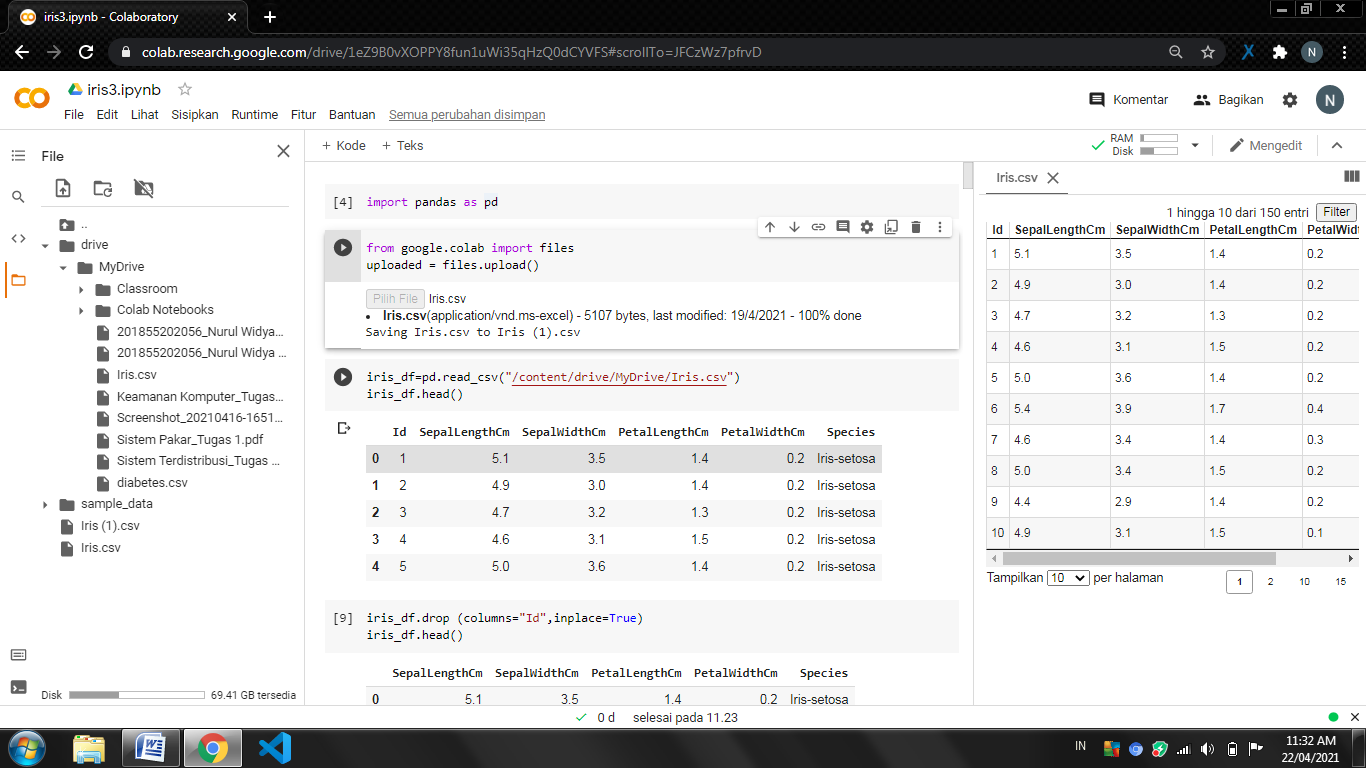
Nama : Nurul Widya Har

Nim : 201855202056

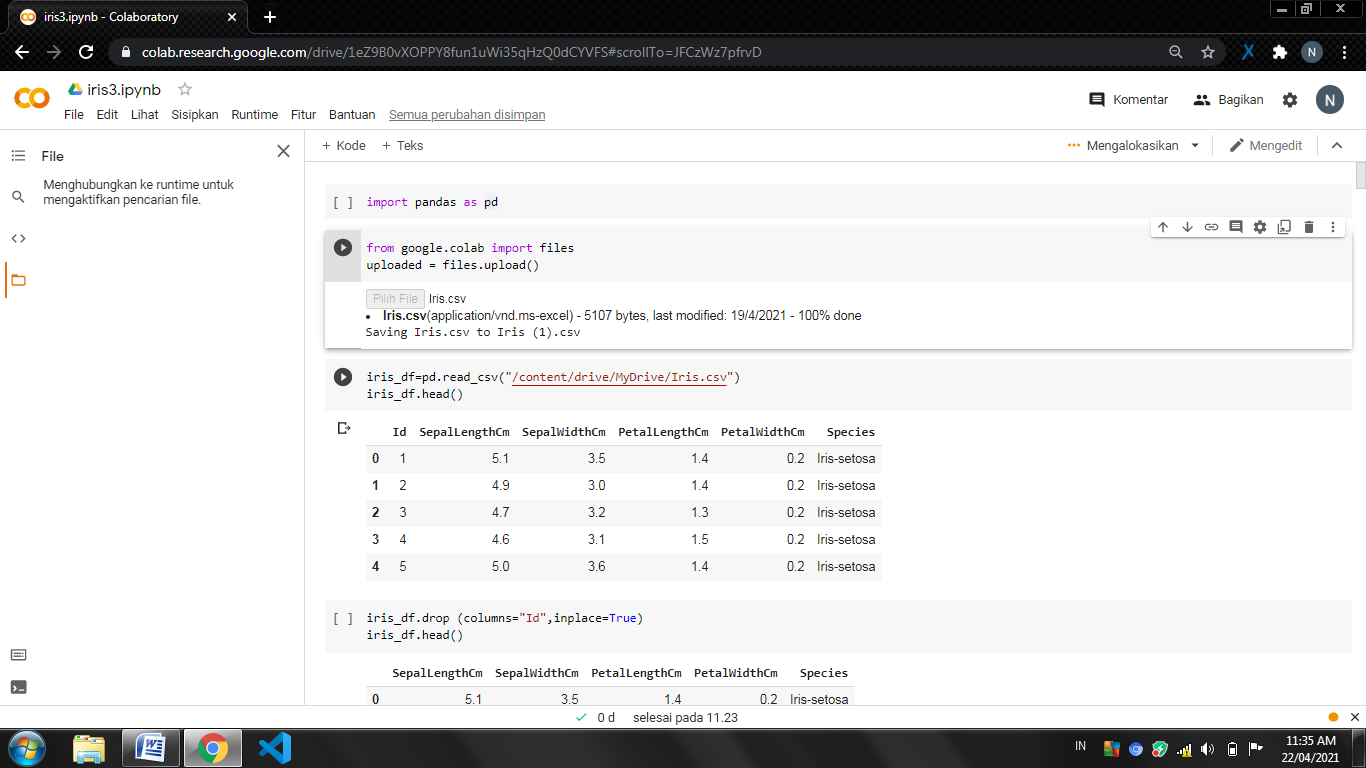
Tugas : 3

Simple Exploratory Data Analysis (EDA) :

1. Menganalisa data

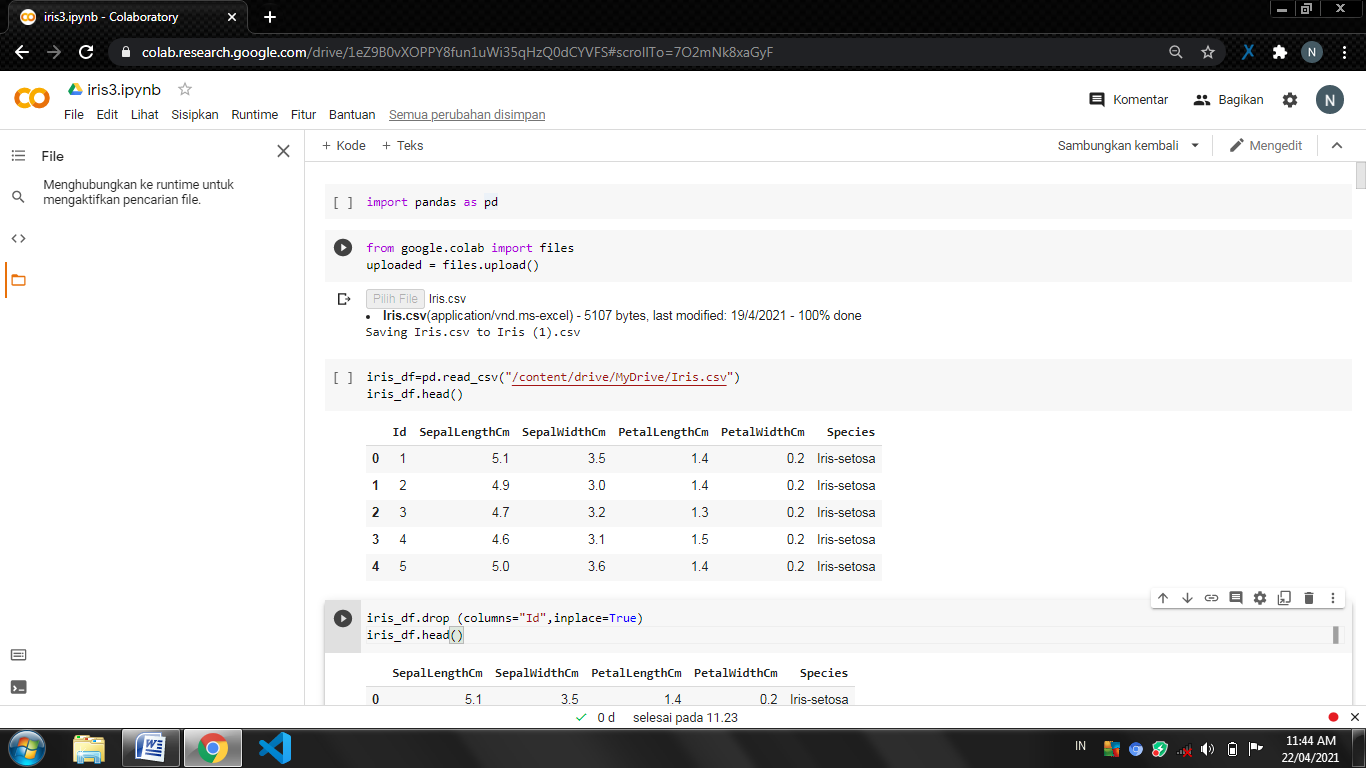


1. Upload dataset pada google colab

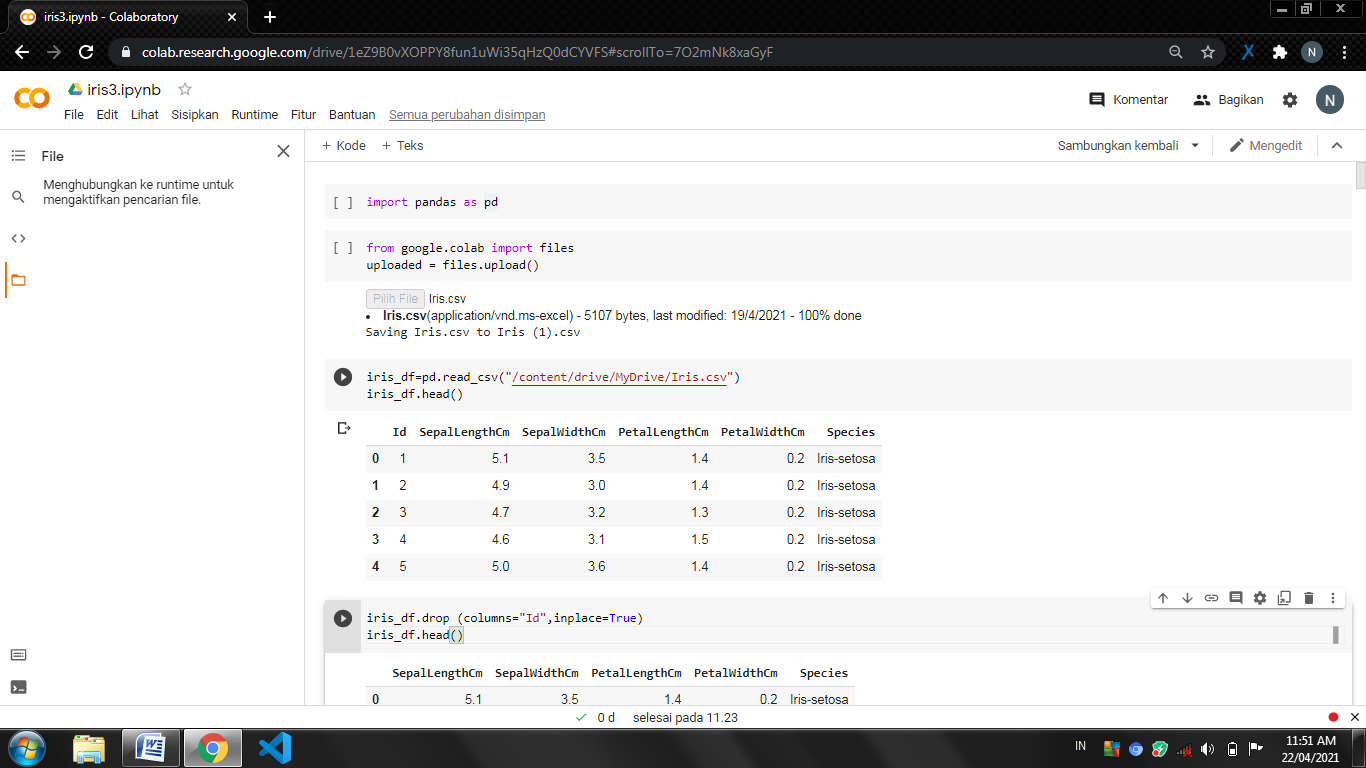


1. Memuat file csv

Code :

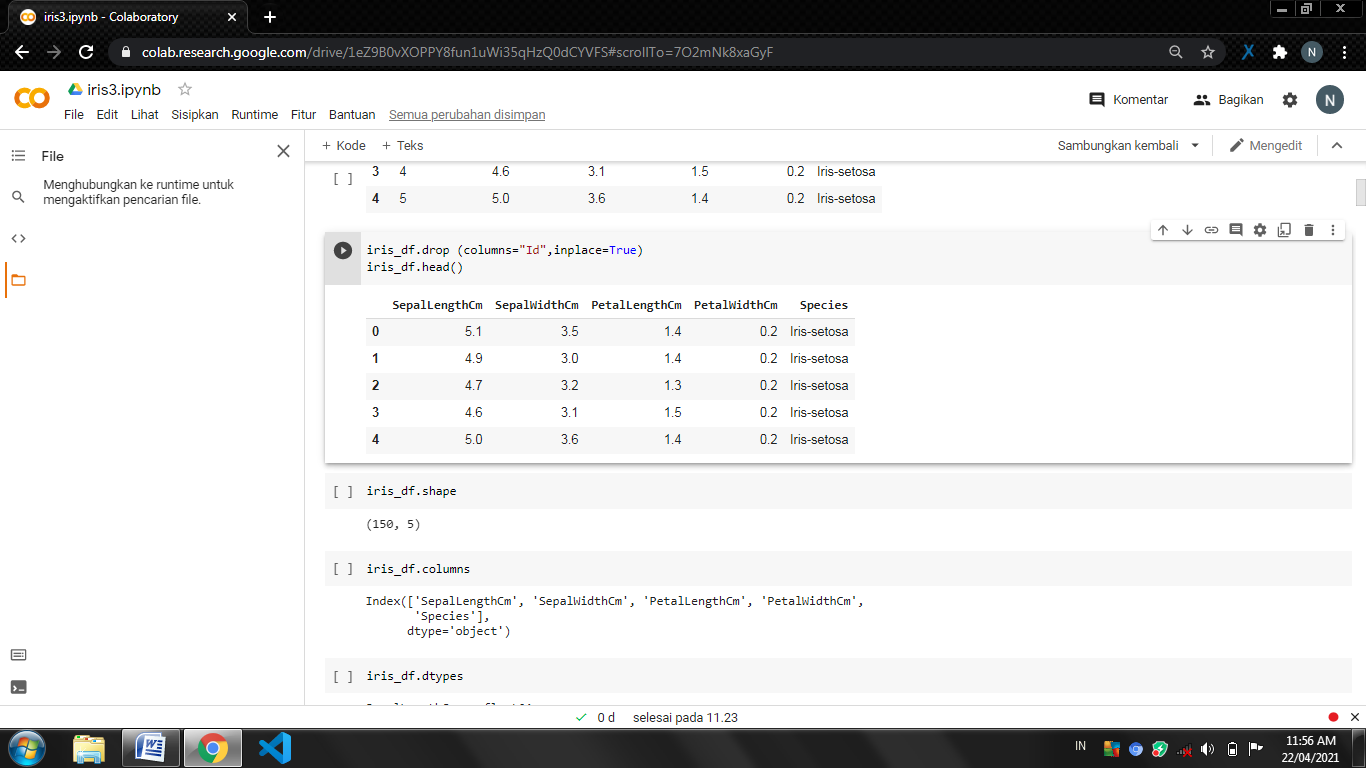


Output :

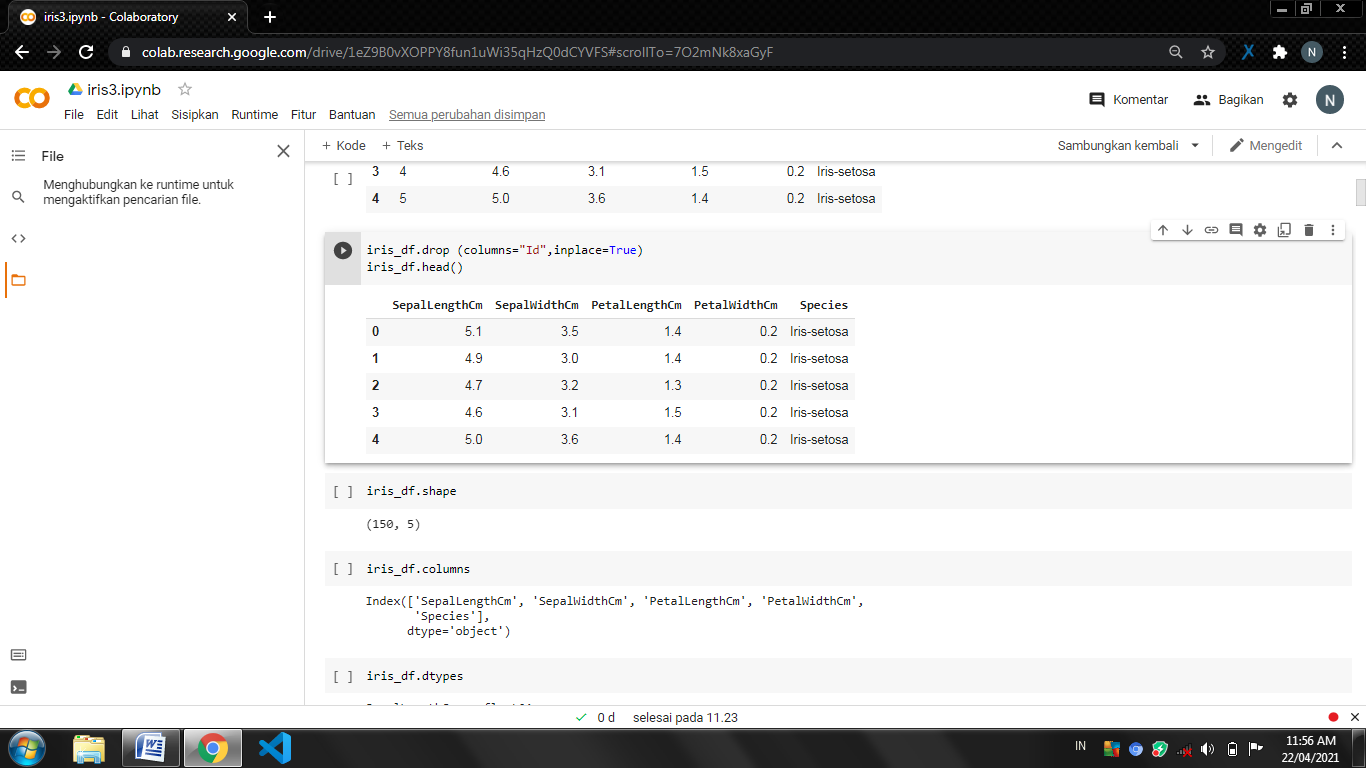


1. Menghapus kolom id

Code :

