

```
!pip install xgboost lightgbm shap lime imbalanced-learn
```

```

275.7/275.7 kB 5.0 MB/s eta 0:00:00
Preparing metadata (setup.py) ... done
Requirement already satisfied: imbalanced-learn in /usr/local/lib/python3.11/dist-packages (0.13.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (from xgboost) (2.2.6)
Requirement already satisfied: nvidia-nccl-cu12 in /usr/local/lib/python3.11/dist-packages (from xgboost) (2.21.5)
Requirement already satisfied: scipy in /usr/local/lib/python3.11/dist-packages (from xgboost) (1.15.3)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (from shap) (1.6.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from shap) (1.5.3)
Requirement already satisfied: tqdm>=4.27.0 in /usr/local/lib/python3.11/dist-packages (from shap) (4.67.1)
Requirement already satisfied: packaging>20.9 in /usr/local/lib/python3.11/dist-packages (from shap) (24.2)
Requirement already satisfied: slicer==0.0.8 in /usr/local/lib/python3.11/dist-packages (from shap) (0.0.8)
Requirement already satisfied: numba>=0.54 in /usr/local/lib/python3.11/dist-packages (from shap) (0.60.0)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.11/dist-packages (from shap) (3.1.1)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.11/dist-packages (from shap) (4.13.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (from lime) (3.10.0)
Requirement already satisfied: scikit-image>=0.12 in /usr/local/lib/python3.11/dist-packages (from lime) (0.25.2)
Requirement already satisfied: sklearn-compat<1,>=0.1 in /usr/local/lib/python3.11/dist-packages (from imbalanced-learn) (0.1.3)
Requirement already satisfied: joblib<2,>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from imbalanced-learn) (1.5.0)
Requirement already satisfied: threadpoolctl<4,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from imbalanced-learn) (3.6.0)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.11/dist-packages (from numba>=0.54->shap) (0.43.0)
Collecting numpy (from xgboost)
  Downloading numpy-2.0.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (60 kB)
    60.9/60.9 kB 3.5 MB/s eta 0:00:00
Requirement already satisfied: networkx>=3.0 in /usr/local/lib/python3.11/dist-packages (from scikit-image>=0.12->lime) (3.4.2)
Requirement already satisfied: pillow>=10.1 in /usr/local/lib/python3.11/dist-packages (from scikit-image>=0.12->lime) (11.2.1)
Requirement already satisfied: imageio!=2.35.0,>=2.33 in /usr/local/lib/python3.11/dist-packages (from scikit-image>=0.12->lime) (2025.0.1)
Requirement already satisfied: tifffile>=2022.8.12 in /usr/local/lib/python3.11/dist-packages (from scikit-image>=0.12->lime) (2025.0.1)
Requirement already satisfied: lazy-loader>=0.4 in /usr/local/lib/python3.11/dist-packages (from scikit-image>=0.12->lime) (0.4)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->lime) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib->lime) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib->lime) (4.58.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->lime) (1.4.8)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib->lime) (3.2.3)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.11/dist-packages (from matplotlib->lime) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->shap) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.7->matplotlib->lime) (1.17.0)
  Downloading numpy-2.0.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (19.5 MB)
    19.5/19.5 MB 22.4 MB/s eta 0:00:00
Building wheels for collected packages: lime
  Building wheel for lime (setup.py) ... done
  Created wheel for lime: filename=lime-0.2.0.1-py3-none-any.whl size=283834 sha256=6921ff8dd673a6307cd8e86f3430b902e7f2f89ac0102f8e8e8e8e8e8e8e8e8e8
  Stored in directory: /root/.cache/pip/wheels/85/fa/a3/9c2d44c9f3cd77cf4e533b58900b2bf4487f2a17e8ec212a3d
Successfully built lime
Installing collected packages: numpy, lime
  Attempting uninstall: numpy
    Found existing installation: numpy 2.2.6
    Uninstalling numpy-2.2.6:
      Successfully uninstalled numpy-2.2.6
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.
google-colab 1.0.0 requires pandas==2.2.2, but you have pandas 1.5.3 which is incompatible.
dask-expr 1.1.21 requires pandas>=2, but you have pandas 1.5.3 which is incompatible.
plotnine 0.14.5 requires pandas>=2.2.0, but you have pandas 1.5.3 which is incompatible.
xarray 2025.3.1 requires pandas>=2.1, but you have pandas 1.5.3 which is incompatible.
dask-cudf-cu12 25.2.2 requires pandas<2.2.4dev0,>=2.0, but you have pandas 1.5.3 which is incompatible.
cudf-cu12 25.2.1 requires pandas<2.2.4dev0,>=2.0, but you have pandas 1.5.3 which is incompatible.
mizani 0.13.5 requires pandas>=2.2.0, but you have pandas 1.5.3 which is incompatible.
Successfully installed lime-0.2.0.1 numpy-2.0.2

```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
import os
```

```
folder_path = '/content/drive/My Drive/Research paper dataset'
os.listdir(folder_path)
```

```
['diabetes_binary_health_indicators_BRFSS2015.csv',
 'diabetes_012_health_indicators_BRFSS2015.csv',
 'diabetes_binary_5050split_health_indicators_BRFSS2015.csv']
```

```
['Diabetes_BRFSS.csv']
```

```
['Diabetes_BRFSS.csv']
```

```
file_path = '/content/drive/My Drive/Research paper dataset/Diabetes_BRFSS.csv'
```

```
import os
```

```
folder_path = r'/content/drive/My Drive/Research paper dataset'
print("Files in folder:", os.listdir(folder_path))
```

Files in folder: ['diabetes_binary_health_indicators_BRFSS2015.csv', 'diabetes_012_health_indicators_BRFSS2015.csv', 'diabetes_binar

```
# FIRST RUN THIS CELL TO SET UP ENVIRONMENT
# Then RESTART RUNTIME before running the next cell
```

```
!pip uninstall -y numpy pandas
!pip install numpy==1.23.5 pandas==1.5.3
print("Please go to Runtime > Restart runtime now, then run the next cell")
```

Found existing installation: numpy 2.0.2
Uninstalling numpy-2.0.2:
Successfully uninstalled numpy-2.0.2
Found existing installation: pandas 1.5.3
Uninstalling pandas-1.5.3:
Successfully uninstalled pandas-1.5.3
Collecting numpy==1.23.5
Downloading numpy-1.23.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (2.3 kB)
Collecting pandas==1.5.3
Using cached pandas-1.5.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (11 kB)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.11/dist-packages (from pandas==1.5.3) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas==1.5.3) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.1->pandas==1.5.3) (1.16.0)
Downloading numpy-1.23.5-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.1 MB)
17.1/17.1 MB 19.7 MB/s eta 0:00:00
Using cached pandas-1.5.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.0 MB)
Installing collected packages: numpy, pandas
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the sou
google-colab 1.0.0 requires pandas==2.2.2, but you have pandas 1.5.3 which is incompatible.
imbalanced-learn 0.13.0 requires numpy<3,>=1.24.3, but you have numpy 1.23.5 which is incompatible.
jaxlib 0.5.1 requires numpy>=1.25, but you have numpy 1.23.5 which is incompatible.
dask-expr 1.1.21 requires pandas>=2, but you have pandas 1.5.3 which is incompatible.
plotnine 0.14.5 requires pandas>=2.2.0, but you have pandas 1.5.3 which is incompatible.
pymc 5.22.0 requires numpy>=1.25.0, but you have numpy 1.23.5 which is incompatible.
xarray 2025.3.1 requires numpy>=1.24, but you have numpy 1.23.5 which is incompatible.
xarray 2025.3.1 requires pandas>=2.1, but you have pandas 1.5.3 which is incompatible.
dask-cudf-cu12 25.2.2 requires pandas<2.2.4dev0,>=2.0, but you have pandas 1.5.3 which is incompatible.
cudf-cu12 25.2.1 requires pandas<2.2.4dev0,>=2.0, but you have pandas 1.5.3 which is incompatible.
mizani 0.13.5 requires pandas>=2.2.0, but you have pandas 1.5.3 which is incompatible.
blosc2 3.3.3 requires numpy>=1.26, but you have numpy 1.23.5 which is incompatible.
scikit-image 0.25.2 requires numpy>=1.24, but you have numpy 1.23.5 which is incompatible.
chex 0.1.89 requires numpy>=1.24.1, but you have numpy 1.23.5 which is incompatible.
jax 0.5.2 requires numpy>=1.25, but you have numpy 1.23.5 which is incompatible.
albucore 0.0.24 requires numpy>=1.24.4, but you have numpy 1.23.5 which is incompatible.
bigframes 2.4.0 requires numpy>=1.24.0, but you have numpy 1.23.5 which is incompatible.
db-dtypes 1.4.3 requires numpy>=1.24.0, but you have numpy 1.23.5 which is incompatible.
tensorflow 2.18.0 requires numpy<2.1.0,>=1.26.0, but you have numpy 1.23.5 which is incompatible.
thinc 8.3.6 requires numpy<3.0.0,>=2.0.0, but you have numpy 1.23.5 which is incompatible.
alumentations 2.0.7 requires numpy>=1.24.4, but you have numpy 1.23.5 which is incompatible.
treescope 0.1.9 requires numpy>=1.25.2, but you have numpy 1.23.5 which is incompatible.
Successfully installed numpy-1.23.5 pandas-1.5.3
WARNING: The following packages were previously imported in this runtime:
[numpy]
You must restart the runtime in order to use newly installed versions.

RESTART SESSION

Please go to Runtime > Restart runtime now, then run the next cell

