Lembar Jawaban UTS
Machine Learning

Nama : **AWANDA BINTANG PAMUNGKAS**

NIM : **5210411366**

Kelas : **C**

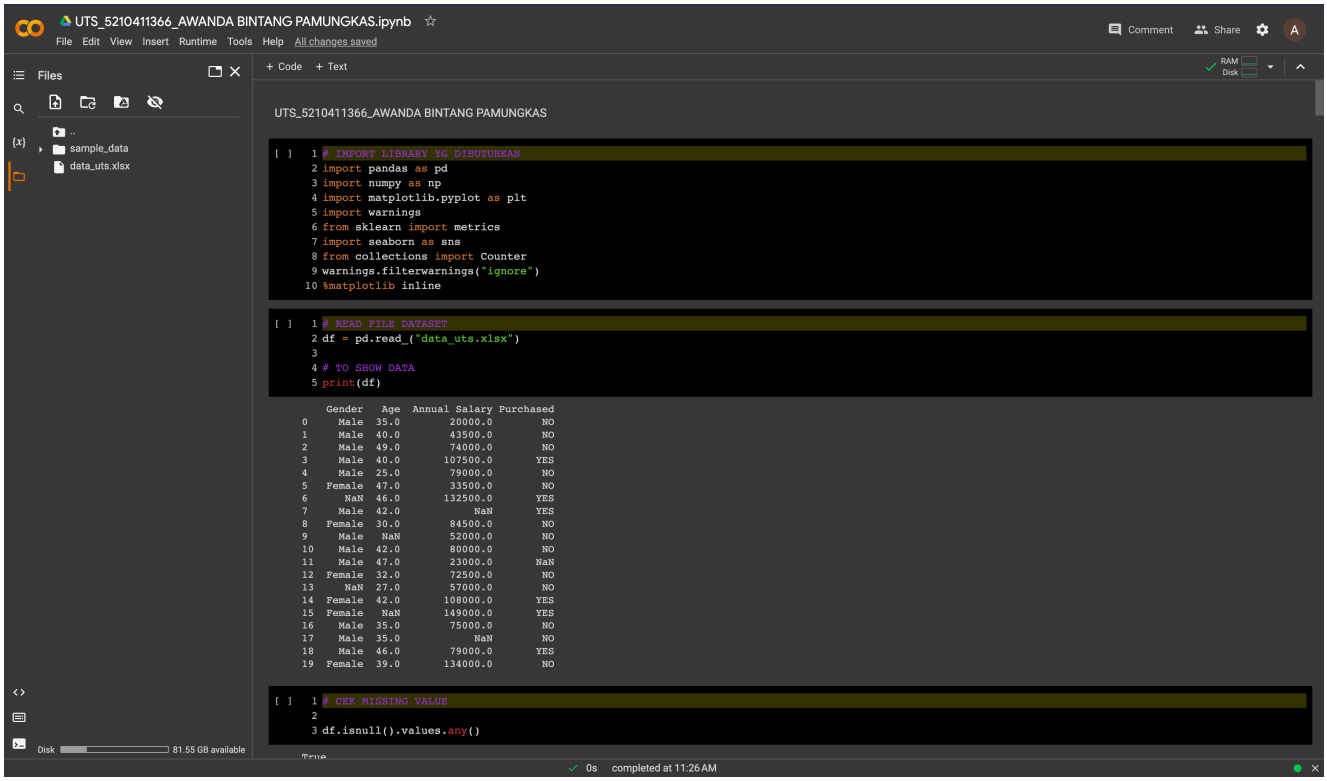
Pernyataan:

Dengan mengerjakan ujian ini, maka saya menyatakan bahwa semua jawaban **SAYA KERJAKAN SENDIRI** tanpa bekerjasama maupun meminta bantuan siapapun. Saya bersedia diberi **nilai E** jika pernyataan ini terbukti salah.

Tanda tangan:

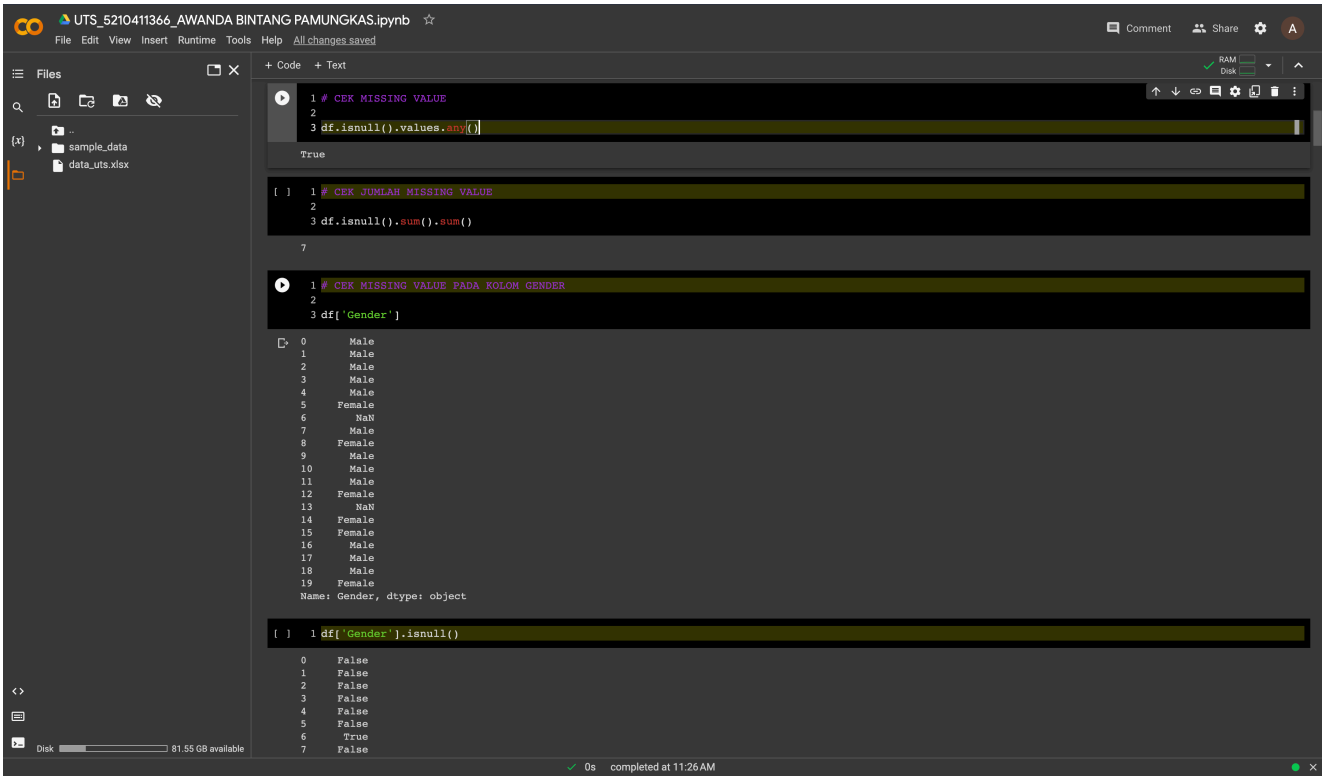
---

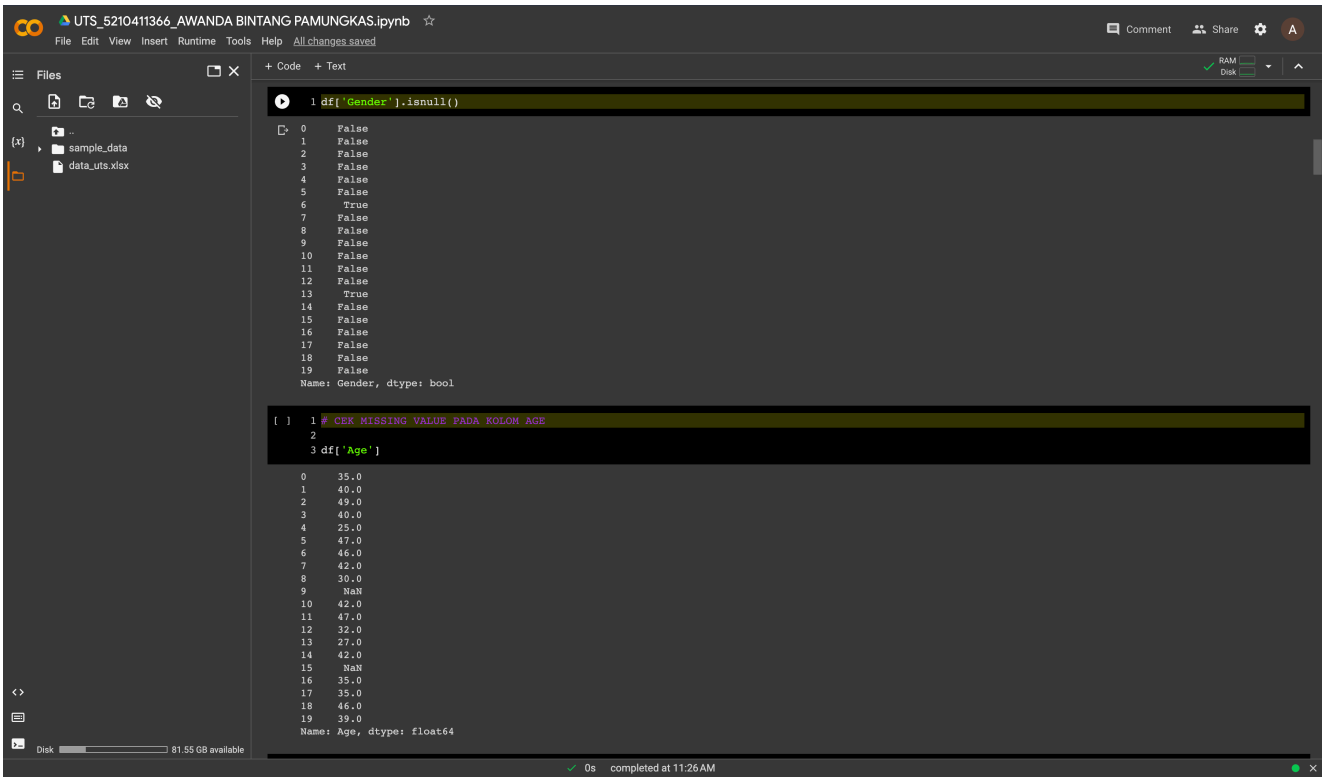
1. Convert data dari table menjadi bentuk data tabular yaitu excel atau CSV



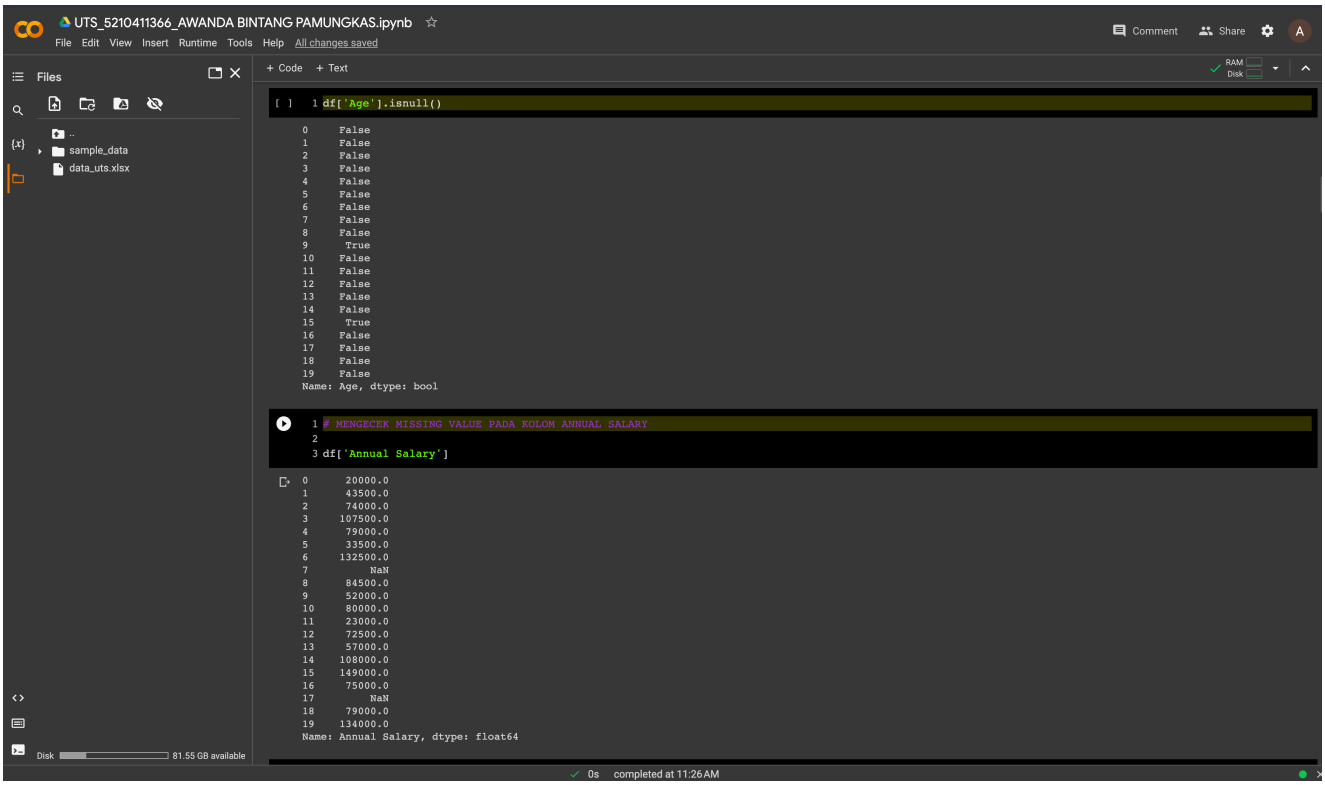
2. Perbaiki data yg kosong atau missing pada setiap kolom