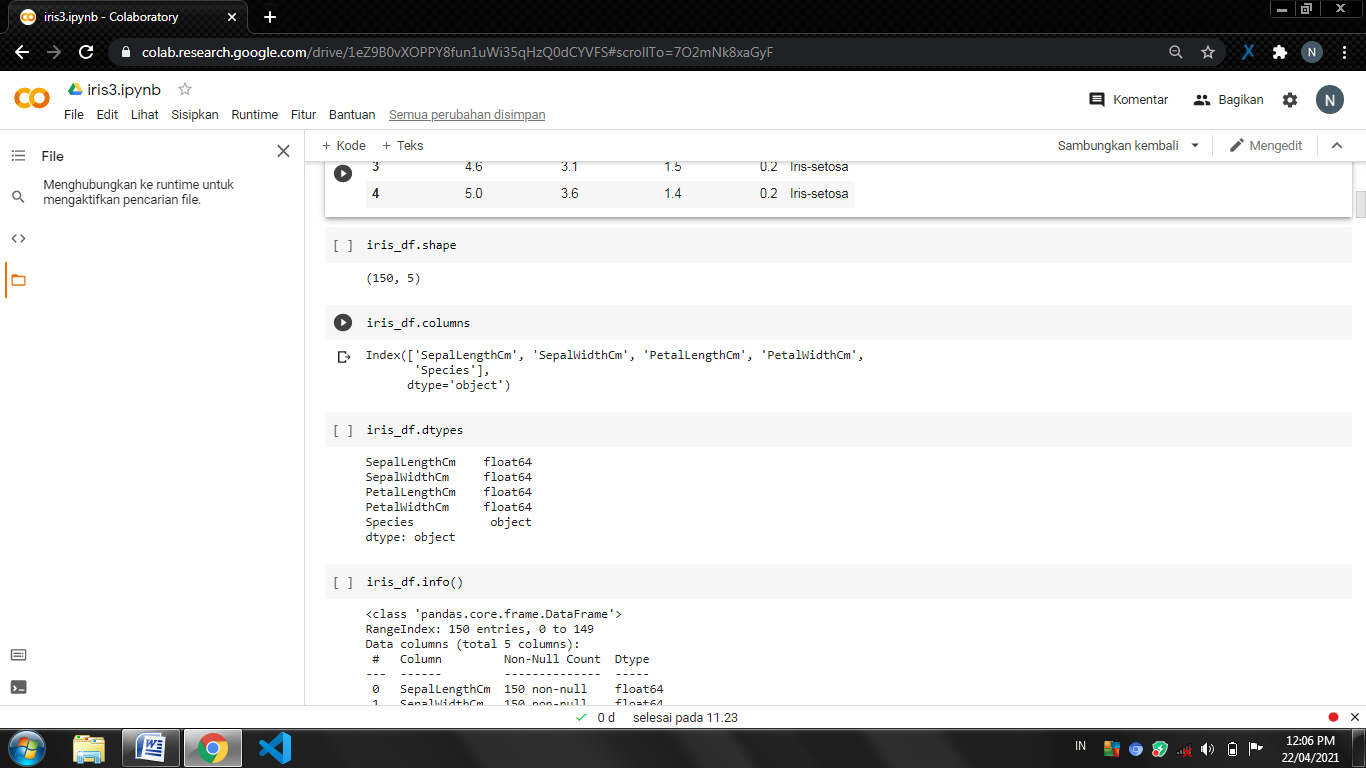


Output :

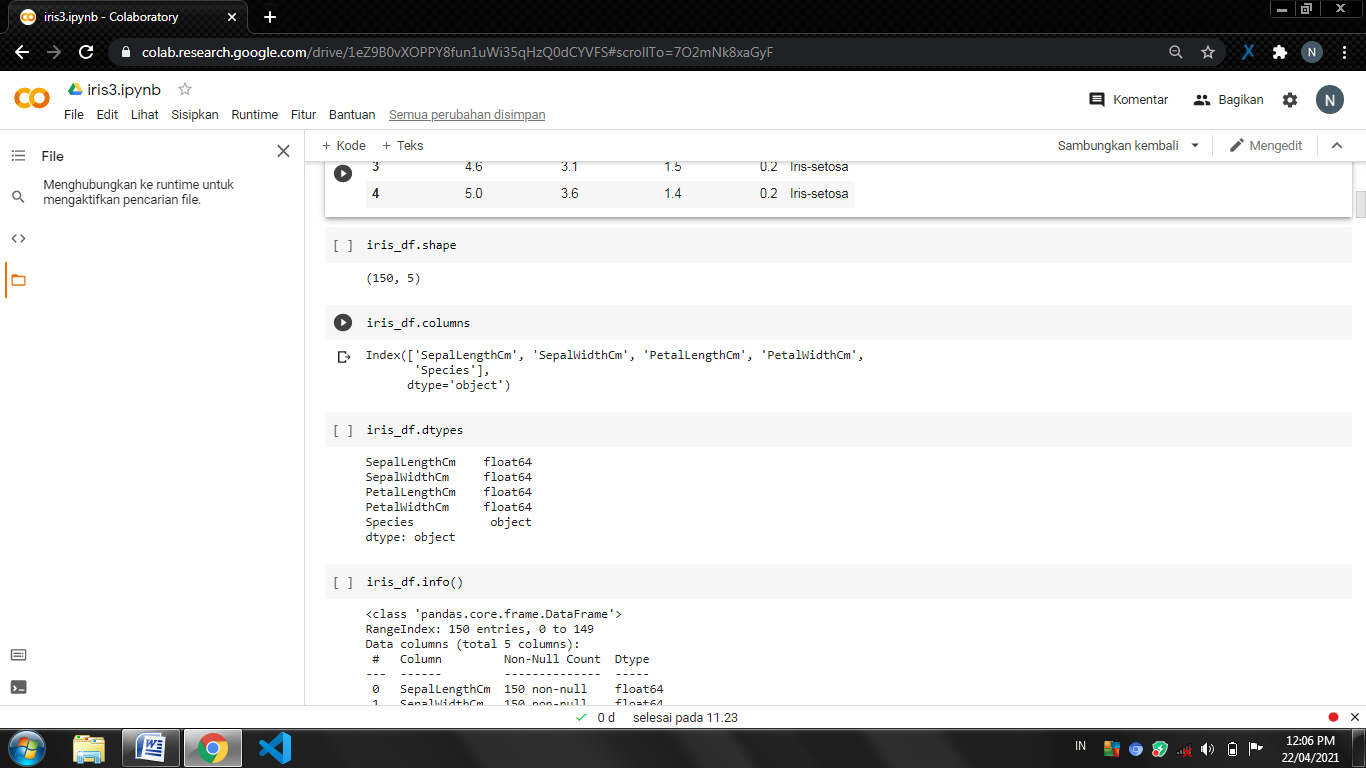


1. Identifikasi Data

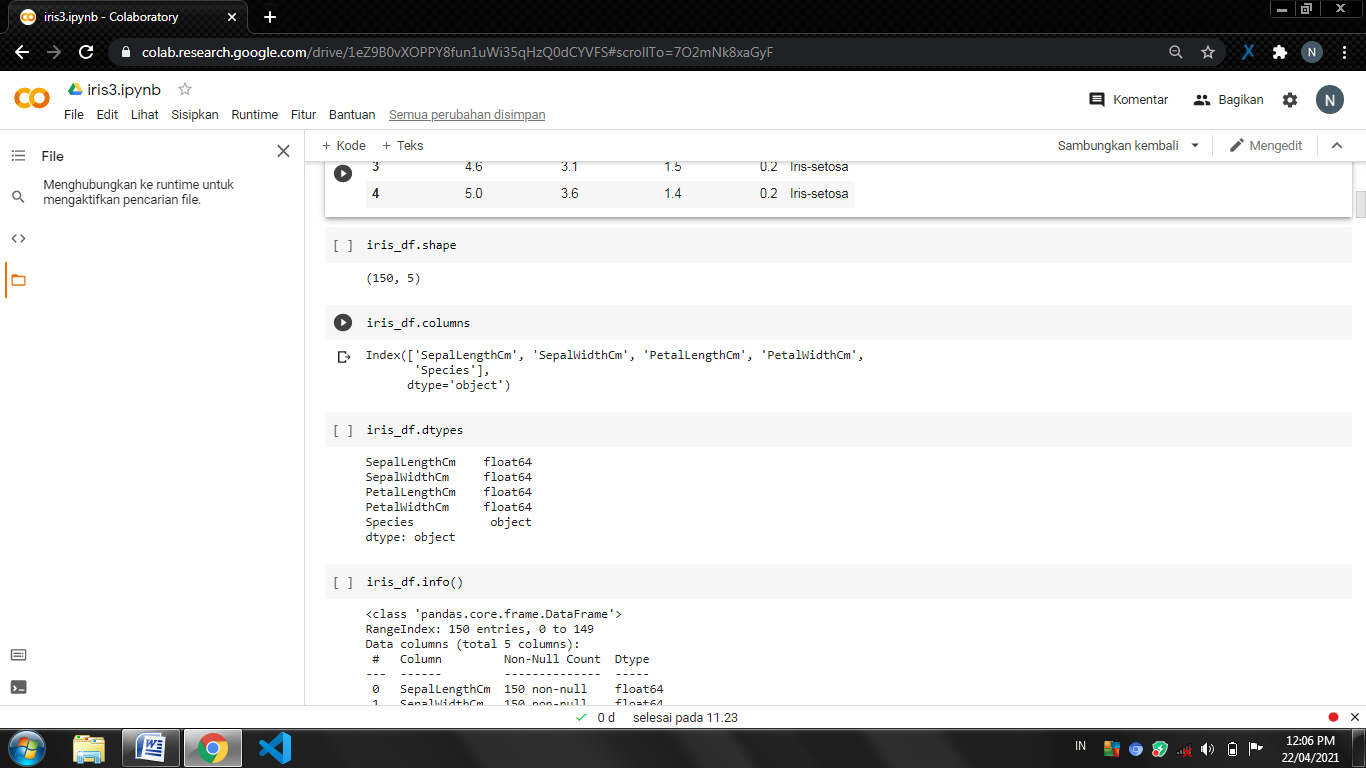
Untuk mengetahui bentuk dataset berupa baris,kolom



