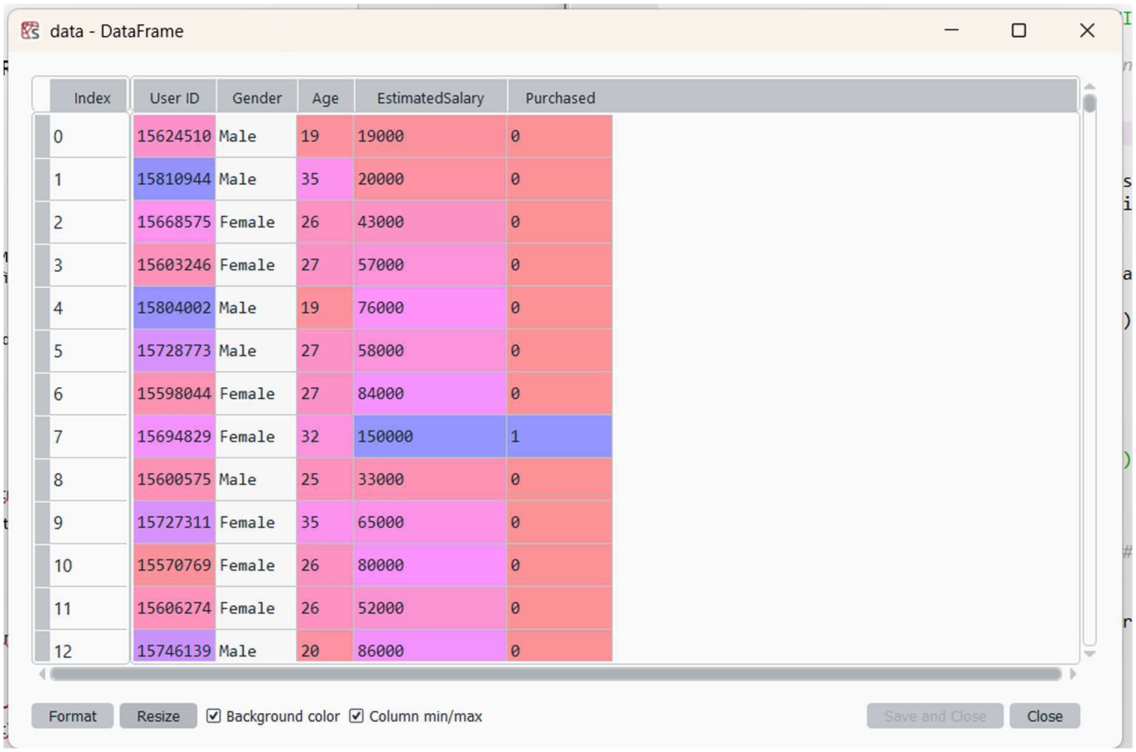
DecisionTree-Classifier

#lets import libraries

import pandas as pd

#lets read the dataset

data=pd.read_csv(r"C:\Users\TharunMahendra\NIT\6.Algorithms\2.Classification\ClassificationData.csv")

