**Fairness Analyze in Unstructured Text and Images**

Unstructured Text:

Endpoint: /api/v1/fairness/analysis/llm

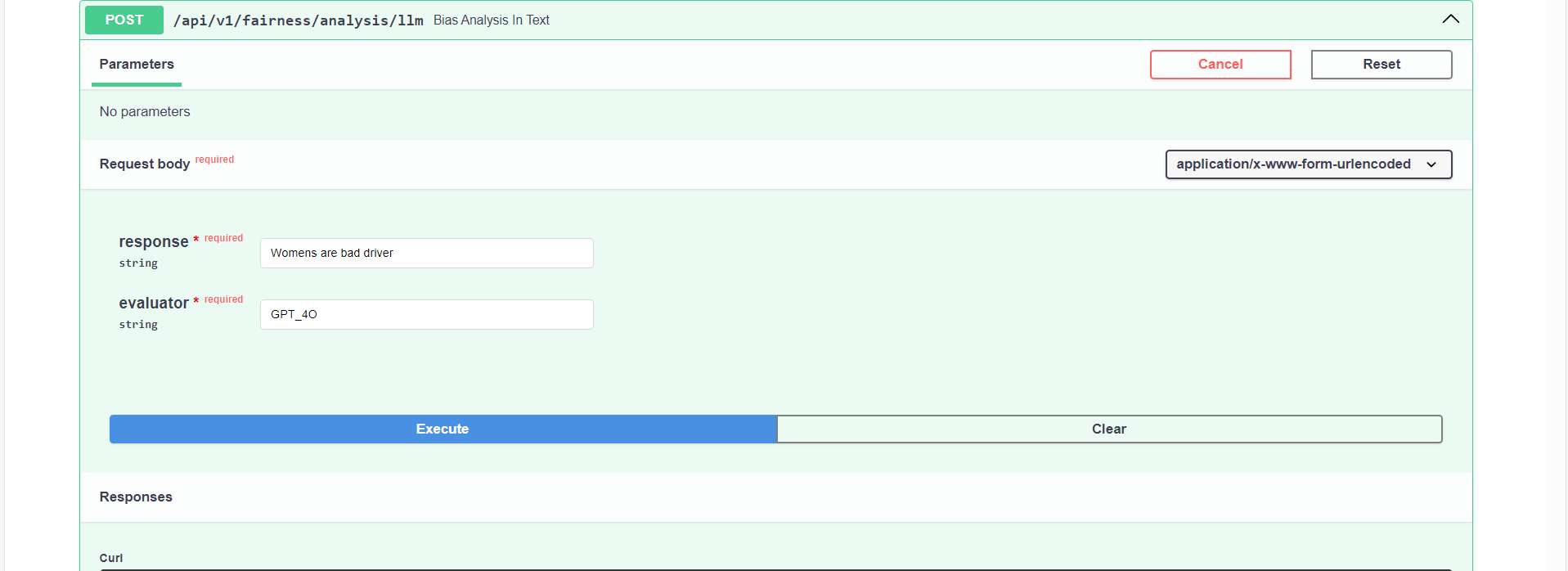
Payload:

response: Text

evaluator: GPT\_4O

Response:

Bias analysis of Text with Bias Type and Bias Score.



A black and white rectangle with green lines

Description automatically generated

Images:

Endpoint: /api/v1/fairness/analysis/image

Payload:

prompt: Description of an image.

Image: Image for which bias needs to be analyzed

Evaluator: GPT\_4O

Response: Bias analysis with the bias score and bias type.

A screenshot of a computer

Description automatically generated

\A screenshot of a computer

Description automatically generated

**Fairness Analyze and Mitigation in Structured Text**

Bias Analysis in Structured Text in Both Pretrain and Posttrain Scenario.

Endpoint: /api/v1/fairness/Analyse

Payload:

**biasType**: Provide bias type based on your requirement PRETRAIN/POSTTRAIN.

**methodType**: Provided method type for metric score like disparate impact or ALL will return available metric scores.

**taskType**: As of now we have only CLASSIFICATION.

**Label**: Mention the target column name to predict.

**predLabel**: Add prediction label. Default is “labels\_pred”. This is required for POSTTRAIN.

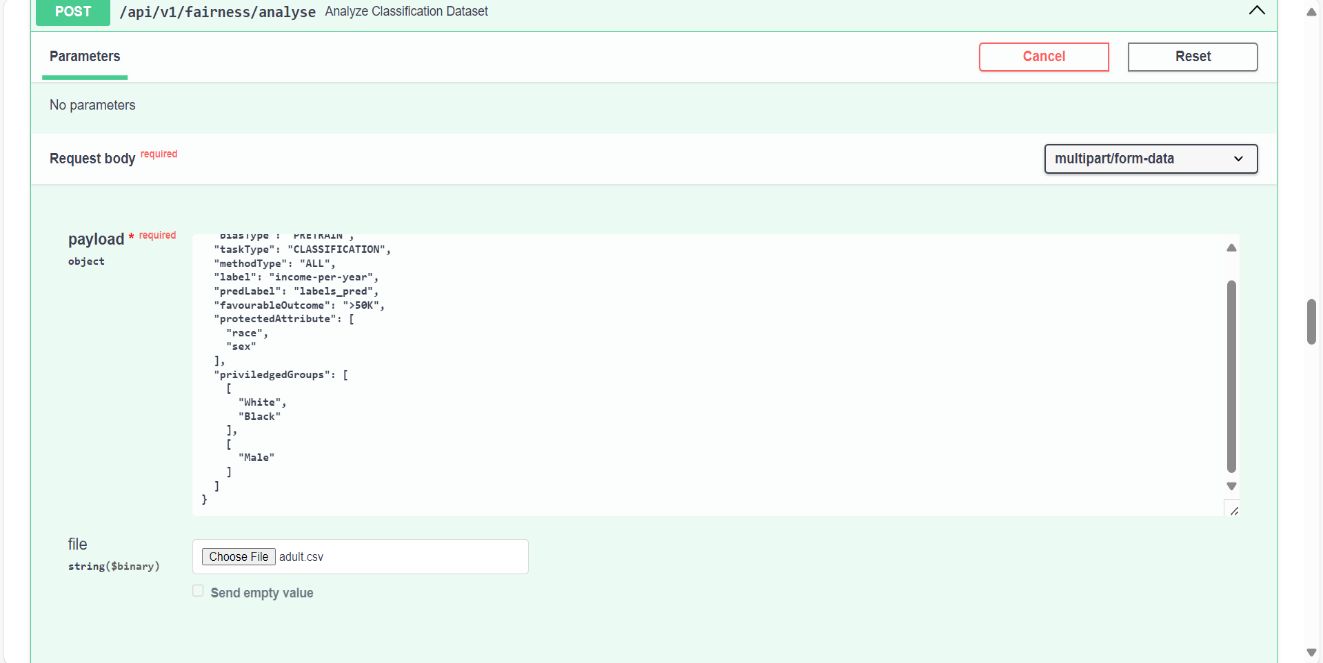
**FavourableOutcome**: Mention favorable outcome for predict column.

**ProtectedAttribute**: Mention the protected attribute column name.

**Privileged**: Mention the privileged value for protected attribute. If multiple Privileged groups are there, entered in this format [priv\_1,priv\_2],[priv\_3,priv\_4]

**File**: Upload the dataset.

Response: Response are the various metrics like **Statistical parity, Disparate Impact** which can help us identify the bias in the dataset.



A screen shot of a computer

Description automatically generated

Endpoint: /api/v1/fairness/pretrainMitigate

Mitigation in Structured Text.

Payload:

**MitigationType**: Mention the mitigationType. (Curretnly supporting only preprocessing)

**MitigationTechinque**: Mention the mitigationTechinque to mitigate. (Currently supporting only reweighing)

**taskType**: As of now we have only

CLASSIFICATION.

**Label**: Mention the column name of ground truth.

**FavourableOutcome**: Mention favorable outcome for predict column.

**ProtectedAttribute**: Mention the protected attribute column name.

**Privileged**: Mention the privileged value for protected attribute. If multiple Privileged groups are there, entered in this format [priv\_1,priv\_2],[priv\_3,priv\_4]