MISSING VALUE PADA KOLOM “GENDER”

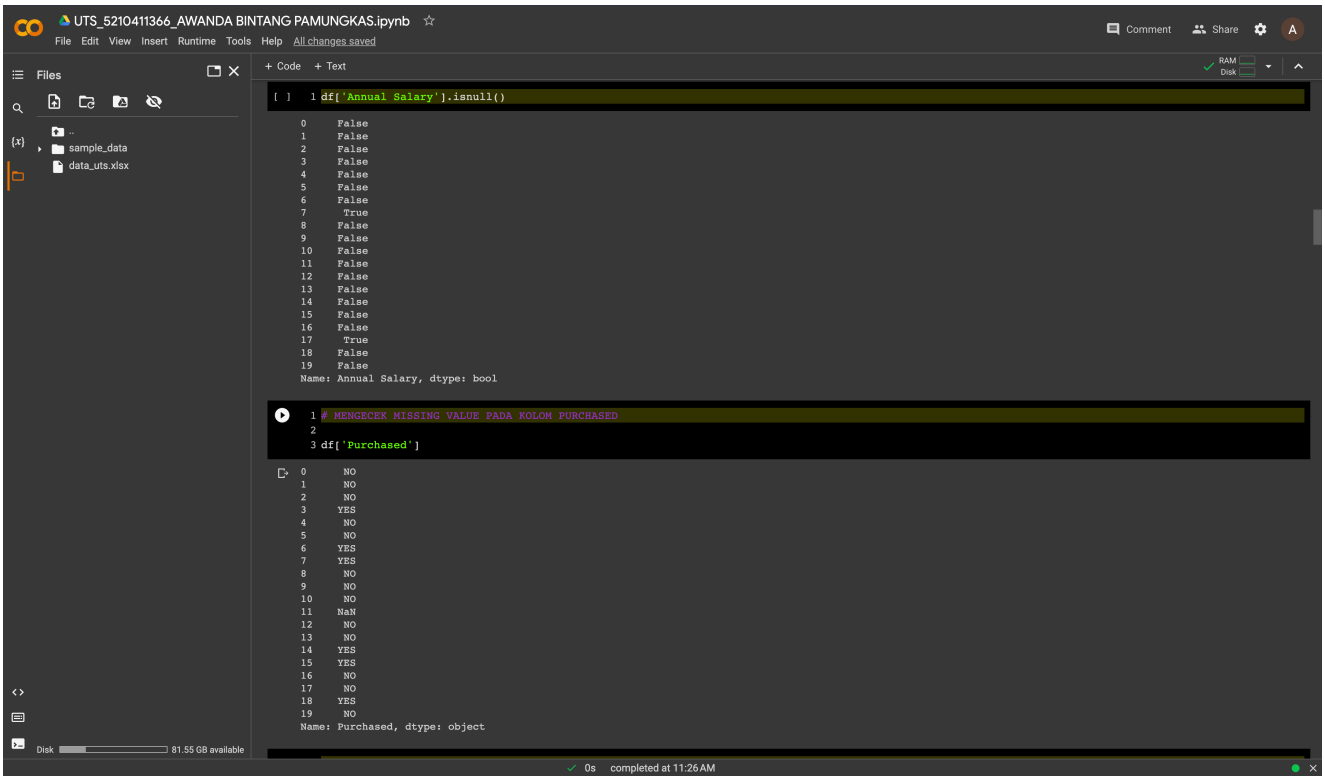




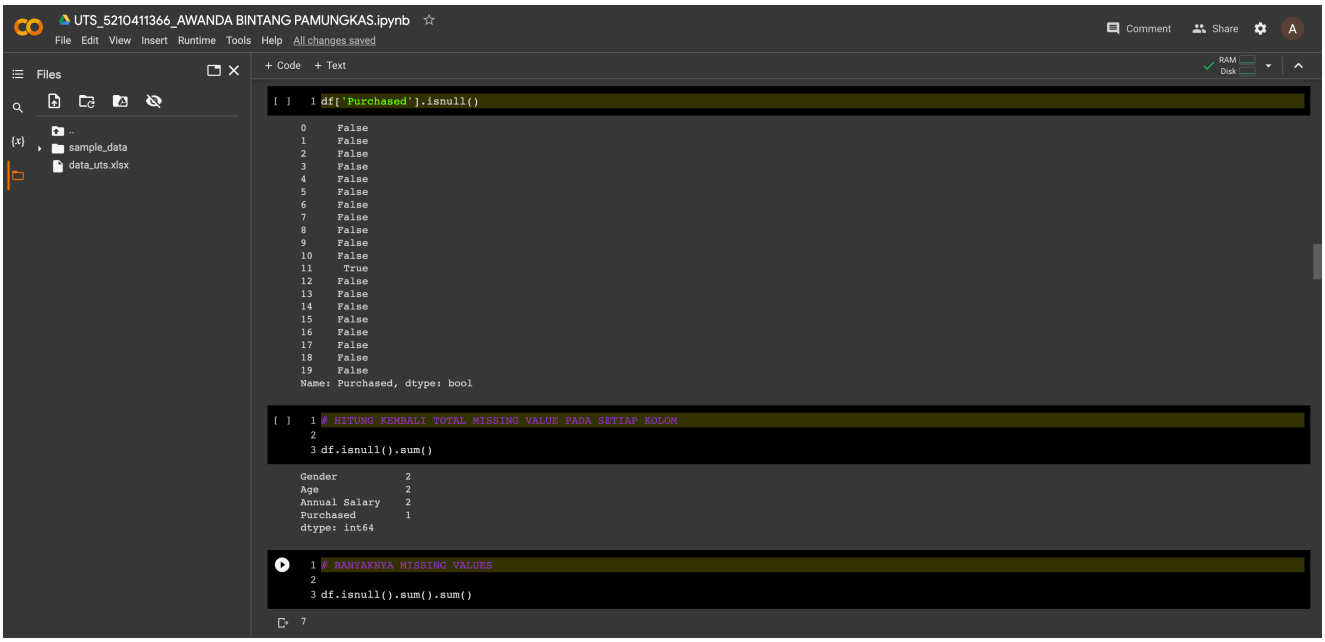
MISSING VALUE PADA KOLOM “AGE”



