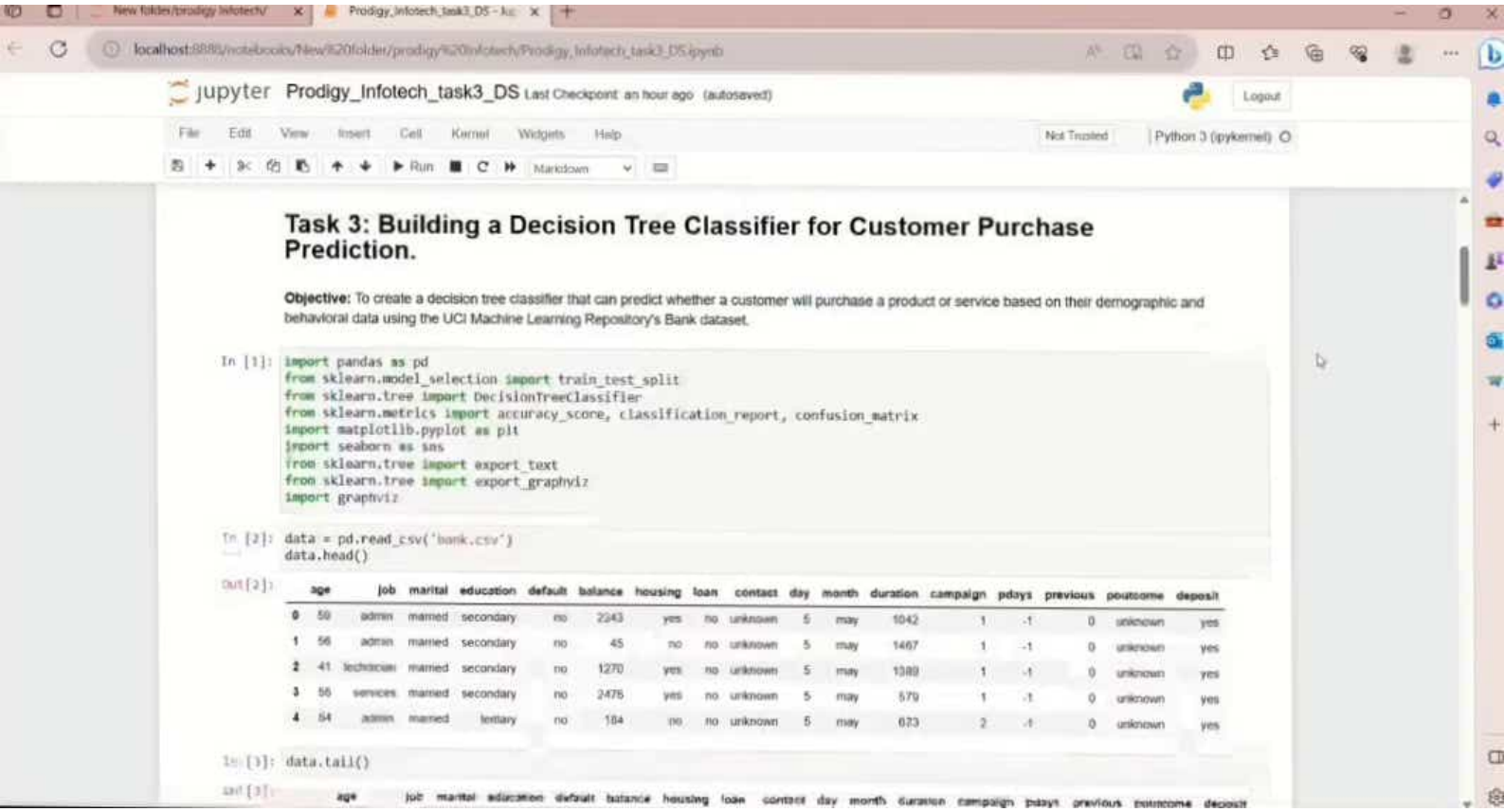
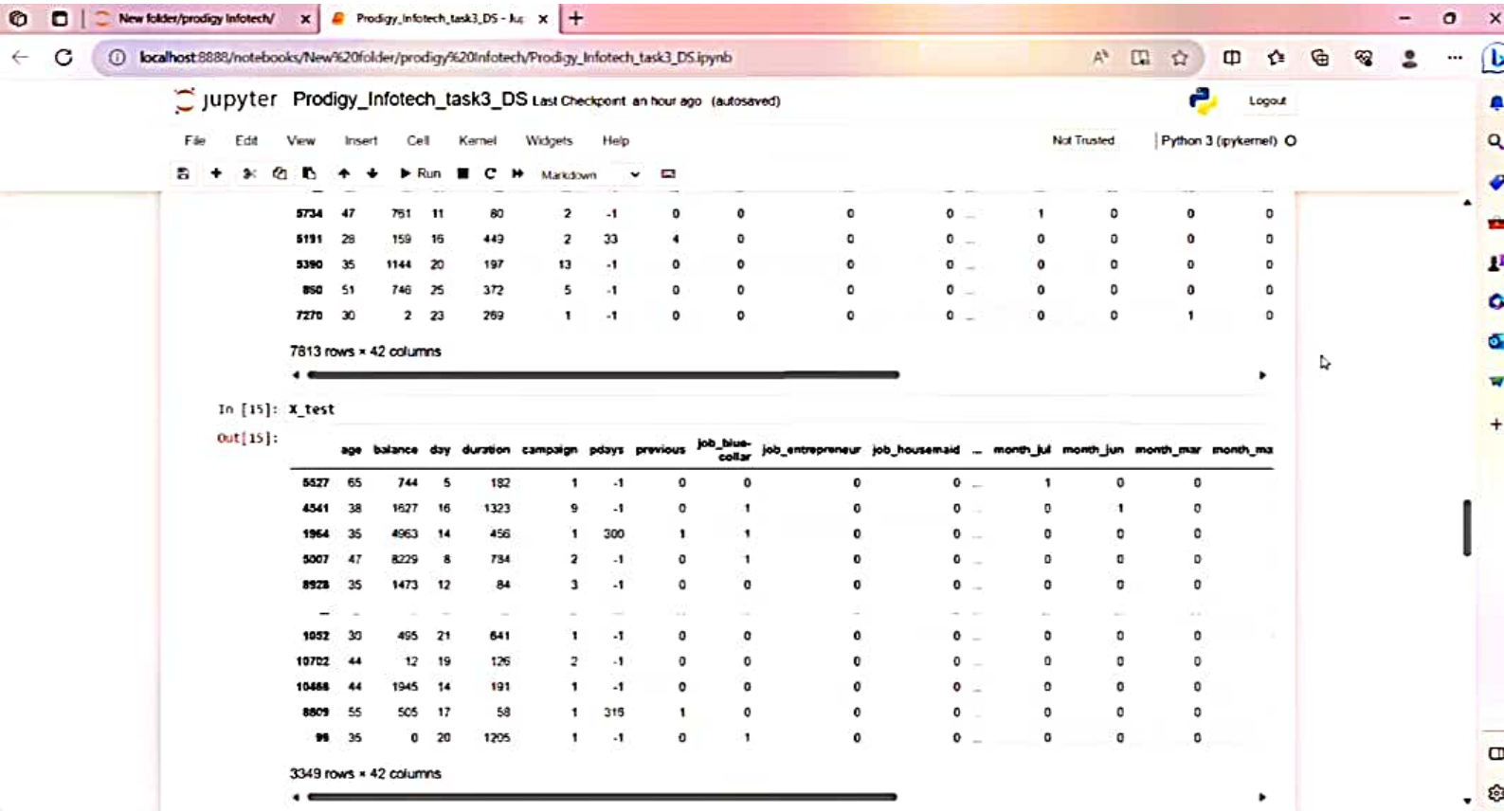

jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Markdown
Task 3: Building a Decision Tree Classifier for Customer Purchase Prediction.
Objective: To create a decision tree classifier that can predict whether a customer will purchase a product or service based on their demographic and behavioral data using the UCI Machine Learning Repository's Bank dataset.
In [1]:
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.tree import export_text
from sklearn.tree import export_graphviz
import graphviz
In [2]:
data = pd.read_csv('bank.csv')
data.head()
Out[2]:
age job marital education default balance housing loan contact day month duration campaign pdays previous poutcome deposit
0 59 admin. married secondary no 2343 yes no unknown 5 may 1042 1 -1 0 unknown yes
1 56 admin. married secondary no 45 no no unknown 5 may 1467 1 -1 0 unknown yes
2 41 technician married secondary no 1270 yes no unknown 5 may 1389 1 -1 0 unknown yes
3 55 services married secondary no 2476 yes no unknown 5 may 579 1 -1 0 unknown yes
4 54 admin. married tertiary no 184 no no unknown 5 may 673 2 -1 0 unknown yes
In [3]:
data.tail()
Out[3]:
age job marital education default balance housing loan contact day month duration campaign pdays previous poutcome deposit



5734	47	761	11	80	2	-1	0	0	0	0	...	1	0	0	0
5191	28	159	16	449	2	33	4	0	0	0	...	0	0	0	0
5390	35	1144	20	197	13	-1	0	0	0	0	...	0	0	0	0
850	51	746	25	372	5	-1	0	0	0	0	...	0	0	0	0
7270	30	2	23	269	1	-1	0	0	0	0	...	0	0	1	0