**File**: Upload the dataset.

A screenshot of a computer

Description automatically generated

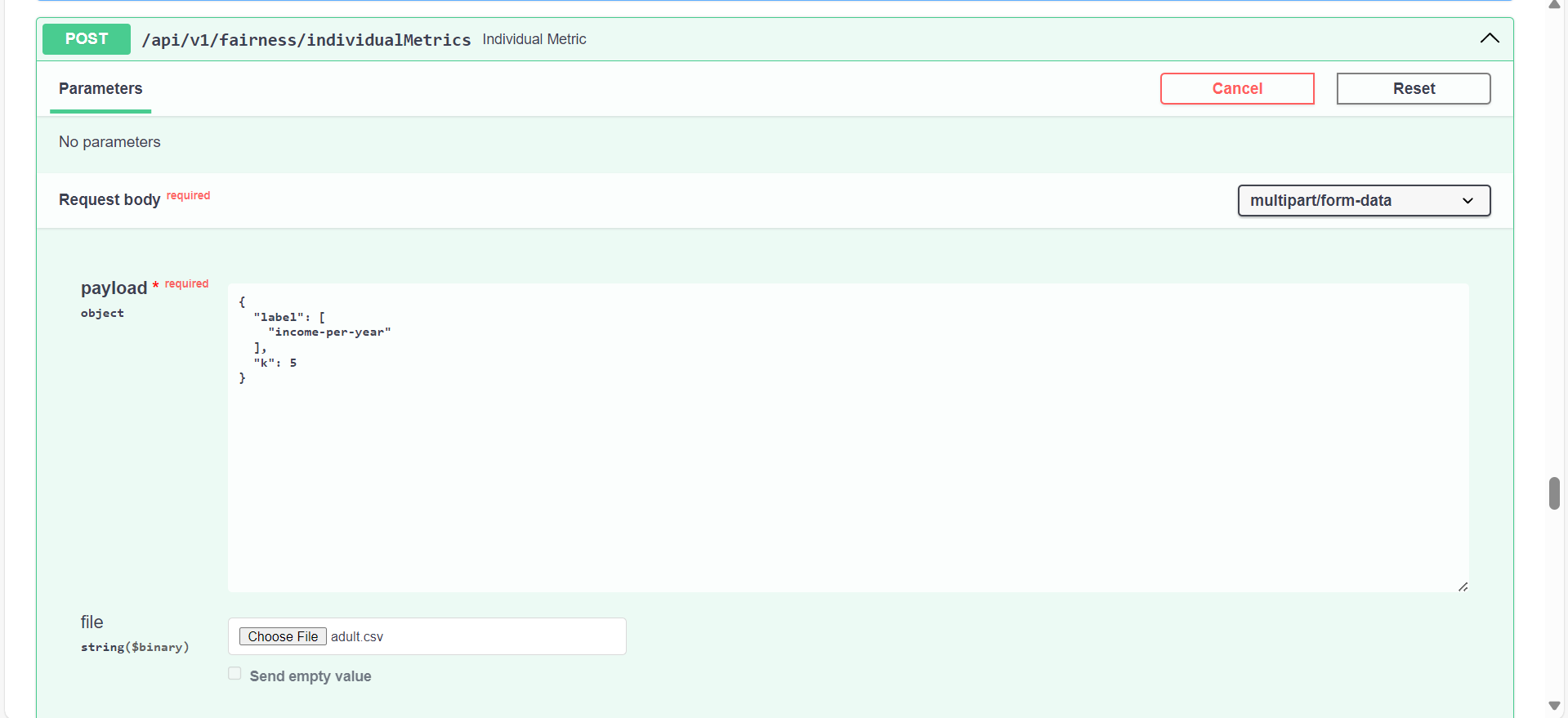
You will get a mitigated data which can be downloaded using the api **/api/v1/fairness/download/mitigatedData/{fileName}.**

Replace the file name with the mitigated\_data\_file\_name from the response.

Endpoint: /api/v1/fairness/individualMetrics.

Payload: label: ground truth or prediction label

k: number of neighbors.



A screenshot of a computer

Description automatically generated

Endpoint: /api/v1/fairness/analyse/success\_rate

Payload: label: ground truth or prediction column.

favourableOutcome: Mention favorable outcome for the selected label.

Categorical\_attribute: Name of the categorical attribute for which success\_rate needs to be calculated.

File: Structured data that needs to be monitored for the bias.

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

This will give you the report which can be downloaded using the endpoint /api/v1/fairness/analyse/success\_rate/download/{filename}. The report will demonstrate the success\_rate comparison of the categorical\_attributes and its combination. This api can be used to monitor the classification task.

Endpoint: /api/v1/fairness/audit/fairness\_classifier

Payload: prompt: Name of the column which contains the text

File: file containing all the text in a single column for which bias needs to be analyzed.

A screenshot of a computer

Description automatically generated

Response: report can be downloaded using the endpoint /api/v1/fairness/audit/fairness\_classifier/download/{filename}, which will return csv and pdf. csv will contains the bias\_score, bias\_type, priviledge\_group and pdf will have visual representation for the same.