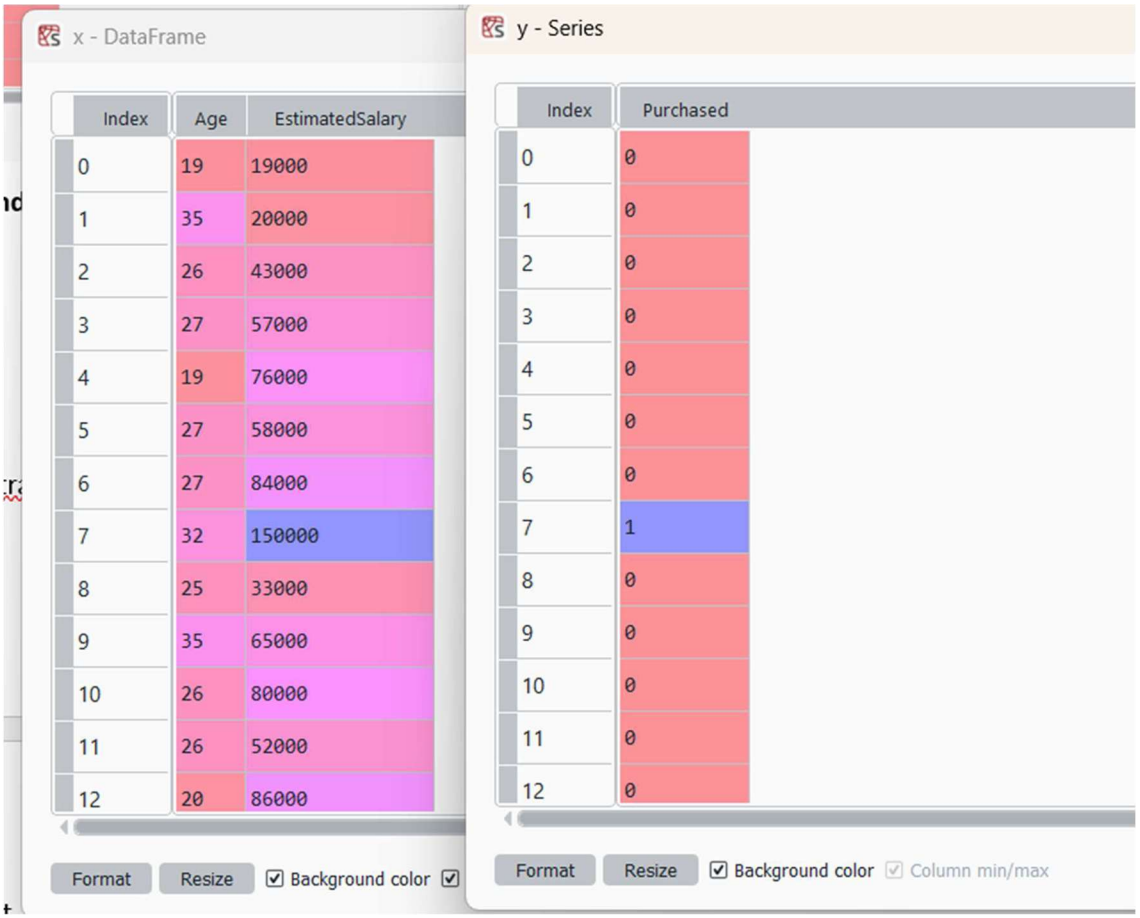


Index	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
5	15728773	Male	27	58000	0
6	15598044	Female	27	84000	0
7	15694829	Female	32	150000	1
8	15600575	Male	25	33000	0
9	15727311	Female	35	65000	0
10	15570769	Female	26	80000	0
11	15606274	Female	26	52000	0
12	15746139	Male	20	86000	0