```
# NOW RUN THIS CELL AFTER RESTARTING RUNTIME
```

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
```

```
# Verify versions
```

```
print(f"NumPy version: {np.__version__}")
print(f"Pandas version: {pd.__version__}")
```

```
# Load dataset (replace with your actual path)
```

```
try:
    df = pd.read_csv('/content/drive/MyDrive/Research paper dataset/diabetes_binary_5050split_health_indicators_BRFSS2015.csv')
    print("Dataset loaded successfully")
```

```

# Check basic info
print("\nDataset info:")
print(df.info())
print("\nClass distribution:")
print(df['Diabetes_binary'].value_counts(normalize=True))

# Split features and target
X = df.drop('Diabetes_binary', axis=1)
y = df['Diabetes_binary']

# Split data (70/15/15 stratified split)
X_train, X_temp, y_train, y_temp = train_test_split(
    X, y,
    test_size=0.3,
    stratify=y,
    random_state=42
)
X_val, X_test, y_val, y_test = train_test_split(
    X_temp, y_temp,
    test_size=0.5,
    stratify=y_temp,
    random_state=42
)


# Apply SMOTE for class imbalance
smote = SMOTE(random_state=42)
X_train_res, y_train_res = smote.fit_resample(X_train, y_train)

# Standardize features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train_res)
X_val_scaled = scaler.transform(X_val)
X_test_scaled = scaler.transform(X_test)

print("\nPreprocessing completed successfully!")
print(f"Train shape: {X_train_scaled.shape}")
print(f"Validation shape: {X_val_scaled.shape}")
print(f"Test shape: {X_test_scaled.shape}")

except Exception as e:
    print(f"Error occurred: {str(e)}")

```

 NumPy version: 1.23.5
 Pandas version: 1.5.3
 Dataset loaded successfully

Dataset info:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 70692 entries, 0 to 70691

Data columns (total 22 columns):

#	Column	Non-Null Count	Dtype
0	Diabetes_binary	70692 non-null	float64
1	HighBP	70692 non-null	float64
2	HighChol	70692 non-null	float64
3	CholCheck	70692 non-null	float64
4	BMI	70692 non-null	float64
5	Smoker	70692 non-null	float64
6	Stroke	70692 non-null	float64
7	HeartDiseaseorAttack	70692 non-null	float64
8	PhysActivity	70692 non-null	float64
9	Fruits	70692 non-null	float64
10	Veggies	70692 non-null	float64
11	HvyAlcoholConsump	70692 non-null	float64
12	AnyHealthcare	70692 non-null	float64
13	NoDocbcCost	70692 non-null	float64
14	GenHlth	70692 non-null	float64
15	MentHlth	70692 non-null	float64
16	PhysHlth	70692 non-null	float64
17	DiffWalk	70692 non-null	float64
18	Sex	70692 non-null	float64
19	Age	70692 non-null	float64
20	Education	70692 non-null	float64
21	Income	70692 non-null	float64

dtypes: float64(22)
 memory usage: 11.9 MB
 None

Class distribution:

```
0.0    0.5
1.0    0.5
Name: Diabetes_binary, dtype: float64
```

Preprocessing completed successfully!

```
Train shape: (49484, 21)
Validation shape: (10604, 21)
```

Test shape: (10604, 21)

```
from sklearn.ensemble import RandomForestClassifier, VotingClassifier
from xgboost import XGBClassifier
from lightgbm import LGBMClassifier
from sklearn.metrics import accuracy_score, roc_auc_score, classification_report
```

```
# Initialize individual models
rf = RandomForestClassifier(n_estimators=200, max_depth=15, min_samples_split=5,
                           random_state=42, class_weight='balanced')
xgb = XGBClassifier(n_estimators=150, max_depth=8, learning_rate=0.1,
                   subsample=0.8, colsample_bytree=0.8, random_state=42)
lgbm = LGBMClassifier(n_estimators=150, max_depth=10, learning_rate=0.05,
                     subsample=0.8, colsample_bytree=0.8, random_state=42)
```

```
# Soft voting ensemble
ensemble = VotingClassifier(estimators=[
    ('rf', rf),
    ('xgb', xgb),
    ('lgbm', lgbm)],
    voting='soft')
```

```
# Train ensemble
ensemble.fit(X_train_scaled, y_train_res)
```

```
# Evaluate on validation set
val_pred = ensemble.predict(X_val_scaled)
val_proba = ensemble.predict_proba(X_val_scaled)[: , 1]
```

```
print(f"Validation Accuracy: {accuracy_score(y_val, val_pred):.4f}")
print(f"Validation ROC-AUC: {roc_auc_score(y_val, val_proba):.4f}")
print(classification_report(y_val, val_pred))
```

```
⚠ /usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:865: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finite'
_X, _y = _LGBMCheckXY(X, y, accept_sparse=True, force_all_finite=False, ensure_min_samples=2)
[LightGBM] [Info] Number of positive: 24742, number of negative: 24742
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.009752 seconds.
You can set `force_row_wise=true` to remove the overhead.
And if memory is not enough, you can set `force_col_wise=true`.
[LightGBM] [Info] Total Bins 213
[LightGBM] [Info] Number of data points in the train set: 49484, number of used features: 21
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.500000 -> initscore=0.000000
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finite'
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finite'
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
Validation Accuracy: 0.7531
Validation ROC-AUC: 0.8290
```