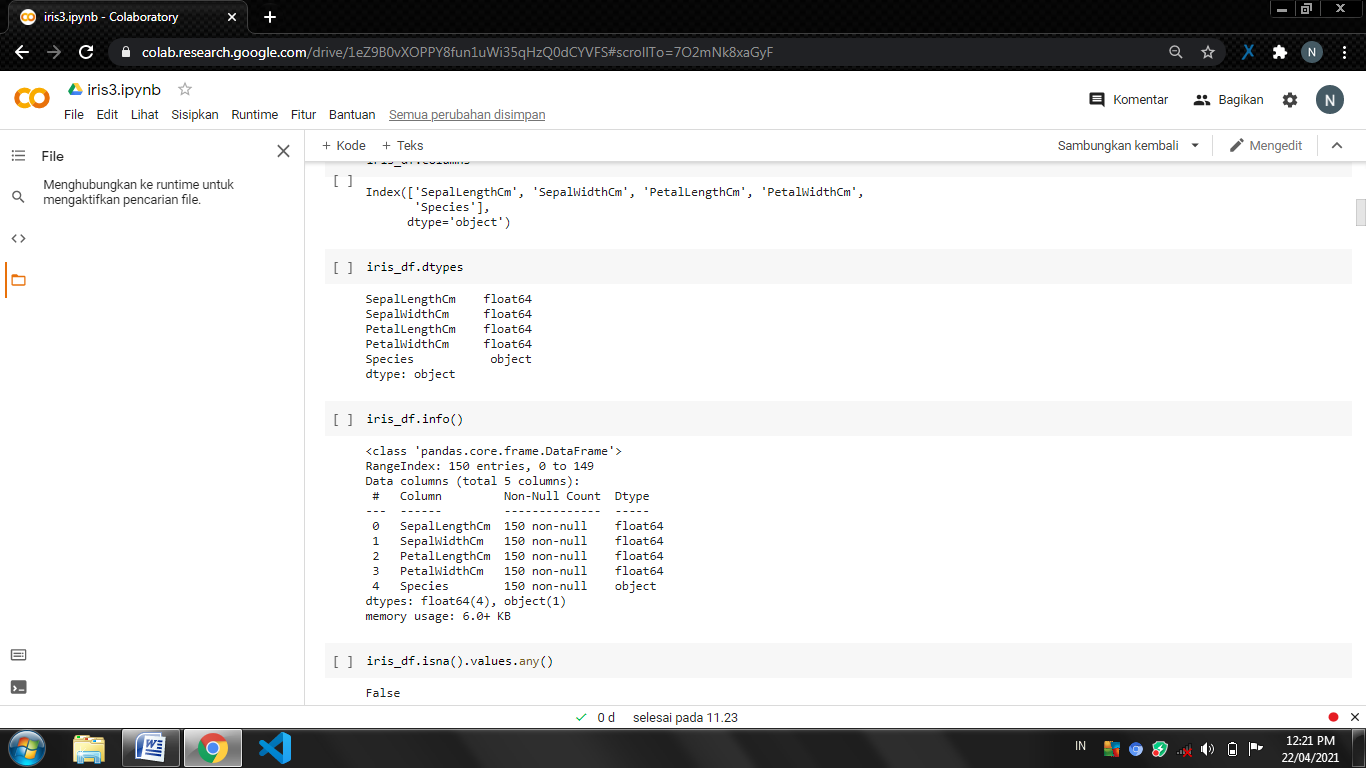
1. Untuk mengetahui daftar kolom



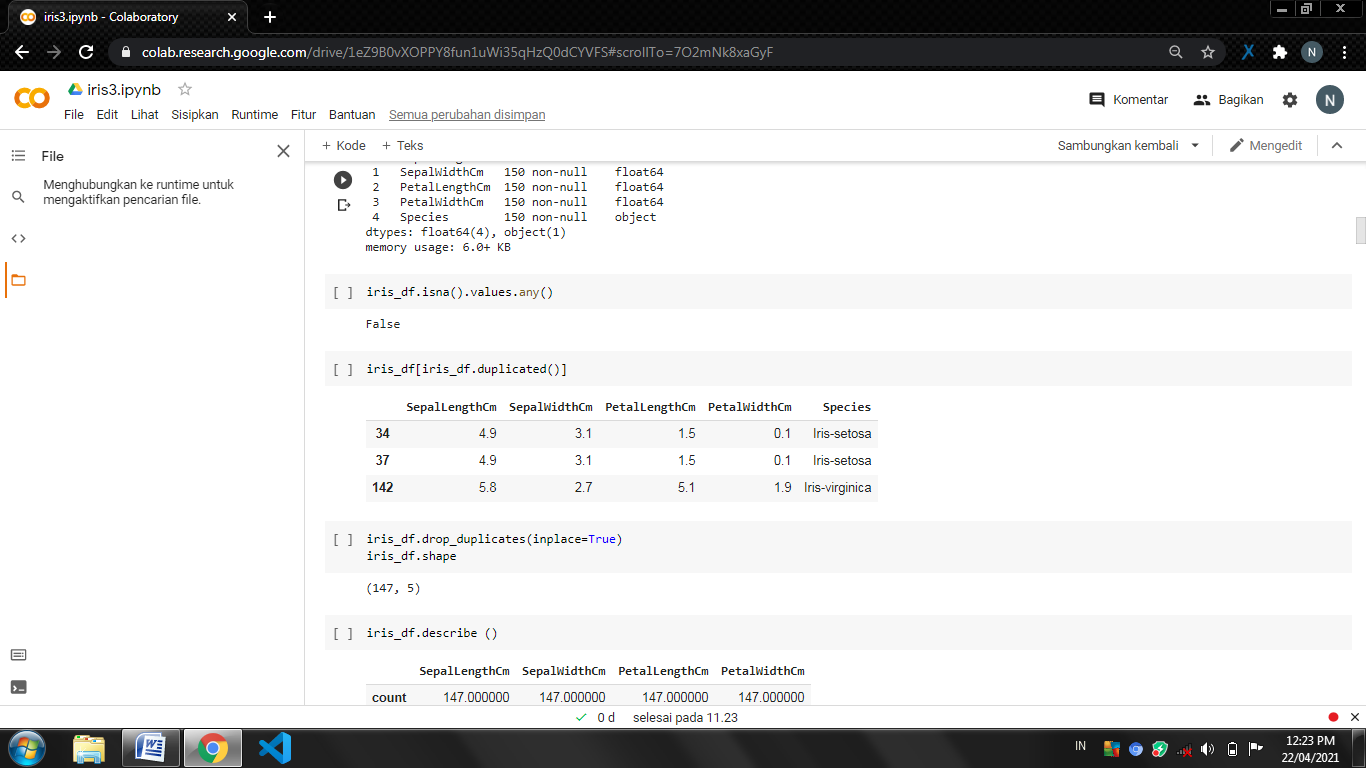
1. Identifikasi data disetiap kolom



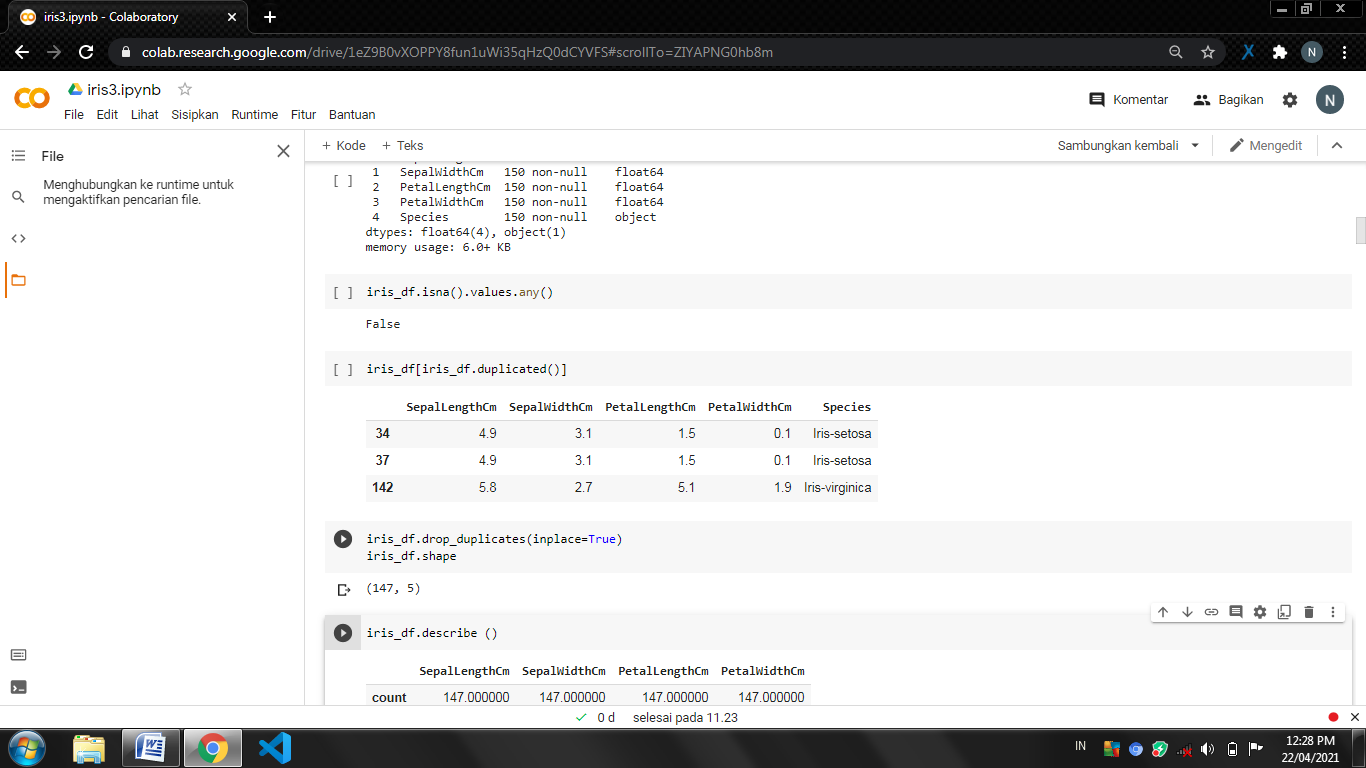
1. Informasi dataset



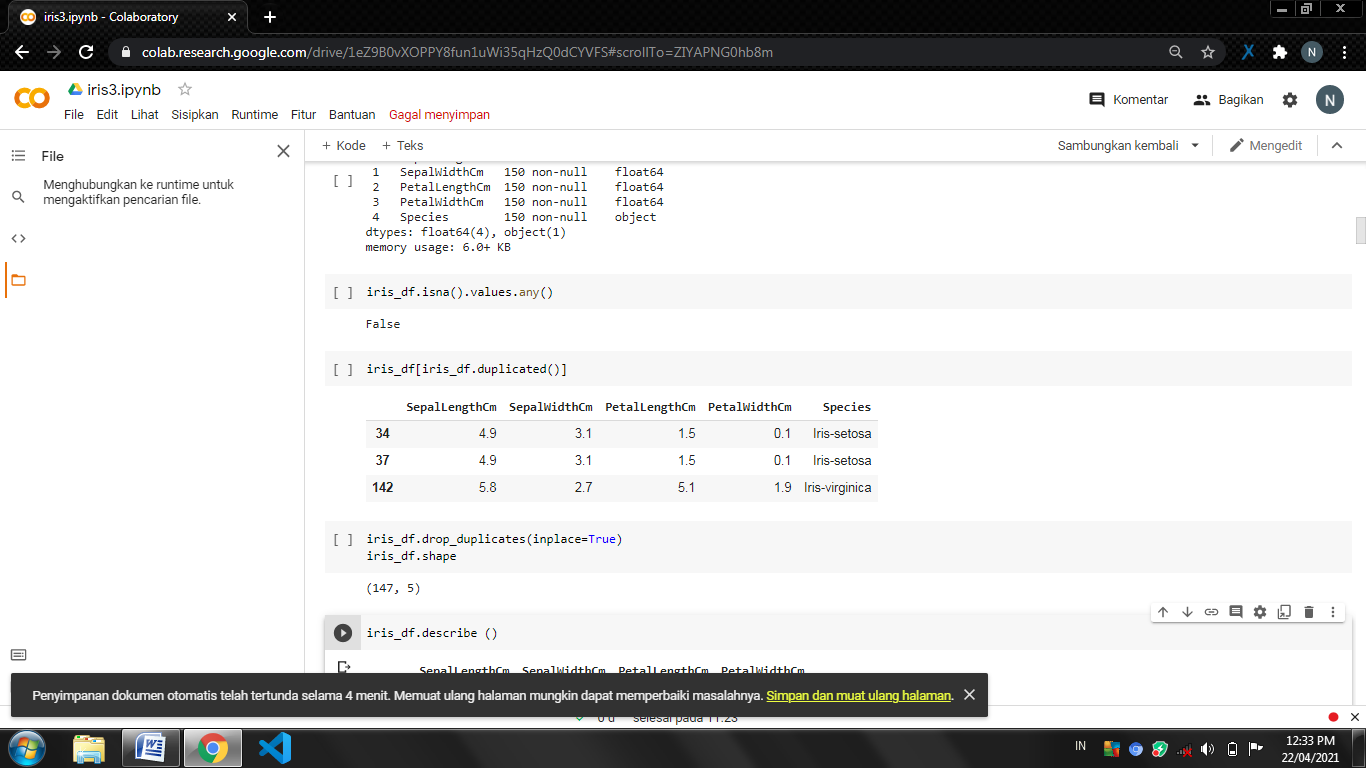
1. Mengidentifikasi nilai kosong



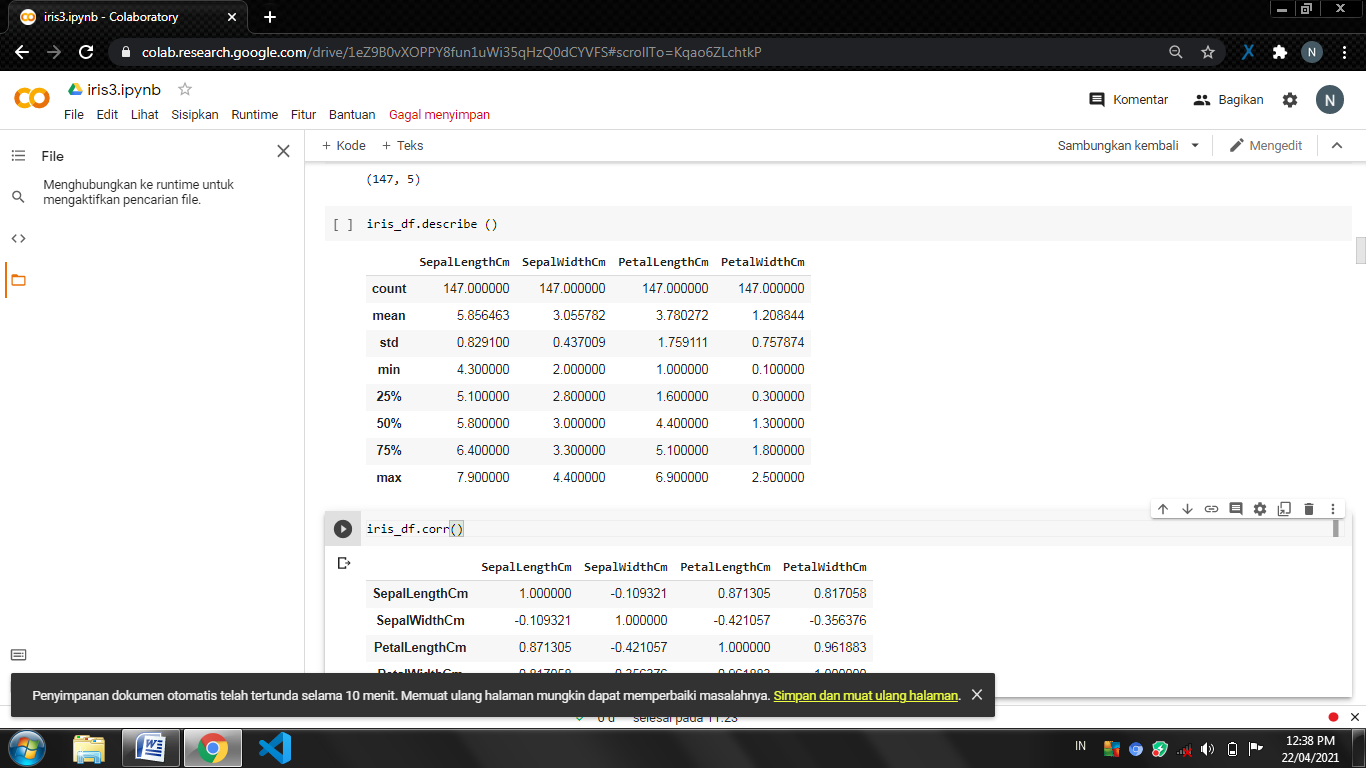
1. Mencari nilai duplikasi pada data



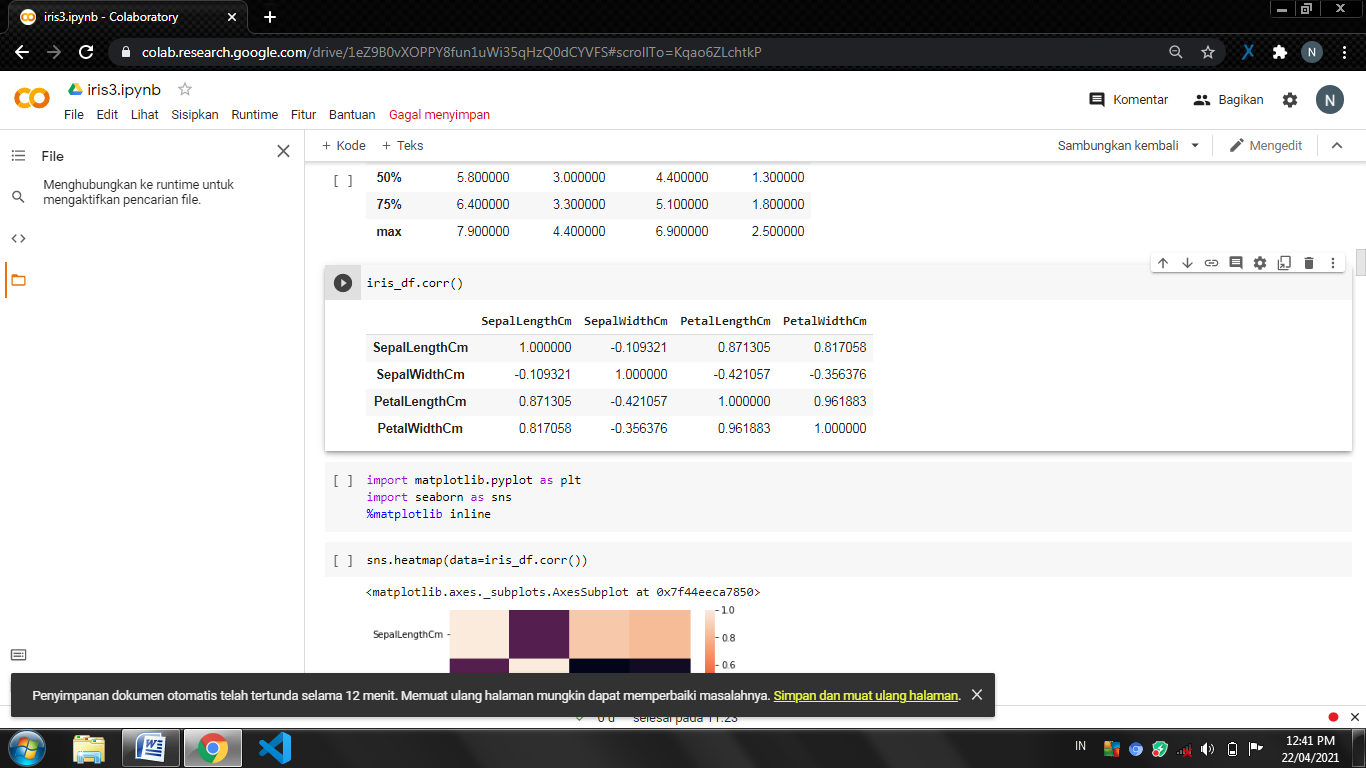
1. Mengetahui jumlah duplikasi pada data