MISSING VALUE PADA KOLOM “ANNUAL SALARY”

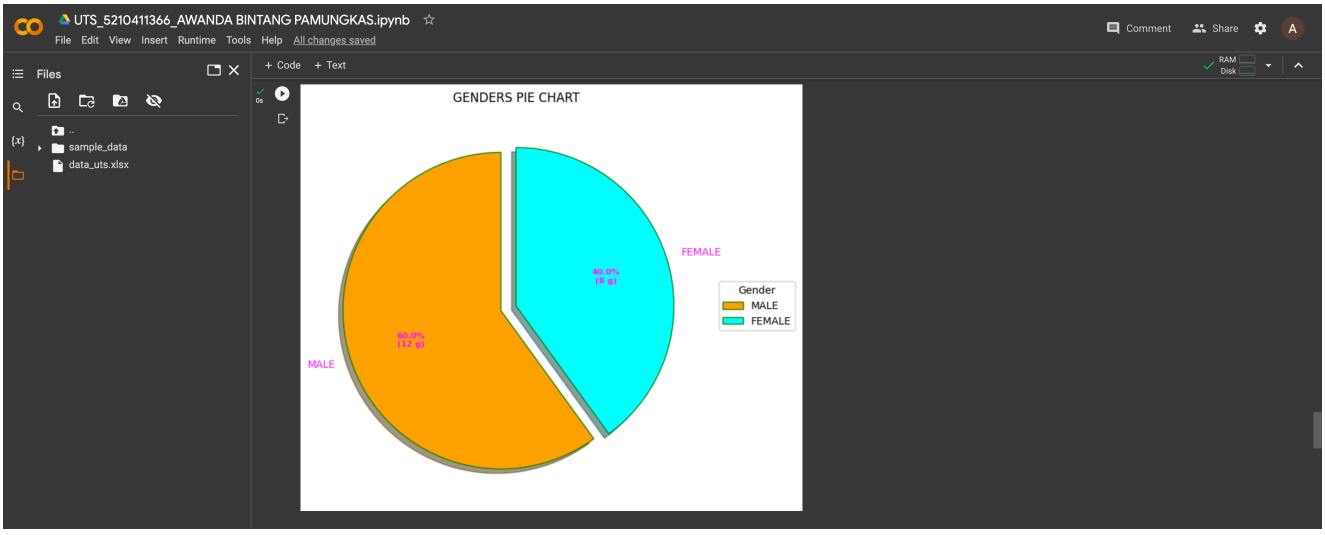


MISSING VALUE PADA KOLOM “PURCHASED”



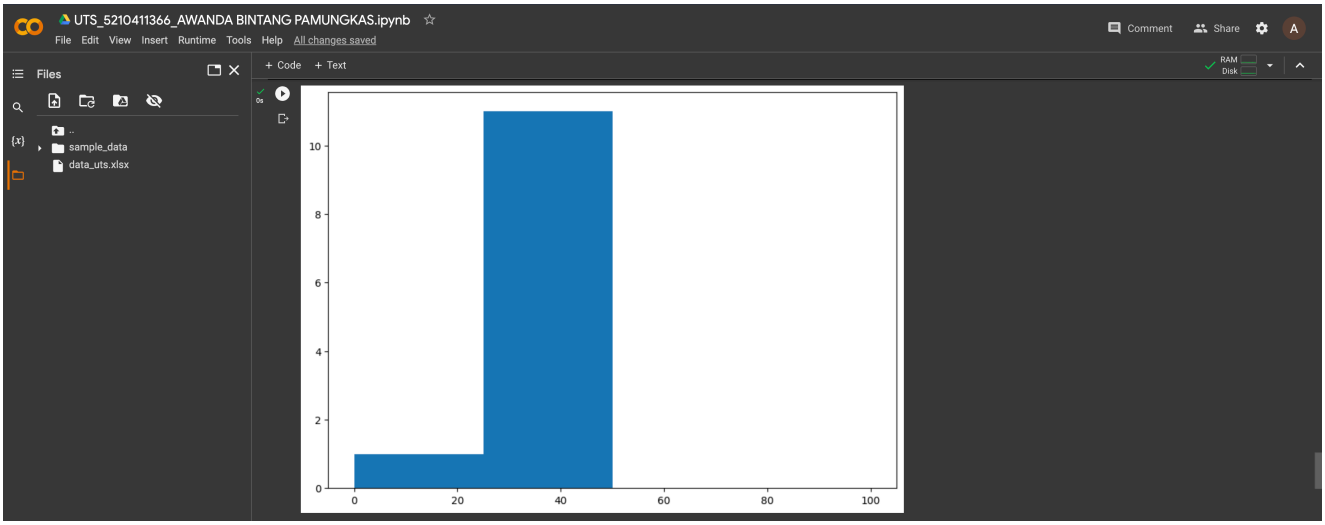
3. Visualisasikan setiap kolom dengan diagram yang cocok dan berikan alasan kenapa diagram itu cocok dengan kolom yang divisualisasikan