	precision	recall	f1-score	support
0.0	0.78	0.70	0.74	5302
1.0	0.73	0.81	0.77	5302
accuracy			0.75	10604
macro avg	0.76	0.75	0.75	10604
weighted avg	0.76	0.75	0.75	10604

```
# Final evaluation on test set
test_pred = ensemble.predict(X_test_scaled)
test_proba = ensemble.predict_proba(X_test_scaled)[: , 1]

print("\nTest Set Performance:")
print(f"Accuracy: {accuracy_score(y_test, test_pred):.4f}")
print(f"ROC-AUC: {roc_auc_score(y_test, test_proba):.4f}")
print(classification_report(y_test, test_pred))
```

```
# Feature names
feature_names = X.columns.tolist()
```

```
⚠ /usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finite'
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)

Test Set Performance:
Accuracy: 0.7507
ROC-AUC: 0.8288
```

	precision	recall	f1-score	support
0.0	0.78	0.71	0.74	5302
1.0	0.73	0.80	0.76	5302
accuracy			0.75	10604
macro avg	0.75	0.75	0.75	10604

weighted avg 0.75 0.75 0.75 10604

/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)

```
import shap

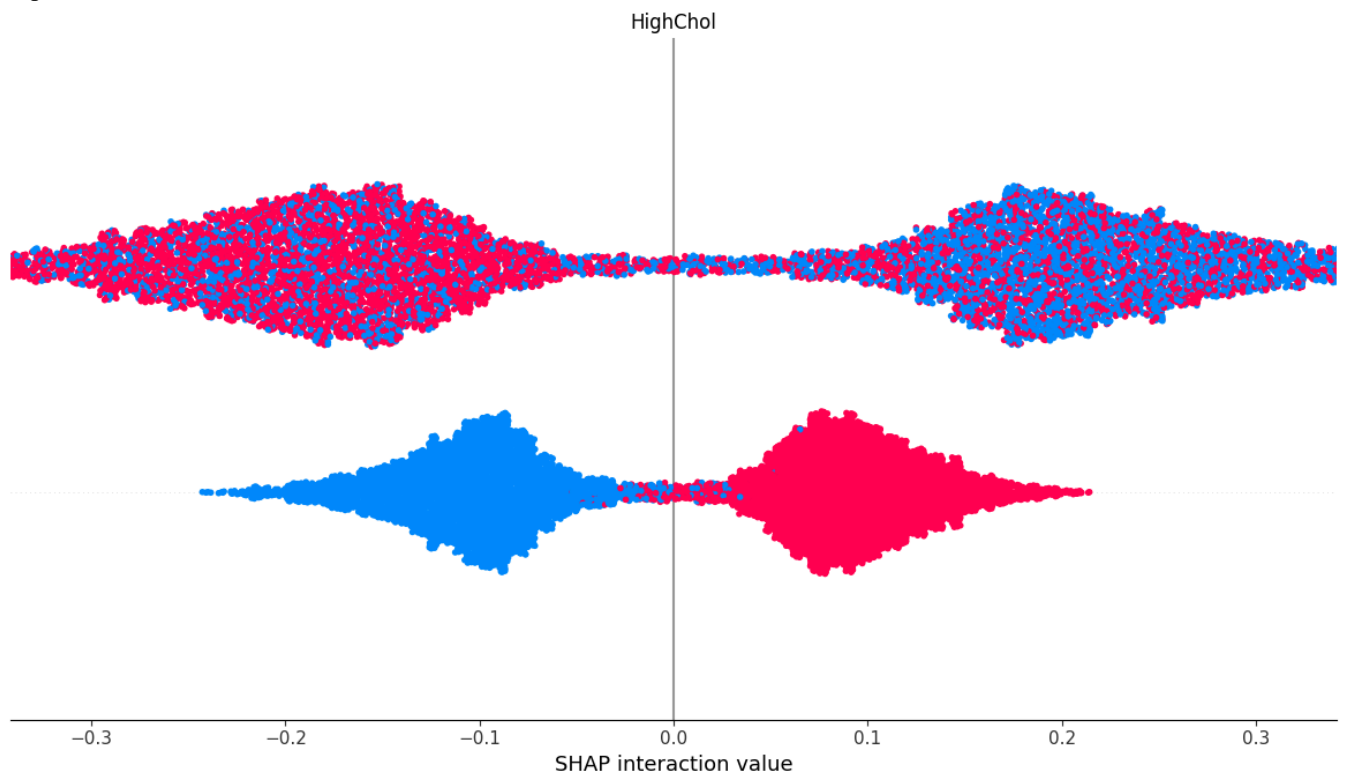
# Choose one of the base estimators from your ensemble
base_model = ensemble.estimators_[0][1] # Gets the RandomForest model

# Initialize TreeExplainer
explainer = shap.TreeExplainer(base_model)

# Get SHAP values
shap_values = explainer.shap_values(X_test_scaled)

# Plot summary
shap.summary_plot(shap_values, X_test_scaled, feature_names=feature_names)
```