1. Deskripsi statistik dataset

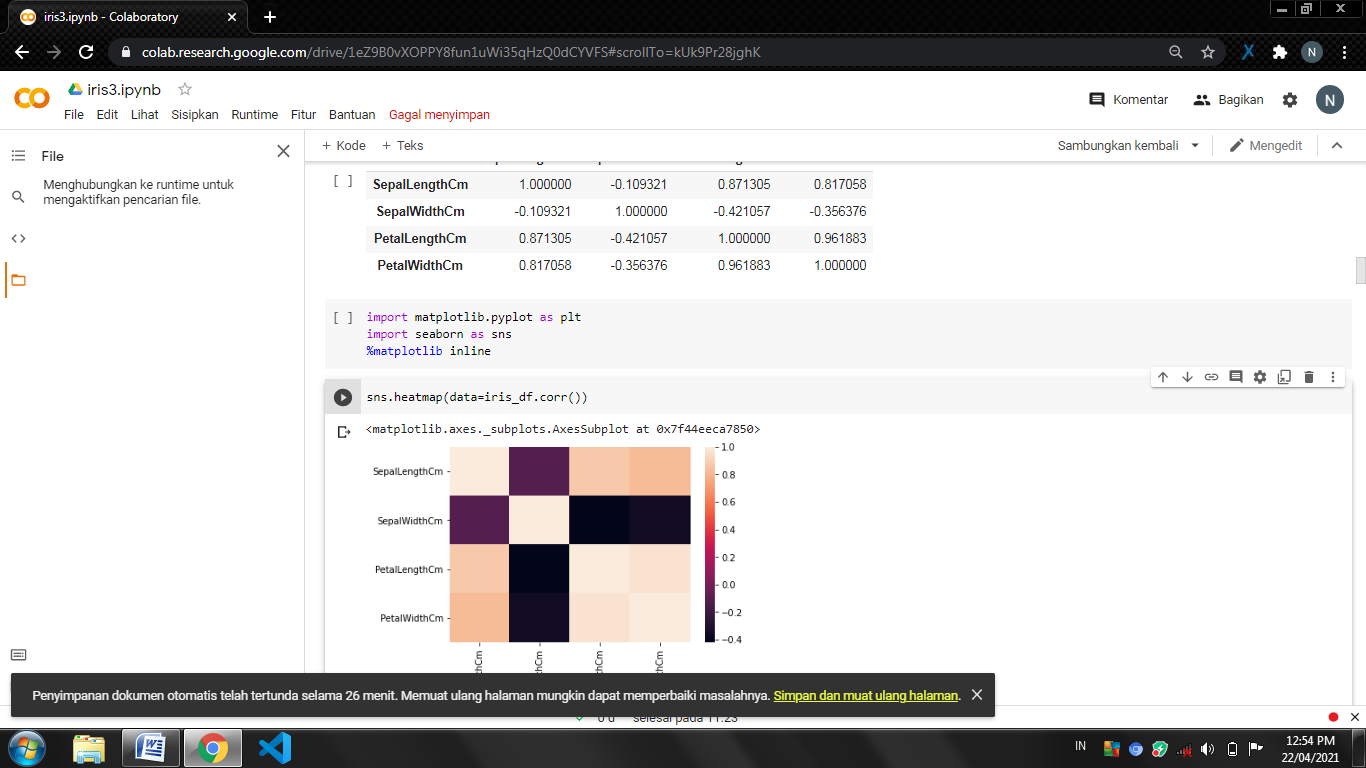


1. Korelasi antar kolom



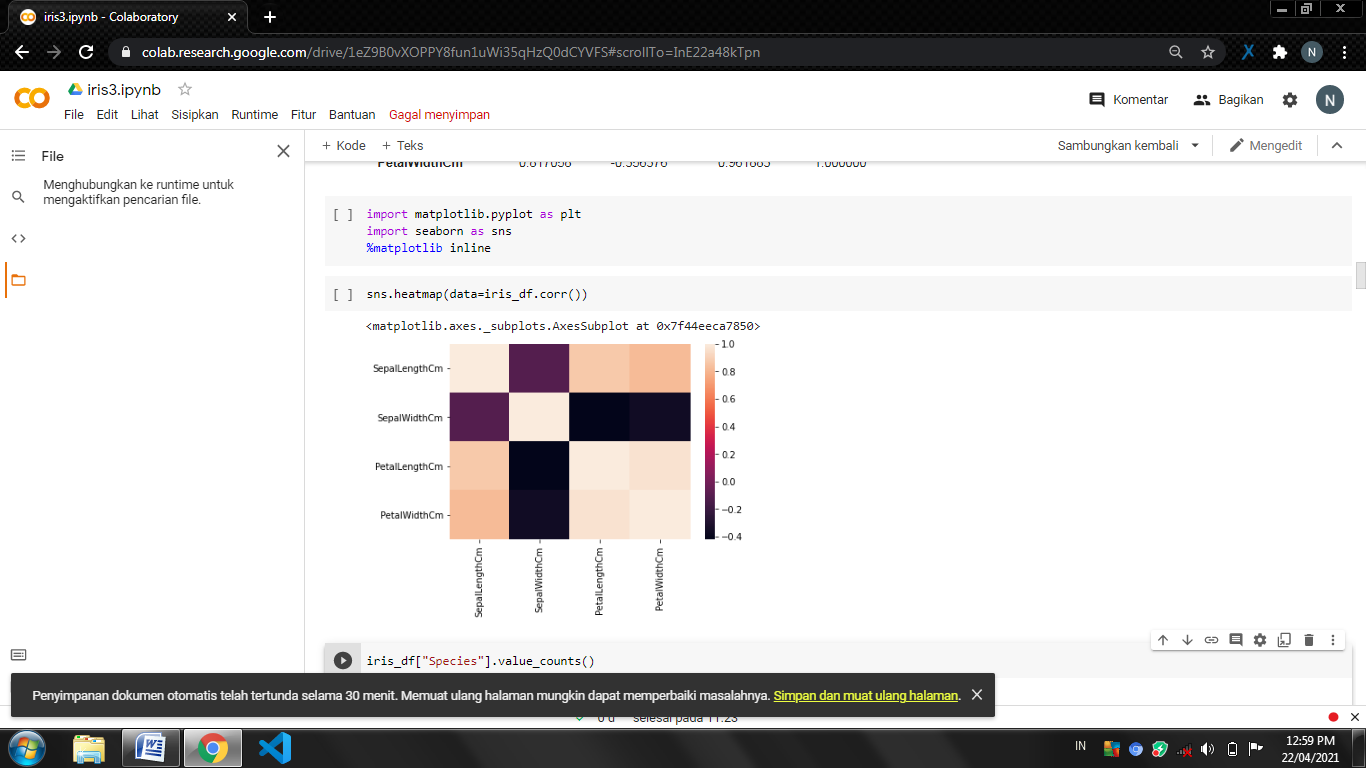
Iris Dataset : Data Visualisation

1. Visualisasi data menggunakan dua modul (matplotlib dan seaborn)



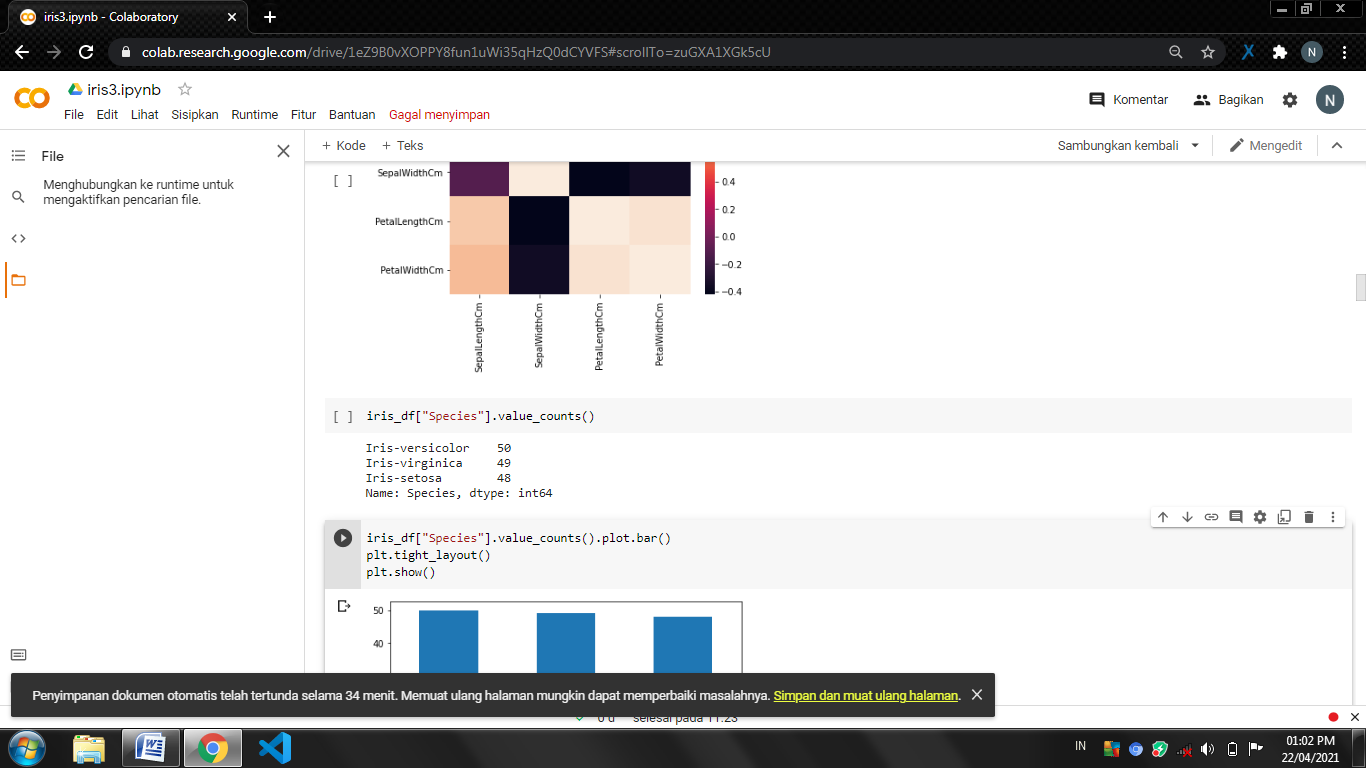
1. Heatmap

Visualisasi yang paling cocok untuk menvisualisasikan corelation matricks

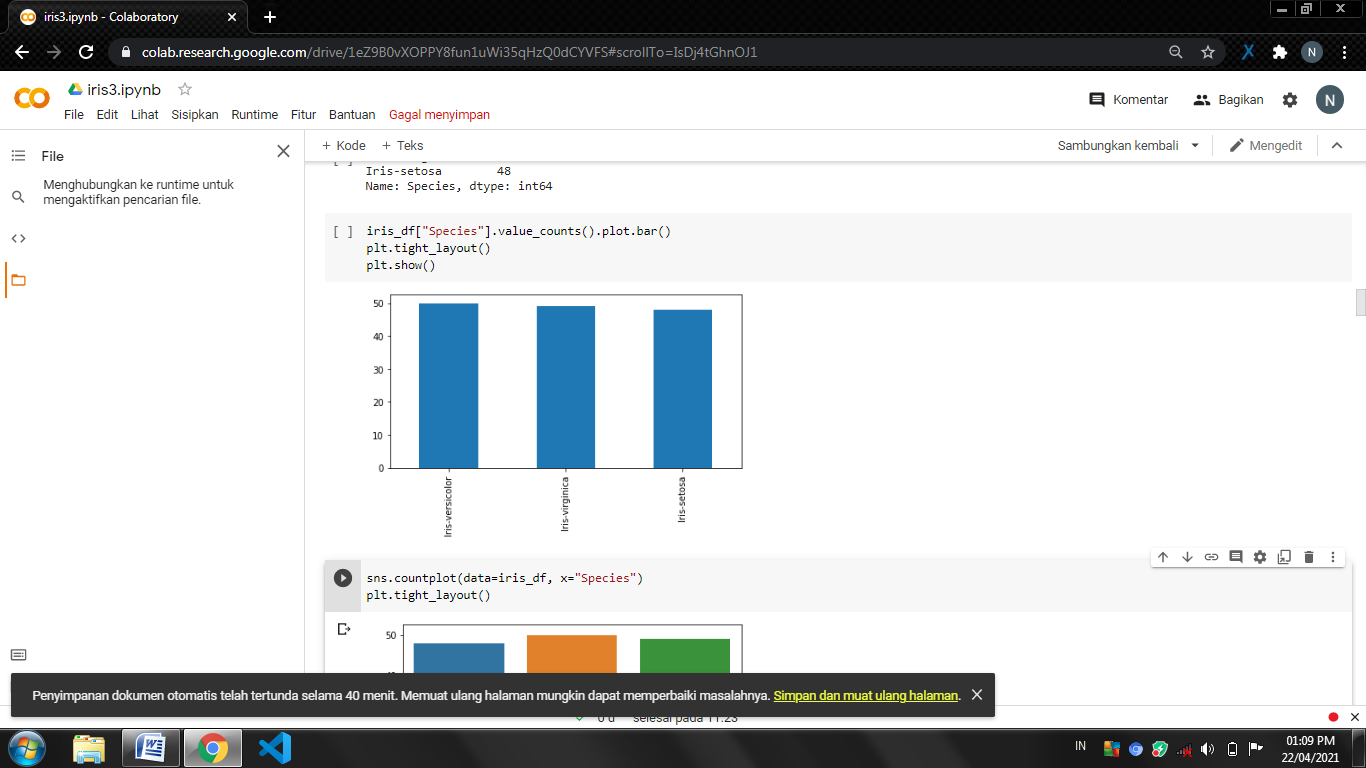


1. Bar Plot

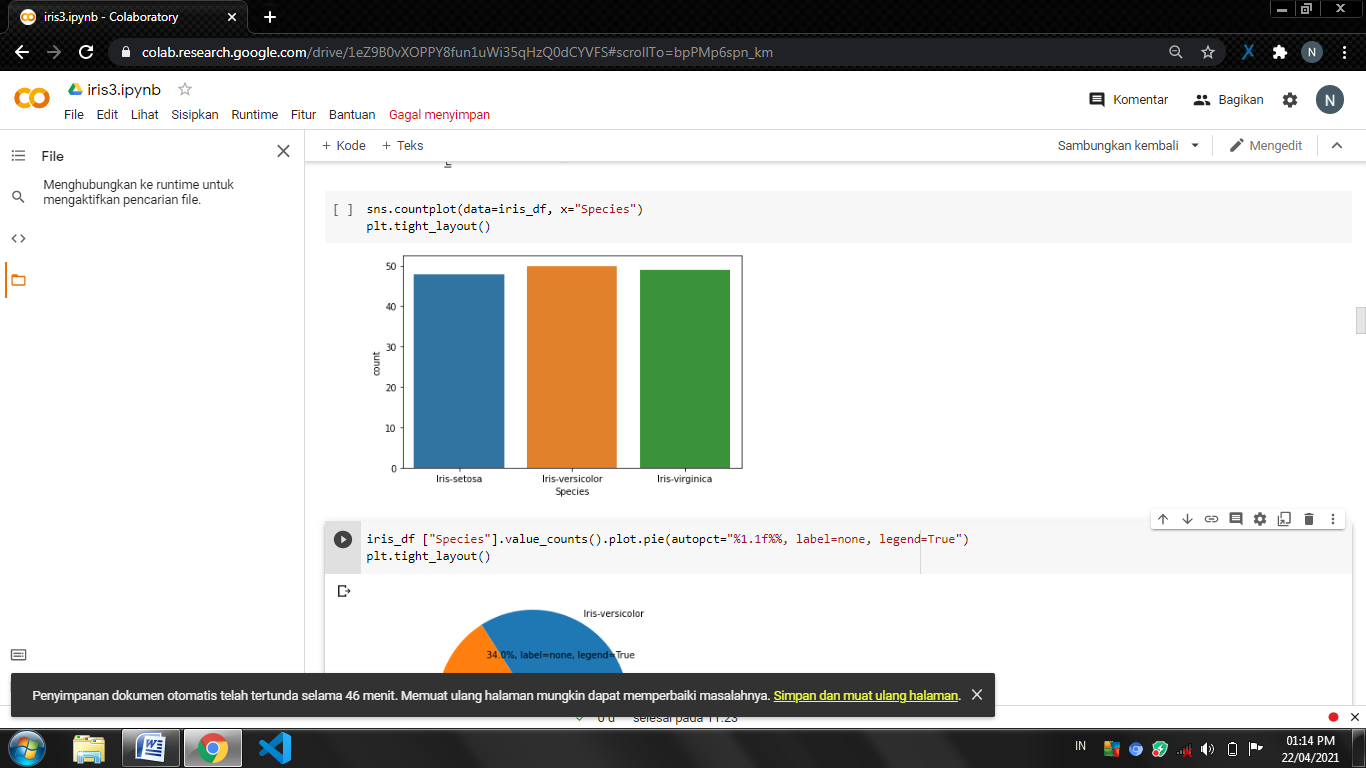
Menghitung jumlah species



Data di atas dapat divisualisasikan dengan menggunakan bar plot, caranya panggil dataset yang ingin divisualisasikan lalu di tambahkan *.plot.bar*

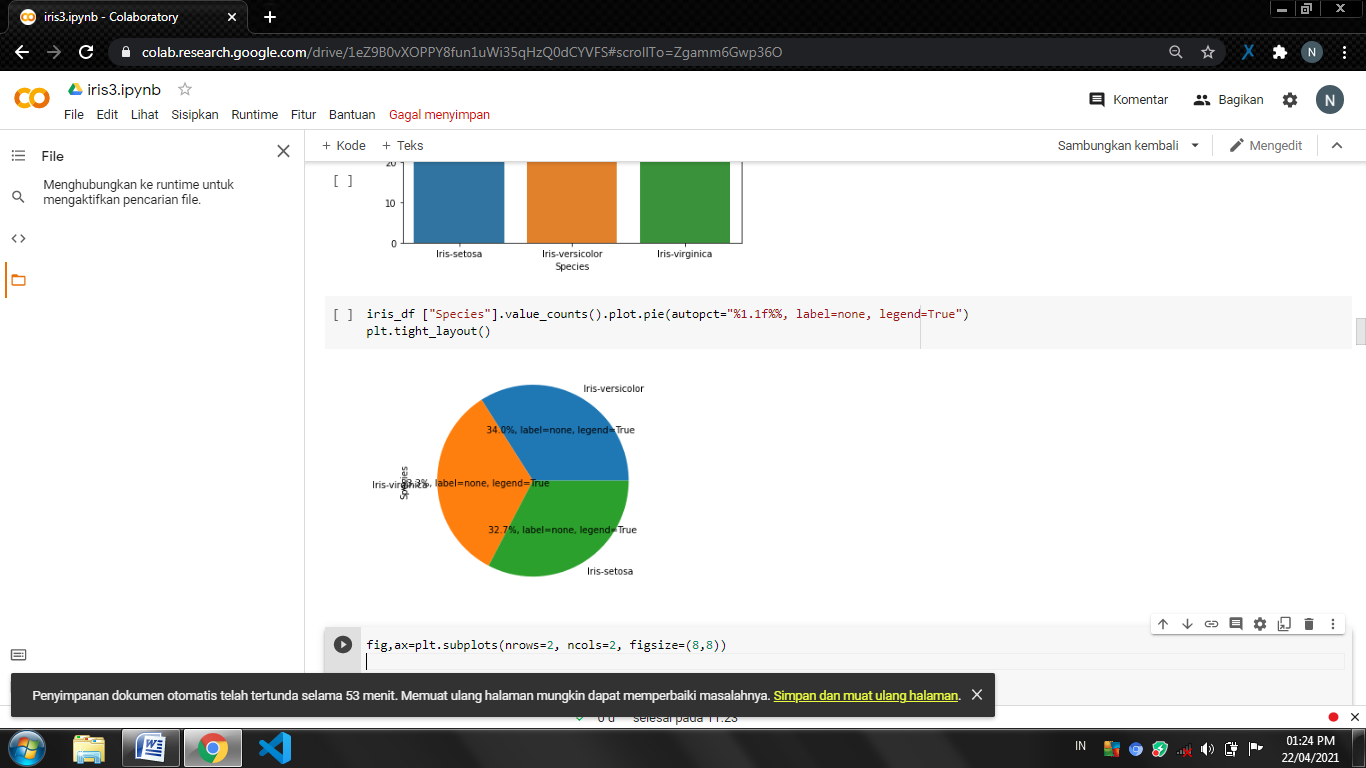


1. Melakukan bar plot bisa juga menggunakan seaborn dengan memanggil *sns.countplot*



1. Pie Chart

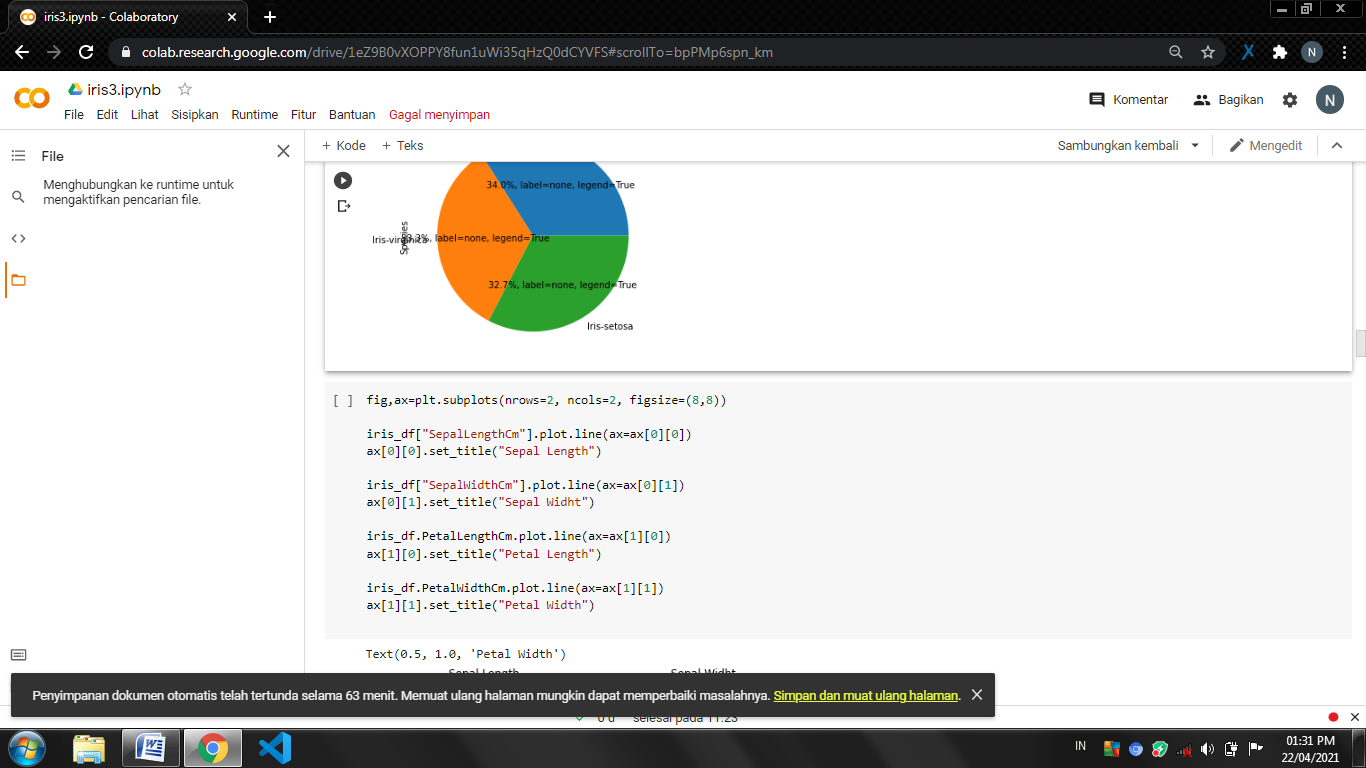
Melakukan visualisasi menggunakan pie chart, dengan memanfaatkan presentase



