

The PIMA Indian Diabetes dataset is a widely used dataset for practicing machine learning and data analysis. It contains various health-related features to predict the onset of diabetes in PIMA Indian women.

Machine Learning for Predictive Modeling (Using Scikit-Learn):

```
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Assuming 'pima_data.csv' is the dataset file
pima_data = pd.read_csv('diabetes.csv')

# Split data into features (X) and target (y)
X = pima_data.drop('Outcome', axis=1)
y = pima_data['Outcome']

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                    random_state=42)

# Initialize and train a Random Forest Classifier
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Predict outcomes on the test set
y_pred = model.predict(X_test)
```

```
# Evaluate the model

accuracy = accuracy_score(y_test, y_pred)

report = classification_report(y_test, y_pred)

print(f'Accuracy: {accuracy}')

print(f'Classification Report:\n{report}')
```

OUTPUT:

Classification Report:					
	precision	recall	f1-score	support	
0	0.79	0.78	0.78	99	
1	0.61	0.62	0.61	55	
accuracy			0.72	154	
macro avg	0.70	0.70	0.70	154	
weighted avg	0.72	0.72	0.72	154	

Data Visualization with Plotly:

```
import plotly.express as px

import pandas as pd

pima_data = pd.read_csv("diabetes.csv")

fig = px.scatter(pima_data, x='BMI', y='Age', color='Outcome',

                 labels={'BMI': 'Body Mass Index', 'Age': 'Age'},

                 title='BMI vs. Age with Outcome')

fig.show()
```

OUTPUT:

