



# CDW Documentation

## Responsible AI Test

---

# Responsible AI Test

## Purpose

Evaluate the Responsible AI dashboard and see what it does.

## Test Process

Here's a structured list of **Responsible AI Dashboard Deployment Steps** using the corrected scripts. Each step includes:

- **Step Number & Action**
- **Purpose**
- **Expected Result**

---

### Step 1: Install Required Packages

```
pip install --upgrade raiutils raiwidgets responsibleai ipywidgets
```

#### Purpose:

Install the Python packages required to run Responsible AI analysis and render the dashboard.

#### Expected Result:

Packages are installed without errors; dashboard widgets can render in the notebook (after kernel restart).

---

### Step 2: Load and Preprocess the Dataset

```
from sklearn.datasets import fetch_openml
import pandas as pd

data = fetch_openml(name='adult', version=2, as_frame=True)
df = data.frame.dropna()
```

#### Purpose:

Load a well-known classification dataset (Adult Census Income) and remove any missing values to avoid downstream errors.

#### Expected Result:

A clean DataFrame with no null values is loaded.

### □ Step 3: Split Dataset into Train and Test Sets

```
from sklearn.model_selection import train_test_split

target_column = 'class'
X = df.drop(columns=[target_column])
y = df[target_column]

X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y,
random_state=42)
```

#### □ Purpose:

Separate features and target, then split into training/testing sets for model training and evaluation.

#### □ Expected Result:

X\_train, X\_test, y\_train, y\_test variables created and stratified properly.

### □ Step 4: Define Preprocessing and Train a Model

```
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.ensemble import RandomForestClassifier

categorical_cols = X_train.select_dtypes(include=['object',
'category']).columns.tolist()
numerical_cols = X_train.select_dtypes(include=['int64',
'float64']).columns.tolist()

preprocessor = ColumnTransformer([
    ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_cols),
    ('num', StandardScaler(), numerical_cols)
])

clf = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('classifier', RandomForestClassifier(n_estimators=10, random_state=42))
])

clf.fit(X_train, y_train)
```

#### □ Purpose:

Build a model pipeline that encodes categorical features, scales numeric ones, and trains a classifier.

#### □ Expected Result:

Pipeline is trained successfully on the training data without conversion errors.

## □ Step 5: Prepare Data for RAIInsights

```
# Ensure target column is a supported type
y_train_clean = y_train.astype(str)
y_test_clean = y_test.astype(str)

train_data = X_train.copy()
train_data[target_column] = y_train_clean

test_data = X_test.copy()
test_data[target_column] = y_test_clean
```

### □ Purpose:

Re-attach the target column (as string) to the feature DataFrames — required for RAIInsights.

### □ Expected Result:

train\_data and test\_data DataFrames contain all required columns including the target.

## □ Step 6: Initialize the Responsible AI Insights Object

```
from responsibleai import RAIInsights, FeatureMetadata

feature_metadata = FeatureMetadata(categorical_features=categorical_cols)

rai_insights = RAIInsights(
    model=clf,
    train=train_data,
    test=test_data,
    target_column=target_column,
    task_type="classification",
    feature_metadata=feature_metadata
)
```

### □ Purpose:

Create a RAIInsights object that acts as the core engine for the Responsible AI dashboard.

### □ Expected Result:

RAIInsights object is initialized successfully and ready for configuration.

## □ Step 7: Add Responsible AI Analysis Tools

```
rai_insights.explainer.add()
rai_insights.error_analysis.add()
```

```
rai_insights.counterfactual.add(total_CFs=5, desired_class='opposite')
rai_insights.causal.add(treatment_features=categorical_cols)
```

□ **Purpose:**

Attach various tools (explanation, error analysis, counterfactuals, causal inference) to the insights engine.

□ **Expected Result:**

No errors thrown; tools are queued for computation.

---

□ **Step 8: Compute Insights**

```
rai_insights.compute()
```

□ **Purpose:**

Run analysis for all selected tools. This step may take a minute or more.

□ **Expected Result:**

Tool outputs are generated for the first 5,000 rows of the test set.

---

□ **Step 9: Launch the Responsible AI Dashboard**

```
from raiwidgets import ResponsibleAIDashboard
ResponsibleAIDashboard(rai_insights)
```

□ **Purpose:**

Open an interactive dashboard to explore insights such as feature importance, what-if analysis, and error breakdowns.

□ **Expected Result:**

A dashboard is displayed inside the notebook. Interactive plots and controls are available for analysis.