↗ <Figure size 640x480 with 0 Axes>



```
# Create a predict function wrapper for the ensemble
def ensemble_predict(data):
    return ensemble.predict_proba(data)

# Initialize KernelExplainer (slower but more general)
explainer = shap.KernelExplainer(ensemble_predict, X_train_scaled[:100]) # Use subset for speed

# Calculate SHAP values
shap_values = explainer.shap_values(X_test_scaled[:100]) # Use subset for demo

# Plot results
shap.summary_plot(shap_values, X_test_scaled[:100], feature_names=feature_names)
```

https://colab.research.google.com/drive/1UoR6-x81z6l_wl6cWc_XmYiHnuieQzzc?usp=chrome_ntp#scrollTo=L16ZK6aZ9Tud&printMode=true 6/19

https://colab.research.google.com/drive/1UoR6-x81z6l_wl6cWc_XmYiHnuieQzzc?usp=chrome_ntp#scrollTo=L16ZK6aZ9Tud&printMode=true 7/19

https://colab.research.google.com/drive/1UoR6-x81z6l_wl6cWc_XmYiHnuieQzzc?usp=chrome_ntp#scrollTo=L16ZK6aZ9Tud&printMode=true 8/19

https://colab.research.google.com/drive/1UoR6-x81z6l_wl6cWc_XmYiHnuieQzzc?usp=chrome_ntp#scrollTo=L16ZK6aZ9Tud&printMode=true 9/19

[illegible]

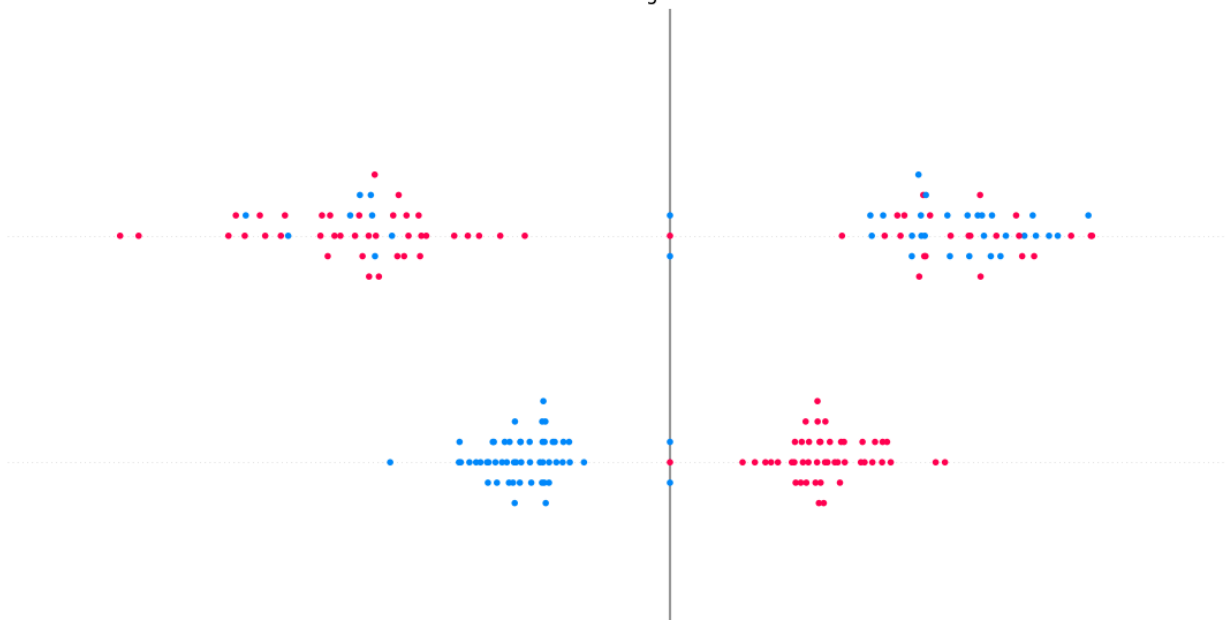
https://colab.research.google.com/drive/1UoR6-x81z6l_wl6cWc_XmYiHnueQzzc?usp=chrome_ntp#scrollTo=L16ZK6aZ9Tud&printMode=true 11/19