#lets divide them into dependent & independent

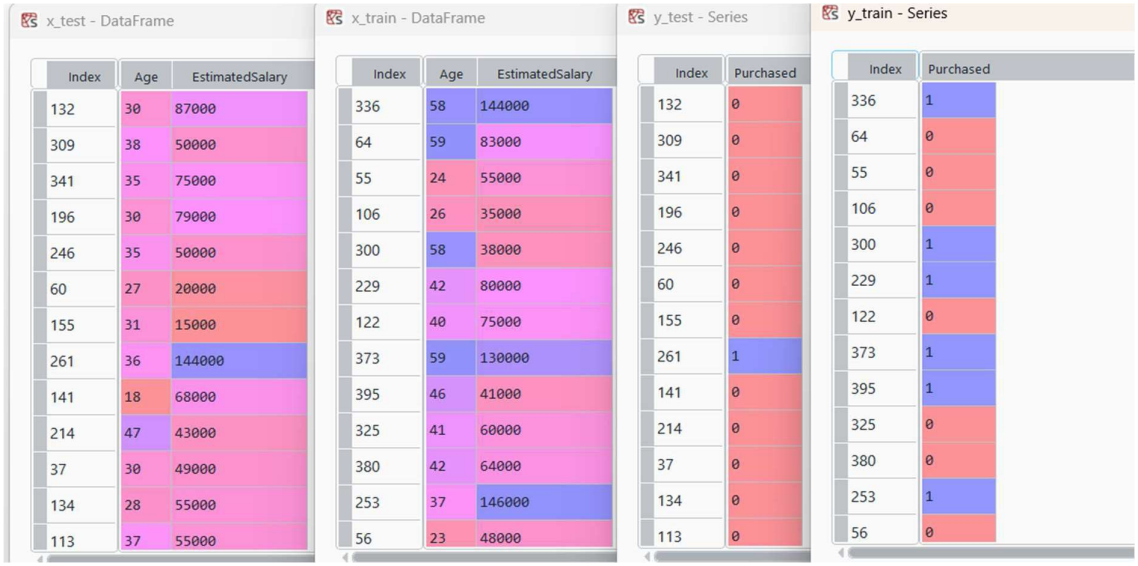
x=data.iloc[:,2:4] #age,salary

y=data.iloc[:,-1] #purchased



#splitting data

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,
random_state=0)
```



x_test - DataFrame

Index	Age	EstimatedSalary
132	30	87000
309	38	50000
341	35	75000
196	30	79000
246	35	50000
60	27	20000
155	31	15000
261	36	144000
141	18	68000
214	47	43000
37	30	49000
134	28	55000
113	37	55000

x_train - DataFrame

Index	Age	EstimatedSalary
336	58	144000
64	59	83000
55	24	55000
106	26	35000
300	58	38000
229	42	80000
122	40	75000
373	59	130000
395	46	41000
325	41	60000
380	42	64000
253	37	146000
56	23	48000

y_test - Series

Index	Purchased
132	0
309	0
341	0
196	0
246	0
60	0
155	0
261	1
141	0
214	0
37	0
134	0
113	0

y_train - Series

Index	Purchased
336	1
64	0
55	0
106	0
300	1
229	1
122	0
373	1
395	1
325	0
380	0
253	1
56	0

#model building

```
from sklearn.tree import DecisionTreeClassifier
model=DecisionTreeClassifier(random_state=0,criterion='entropy')
model.fit(x_train,y_train)
```

#prediction

```
y_pred=model.predict(x_test)
```

y_test - Series

Index	Purchased
132	0
309	0
341	0
196	0
246	0
60	0
155	0
261	1
141	0
214	0
37	0
134	0
113	0

y_pred - NumPy object array

	0
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	1
8	0
9	0
10	0
11	0
12	0

#confusion Matrix

from sklearn.metrics import confusion_matrix

cm=confusion_matrix(y_test, y_pred)

cm - NumPy object array

	0	1
0	53	5
1	3	19

from sklearn.metrics import accuracy_score

ac=**accuracy_score**(y_test,y_pred) **->0.9**

from sklearn.metrics import classification_report

cr=classification_report(y_test,y_pred)

Text editor - cr

	precision	recall	f1-score	support
0	0.95	0.91	0.93	58
1	0.79	0.86	0.83	22
accuracy			0.90	80
macro avg	0.87	0.89	0.88	80
weighted avg	0.90	0.90	0.90	80

bias=model.score(x_train,y_train) **-> 0.996875**

variance=model.score(x_test,y_test)-**> 0.9**

Deployment Code

```
# importing libraries
import streamlit as st
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics
import(confusion_matrix,accuracy_score,classification_report,roc_curve,roc_auc_score)
import seaborn as sns


st.title("Decision Tree Classifier")


# Uploading File
file=st.file_uploader('Upload Your File for Model Building',type=['csv'])
if file is not None:
    # Load
    data=pd.read_csv(file)
    st.write('- Preview')
    st.dataframe(data.head())


    # FeatureSelection
    x=data.iloc[:,2:4]
    y=data.iloc[:,-1]
```

```
# SplittingData
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,random_
state=0)

