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
5. Model explanations

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

The figure below shows the distribution of values of `target`.
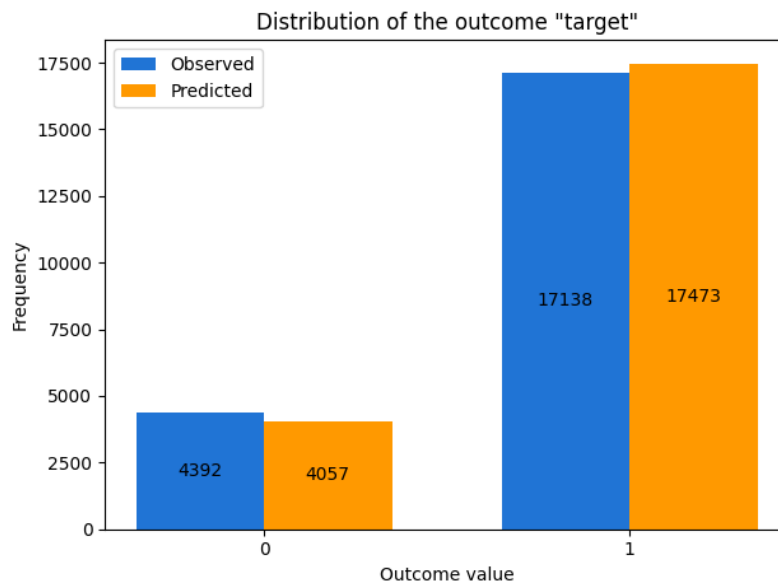
Facet: You chose the column `doorstep` in the input data as the facet. `doorstep` consisted of values `[0.0, 1.0]`. Bias metrics were computed by comparing the inputs `doorstep = (0.0, 1.0]` with all other inputs.

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

Distribution of the facet "doorstep"
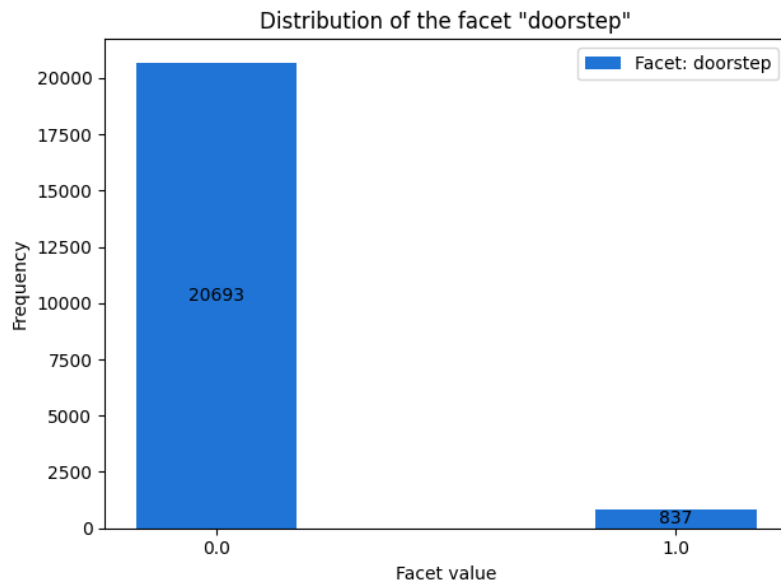
High level model performance

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 (target = 1) but negative predicted label (target != 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:
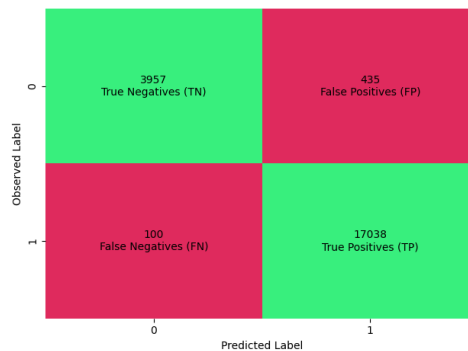
Here are metrics showing the model performance.

Metric	Description	Value
Accuracy	Proportion of inputs assigned the correct predicted label by the model.	0.975
Proportion of Positive Predictions in Labels	Proportion of input assigned in positive predicted label.	0.812
Proportion of Negative Predictions in Labels	Proportion of input assigned the negative predicted label.	0.188
True Positive Rate / Recall	Proportion of inputs with positive observed label correctly assigned the positive predicted label.	0.994
True Negative Rate / Specificity	Proportion of inputs with negative observed label correctly assigned the negative predicted label.	0.901
Acceptance Rate / Precision	Proportion of inputs with positive predicted label that actually have a positive observed label.	0.975
Rejection Rate	Proportion of inputs with negative predicted label that actually have a negative observed label.	0.975
Conditional Acceptance	Ratio between the positive observed labels and positive predicted labels.	0.981
Conditional Rejection	Ratio between the negative observed labels and negative predicted labels.	1.083
F1 Score	Harmonic mean of precision and recall.	0.985

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.
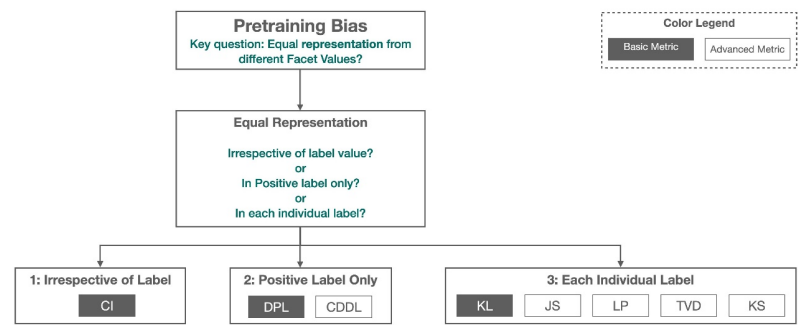

The metric values along with an informal description of what they mean are shown below. For mathematical formulas and examples, see the [Measure Pretraining Bias](https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-measure-data-bias.html) section of the AWS documentation.
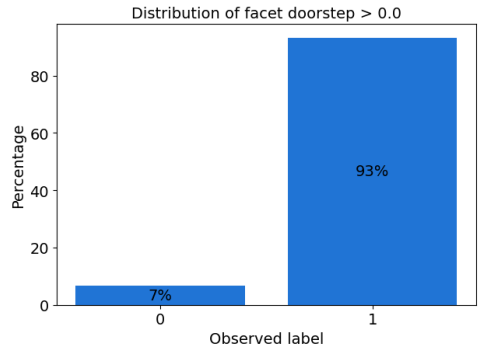
We computed the bias metrics for the label target using label value(s)/threshold target = 1 for the following facets:

- Facet column: **doorstep**

The pie chart shows the distribution of facet column doorstep in your data.
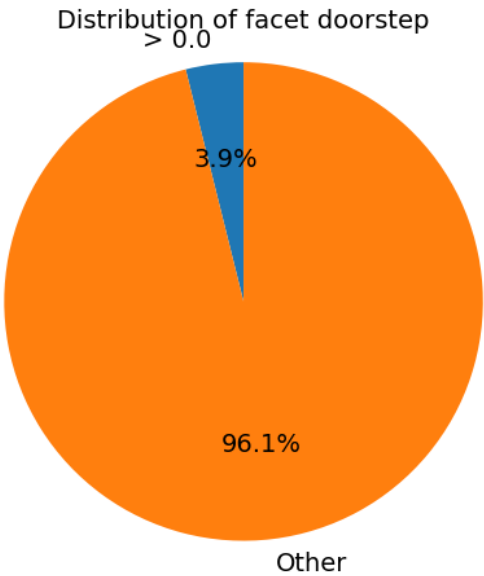
The bar plot(s) below show the distribution of facet column doorstep in your data.