Genders



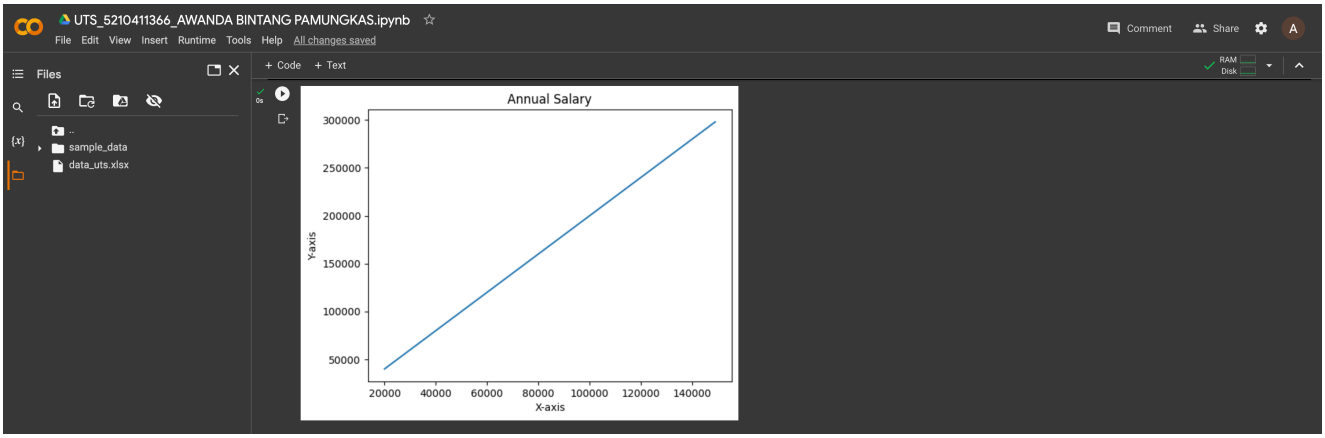
PENJELASAN : Kolom "Gender": Diagram pie atau bar chart dapat digunakan untuk memvisualisasikan proporsi atau jumlah individu berdasarkan jenis kelamin. Pie chart cocok jika hanya ada dua jenis kelamin yang terlibat, sementara bar chart lebih cocok jika lebih dari dua jenis kelamin yang terlibat.

Age



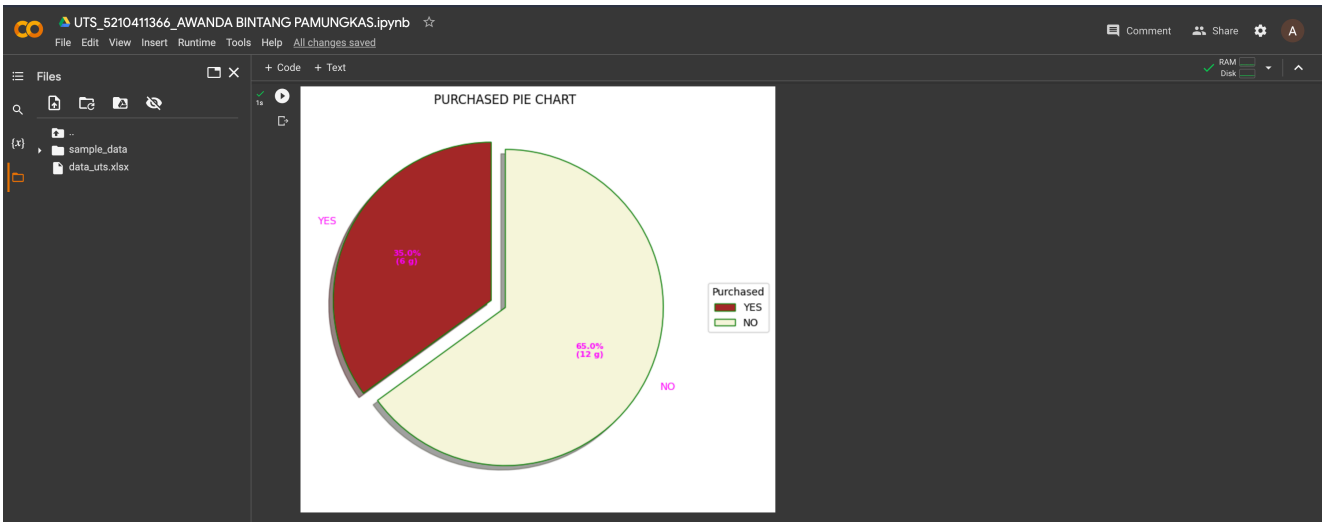
Kolom "Age": Histogram atau box plot dapat digunakan untuk memvisualisasikan distribusi usia dalam dataset. Histogram cocok untuk melihat frekuensi usia dalam interval tertentu, sementara box plot cocok untuk melihat kuartil, rentang, dan outlier dalam distribusi usia.

Annual Salary



Kolom "AnnualSalary": Box plot atau line chart dapat digunakan untuk memvisualisasikan distribusi atau tren gaji tahunan dalam dataset. Box plot cocok untuk melihat kuartil, rentang, dan outlier dalam distribusi gaji, sementara line chart cocok untuk melihat tren gaji dari waktu ke waktu.