# model building
model=DecisionTreeClassifier()
model.fit(x_train,y_train)

# prediction
y_pred=model.predict(x_test)
y_prob=model.predict_proba(x_test)[:,1]

# Metrics
st.subheader("Confusion Matrix")
cm = confusion_matrix(y_test, y_pred)
fig_cm, ax = plt.subplots()
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", ax=ax)
st.pyplot(fig_cm)

ac = accuracy_score(y_test, y_pred)
st.write(f"**Accuracy:** {ac:.2f}")

st.subheader("Classification Report")
st.text(classification_report(y_test, y_pred))
```

```
    st.write(f"**Training Accuracy (Bias):** {model.score(x_train, y_train):.2f}")
    st.write(f"**Testing Accuracy (Variance):** {model.score(x_test, y_test):.2f}")

    # ROC Curve and AUC
    fpr, tpr, _ = roc_curve(y_test, y_prob)
    auc_score = roc_auc_score(y_test, y_prob)
    st.write(f"**AUC Score:** {auc_score:.2f}")

    st.subheader("ROC Curve")
    fig_roc, ax = plt.subplots()
    ax.plot(fpr, tpr, color="blue", label=f"ROC curve (AUC = {auc_score:.2f})")
    ax.plot([0, 1], [0, 1], color="gray", linestyle="--")
    ax.set_xlabel("False Positive Rate")
    ax.set_ylabel("True Positive Rate")
    ax.set_title("ROC Curve")
    ax.legend(loc="lower right")
    st.pyplot(fig_roc)
```

Deployment code

localhost:8501

Deploy

Decision Tree Classifier

Upload Your File for Model Building

Drag and drop file here
Limit 200MB per file • CSV

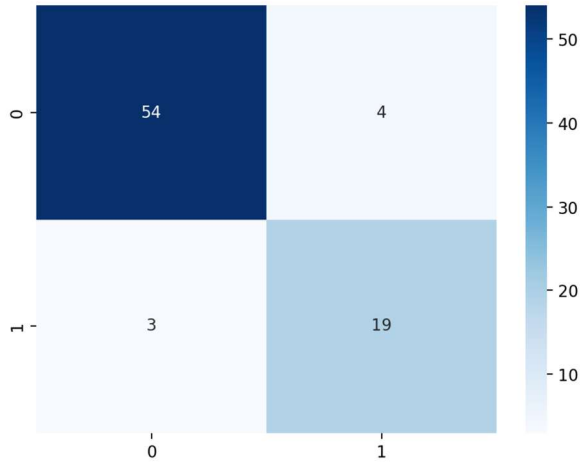
Browse files

ClassificationData.csv 10.7KB

- Preview

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15,624,510	Male	19	19,000	0
1	15,810,944	Male	35	20,000	0
2	15,668,575	Female	26	43,000	0
3	15,603,246	Female	27	57,000	0
4	15,804,002	Male	19	76,000	0

Confusion Matrix



Accuracy: 0.91

Classification Report

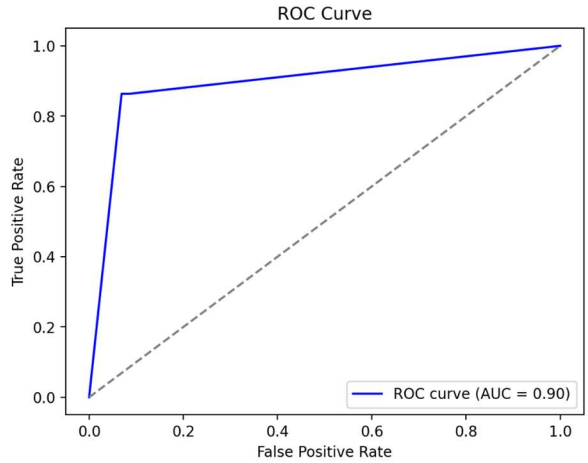
	precision	recall	f1-score	support
0	0.95	0.93	0.94	58
1	0.83	0.86	0.84	22
accuracy			0.91	80
macro avg	0.89	0.90	0.89	80
weighted avg	0.91	0.91	0.91	80

Training Accuracy (Bias): 1.00

Testing Accuracy (Variance): 0.91

AUC Score: 0.90

ROC Curve


ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (AUC = 0.90)