Download

```
NOTE: Due to the way that these URLs are deployed, this step will fail because the notebook sends the wrong headers and this is expected. You have to either pull the notebook local and use it from the terminal or register the dashboard/dataset and review it through the portal.
```

## Output

### Error analysis

**Tree map** **Heat map** **Feature list**

The tree visualization uses the mutual information between each feature and the error to best separate error instances from accurate instances hierarchically in the data. This simplifies the process of discovering and highlighting common feature patterns. To find important feature patterns, look for nodes with a stronger red color (i.e., high error rate) and a higher fit line (i.e., high error coverage). To edit the list of features being used in the tree, click on "Feature list". Use the "Select metric" dropdown menu to learn more about your error and success model performance. Please note that this metric selection will impact the way your error tree is generated.

**Basic Information**  
 All data  
 All data (0 items)

**Instances in global cohort**  
 Total: 200  
 Correct: 135  
 Incorrect: 74

**Instances in the selected cohort**  
 Total: 200  
 Correct: 135  
 Incorrect: 74

**Predictor path (Items)**

### Model overview

Evaluate the performance of your model by exploring the distribution of your predictor values and the values of your model performance metrics. Use the "Feature cohorts" tab to investigate your model's behavior in a comparative analysis of its performance across different groups of model inputs. Use the "Feature cohorts" tab to investigate your model by looking at a comparative analysis of its performance across sensitive/non-sensitive feature subgroups (e.g., performance across different genders, income levels).

**Feature cohorts** **Feature cohorts**

**Metrics:**  
 Accuracy score, F1 score, Precision score, Recall score, AUC, ROC

Cohort	Sample size	Accuracy score	F1 score	Precision score	Recall score	AUC
All data	200	0.69	0.69	0.69	0.69	0.69

**Probability distribution** **Metric distributions** **Confusion matrix**

Top option chart  Choose cohorts

### Data analysis

**Table view** **Chart view**

View the dataset in a table format for all features and rows.

Index	trust	product	age	workless	salary	education	education-num	sex
0	1	1	41	None	20500	Bachelor	10	Dir
10	1	1	34	None	20500	Some college	10	Ma
11	1	1	41	None	18000	HS grad	9	Ma
12	1	1	42	None	17500	Prof school	10	Ma
13	1	1	41	Local gov	17000	Assoc voc	10	Ma
16	1	1	35	None	16000	HS grad	9	Ma
17	1	1	38	None	20000	Assoc voc	10	Ma
18	1	1	33	None	12000	HS grad	9	Ma

### Data analysis

**Table view** **Chart view**

View the dataset in a table format for all features and rows.

Index	trust	product	age	workless	salary	education	education-num	sex
0	1	1	41	None	20500	Bachelor	10	Dir
10	1	1	34	None	20500	Some college	10	Ma
11	1	1	41	None	18000	HS grad	9	Ma
12	1	1	42	None	17500	Prof school	10	Ma
13	1	1	41	Local gov	17000	Assoc voc	10	Ma
16	1	1	35	None	16000	HS grad	9	Ma
17	1	1	38	None	20000	Assoc voc	10	Ma
18	1	1	33	None	12000	HS grad	9	Ma

### Feature importances

**Aggregate feature importance** **Individual feature importance**

Explore the top-4 important features that impact your overall model predictions (i.e., global explanations). Use the slider to show descending feature importances. All values feature importances are shown side by side and can be toggled off by selecting the values in the legend. Click any of the features in the graph to see a density plot below of how values of the selected feature affect prediction.

**Top 4 features by their importance**

**Sort by cohort**  
 All data

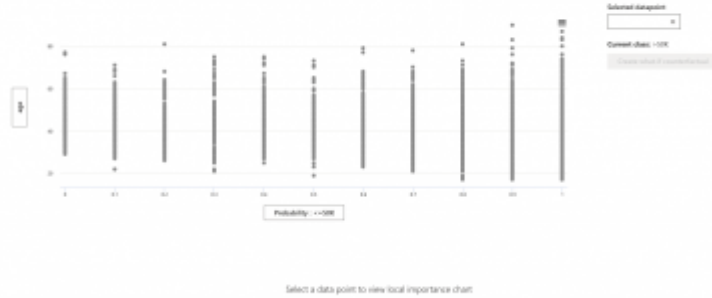
**Chart type**  
 Bar

**Class importance**  
 Average of absolute

**View dependence plot for:**  
 Select feature: workless  
 Select a dataset cohort: All data

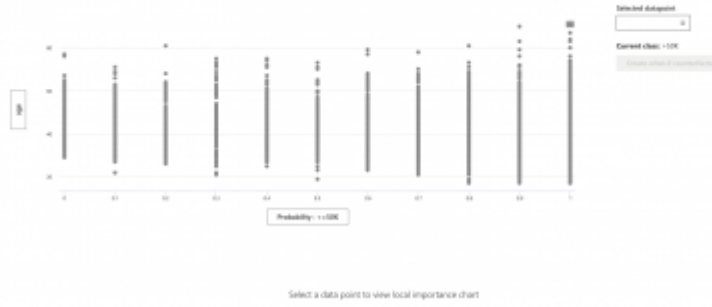
### Counterfactuals

What if allow you to perturb features for any input and observe how the model's prediction changes. You can perturb features manually or specify the desired prediction (e.g. class) for a dataset to see a list of closest data points to the original input that would lead to the desired prediction. Also known as prediction counterfactuals, you can use them for exploring the relationships learned by the model, understanding important, necessary features for the model's predictions, or adding edge cases for the model. To start, choose input points from the data table or scatter plot.