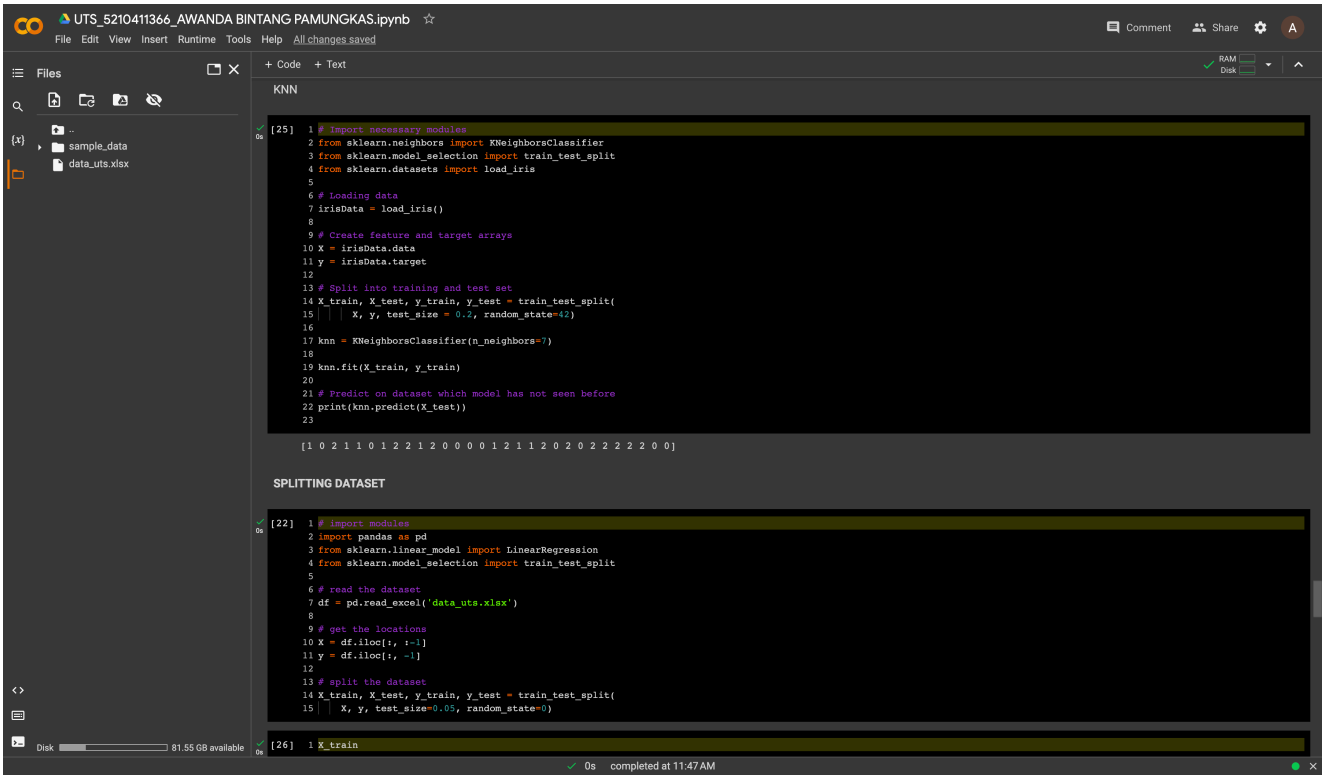
Purchased

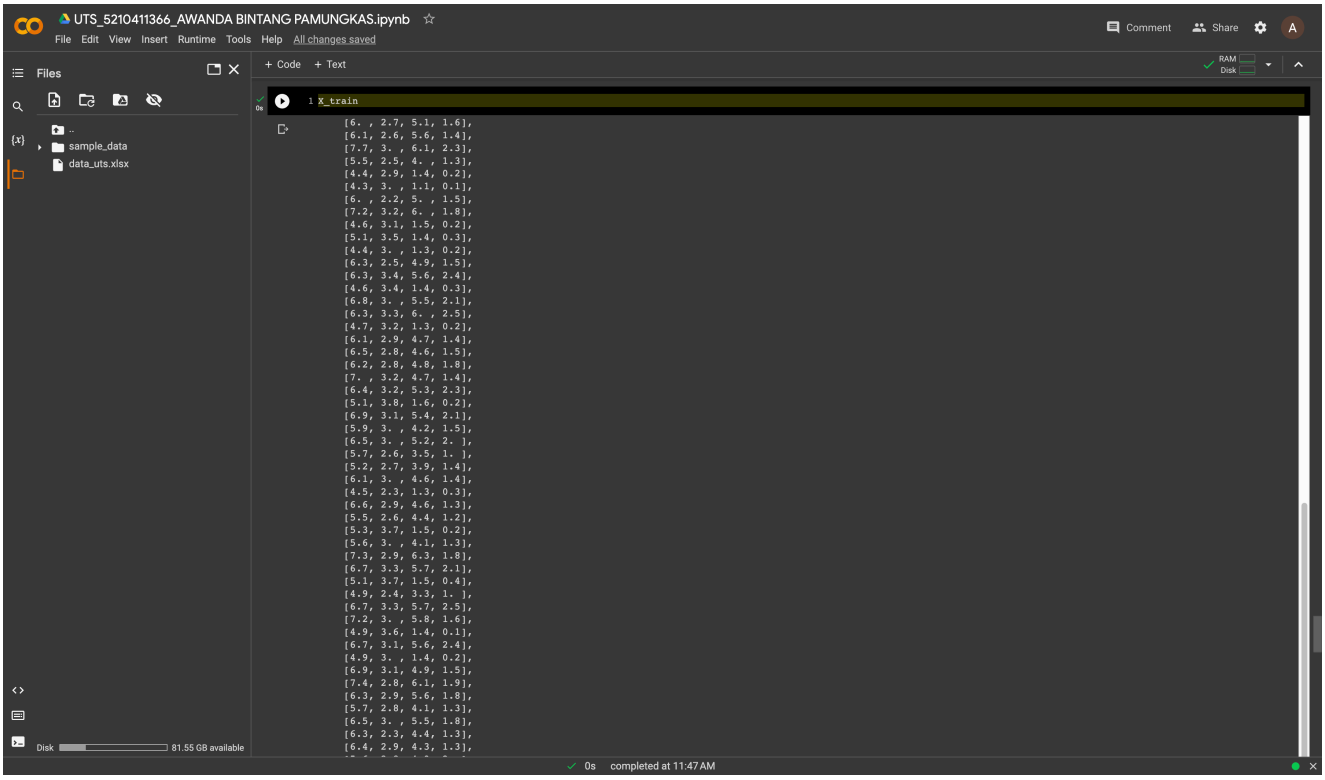


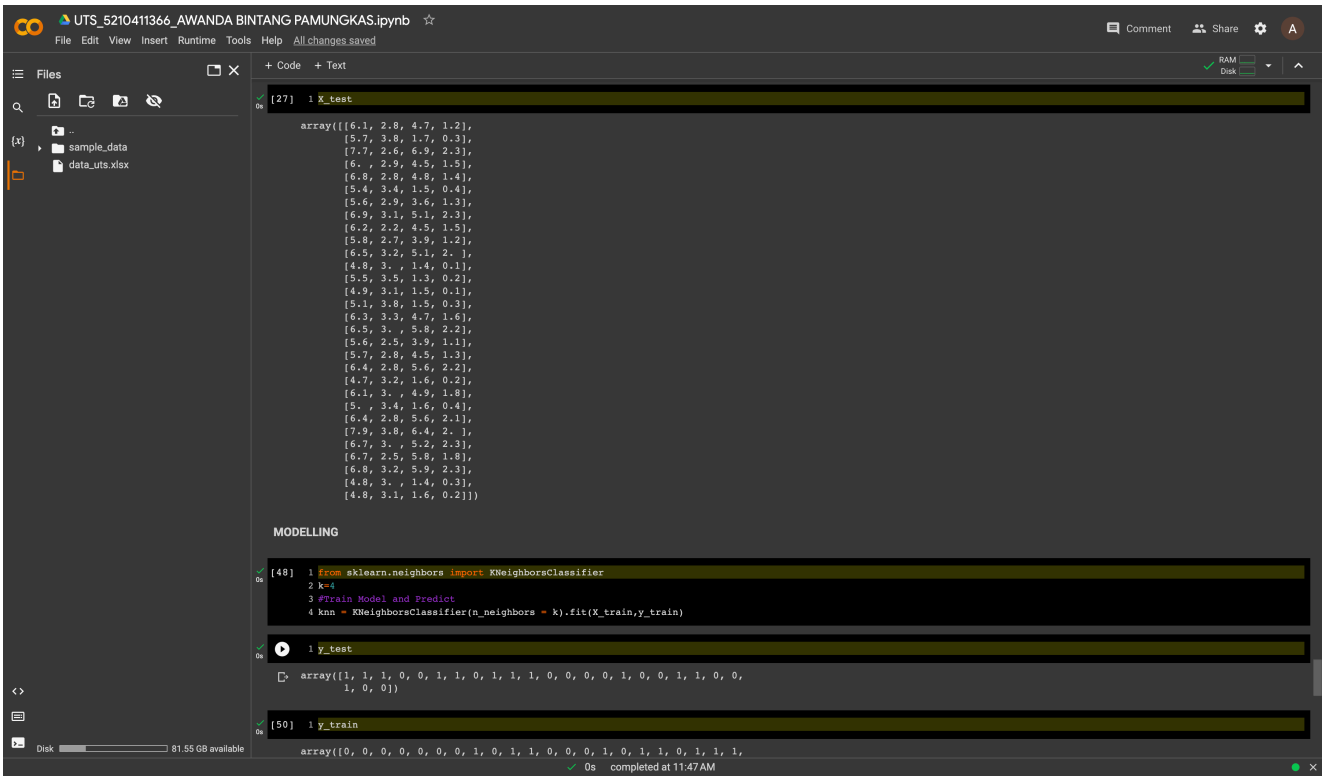
Kolom "Purchased": Bar chart atau pie chart dapat digunakan untuk memvisualisasikan jumlah pembelian atau proporsi individu yang melakukan pembelian. Bar chart cocok jika ingin mengetahui jumlah pembelian berdasarkan kategori tertentu, seperti jenis produk atau waktu, sementara pie chart cocok jika ingin mengetahui proporsi pembelian dalam dua kategori atau lebih.