Facet Value(s)/Threshold: doorstep > 0.0

Metric	Description	Value	Error
Conditional Demographic Disparity in Labels (CDDL)	Measures maximum divergence between the observed label distributions for facet values doorstep > 0.0 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 doorstep > 0.0 and rest of the inputs.	0.922	None
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values doorstep > 0.0 and rest of the inputs.	-0.143	None
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values doorstep > 0.0 and rest of the inputs diverge from each other entropically.	0.022	None
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values doorstep > 0.0 and rest of the inputs diverge from each other entropically.	0.108	None
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values doorstep > 0.0 and rest of the inputs in the dataset.	0.143	None
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values doorstep > 0.0 rest of the inputs in the dataset.	0.202	None
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values doorstep > 0.0 and rest of the inputs in the dataset.	0.143	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.
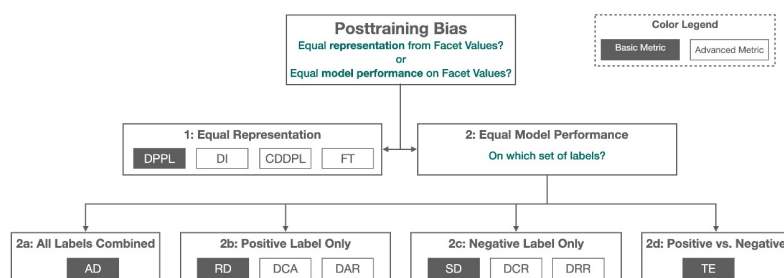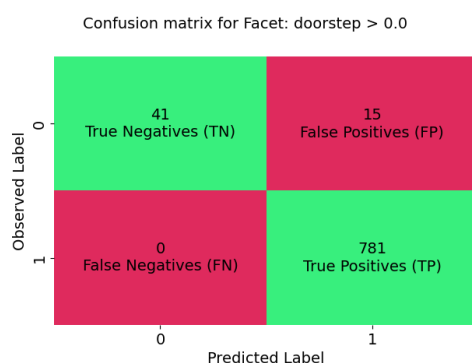

Bias can also result form imbalances in the model outcomes even when the facet value is not considered. The metric computing these imbalances is GE. The metric values along with an informal description of what they mean are shown below. For mathematical formulas and examples, see the [Measure Posttraining 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 target using label value(s)/threshold target = 1 for the following facets:

- Facet column: **doorstep**

Facet Value(s)/Threshold: doorstep > 0.0

Confusion matrix for Facet: doorstep > 0.0

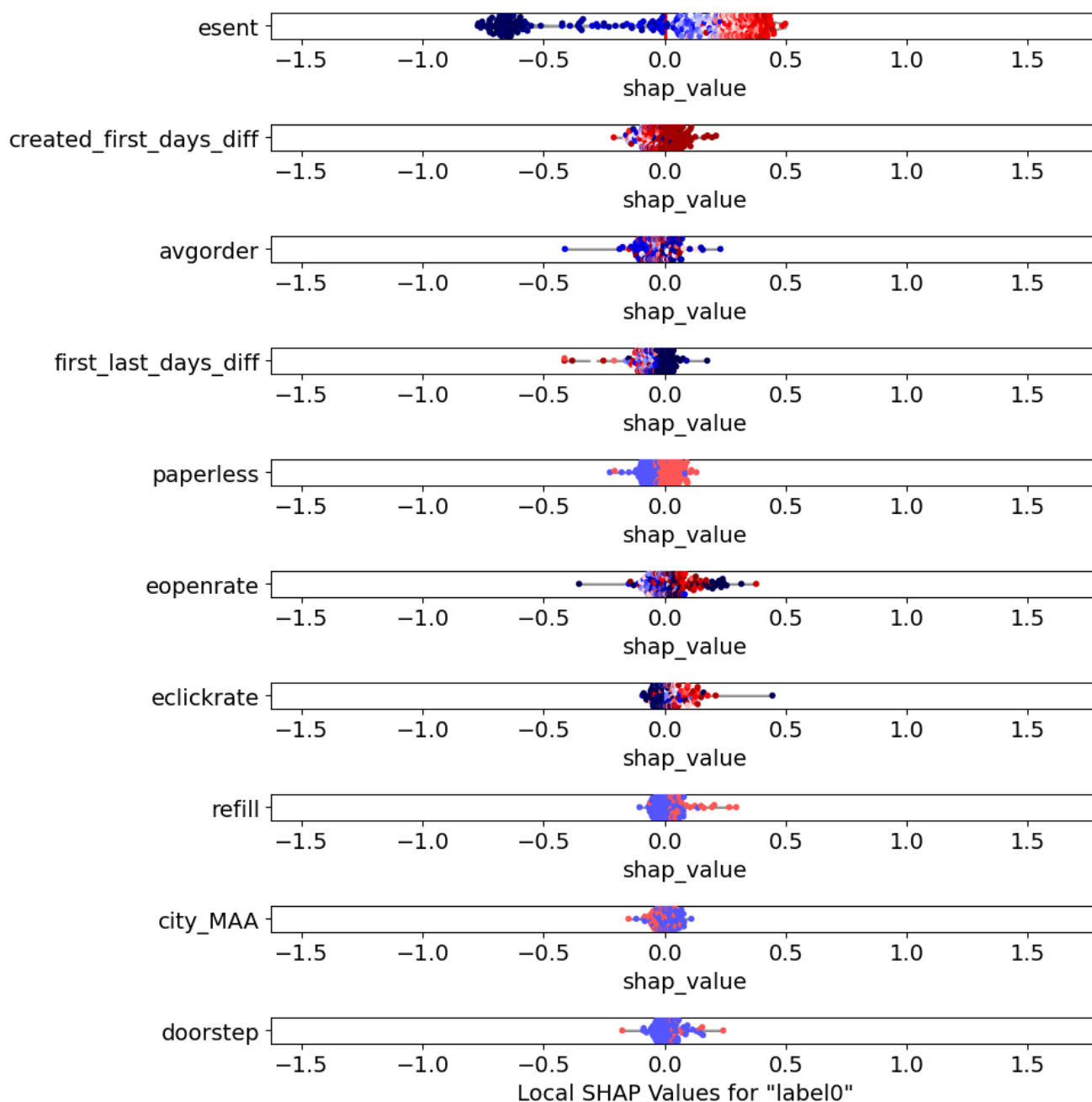| Metric	Description	Value	Error
[Accuracy Difference (AD)](#)	Measures the difference between the prediction accuracy for facet values `doorstep > 0.0` and rest of the inputs.	-0.007	None
[Conditional Demographic Disparity in Predicted Labels (CDDPL)](#)	Measures the disparity of predicted labels between facet values `doorstep > 0.0` 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 `doorstep > 0.0` and rest of the inputs.	-0.006	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 `doorstep > 0.0` and rest of the inputs for predicted positive outcomes (acceptances).	-0.000	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 `doorstep > 0.0` and rest of the inputs for negative outcomes (rejections).	0.286	None
[Disparate Impact (DI)](#)	Measures the ratio of proportions of the predicted labels for facet values `doorstep > 0.0` and rest of the inputs.	1.180	None
[Difference in Positive Proportions in Predicted Labels (DPPL)](#)	Measures the difference in the proportion of positive predictions between facet values `doorstep > 0.0` and rest of the inputs.	-0.145	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 `doorstep > 0.0` and rest of the inputs.	0.025	None
[Counterfactual Fliptest (FT)](#)	Examines each input with facet value `doorstep > 0.0` and assesses whether similar members from rest of the inputs have different model predictions.	0.001	None
[Generalized entropy (GE)](#)	Measures the inequality in benefits b assigned to each input by the model predictions.	0.012	None
[Recall Difference (RD)](#)	Measures the difference between the recall, aka true positive rate, of the model for facet values `doorstep > 0.0` and rest of the inputs.	-0.006	None
[Specificity difference (SD)](#)	Measures the difference between the specificity, aka true negative rate, of the model for facet values `doorstep > 0.0` and rest of the inputs.	-0.171	None
[Treatment Equality (TE)](#)	Measures the difference in the ratio of false positives to false negatives between facet values `doorstep > 0.0` and rest of the inputs.	-0.238	None

Model explanations

SHAP Explanations for individual labels

The model has 21 features.