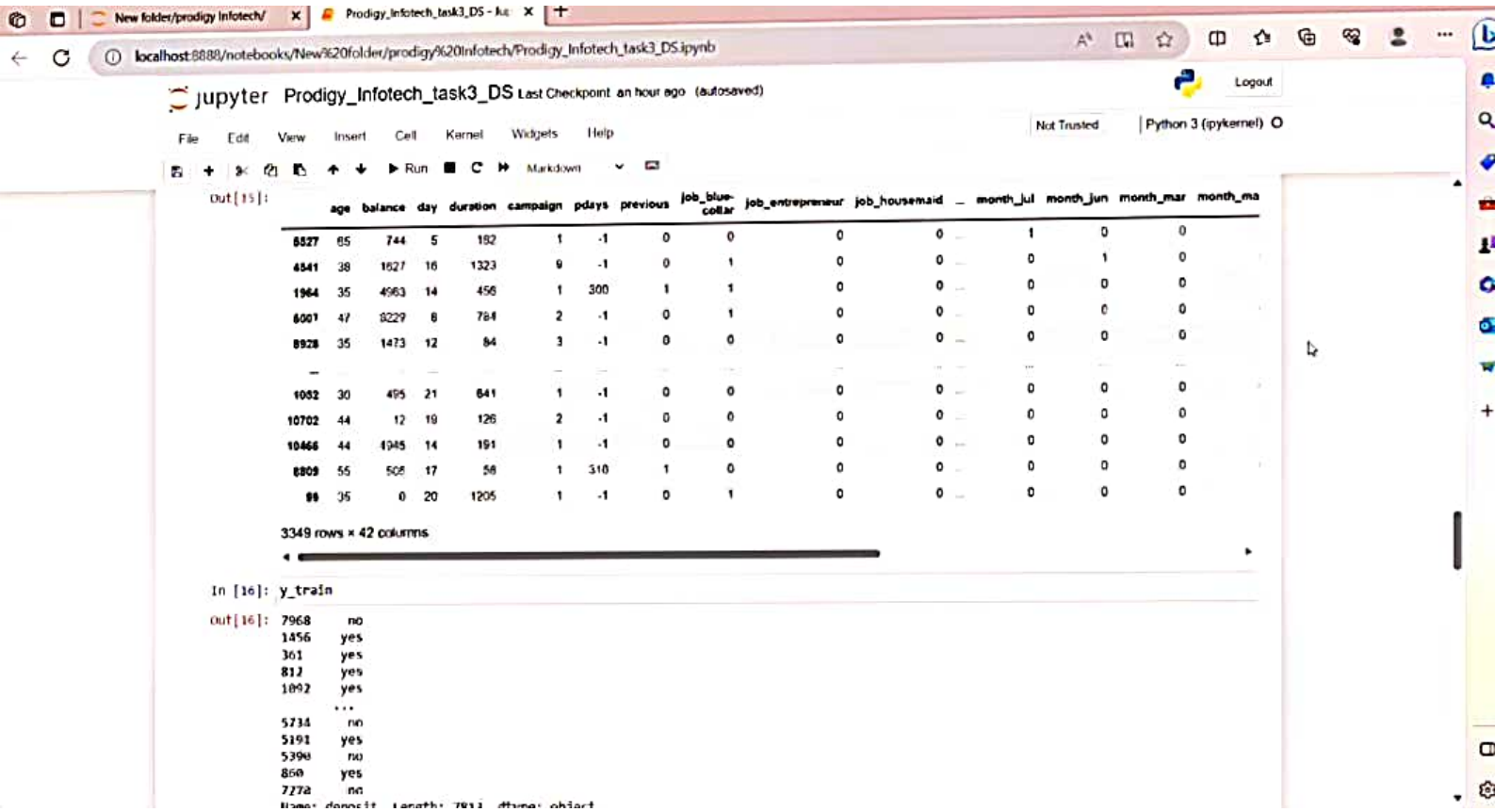
7813 rows × 42 columns

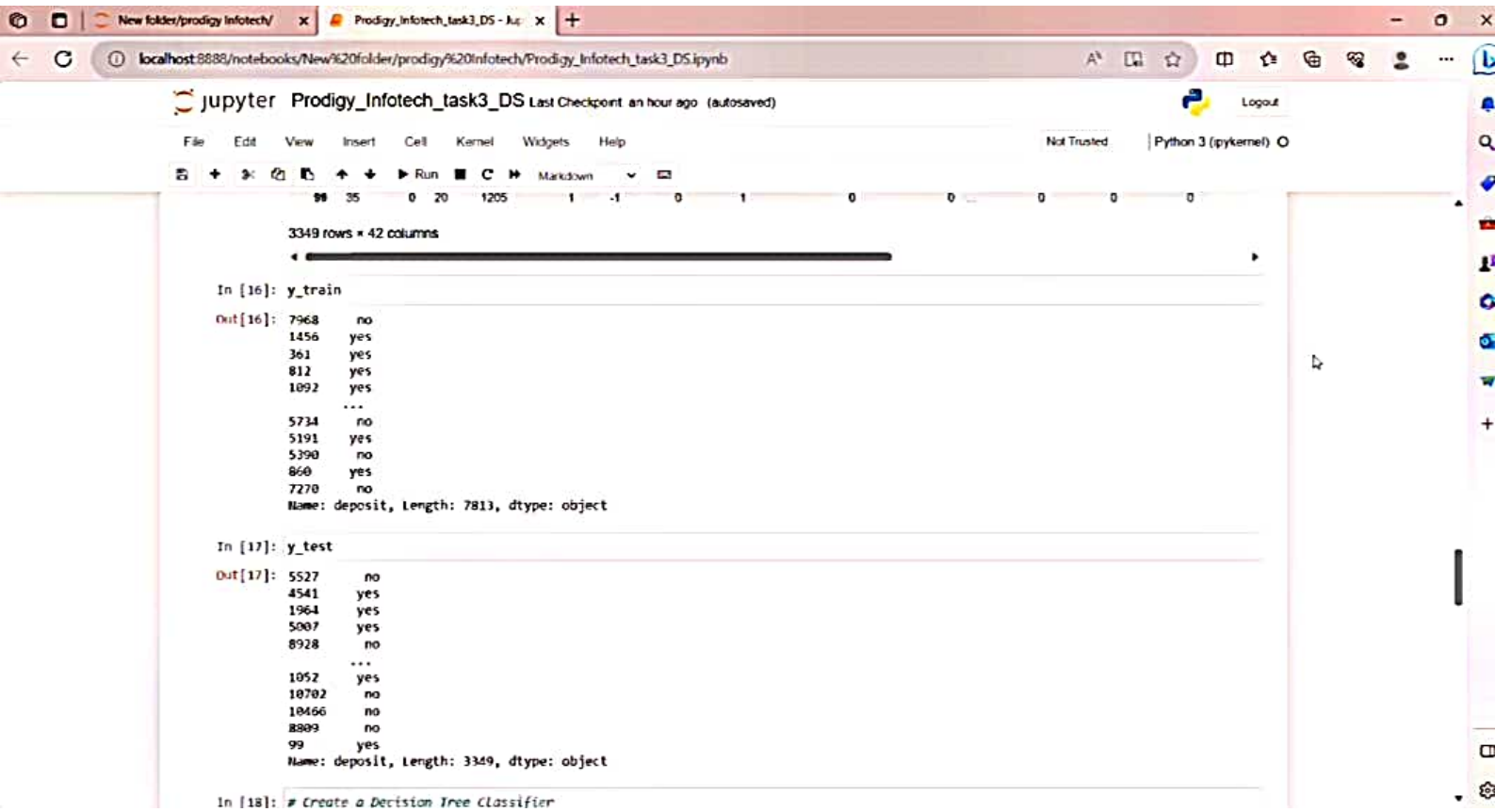
