_category = Sleep	
Distance (TVD) acet Value(s)/Thresl	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset. nold: product_category = Sleep Description Measures the imbalance in the number of inputs with facet values product_category = Sleep and	Value
Distance (TVD) acet Value(s)/Thresl Metric Class Imbalance (CI) Difference in Proportions of Labels	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset. nold: product_category = Sleep Description Measures the imbalance in the number of inputs with facet values product_category = Sleep and rest of the inputs. Measures the imbalance of positive observed labels between facet values product_category = Sleep	Value 0.889
Distance (TVD) acet Value(s)/Thresl Metric Class Imbalance (CI) Difference in Proportions of Labels (DPL) Jensen-Shannon	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset. nold: product_category = Sleep Description Measures the imbalance in the number of inputs with facet values product_category = Sleep and rest of the inputs. Measures the imbalance of positive observed labels between facet values product_category = Sleep and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Sleep and	Value 0.889 0.000
Distance (TVD) acet Value(s)/Thresl Metric Class Imbalance (CI) Difference in Proportions of Labels (DPL) Jensen-Shannon Divergence (JS) Kullback-Leibler	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset. nold: product_category = Sleep Description Measures the imbalance in the number of inputs with facet values product_category = Sleep and rest of the inputs. Measures the imbalance of positive observed labels between facet values product_category = Sleep and rest of the inputs. Measures how much the observed label distributions of facet values product_category = Sleep and rest of the inputs diverge from each other entropically. Measures how much the observed label distributions of facet values product_category = Sleep and rest of the inputs diverge from each other entropically.	Value 0.889 0.000 0.000

Facet Value(s)/Threshold: product_category = Sweaters

Facet Value(s)/Threshold: product_category = Swim

Total Variation Distance (TVD)

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Sweaters and rest of the inputs.	0.889
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values <code>product_category = Sweaters</code> and rest of the inputs.	0.000
<u>Jensen-Shannon</u> <u>Divergence (JS)</u>	Measures how much the observed label distributions of facet values <code>product_category = Sweaters</code> and rest of the inputs diverge from each other entropically.	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values <code>product_category = Sweaters</code> and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Sweaters and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Sweaters rest of the inputs in the dataset.	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Sweaters and rest of the inputs in the dataset.	0.000

Measures half of the L1-norm difference between the observed label distributions associated with facet values $product_category = Sleep$ and rest of the inputs in the dataset.

0.000

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Swim and rest of the inputs.	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Swim and rest of the inputs.	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values <code>product_category = Swim and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values <code>product_category = Swim and rest of the inputs diverge from each other entropically.</code>	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Swim and rest of the inputs in the dataset.	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $ product_category = Swim \\ rest of the inputs in the dataset. $	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Swim and rest of the inputs in the dataset.	0.000
Facet Value(s)/Thresh	nold: product_category = Trend	

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values $ \begin{array}{ccc} \text{product_category} = \text{Trend} & \text{and} \\ & \text{rest of the inputs.} \end{array} $	0.889
<u>Difference in</u> <u>Proportions of Labels</u> (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Trend and rest of the inputs.	0.000
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values <code>product_category = Trend and rest of the inputs diverge from each other entropically.</code>	0.000
<u>Kullback-Leibler</u> <u>Divergence (KL)</u>	Measures how much the observed label distributions of facet values <code>product_category = Trend</code> and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values $product_category = Trend \ and rest of the inputs in the dataset.$	0.000
<u>Lp-norm (LP)</u>	Measures a p-norm difference between the observed label distributions associated with facet values $ product_category = Trend rest of the inputs in the dataset. $	0.000
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Trend and rest of the inputs in the dataset.	0.000

Appendix: Analysis Configuration Parameters

```
"dataset_type": "text/csv",
"headers": [
  "sentiment",
  "review_body",
  "product_category"
"label": "sentiment",
"label_values_or_threshold": [
  1
],
"facet": [
     "name_or_index": "product_category"
],
"methods": {
  "pre_training_bias": {
    "methods": [
       "CI",
       "DPL",
       "KL",
       "JS",
       "LP",
```

```
"TVD",

"KS"

]

},

"report": {

    "name": "report",

    "title": "Analysis Report"

}

}
```