We computed KernelSHAP explanations on the dataset. For each label, we display the `10` features with the greatest feature attribution. You have `1` label(s).
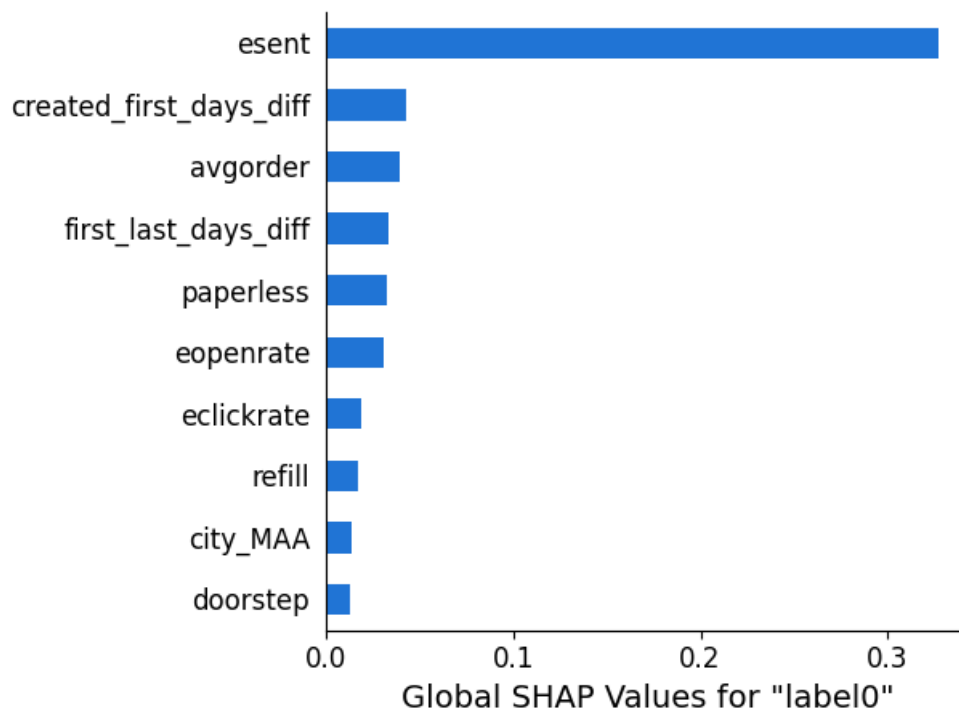
In the chart below, each point in the plot denotes an individual instance being explained. x-axis shows the SHAP value for the corresponding instance and feature. The red-blue color scale shows the value of the feature itself. Red indicates higher values whereas blue indicates lower values.

esent — shap_value

created_first_days_diff — shap_value

avgorder — shap_value

first_last_days_diff — shap_value

paperless — shap_value

eopenrate — shap_value

eclickrate — shap_value

refill — shap_value

city_MAA — shap_value

doorstep — Local SHAP Values for "label0"

Appendix: Analysis Configuration Parameters

```
{
    "dataset_type": "text/csv",
    "headers": [
        "target",
        "esent",
        "eopenrate",
        "eclickrate",
        "avgorder",
        "ordfreq",
        "paperless",
        "refill",
        "doorstep",
        "first_last_days_diff",
        "created_first_days_diff",
        "favday_Friday",
        "favday_Monday",
        "favday_Saturday",
        "favday_Sunday",
```

```json
        "favday_Thursday",
        "favday_Tuesday",
        "favday_Wednesday",
        "city_BLR",
        "city_BOM",
        "city_DEL",
        "city_MAA"
    ],
    "label": 0,
    "label_values_or_threshold": [
        1
    ],
    "facet": [
        {
            "name_or_index": "doorstep",
            "value_or_threshold": [
                0
            ]
        }
    ],
    "predictor": {
        "model_name": "pipelines-knj2s49bhg9v-ChurnCreateModel-bhwLz0t2Qn",
        "instance_type": "ml.c5.xlarge",
        "initial_instance_count": 1,
        "accept_type": "text/csv"
    },
    "probability_threshold": 0.5,
    "methods": {
        "pre_training_bias": {
            "methods": "all"
        },
        "post_training_bias": {
            "methods": "all"
        },
        "shap": {
            "baseline": "s3://churningbucket/input/baseline/baseline.csv",
            "num_samples": 50,
            "agg_method": "mean_abs",
            "save_local_shap_values": true
        },
        "report": {
            "name": "report",
            "title": "Analysis Report"
        }
    }
}
```