In [15]: `X_test`

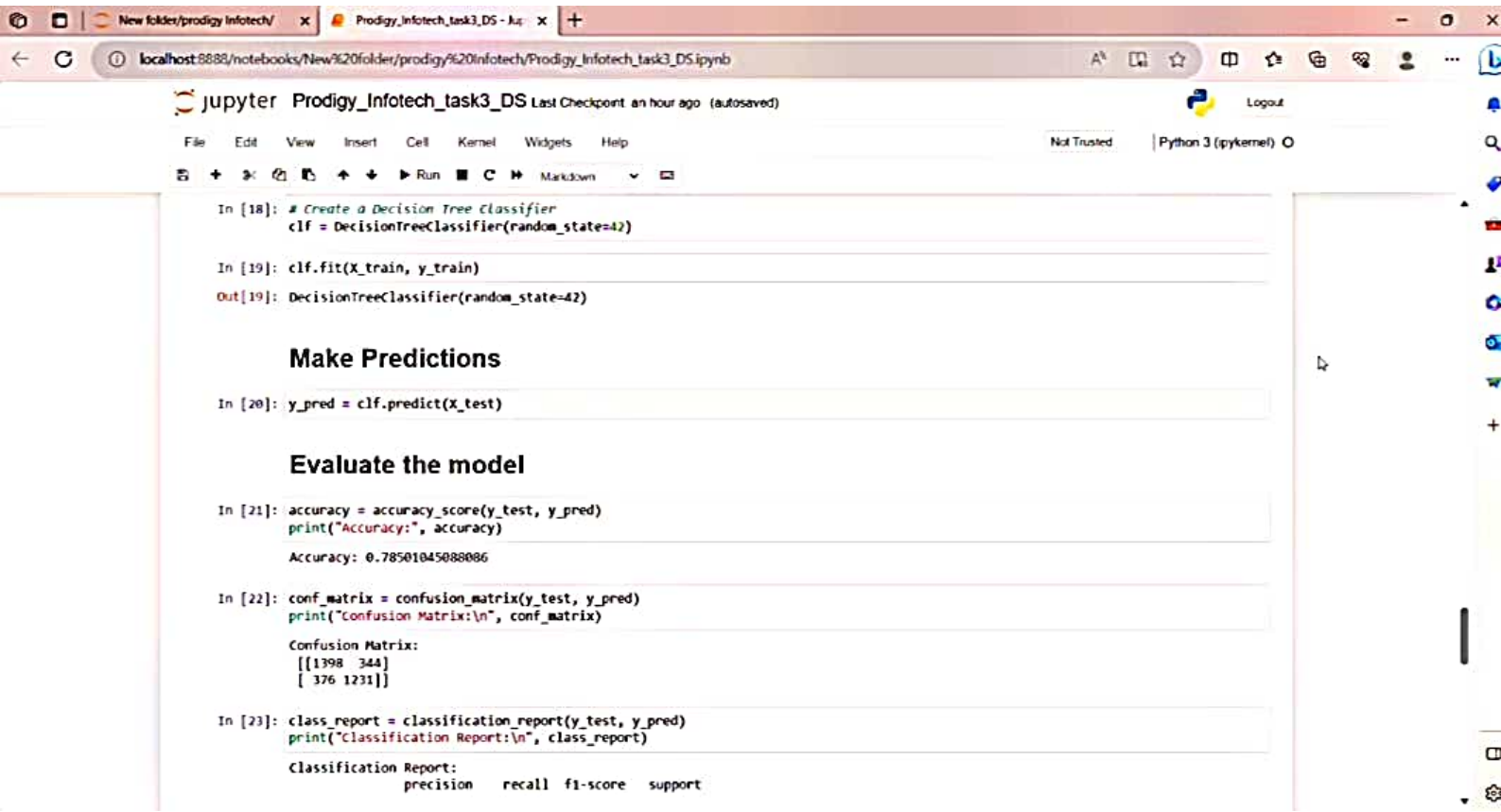
Out[15]:

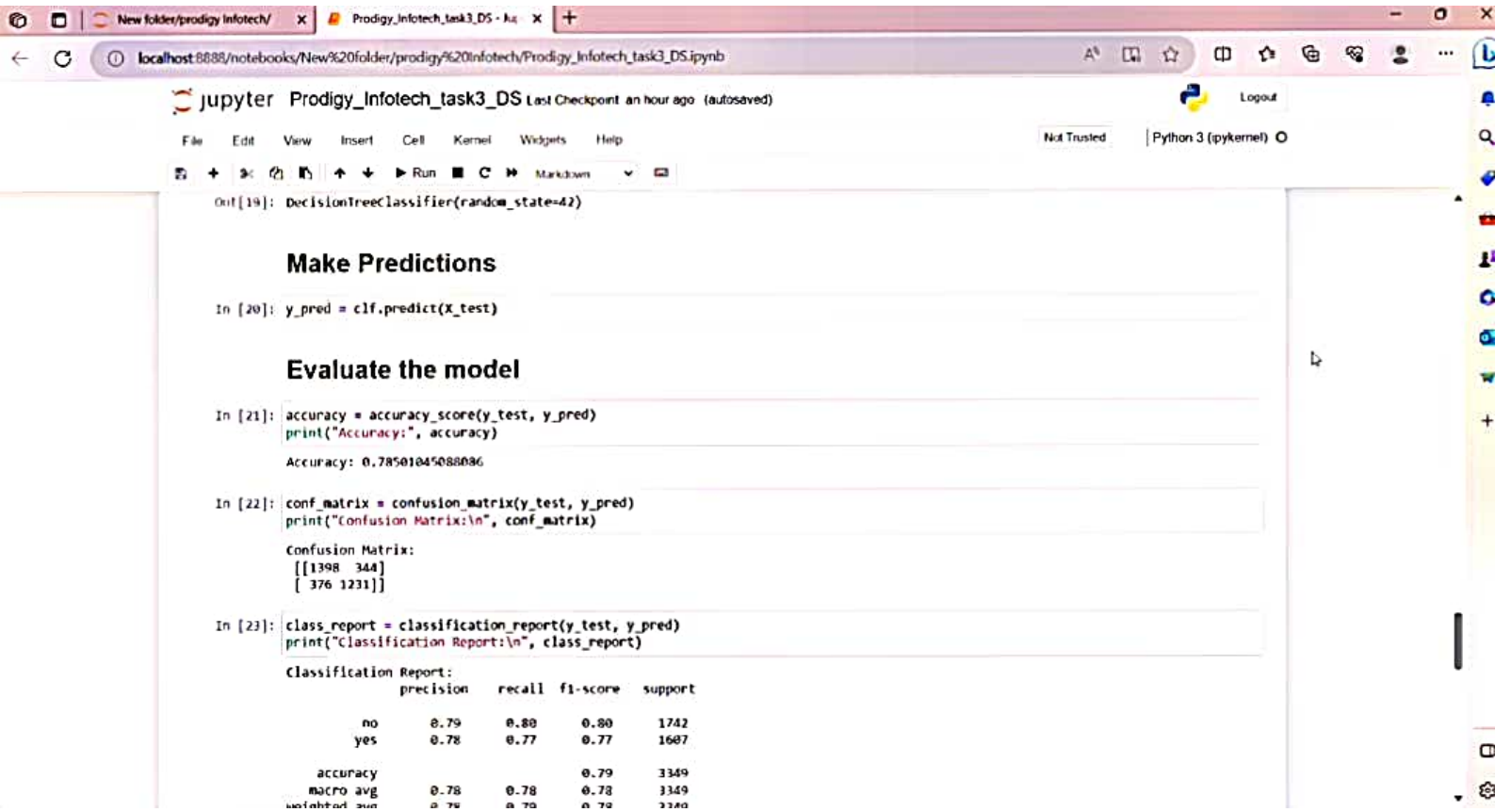
	age	balance	day	duration	campaign	pdays	previous	job_blue-collar	job_entrepreneur	job_housemaid	...	month_jul	month_jun	month_mar	month_ma
5527	65	744	5	182	1	-1	0	0	0	0	...	1	0	0	
4541	38	1627	16	1323	9	-1	0	1	0	0	...	0	1	0	
1964	35	4963	14	456	1	300	1	1	0	0	...	0	0	0	
5007	47	8229	8	734	2	-1	0	1	0	0	...	0	0	0	
8928	35	1473	12	84	3	-1	0	0	0	0	...	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
1052	30	495	21	641	1	-1	0	0	0	0	...	0	0	0	
10702	44	12	19	126	2	-1	0	0	0	0	...	0	0	0	
10468	44	1945	14	191	1	-1	0	0	0	0	...	0	0	0	
8809	55	505	17	58	1	315	1	0	0	0	...	0	0	0	
98	35	0	20	1205	1	-1	0	1	0	0	...	0	0	0	