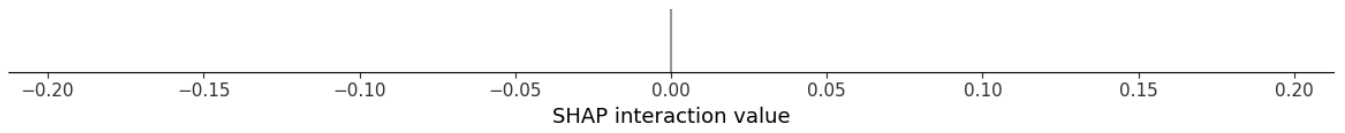
```

X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/basic.py:1278: UserWarning: Usage of np.ndarray subset (sliced data) is not recomme
ptr_data, type_ptr_data, _ = _c_float_array(data)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
<Figure size 640x480 with 0 Axes>

```

HighChol





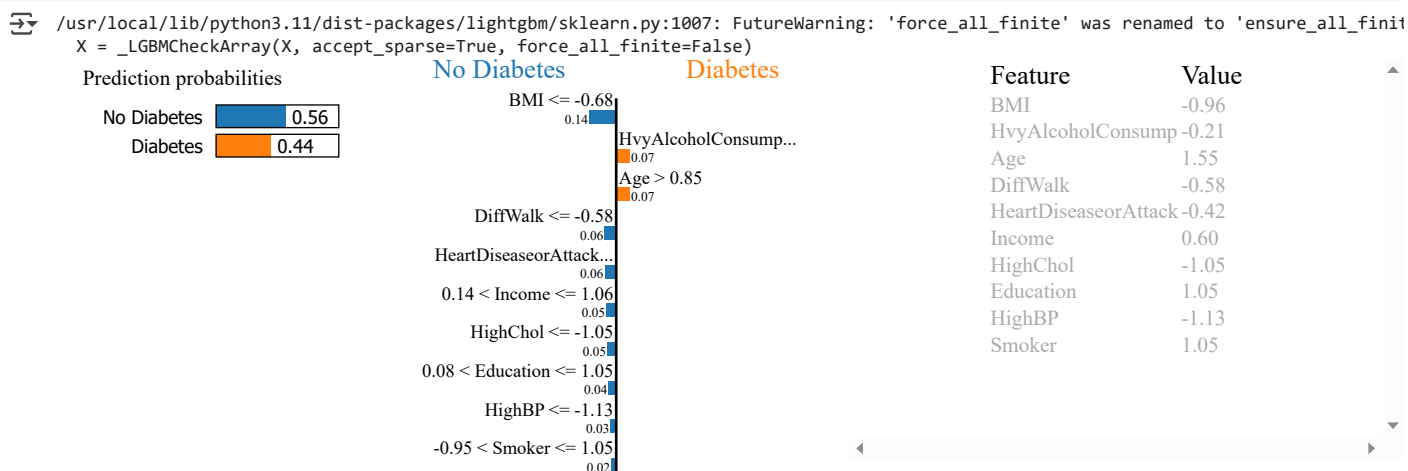
```
import lime
import lime.lime_tabular

# Initialize LIME explainer
explainer_lime = lime.lime_tabular.LimeTabularExplainer(
    X_train_scaled,
    feature_names=feature_names,
    class_names=['No Diabetes', 'Diabetes'],
    mode='classification')

# Define which instance to explain (e.g., first test instance)
instance_idx = 0 # You can change this to any index you want to explain

# Explain the specific instance
exp = explainer_lime.explain_instance(
    X_test_scaled[instance_idx],
    ensemble.predict_proba,
    num_features=10)

# Visualize the explanation
exp.show_in_notebook()
```



```
# LIME Implementation with Error Fix

# 1. First ensure you have these imports
import lime
import lime.lime_tabular
import numpy as np

# 2. Initialize LIME explainer
explainer_lime = lime.lime_tabular.LimeTabularExplainer(
    training_data=X_train_scaled,
    feature_names=feature_names,
    class_names=['No Diabetes', 'Diabetes'],
    mode='classification',
    discretize_continuous=True,
    random_state=42
)

# 3. Select an instance to explain (e.g., a diabetic case)
# Find index of a positive case if you want to explain a diabetic prediction
diabetic_indices = np.where(y_test == 1)[0]
instance_idx = diabetic_indices[0] if len(diabetic_indices) > 0 else 0

# 4. Generate explanation
exp = explainer_lime.explain_instance(
    data_row=X_test_scaled[instance_idx],
    predict_fn=ensemble.predict_proba,
    num_features=10,
    top_labels=1
)

# 5. Display the explanation
exp.show_in_notebook()
```