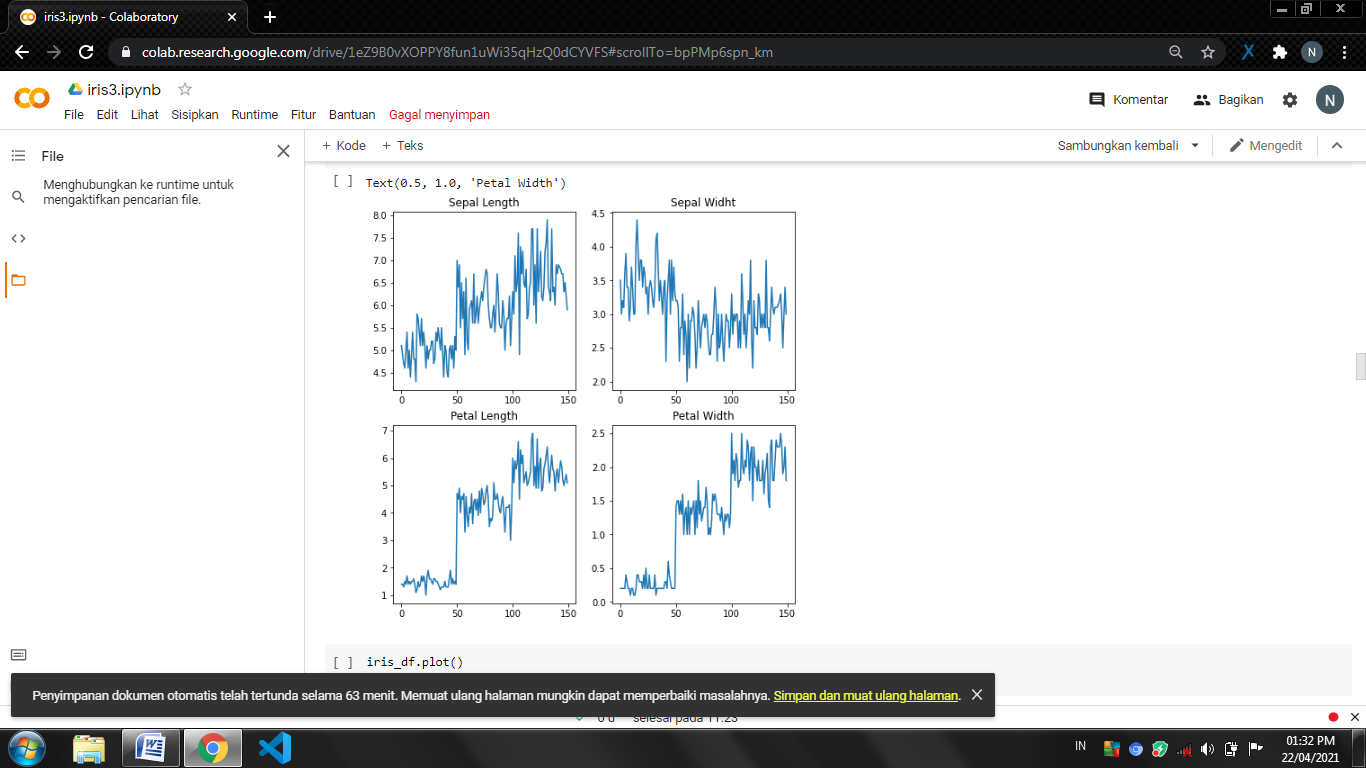
1. Line Plot

Melakukan subplots dengan jumlah baris dan kolom, disini yang akan diploting adalah empat kolom yang ada apada data

Code :

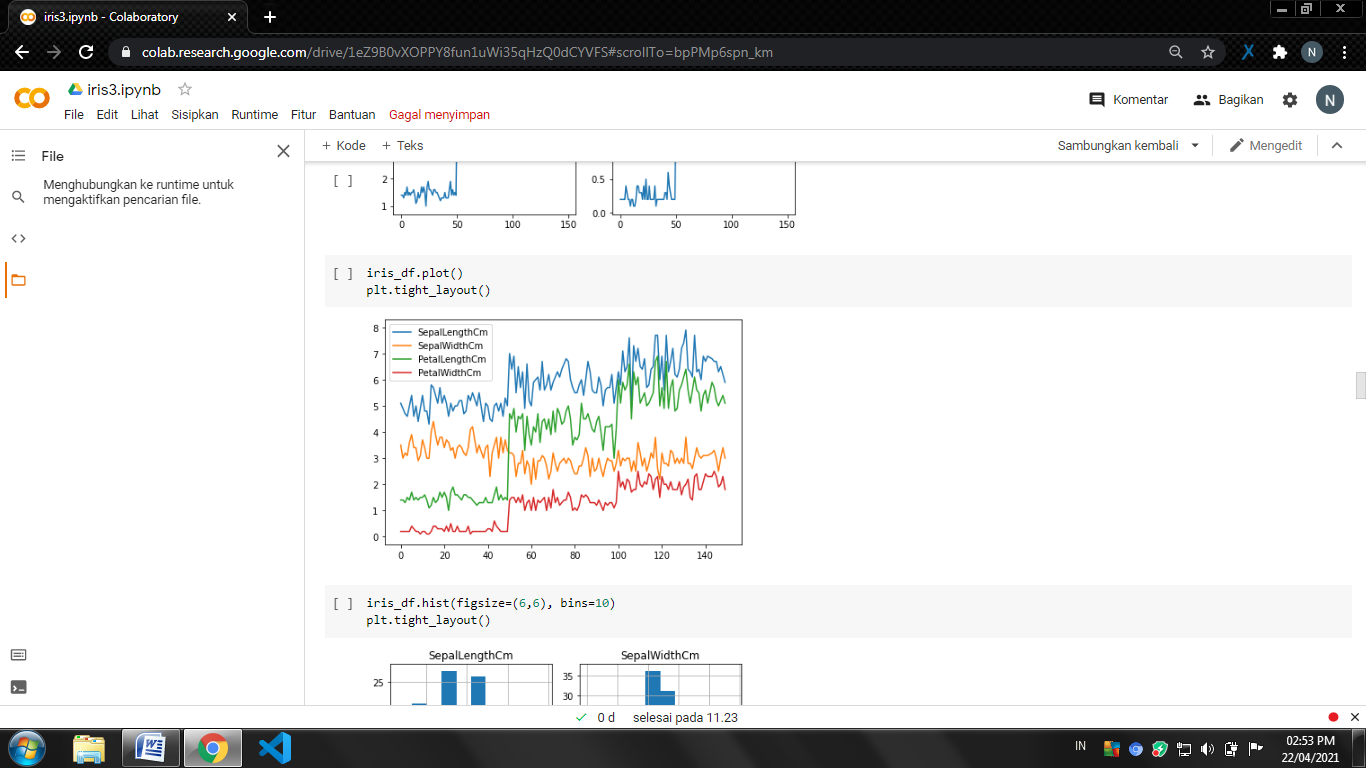


Output :



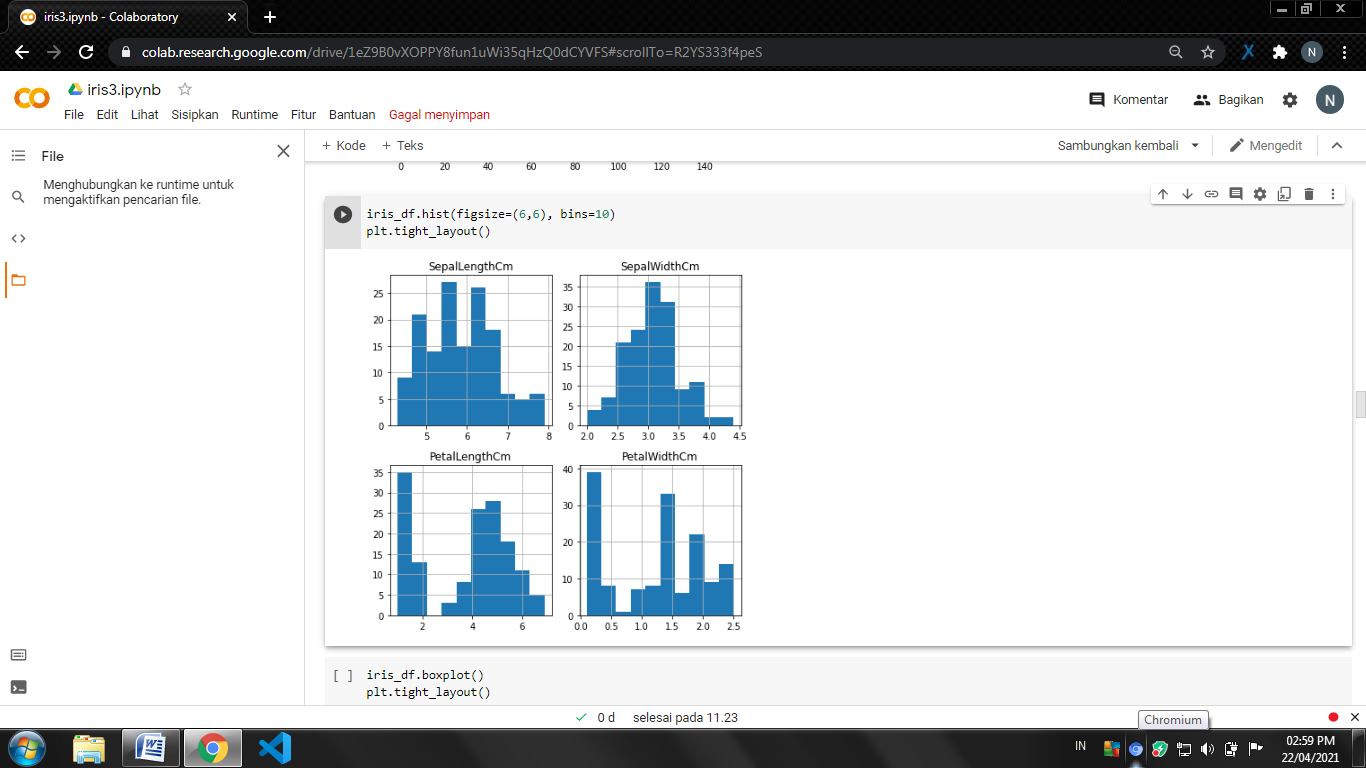
1. Visualisasi baris juga dapat dilakukan dengan menggunakan *iris\_df.plot*

Visualisasi yang dilakukan secara keseluruhan akan menampilkan keempat informasi dalam satu graf yang sama dan setiap kolom diberi keterangan warna