### Counterfactuals

What if allow you to perturb features for any input and observe how the model's prediction changes. You can perturb features manually or specify the desired prediction (e.g. class) for a dataset to see a list of closest data points to the original input that would lead to the desired prediction. Also known as prediction counterfactuals, you can use them for exploring the relationships learned by the model, understanding important, necessary features for the model's predictions, or adding edge cases for the model. To start, choose input points from the data table or scatter plot.



### Causal analysis

The overall causal effects across all data

Aggregate causal effects Individual causal effect Treatment policy

Causal analysis answers "what if" questions about learned relationships that would have changed under different policy choices, such as different pricing strategies for a product or an alternative treatment for a patient. Unlike model predictions that identify important conditions/features, these tools help prioritize the most important causal features that directly affect your outcome of interest. These results identify the causal effect of one feature (typically referred to as a "treatment"), holding other confounding features constant. For best results, make sure that the full dataset contains all available features that may confound with the outcome of interest.

Select aggregate causal effect of each treatment with 95% confidence interval

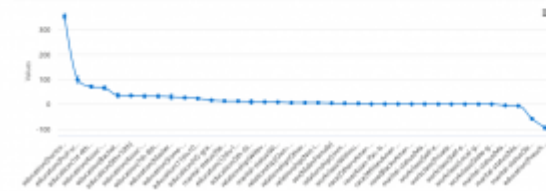
Why is it important to include confounding features?

Feature	Effect estimate	Standard error	Z score	P-value	Confidence interval	Confidence interval upper
education[0]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[1]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[2]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[3]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[4]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[5]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[6]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[7]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[8]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[9]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[10]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[11]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[12]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[13]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[14]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[15]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[16]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[17]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[18]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[19]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[20]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[21]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[22]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[23]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[24]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[25]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[26]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[27]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[28]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[29]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[30]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[31]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[32]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[33]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[34]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[35]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[36]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[37]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[38]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[39]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[40]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[41]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[42]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[43]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[44]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[45]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[46]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[47]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[48]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[49]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[50]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[51]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[52]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[53]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[54]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[55]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[56]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[57]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[58]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[59]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[60]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[61]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[62]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[63]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[64]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[65]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[66]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[67]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[68]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[69]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[70]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[71]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[72]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[73]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[74]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[75]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[76]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[77]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[78]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[79]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[80]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[81]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[82]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[83]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[84]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[85]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[86]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[87]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[88]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[89]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[90]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[91]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[92]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[93]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[94]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[95]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[96]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[97]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[98]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1
education[99]	0.023e+1	1.80e+0	4.30e+0	0.00e+0	0.02e+1	0.03e+1

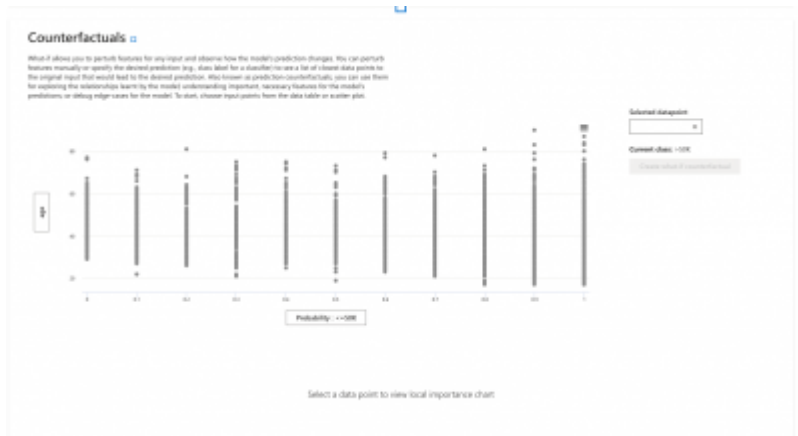
**Confidence intervals:** On average in this sample, increasing this feature by 1 unit will cause the probability of class/label "+10%" to increase by 0 units.

**Binary treatments:** On average in this sample, turning on this feature will cause the probability of class/label "+10%" to increase by 0 units.

A linear fit logistic regression fit is fitted with fit to predict y from X(1) and a least-squares regression fit is used to compute the average contribution of the non-binary features to the non-probability labels. Learn more about [Responsible Machine Learning](#).







## AI Knowledge