3349 rows × 42 columns


New folder/prodigy Infotech/
Prodigy_Infotech_task3_DS - Ju
localhost:8888/notebooks/New%20folder/prodigy%20Infotech/Prodigy_Infotech_task3_DS.ipynb
jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Markdown
Out[15]:
age balance day duration campaign pdays previous job_blue-collar job_entrepreneur job_housemaid ... month_jul month_jun month_mar month_ma
8527 65 744 5 152 1 -1 0 0 0 0 ... 1 0 0
4541 38 1627 16 1323 9 -1 0 1 0 0 ... 0 1 0
1964 35 4963 14 456 1 300 1 1 0 0 ... 0 0 0
6001 47 9229 8 784 2 -1 0 1 0 0 ... 0 0 0
8928 35 1473 12 84 3 -1 0 0 0 0 ... 0 0 0
1052 30 495 21 641 1 -1 0 0 0 0 ... 0 0 0
10702 44 12 19 126 2 -1 0 0 0 0 ... 0 0 0
10466 44 1945 14 191 1 -1 0 0 0 0 ... 0 0 0
8809 55 505 17 58 1 310 1 0 0 0 ... 0 0 0
99 35 0 20 1205 1 -1 0 1 0 0 ... 0 0 0
3349 rows × 42 columns
In [16]: y_train
Out[16]: 7968 no
1456 yes
361 yes
812 yes
1092 yes
5734 no
5191 yes
5398 no
860 yes
7272 no


New folder/prodigy infotech/
Prodigy_Infotech_task3_DS - Ju
localhost:8888/notebooks/New%20folder/prodigy%20infotech/Prodigy_Infotech_task3_DS.ipynb
jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Markdown
98 35 0 20 1205 1 -1 0 1 0 0 0 0 0
3349 rows × 42 columns
In [16]: y_train
Out[16]: 7968 no
1456 yes
361 yes
812 yes
1092 yes
...
5734 no
5191 yes
5390 no
860 yes
7270 no
Name: deposit, Length: 7813, dtype: object
In [17]: y_test
Out[17]: 5527 no
4541 yes
1964 yes
5007 yes
8928 no
...
1052 yes
10702 no
10466 no
8809 no
99 yes
Name: deposit, Length: 3349, dtype: object
In [18]: # Create a Decision Tree Classifier


New folder/prodigy infotech/
Prodigy_Infotech_task3_DS - Ju
localhost:8888/notebooks/New%20folder/prodigy%20infotech/Prodigy_Infotech_task3_DS.ipynb
jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Markdown
In [18]: # Create a Decision Tree Classifier
clf = DecisionTreeClassifier(random_state=42)
In [19]: clf.fit(X_train, y_train)
Out[19]: DecisionTreeClassifier(random_state=42)
Make Predictions
In [20]: y_pred = clf.predict(X_test)
Evaluate the model
In [21]: accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
Accuracy: 0.78501045088086
In [22]: conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", conf_matrix)
Confusion Matrix:
[[1398 344]
[376 1231]]
In [23]: class_report = classification_report(y_test, y_pred)
print("Classification Report:\n", class_report)
Classification Report:
precision recall f1-score support


New folder/prodigy Infotech/
Prodigy_Infotech_task3_DS - Ju
localhost:8888/notebooks/New%20folder/prodigy%20Infotech/Prodigy_Infotech_task3_DS.ipynb
jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Markdown
Out[19]: DecisionTreeClassifier(random_state=42)
Make Predictions
In [20]: y_pred = clf.predict(X_test)
Evaluate the model
In [21]: accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
Accuracy: 0.7850104508806
In [22]: conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", conf_matrix)
Confusion Matrix:
[[1398 344]
[376 1231]]
In [23]: class_report = classification_report(y_test, y_pred)
print("Classification Report:\n", class_report)
Classification Report:
precision recall f1-score support
no 0.79 0.80 0.80 1742
yes 0.78 0.77 0.77 1607
accuracy 0.79 3349
macro avg 0.78 0.78 0.78 3349

localhost:8888/notebooks/New%20folder/prodigy%20Infotech/Prodigy_Infotech_task3_DS.ipynb

jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted | Python 3 (ipykernel)

Run Markdown

In [22]:
```
conf_matrix = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", conf_matrix)
```

```
Confusion Matrix:
 [[1398  344]
 [ 376 1231]]
```

In [23]:
```
class_report = classification_report(y_test, y_pred)
print("Classification Report:\n", class_report)
```

```
Classification Report:
              precision    recall  f1-score   support

          no       0.79      0.80      0.80      1742
         yes       0.78      0.77      0.77      1607

    accuracy                           0.79      3349
   macro avg       0.78      0.78      0.78      3349
weighted avg       0.78      0.79      0.78      3349
```