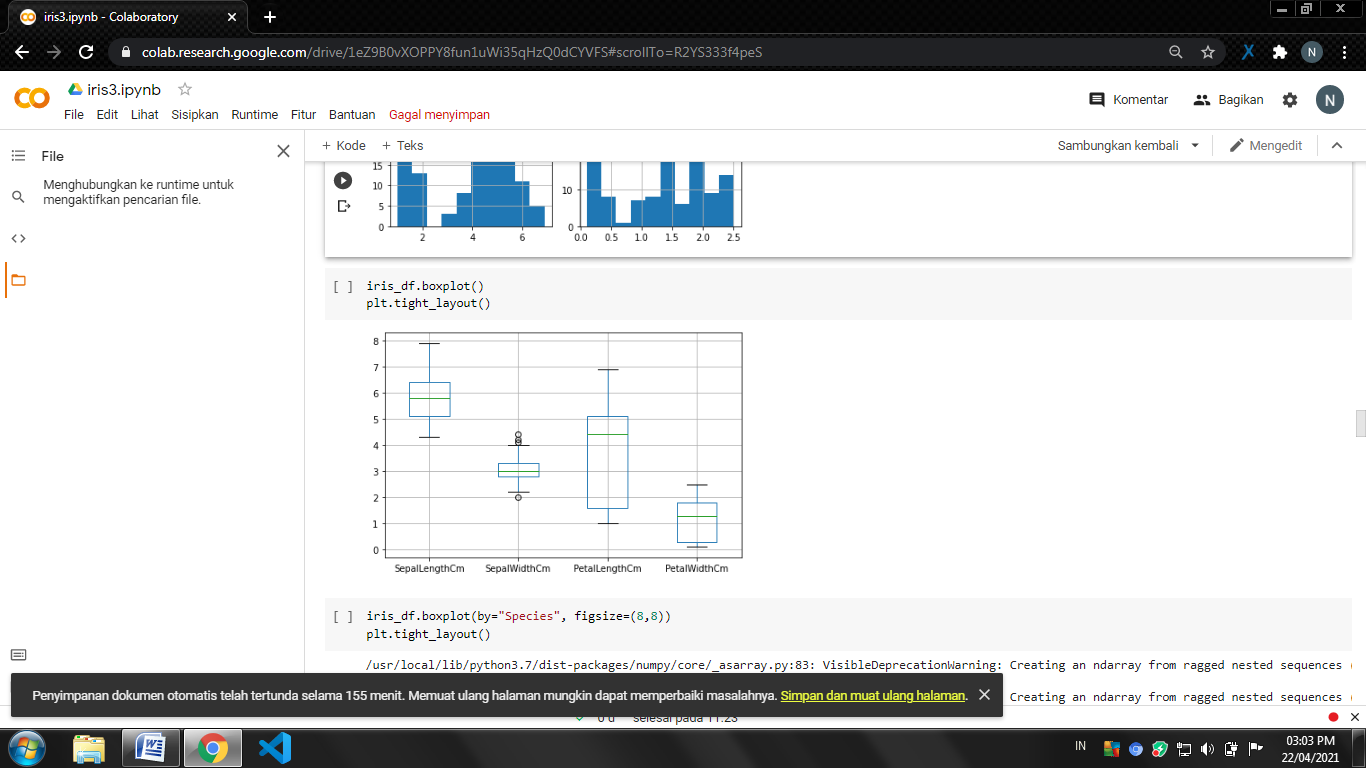
1. Histogram

Caranya dengan memanggil nama data tersebut lalu ditambahkan *.hist*



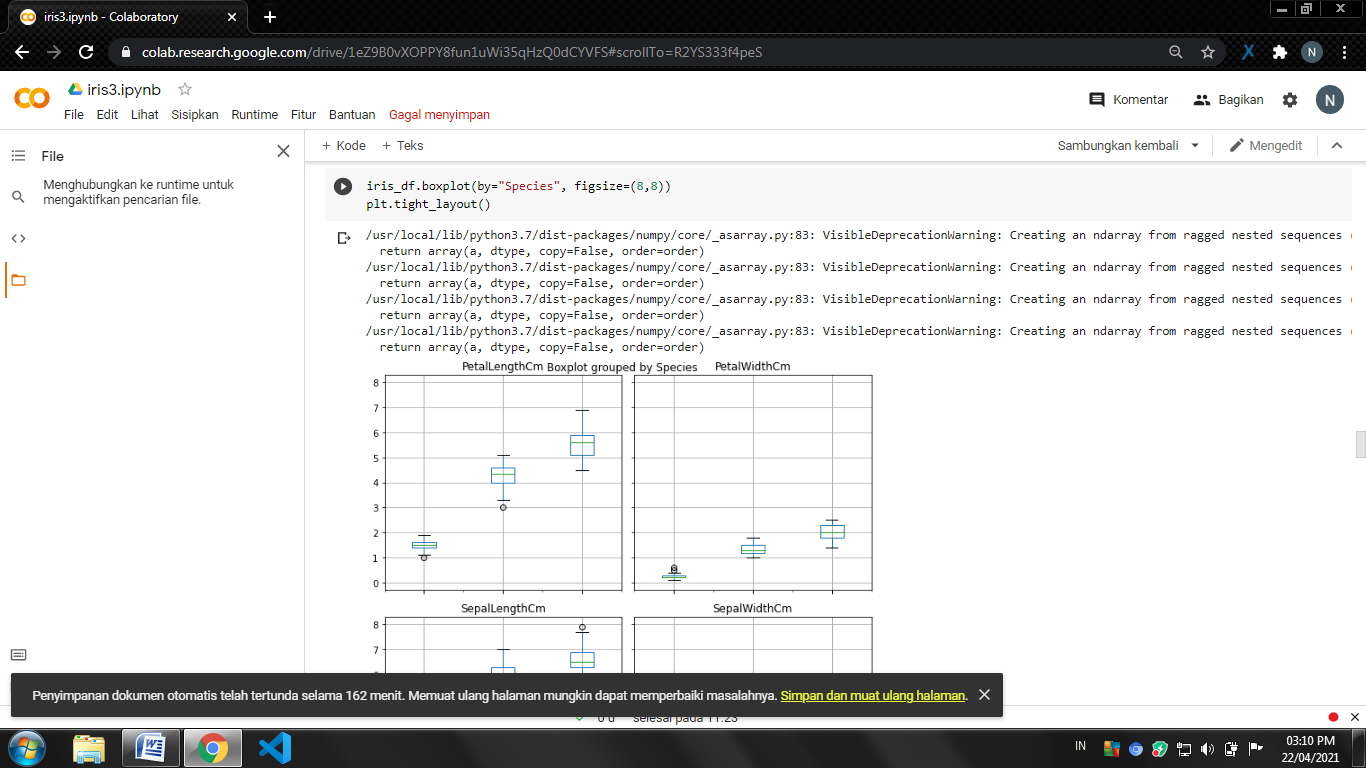
1. Boxplot

Dengan memanggil data lalu ditambahkan *.boxplot*

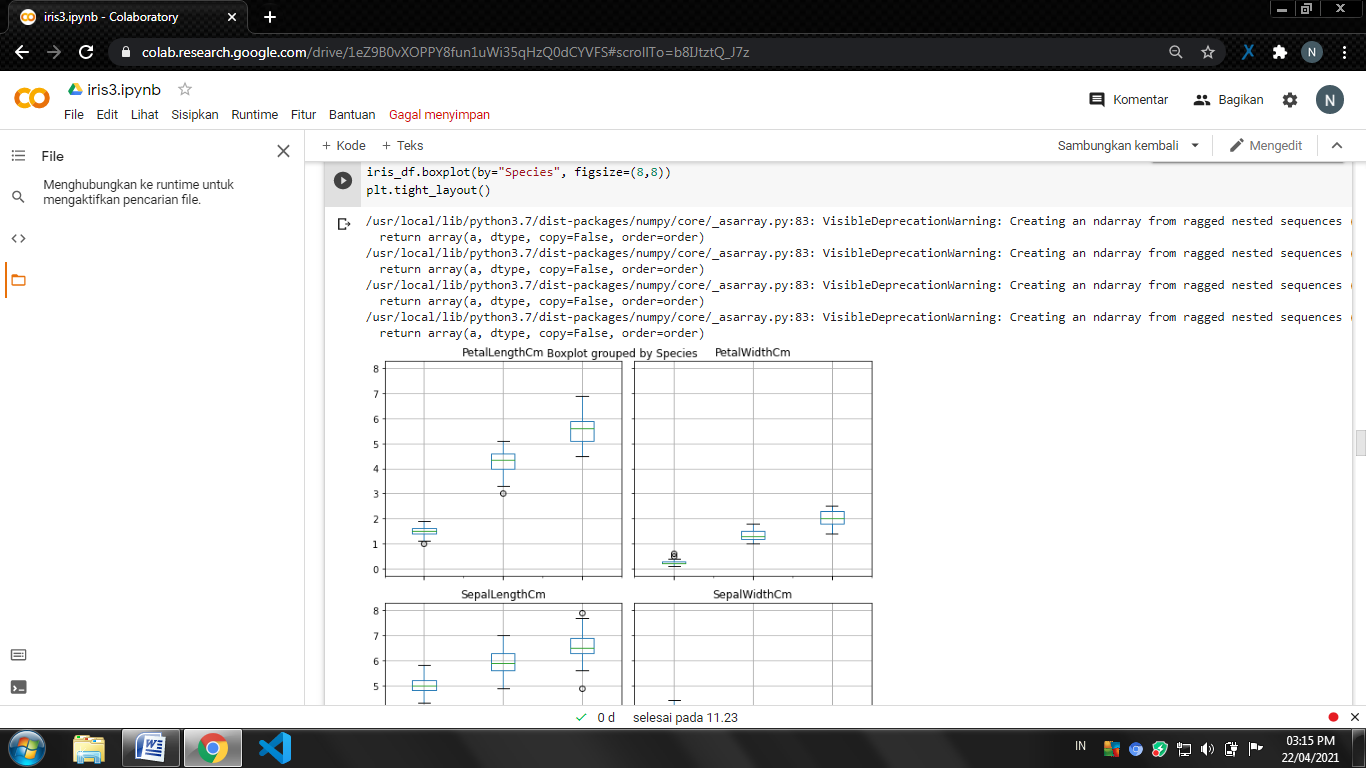


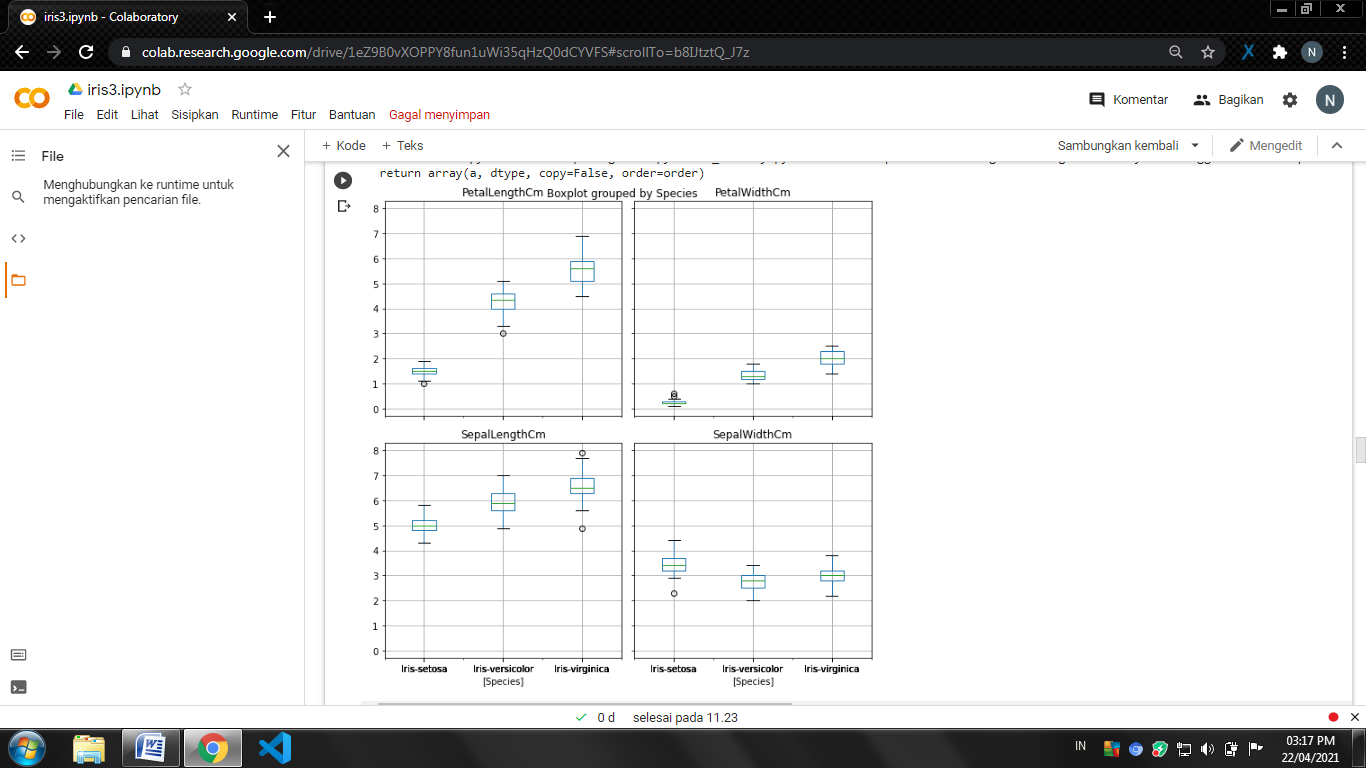
Cara lain utuk menggunakan boxplot dapat memanfaatkan *iris\_df.boxplot by=”Species”*

Code :



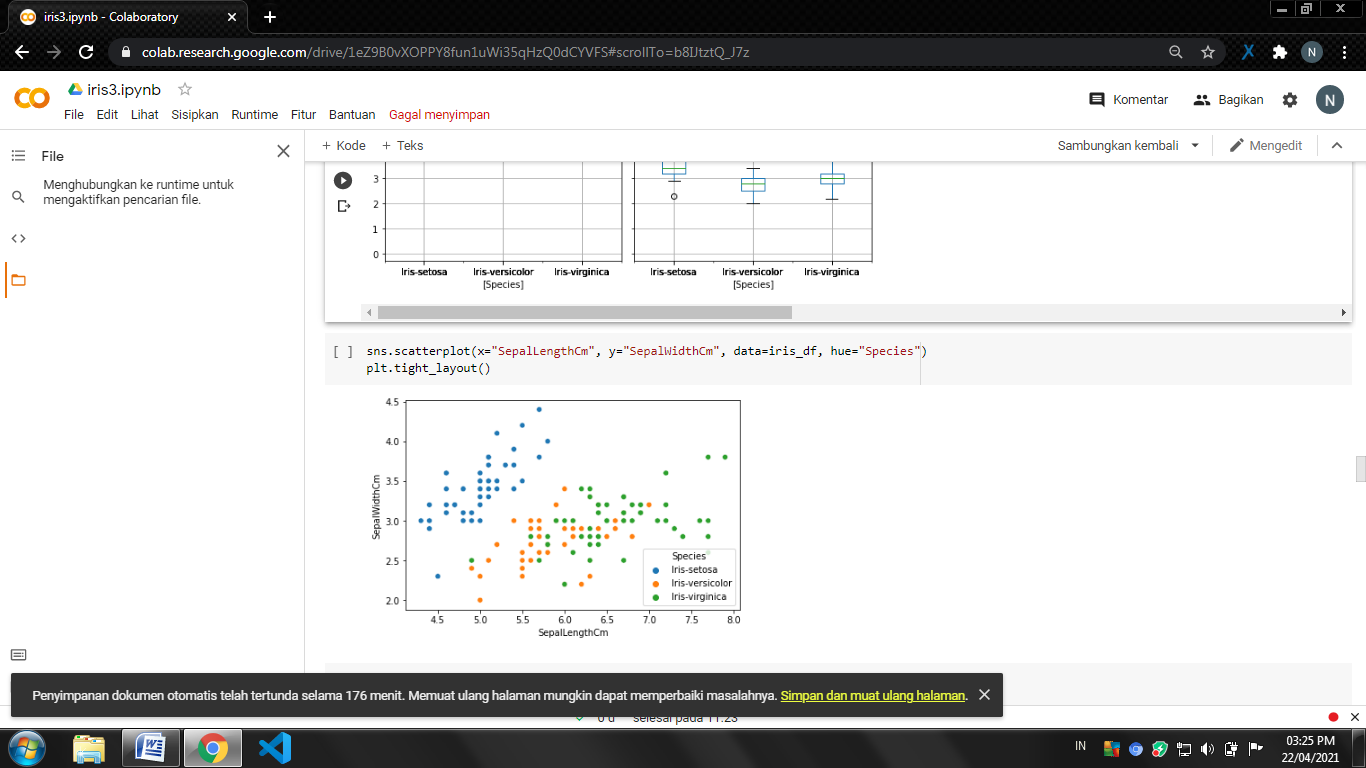
Output :





1. Scatter Plot

Untuk melakukan scatter plot perlu mengatur beberapa parameter yaitu datasetnya dan juga mengatur sumbu x,y untuk membandingkan dua variabel



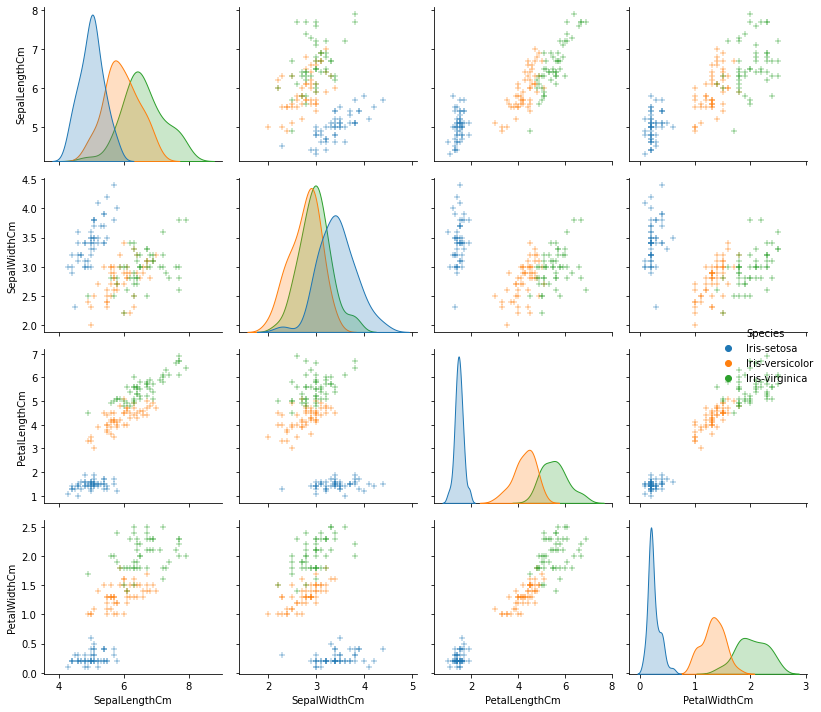
1. Pair Plot

Pair plot digunakan untuk melihat keseluruhan korelasi antar kolom

Code :

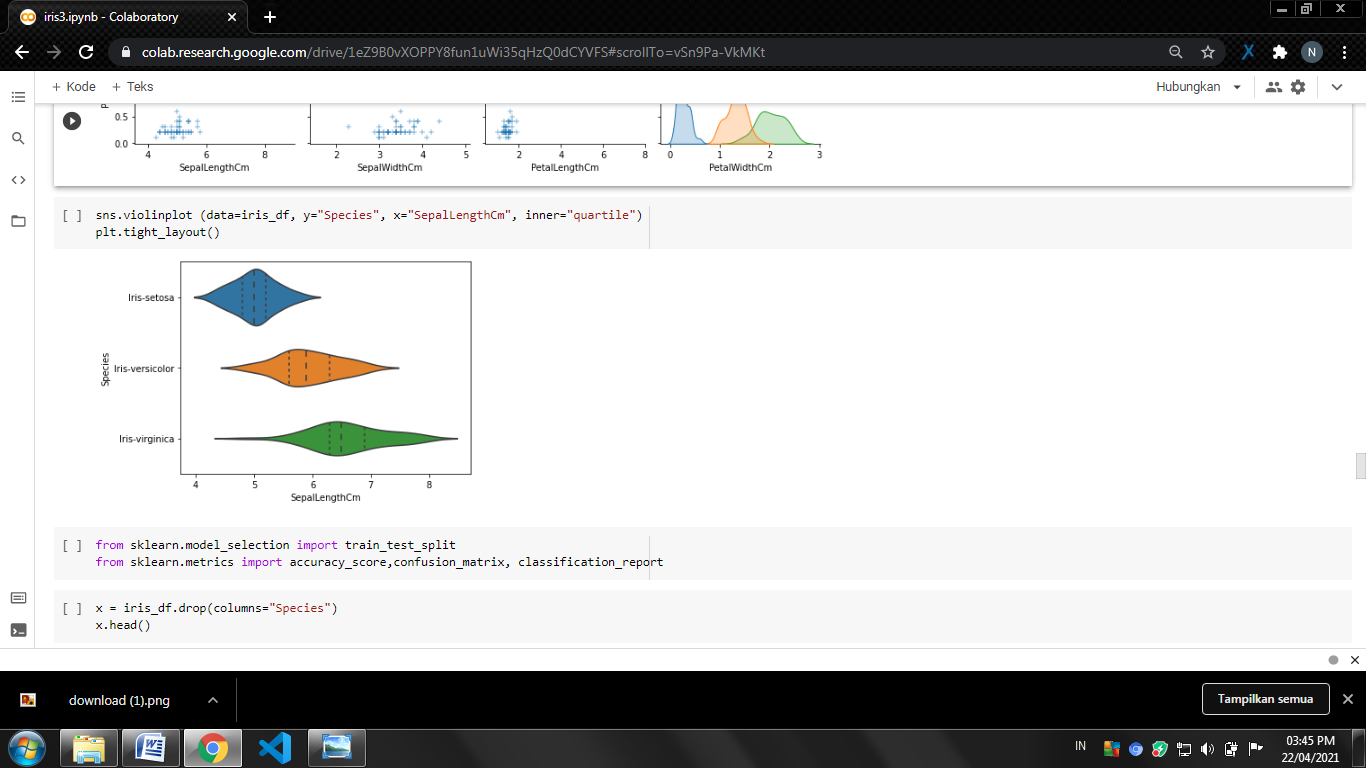


Output :



1. Violin Plot

Mengspesifikasi data dan sumbu x,y



Iris Dataset : Classification Models

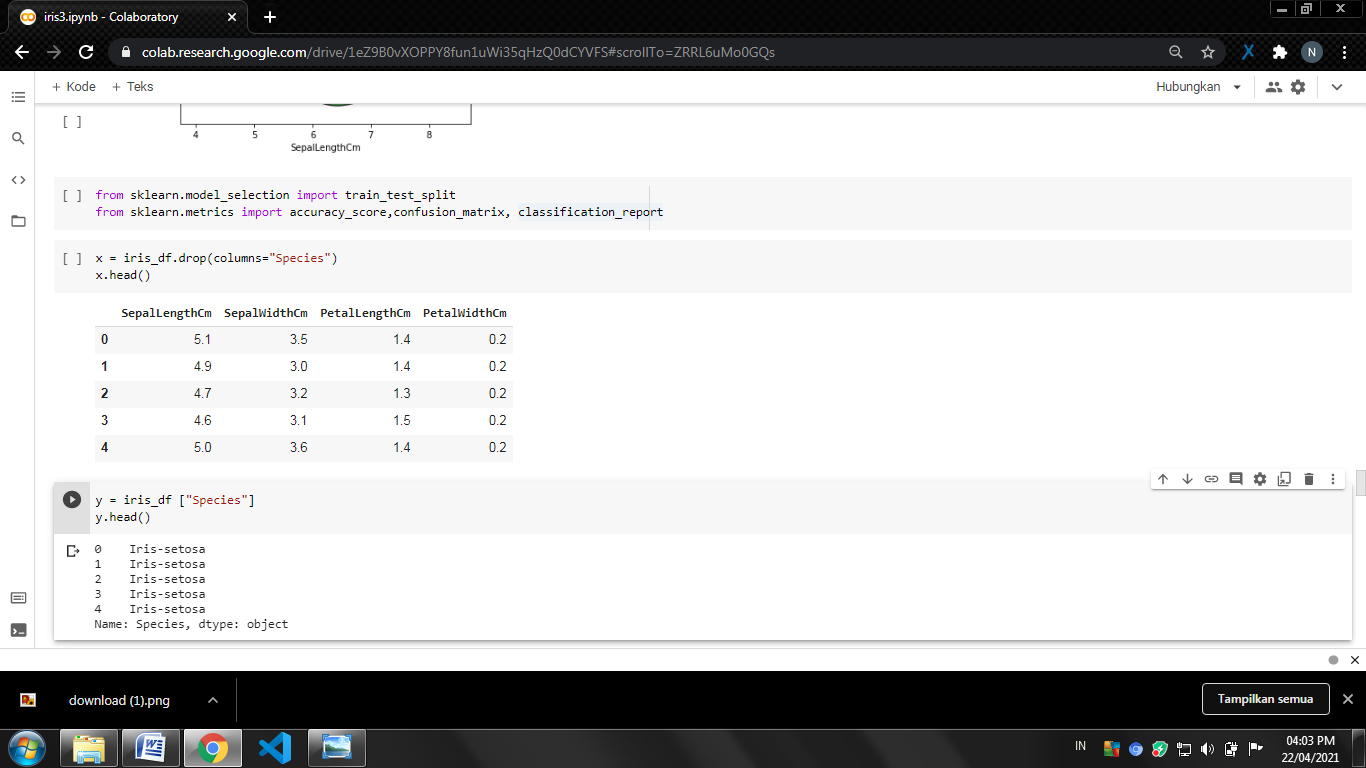
1. Import Modules

Untuk mengakses sklearn menggunakan dua modules yaitu model selection dan metrics

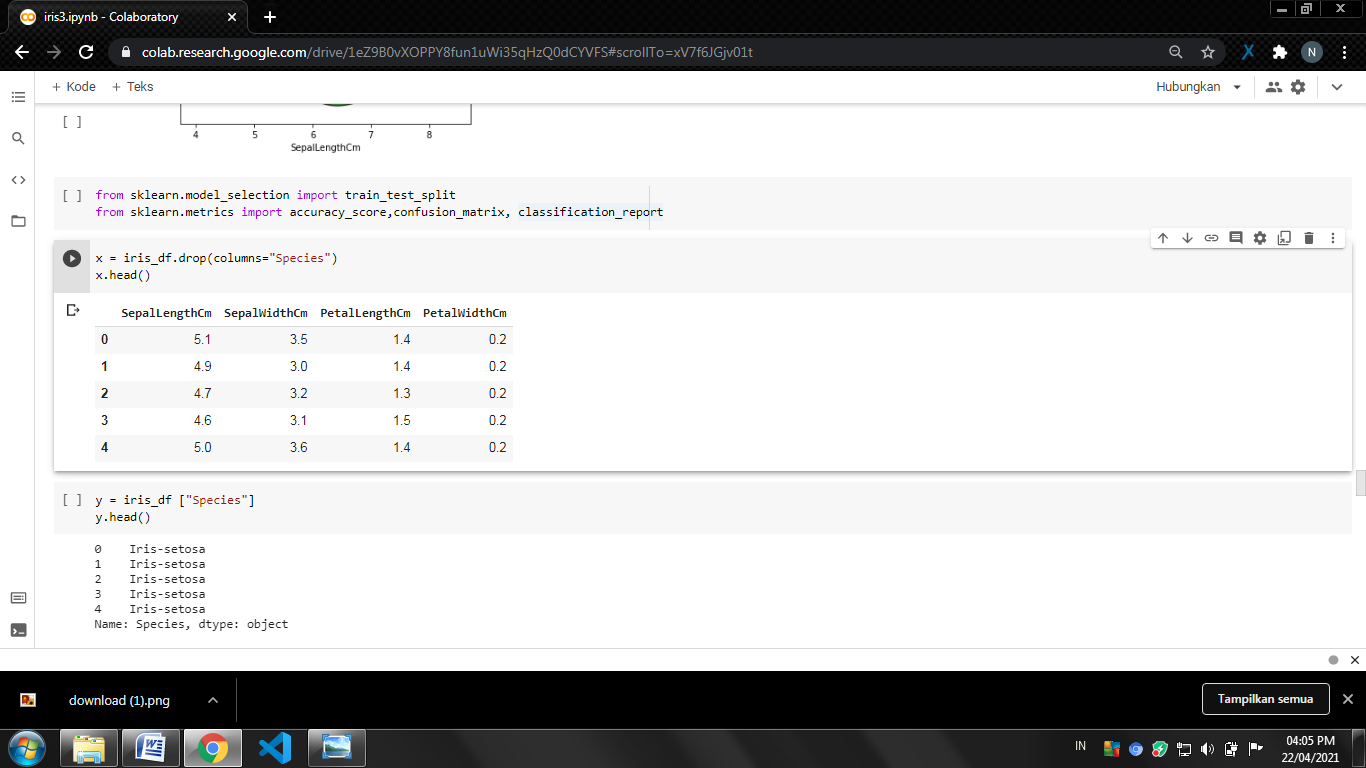


1. Dataset : Features & Class Label

Melakukan features selection dan features extraction, features yang akan di cermati ada empat