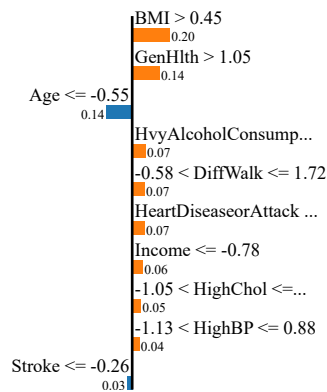
```
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit  
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
```

Prediction probabilities



No Diabetes

Diabetes



Feature	Value
BMI	1.57
GenHlth	1.95
Age	-0.55
HvyAlcoholConsump	-0.21
DiffWalk	1.72
HeartDiseaseorAttack	2.40
Income	-2.16
HighChol	0.95
HighBP	0.88
Stroke	-0.26

```
def explain_with_lime(instance_index):  
    exp = explainer_lime.explain_instance(  
        X_test_scaled[instance_index],  
        ensemble.predict_proba,  
        num_features=10)  
    return exp.show_in_notebook()
```

```
# Then call with any index  
explain_with_lime(0) # First test case  
explain_with_lime(42) # Specific case
```

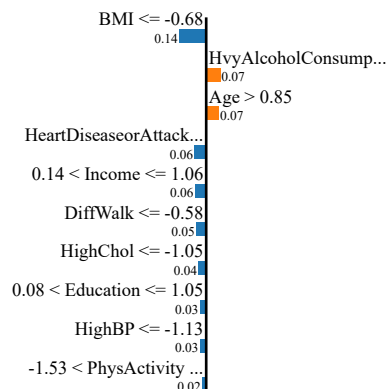
```
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit  
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
```

Prediction probabilities



No Diabetes

Diabetes



Feature	Value
BMI	-0.96
HvyAlcoholConsump	-0.21
Age	1.55
HeartDiseaseorAttack	-0.42
Income	0.60
DiffWalk	-0.58
HighChol	-1.05
Education	1.05
HighBP	-1.13
PhysActivity	0.65

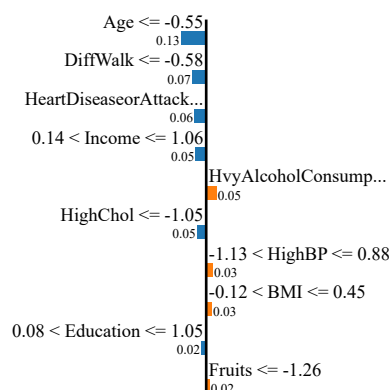
```
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit  
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
```

Prediction probabilities



No Diabetes

Diabetes



Feature	Value
Age	-0.55
DiffWalk	-0.58
HeartDiseaseorAttack	-0.42
Income	0.60
HvyAlcoholConsump	-0.21
HighChol	-1.05
HighBP	0.88
BMI	0.02
Education	1.05
Fruits	-1.26

```
# Save explanation as HTML  
exp.save_to_file('lime_explanation.html')
```

```
from sklearn.inspection import PartialDependenceDisplay

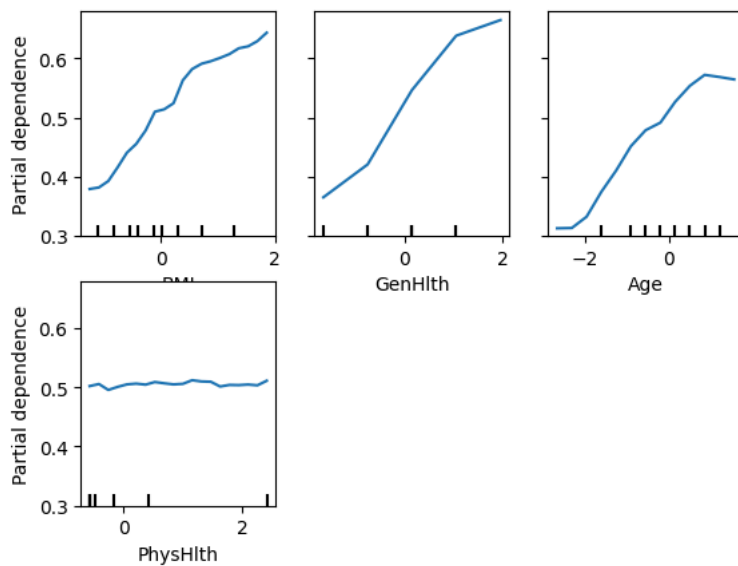
# For top features identified by SHAP
top_features = ['BMI', 'GenHlth', 'Age', 'PhysHlth']
PartialDependenceDisplay.from_estimator(
    ensemble, X_test_scaled, features=top_features,
    feature_names=feature_names, grid_resolution=20)
```

https://colab.research.google.com/drive/1UoR6-x81z6l_wl6cWc_XmYiHnuieQzzc?usp=chrome_ntp#scrollTo=L16ZK6aZ9Tud&printMode=true 17/19

```

/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
/usr/local/lib/python3.11/dist-packages/lightgbm/sklearn.py:1007: FutureWarning: 'force_all_finite' was renamed to 'ensure_all_finit
X = _LGBMCheckArray(X, accept_sparse=True, force_all_finite=False)
<sklearn.inspection._plot.partial_dependence.PartialDependenceDisplay at 0x79dfaabf7bd0>