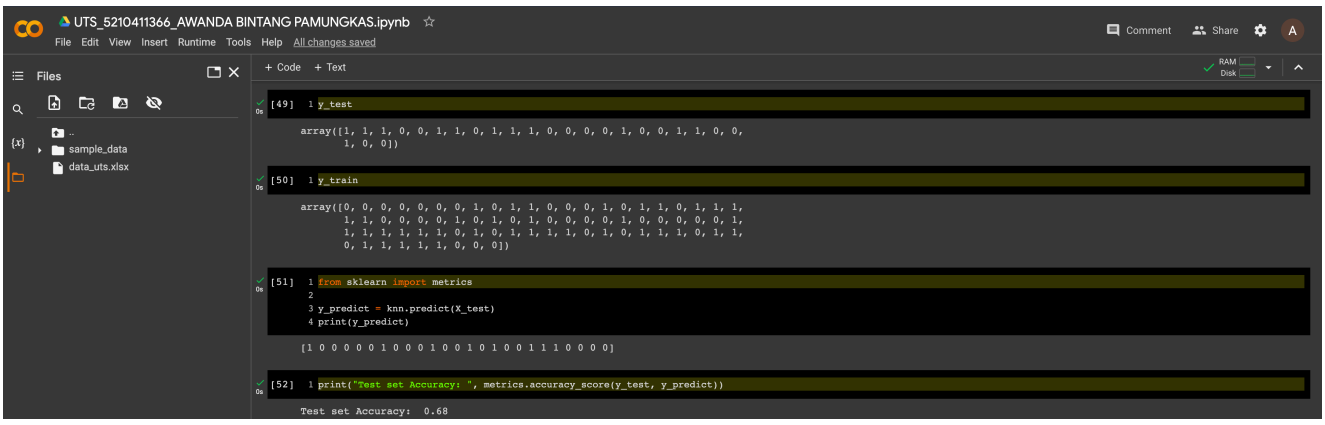
Lakukan proses yang direkomendasikan oleh peneliti tersebut agar data bisa digunakan. Setelah proses pengolahan data selesai. Kemudian peneliti ingin mengklasifikasikan data pembelian modil atau tidak dengan menggunakan metode K-NN. Adapun kriteria penyusunan metode K-NN antara lain:

1. Nilai K sebesar 4

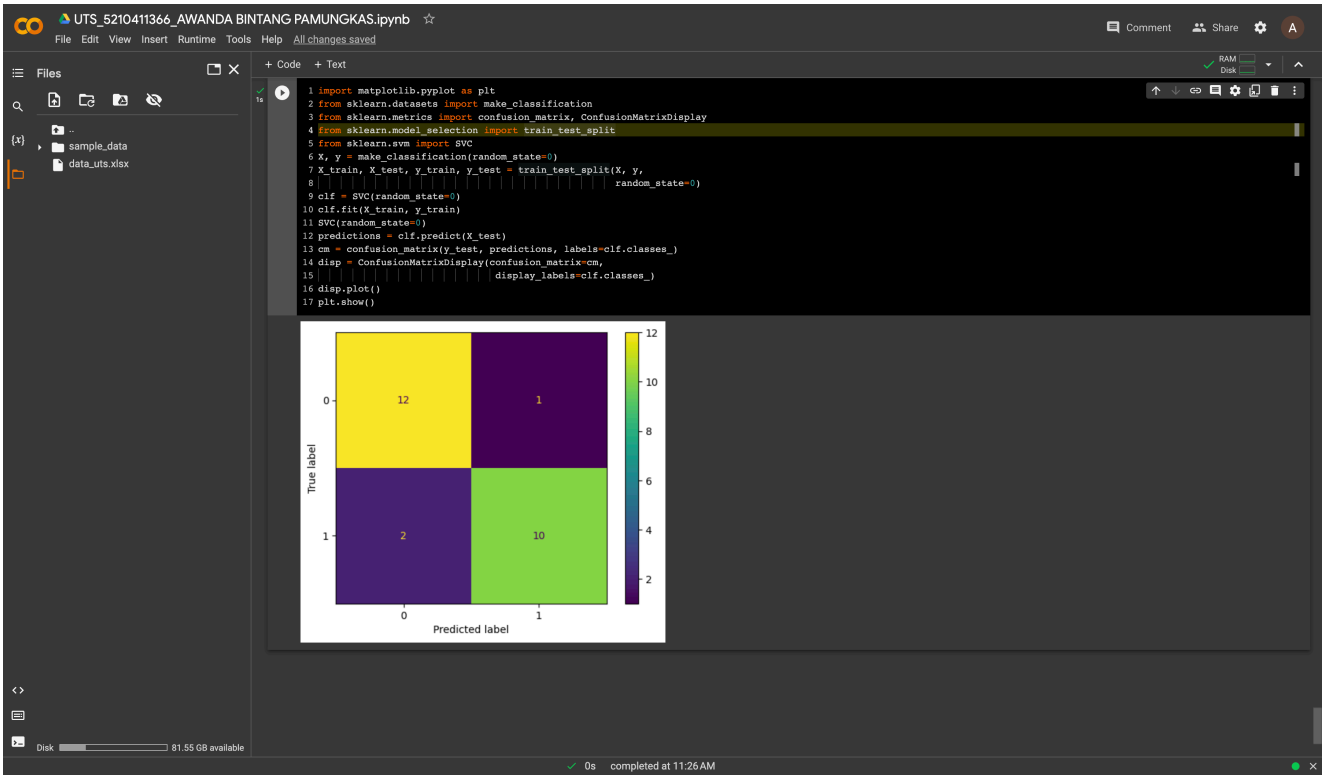



UTS_5210411366_AWANDA BINTANG PAMUNGKAS.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
sample_data
data_uts.xlsx
+ Code + Text
RAM Disk
1 X_train
[6. , 2.7, 5.1, 1.6],
[6.1, 2.6, 5.6, 1.4],
[7.7, 3. , 6.1, 2.3],
[5.5, 2.5, 4. , 1.3],
[4.4, 2.9, 1.4, 0.2],
[4.3, 3. , 1.1, 0.1],
[6. , 2.2, 5. , 1.5],
[7.2, 3.2, 6. , 1.8],
[4.6, 3.1, 1.5, 0.2],
[5.1, 3.5, 1.4, 0.3],
[4.4, 3. , 1.3, 0.2],
[6.3, 2.5, 4.9, 1.5],
[6.3, 3.4, 5.6, 2.4],
[4.6, 3.4, 1.4, 0.3],
[6.8, 3. , 5.5, 2.1],
[6.3, 3.3, 6. , 2.5],
[4.7, 3.2, 1.3, 0.2],
[6.1, 2.9, 4.7, 1.4],
[6.5, 2.8, 4.6, 1.5],
[6.2, 2.8, 4.8, 1.8],
[7. , 3.2, 4.7, 1.4],
[6.4, 3.2, 5.3, 2.3],
[5.1, 3.8, 1.6, 0.2],
[6.9, 3.1, 5.4, 2.1],
[5.9, 3. , 4.2, 1.5],
[6.5, 3. , 5.2, 2.],
[5.7, 2.6, 3.5, 1.],
[5.2, 2.7, 3.9, 1.4],
[6.1, 3. , 4.6, 1.4],
[4.5, 2.3, 1.3, 0.3],
[6.6, 2.9, 4.6, 1.3],
[5.5, 2.6, 4.4, 1.2],
[5.3, 3.7, 1.5, 0.2],
[5.6, 3. , 4.1, 1.3],
[7.3, 2.9, 6.3, 1.8],
[6.7, 3.3, 5.7, 2.1],
[5.1, 3.7, 1.5, 0.4],
[4.9, 2.4, 3.3, 1.],
[6.7, 3.3, 5.7, 2.5],
[7.2, 3. , 5.8, 1.6],
[4.9, 3.6, 1.4, 0.1],
[6.7, 3.1, 5.6, 2.4],
[4.9, 3. , 1.4, 0.2],
[6.9, 3.1, 4.9, 1.5],
[7.4, 2.8, 6.1, 1.9],
[6.3, 2.9, 5.6, 1.8],
[5.7, 2.8, 4.1, 1.3],
[6.5, 3. , 5.5, 1.8],
[6.3, 2.3, 4.4, 1.3],
[6.4, 2.9, 4.3, 1.3],
Disk 81.55 GB available
0s completed at 11:47 AM


UTS_5210411366_AWANDA BINTANG PAMUNGKAS.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
sample_data
data_uts.xlsx
+ Code + Text
RAM Disk
[27] 1 X_test
array([[6.1, 2.8, 4.7, 1.2],
[5.7, 3.8, 1.7, 0.3],
[7.7, 2.6, 6.9, 2.3],
[6. , 2.9, 4.5, 1.5],
[6.8, 2.8, 4.8, 1.4],
[5.4, 3.4, 1.5, 0.4],
[5.6, 2.9, 3.6, 1.3],
[6.9, 3.1, 5.1, 2.3],
[6.2, 2.2, 4.5, 1.5],
[5.8, 2.7, 3.9, 1.2],
[6.5, 3.2, 5.1, 2.],
[4.8, 3. , 1.4, 0.1],
[5.5, 3.5, 1.3, 0.2],
[4.9, 3.1, 1.5, 0.1],
[5.1, 3.8, 1.5, 0.3],
[6.3, 3.3, 4.7, 1.6],
[6.5, 3. , 5.8, 2.2],
[5.6, 2.5, 3.9, 1.1],
[5.7, 2.8, 4.5, 1.3],
[6.4, 2.8, 5.6, 2.2],
[4.7, 3.2, 1.6, 0.2],
[6.1, 3. , 4.9, 1.8],
[5. , 3.4, 1.6, 0.4],
[6.4, 2.8, 5.6, 2.1],
[7.9, 3.8, 6.4, 2.],
[6.7, 3. , 5.2, 2.3],
[6.7, 2.5, 5.8, 1.8],
[6.8, 3.2, 5.9, 2.3],
[4.8, 3. , 1.4, 0.3],
[4.8, 3.1, 1.6, 0.2]])
MODELLING
[48] 1 from sklearn.neighbors import KNeighborsClassifier
2 k=4
3 #Train Model and Predict
4 knn = KNeighborsClassifier(n_neighbors = k).fit(X_train,y_train)
1 y_test
array([1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
1, 0, 0])
[50] 1 y_train
array([0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1,
Disk 81.55 GB available
0s completed at 11:47 AM


UTS_5210411366_AWANDA BINTANG PAMUNGKAS.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
sample_data
data_uts.xlsx
+ Code + Text
RAM Disk
[49] 1 y_test
array([1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
1, 0, 0])
[50] 1 y_train
array([0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1,
1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1,
1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1,
0, 1, 1, 1, 1, 1, 1, 0, 0, 0])
[51] 1 from sklearn import metrics
2
3 y_predict = knn.predict(X_test)
4 print(y_predict)
[1 0 0 0 0 0 1 0 0 0 1 0 0 1 0 1 0 0 1 1 1 0 0 0 0]
[52] 1 print("Test set Accuracy: ", metrics.accuracy_score(y_test, y_predict))
Test set Accuracy: 0.68

3.



Akurasi = TP + TN / TP + FP + FN + TN

= 12 + 10 / 12 + 1 + 2 + 10

= 22 / 25

= 0.88