Label akan dimasukkan ke dalam variabel y



1. Split the dataset into a training set and a testing set

Disini memanfaatkan modules yang namanya *train\_test\_split*, dalam train test split harus menspesifikasikan x,y lalu test size atau ukuran testing dataset

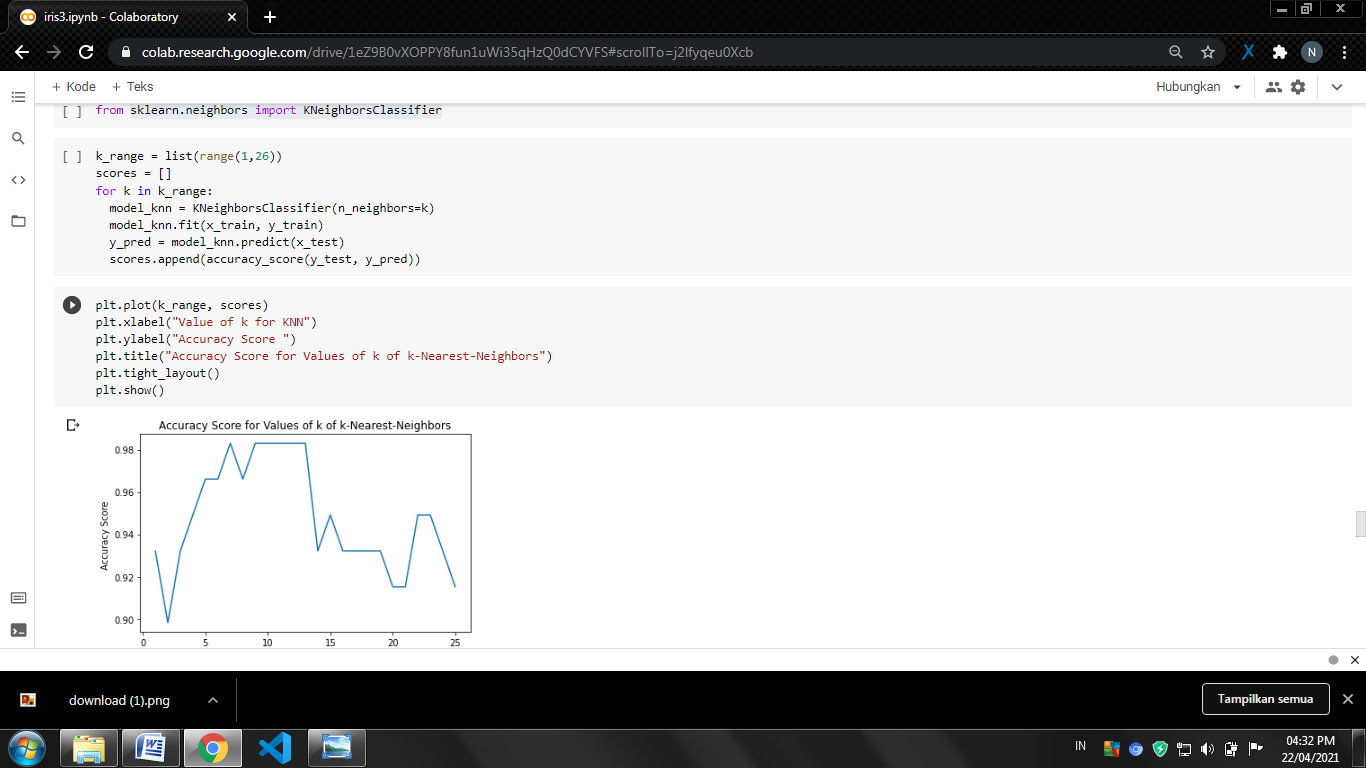


1. K Nearest Neighbors

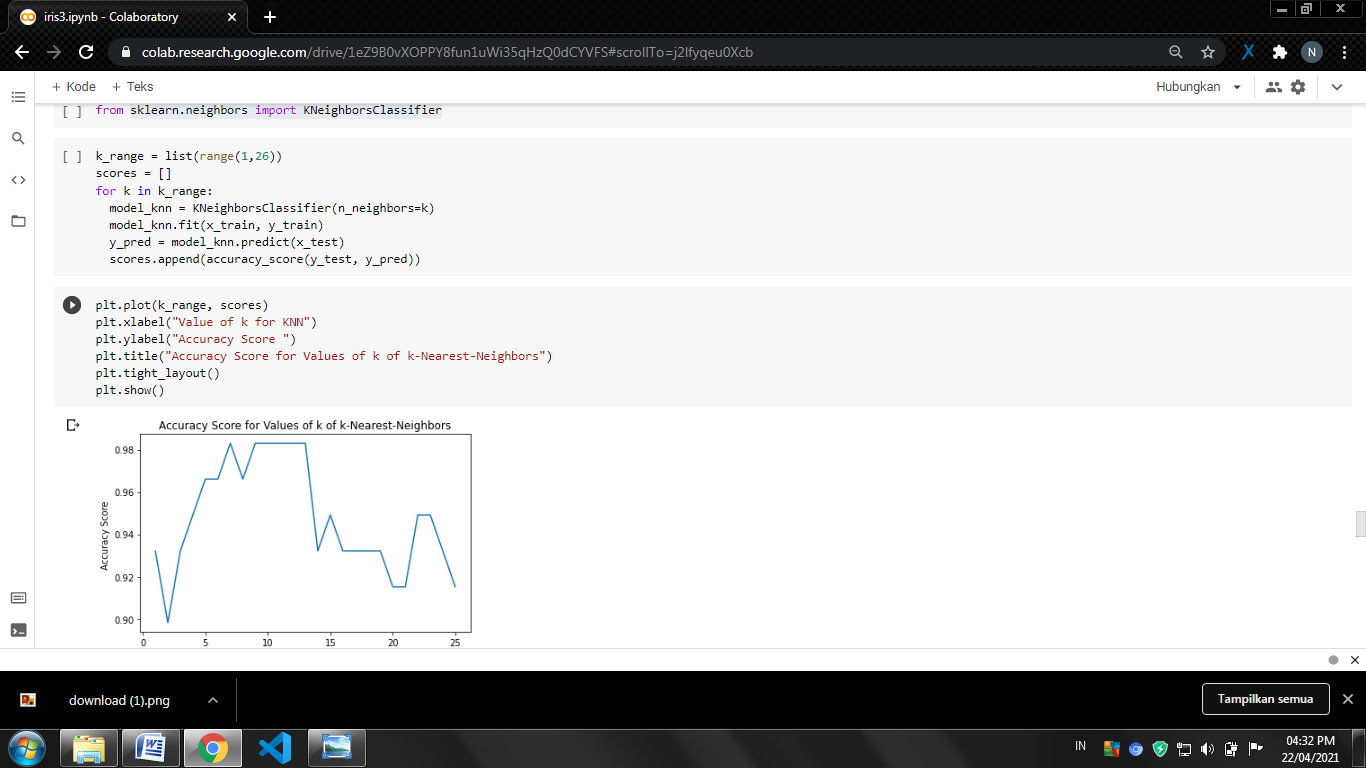
Import *sklearn.neighbors* ke *KNeighborsClassifier*



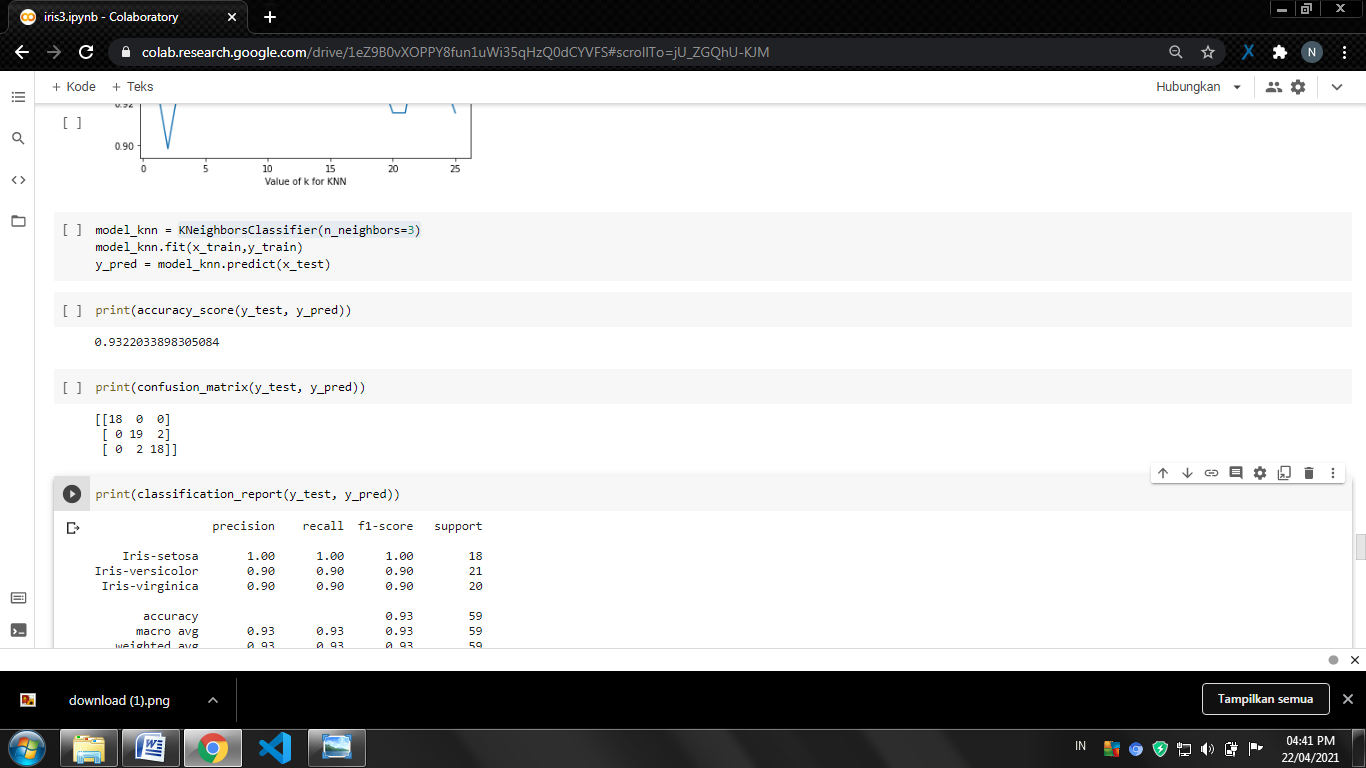
Membuat list, lalu konfigurasi pada algoritma sertakan parameter neighbors



Ploting hasil akurasi skor

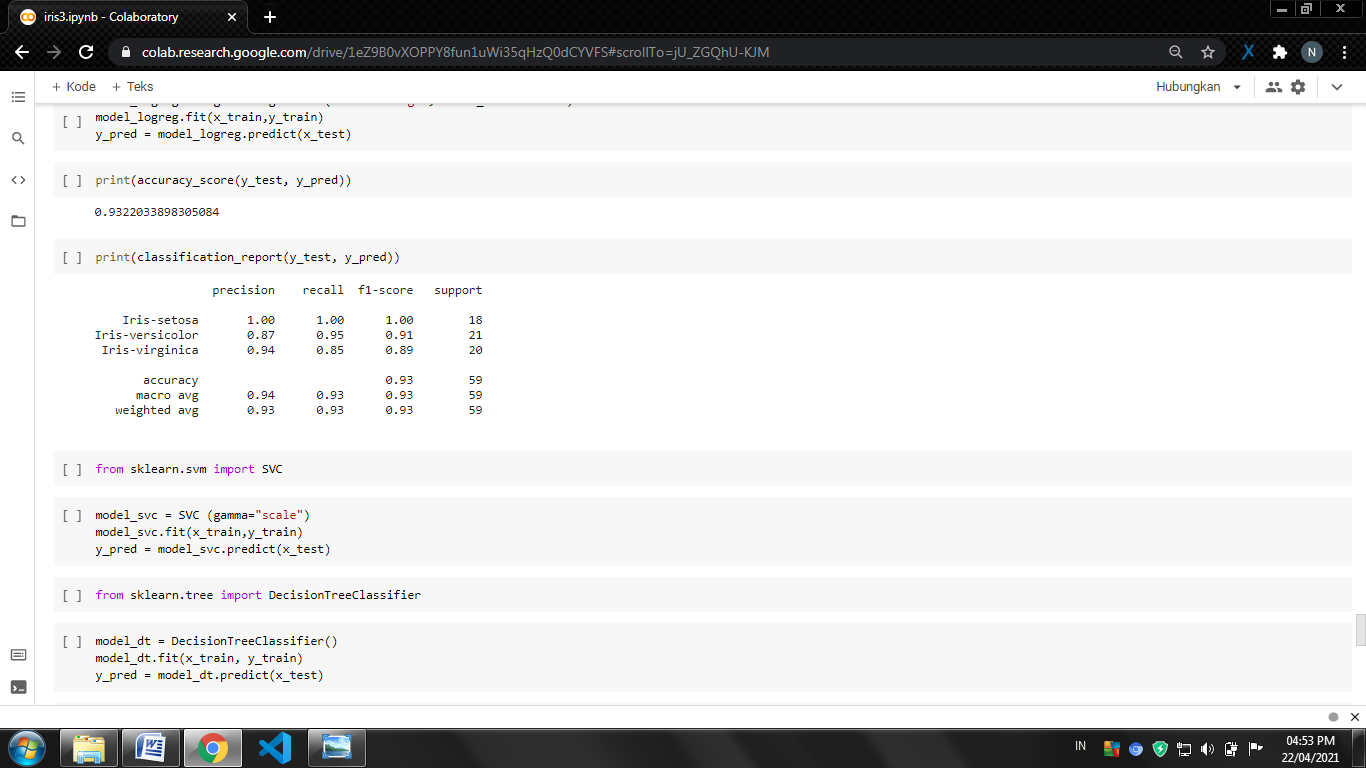


Misalkan jumlah neighbors tiga maka proses training bisa dilakukan dengan memanggil *KNeighborsClassifier(n\_neighbors=3)* lalu kenakan proses testing