```



```
!pip install dice-ml
```

```
Requirement already satisfied: dice-ml in /usr/local/lib/python3.11/dist-packages (0.11)
Requirement already satisfied: jsonschema in /usr/local/lib/python3.11/dist-packages (from dice-ml) (4.23.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (from dice-ml) (1.23.5)
Requirement already satisfied: pandas<2.0.0 in /usr/local/lib/python3.11/dist-packages (from dice-ml) (1.5.3)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (from dice-ml) (1.6.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from dice-ml) (4.67.1)
Requirement already satisfied: raiutils>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from dice-ml) (0.4.2)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.11/dist-packages (from pandas<2.0.0->dice-ml) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas<2.0.0->dice-ml) (2025.2)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from raiutils>=0.4.0->dice-ml) (2.32.3)
Requirement already satisfied: scipy in /usr/local/lib/python3.11/dist-packages (from raiutils>=0.4.0->dice-ml) (1.15.3)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (2023.12.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (0.25.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml) (1.5.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.1->pandas<2.0.0->dice-ml) (1.17.0)
Requirement already satisfied: typing-extensions>=4.4.0 in /usr/local/lib/python3.11/dist-packages (from referencing>=0.28.4->jsonschema->dice-ml) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (3.10.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (2.3.1)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (2025.11.12)
```

```
!pip install dice-ml --upgrade
```

```
Requirement already satisfied: dice-ml in /usr/local/lib/python3.11/dist-packages (0.11)
Requirement already satisfied: jsonschema in /usr/local/lib/python3.11/dist-packages (from dice-ml) (4.23.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (from dice-ml) (1.23.5)
Requirement already satisfied: pandas<2.0.0 in /usr/local/lib/python3.11/dist-packages (from dice-ml) (1.5.3)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (from dice-ml) (1.6.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from dice-ml) (4.67.1)
Requirement already satisfied: raiutils>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from dice-ml) (0.4.2)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.11/dist-packages (from pandas<2.0.0->dice-ml) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas<2.0.0->dice-ml) (2025.2)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from raiutils>=0.4.0->dice-ml) (2.32.3)
Requirement already satisfied: scipy in /usr/local/lib/python3.11/dist-packages (from raiutils>=0.4.0->dice-ml) (1.15.3)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (2023.12.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml) (0.25.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml) (1.5.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.1->pandas<2.0.0->dice-ml) (1.17.0)
Requirement already satisfied: typing-extensions>=4.4.0 in /usr/local/lib/python3.11/dist-packages (from referencing>=0.28.4->jsonschema->dice-ml) (4.13.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (3.10.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (2.3.1)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->raiutils>=0.4.0->dice-ml) (2025.11.12)
```

```
!pip install dice-ml==0.9
```

```
Collecting dice-ml==0.9
  Downloading dice_ml-0.9-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: jsonschema in /usr/local/lib/python3.11/dist-packages (from dice-ml==0.9) (4.23.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (from dice-ml==0.9) (1.23.5)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from dice-ml==0.9) (1.5.3)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (from dice-ml==0.9) (1.6.1)
Requirement already satisfied: h5py in /usr/local/lib/python3.11/dist-packages (from dice-ml==0.9) (3.13.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from dice-ml==0.9) (4.67.1)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml==0.9) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml==0.9) (2023.12.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml==0.9) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from jsonschema->dice-ml==0.9) (0.25.1)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.11/dist-packages (from pandas->dice-ml==0.9) (2.9.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->dice-ml==0.9) (2025.2)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml==0.9) (1.15.3)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml==0.9) (1.5.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn->dice-ml==0.9) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.1->pandas->dice-ml==0.9) (1.17.0)
Requirement already satisfied: typing-extensions>=4.4.0 in /usr/local/lib/python3.11/dist-packages (from referencing>=0.28.4->jsonschema->dice-ml==0.9) (4.13.2)
Downloading dice_ml-0.9-py3-none-any.whl (2.6 MB)
2.6/2.6 MB 2.4 MB/s eta 0:00:00
Installing collected packages: dice-ml
  Attempting uninstall: dice-ml
    Found existing installation: dice-ml 0.11
    Uninstalling dice-ml-0.11:
```