In [24]:
```
confusion = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(confusion, annot=True, fmt="d", cmap="Oranges")
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

New folder/prodigy infotech/ | Prodigy_Infotech_task3_DS - Ju | localhost:8888/notebooks/New%20folder/prodigy%20Infotech/Prodigy_Infotech_task3_DS.ipynb

jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

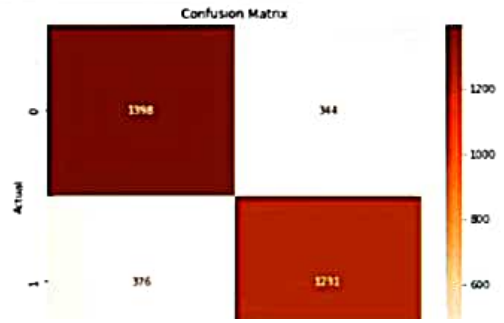
Run Markdown

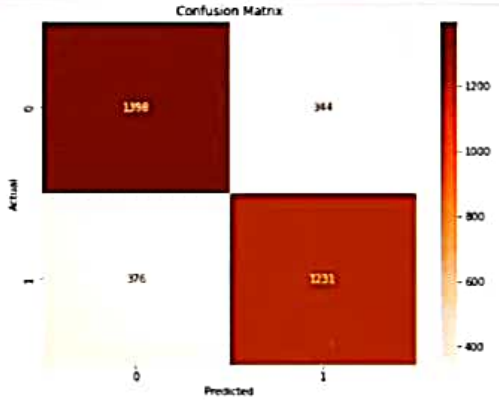
```
Classification Report:
              precision    recall  f1-score   support

          no       0.79      0.80      0.80      1742
         yes       0.78      0.77      0.77      1607

    accuracy                           0.79      3349
   macro avg       0.78      0.78      0.78      3349
weighted avg       0.78      0.79      0.78      3349
```

In [24]:
```
confusion = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(8, 6))
sns.heatmap(confusion, annot=True, fmt="d", cmap="Oranges")
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```

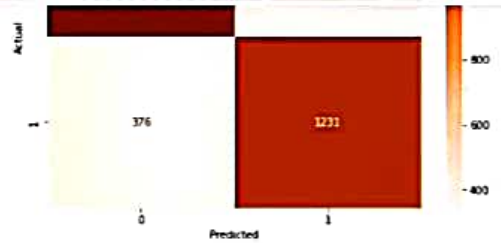




Visualize the tree structure

In [25]:
```
tree_structure = export_text(clf, feature_names=list(X.columns))
print("Decision Tree Structure:")
print(tree_structure)
```

```
Decision Tree Structure:
|--- duration <= 251.50
|   |--- poutcome_success <= 0.50
|   |   |--- duration <= 123.50
|   |   |   |--- month_mar <= 0.50
```



Visualize the tree structure

In [25]:
```
tree_structure = export_text(clf, feature_names=list(X.columns))
print("Decision Tree Structure:")
print(tree_structure)
```

```
Decision Tree Structure:
|--- duration <= 251.50
|   |--- poutcome_success <= 0.50
|   |   |--- duration <= 123.50
|   |   |   |--- month_mar <= 0.50
|   |   |   |   |--- month_oct <= 0.50
|   |   |   |   |   |--- month_feb <= 0.50
|   |   |   |   |   |   |--- job_student <= 0.50
|   |   |   |   |   |   |   |--- day <= 1.50
|   |   |   |   |   |   |   |   |--- age <= 36.50
|   |   |   |   |   |   |   |   |   |--- class: no
|   |   |   |   |   |   |   |   |--- age >  36.50
|   |   |   |   |   |   |   |   |   |--- class: yes
|   |   |   |   |   |   |   |--- day >  1.50
```

jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

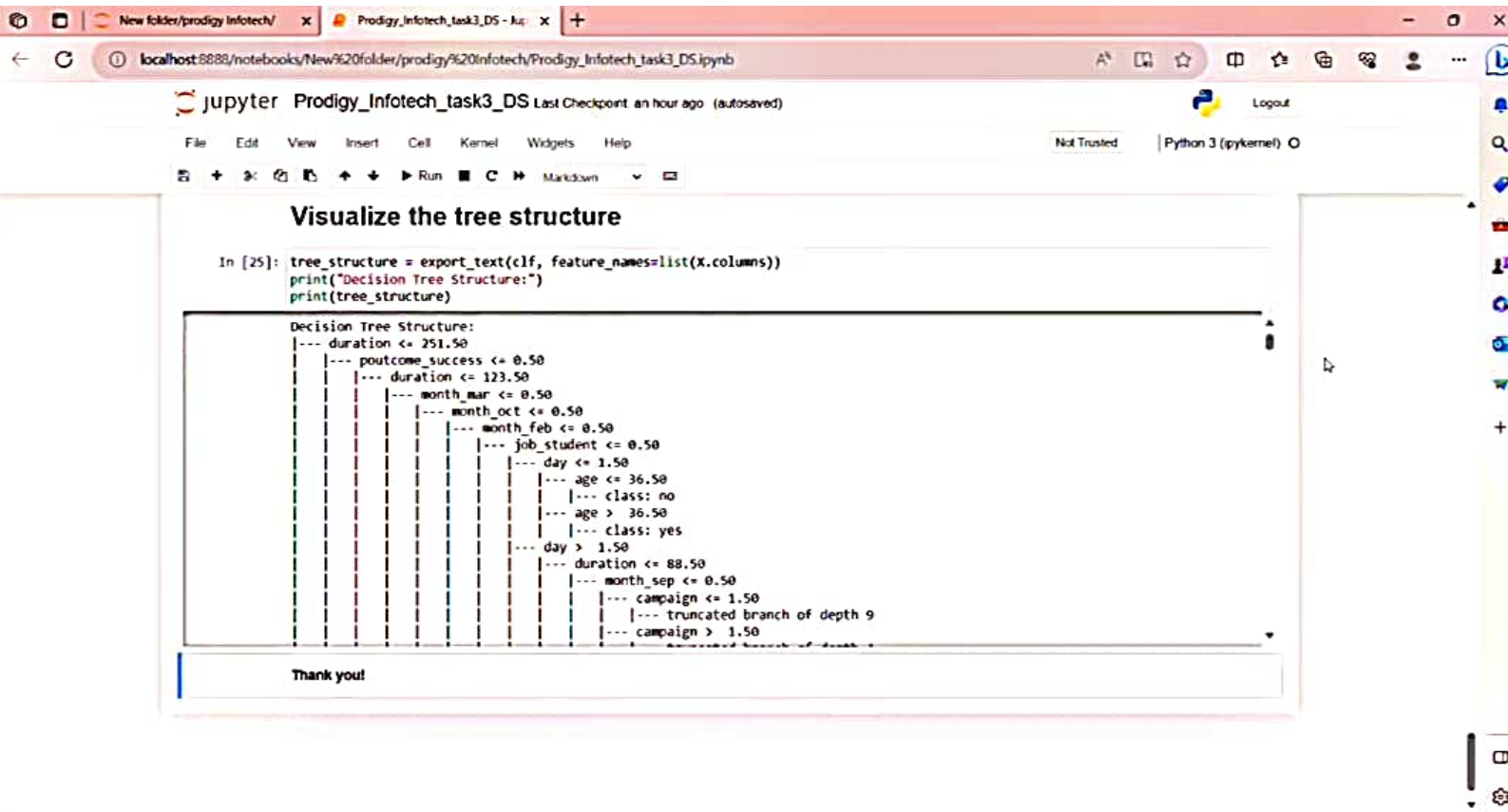
Run Markdown

Visualize the tree structure

In [25]:
```
tree_structure = export_text(clf, feature_names=list(X.columns))
print("Decision Tree Structure:")
print(tree_structure)
```

```
Decision Tree Structure:
|--- duration <= 251.50
|   |--- poutcome_success <= 0.50
|   |   |--- duration <= 123.50
|   |   |   |--- month_mar <= 0.50
|   |   |   |   |--- month_oct <= 0.50
|   |   |   |   |   |--- month_feb <= 0.50
|   |   |   |   |   |   |--- job_student <= 0.50
|   |   |   |   |   |   |   |--- day <= 1.50
|   |   |   |   |   |   |   |   |--- age <= 36.50
|   |   |   |   |   |   |   |   |   |--- class: no
|   |   |   |   |   |   |   |   |--- age >  36.50
|   |   |   |   |   |   |   |   |   |--- class: yes
|   |   |   |   |   |   |   |--- day >  1.50
|   |   |   |   |   |   |   |   |--- duration <= 88.50
|   |   |   |   |   |   |   |   |   |--- month_sep <= 0.50
|   |   |   |   |   |   |   |   |   |   |--- campaign <= 1.50
|   |   |   |   |   |   |   |   |   |   |   |--- truncated branch of depth 9
|   |   |   |   |   |   |   |   |   |   |--- campaign >  1.50
```

Thank you!


New folder/prodigy infotech/
Prodigy_Infotech_task3_DS - Ju
localhost:8888/notebooks/New%20folder/prodigy%20infotech/Prodigy_Infotech_task3_DS.ipynb
jupyter Prodigy_Infotech_task3_DS Last Checkpoint an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Markdown
Visualize the tree structure
In [25]:
tree_structure = export_text(clf, feature_names=list(X.columns))
print("Decision Tree Structure:")
print(tree_structure)
Decision Tree Structure:
|--- duration <= 251.50
| |--- poutcome_success <= 0.50
| | |--- duration <= 123.50
| | | |--- month_mar <= 0.50
| | | | |--- month_oct <= 0.50
| | | | | |--- month_feb <= 0.50
| | | | | | |--- job_student <= 0.50
| | | | | | | |--- day <= 1.50
| | | | | | | | |--- age <= 36.50
| | | | | | | | | |--- class: no
| | | | | | | | |--- age > 36.50
| | | | | | | | | |--- class: yes
| | | | | | | |--- day > 1.50
| | | | | | | | |--- duration <= 88.50
| | | | | | | | | |--- month_sep <= 0.50
| | | | | | | | | | |--- campaign <= 1.50
| | | | | | | | | | | |--- truncated branch of depth 9
| | | | | | | | | | |--- campaign > 1.50
Thank you!