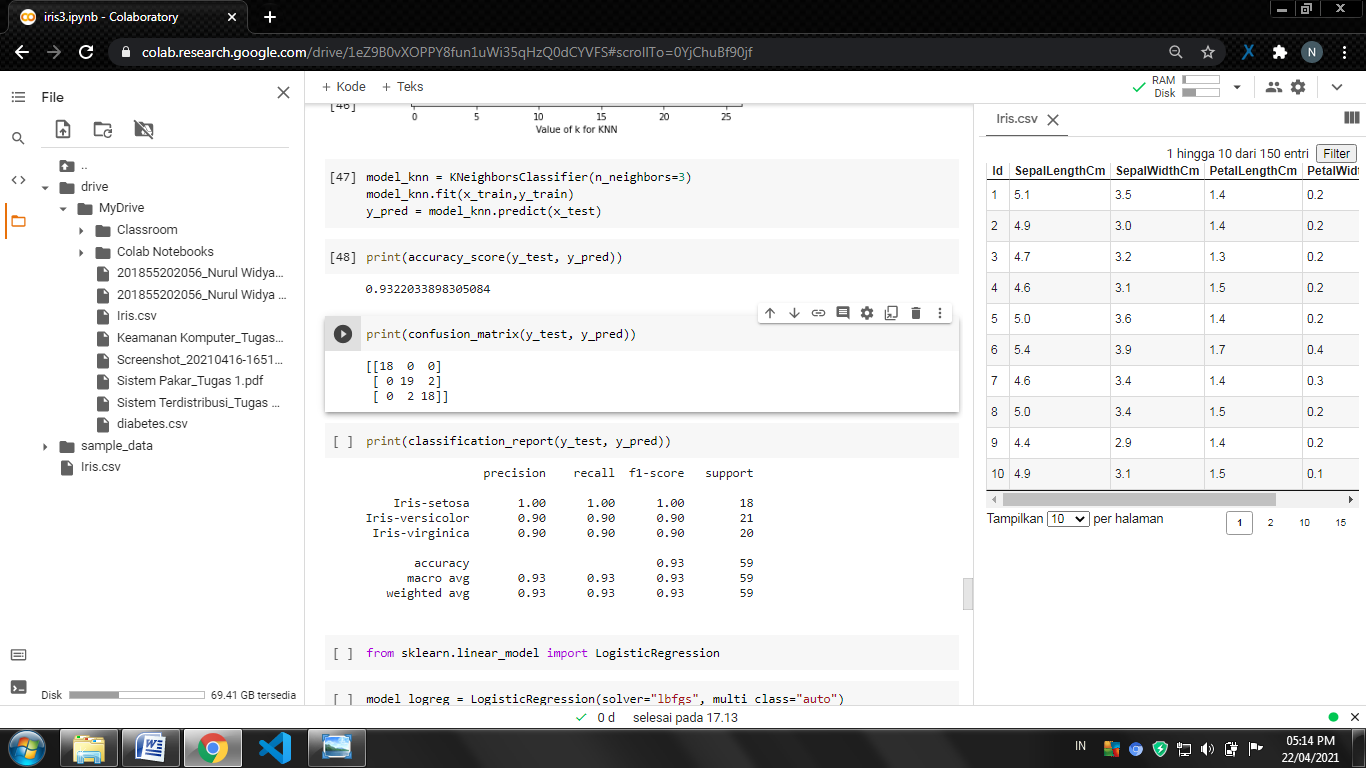


1. Accuracy Score

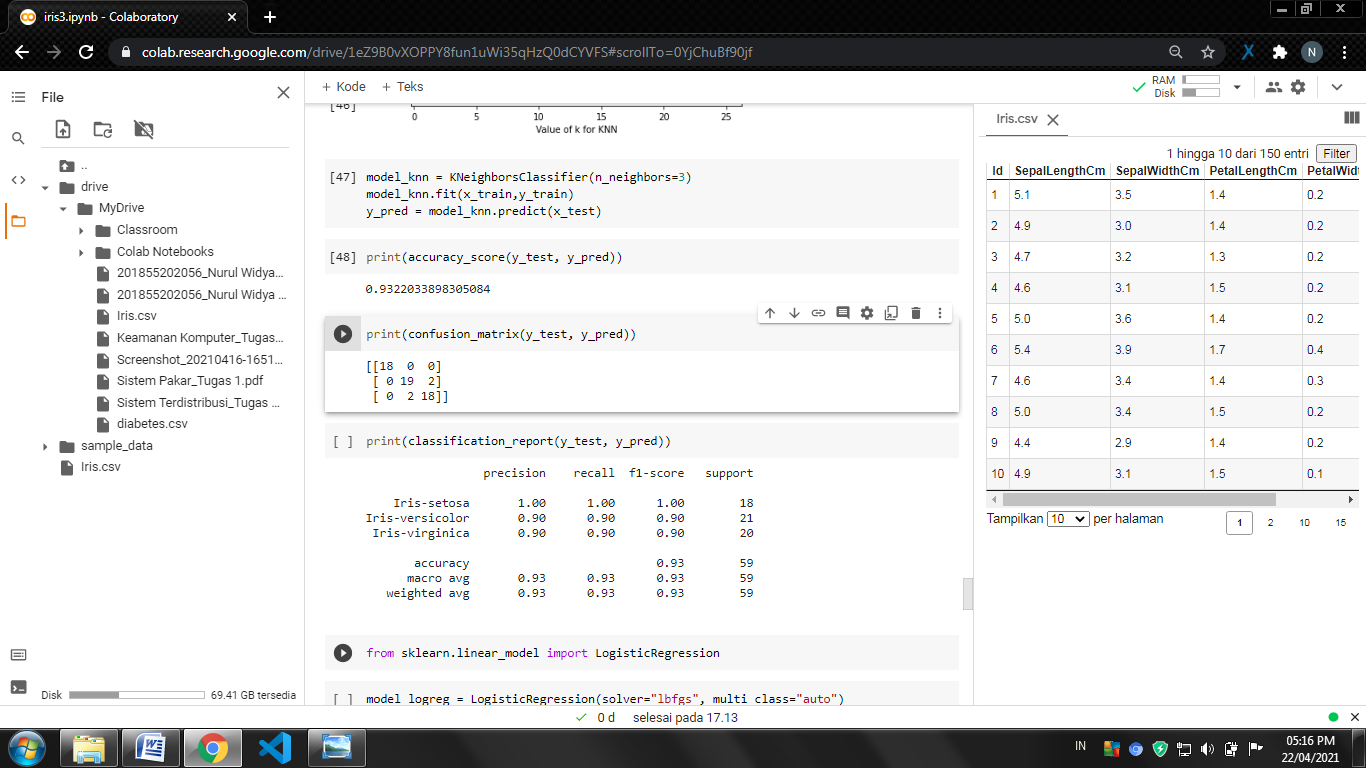
Evaluasi akurasi



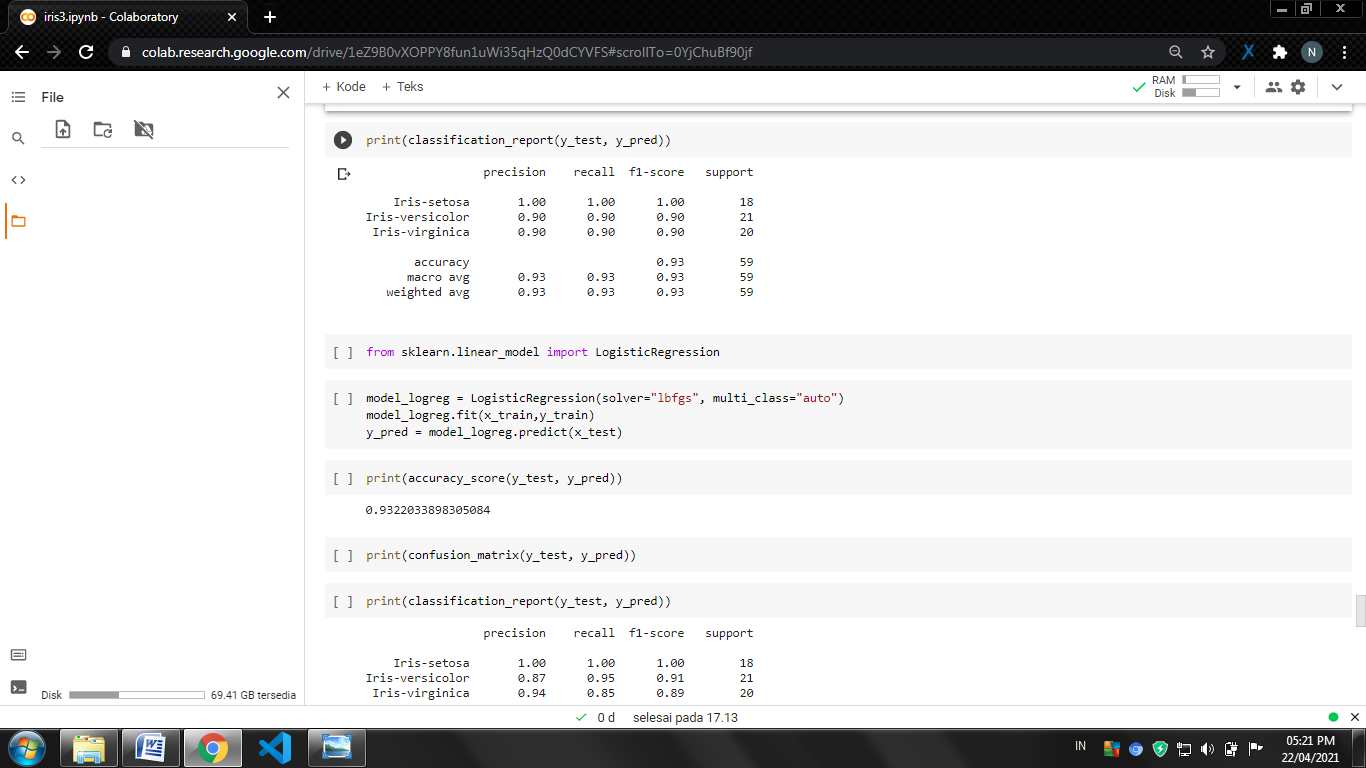
1. Confusion Matrix



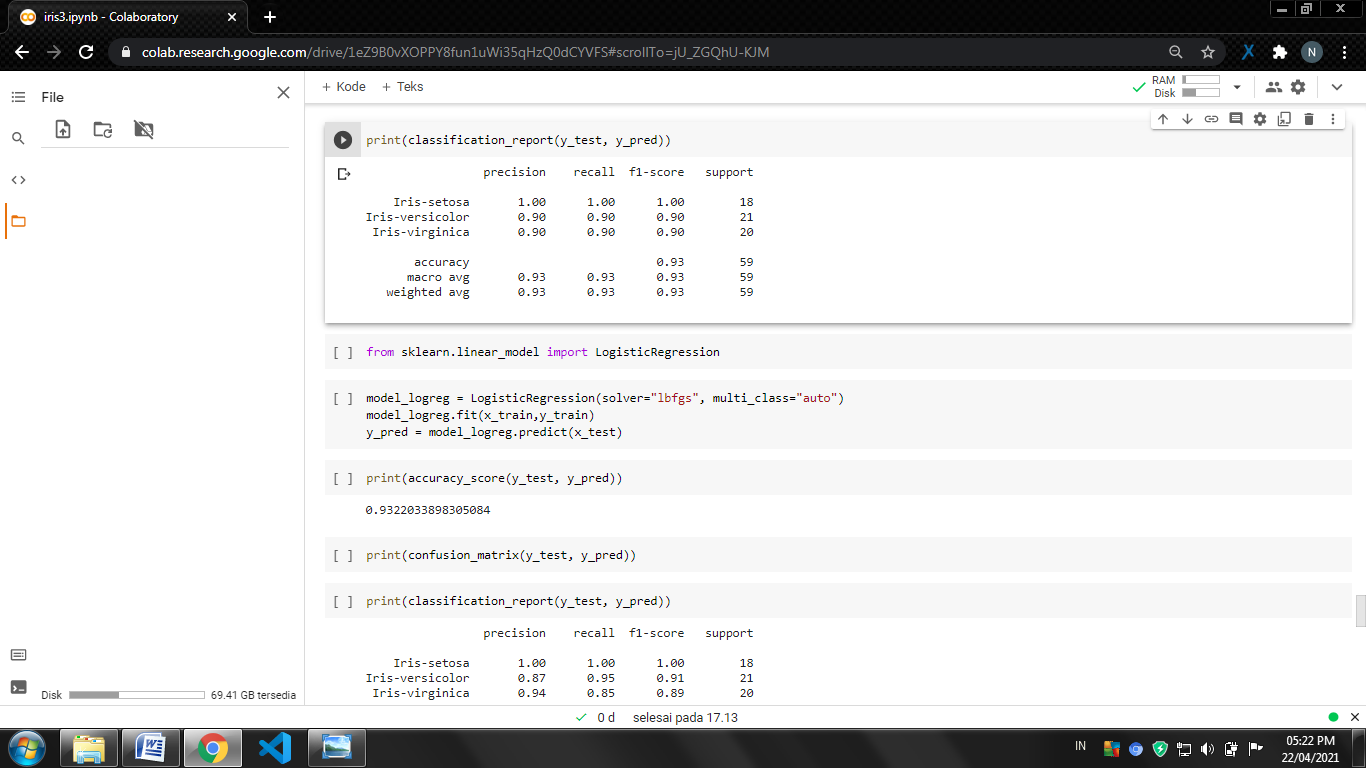
1. Classification Report



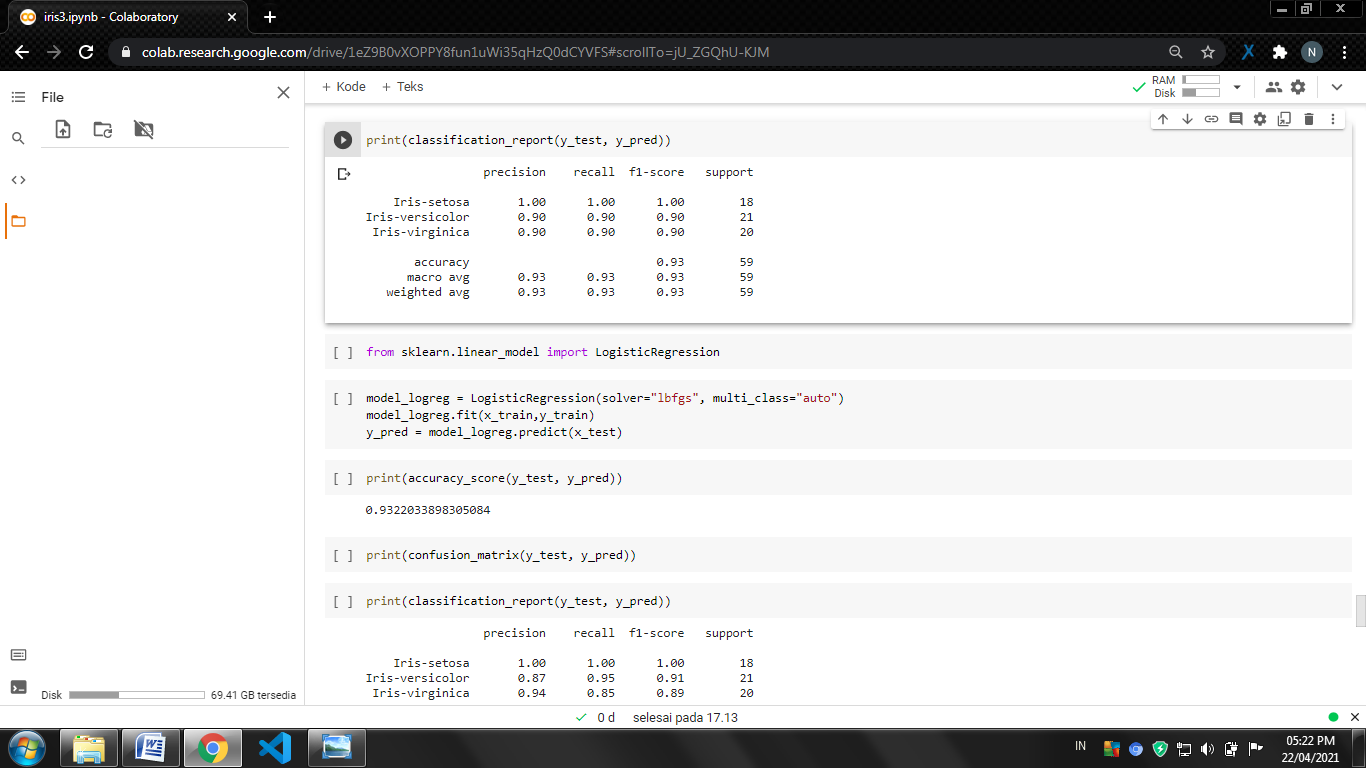
1. Logistic Regression



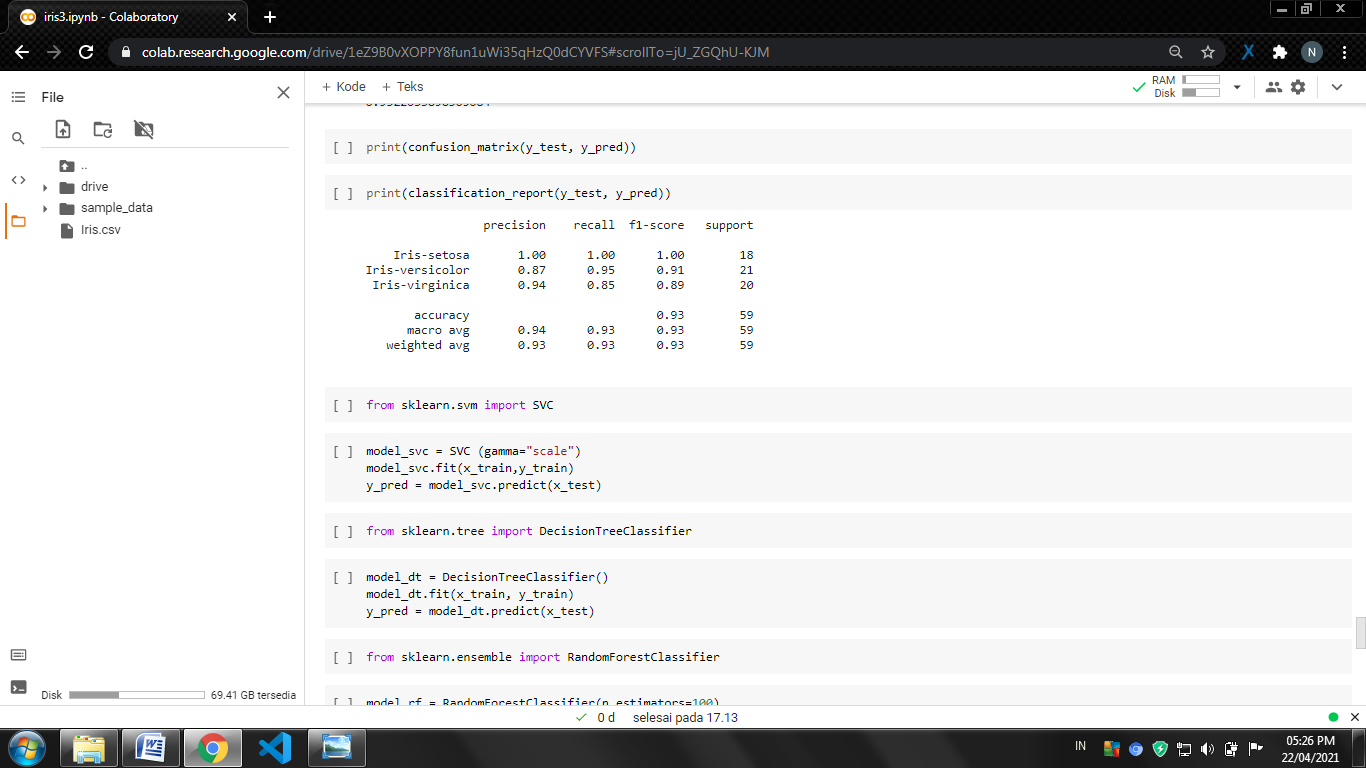
1. Accuracy Score



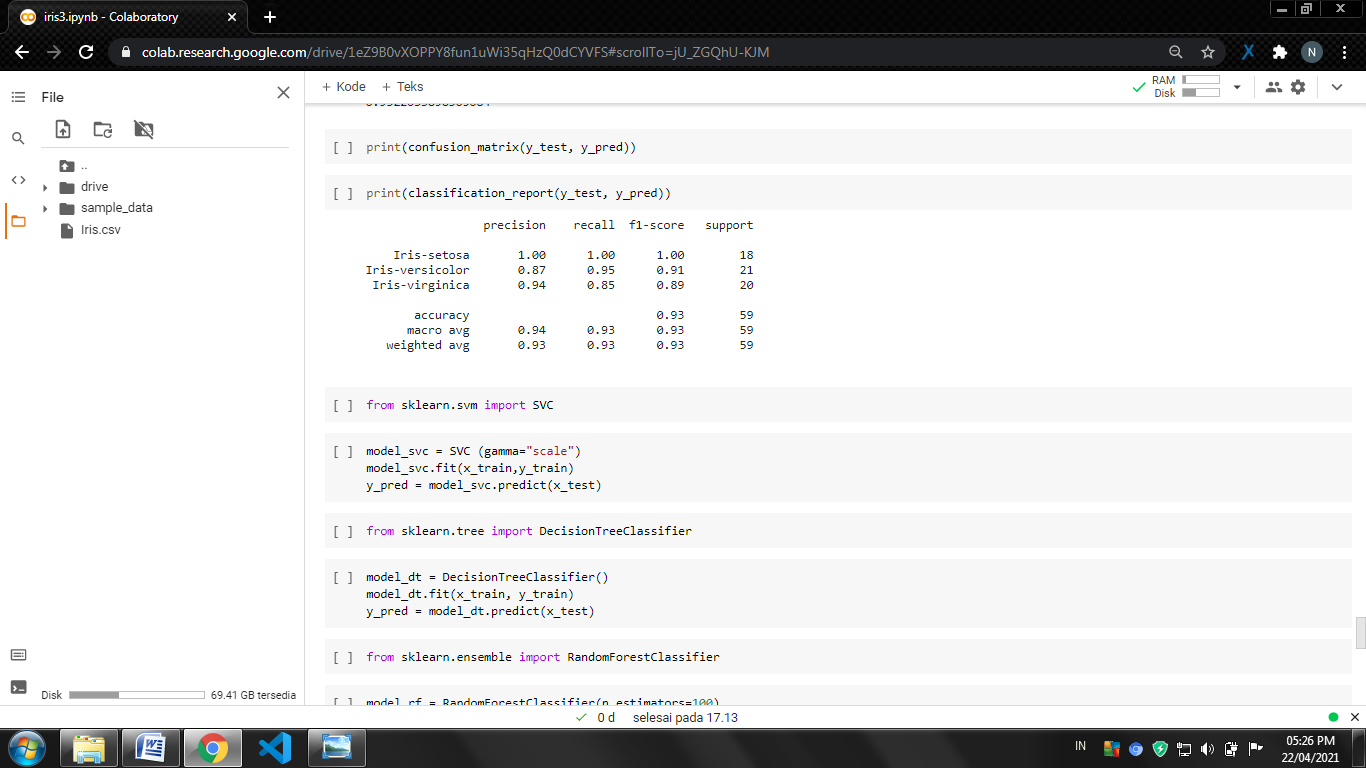
1. Confusion Matrix



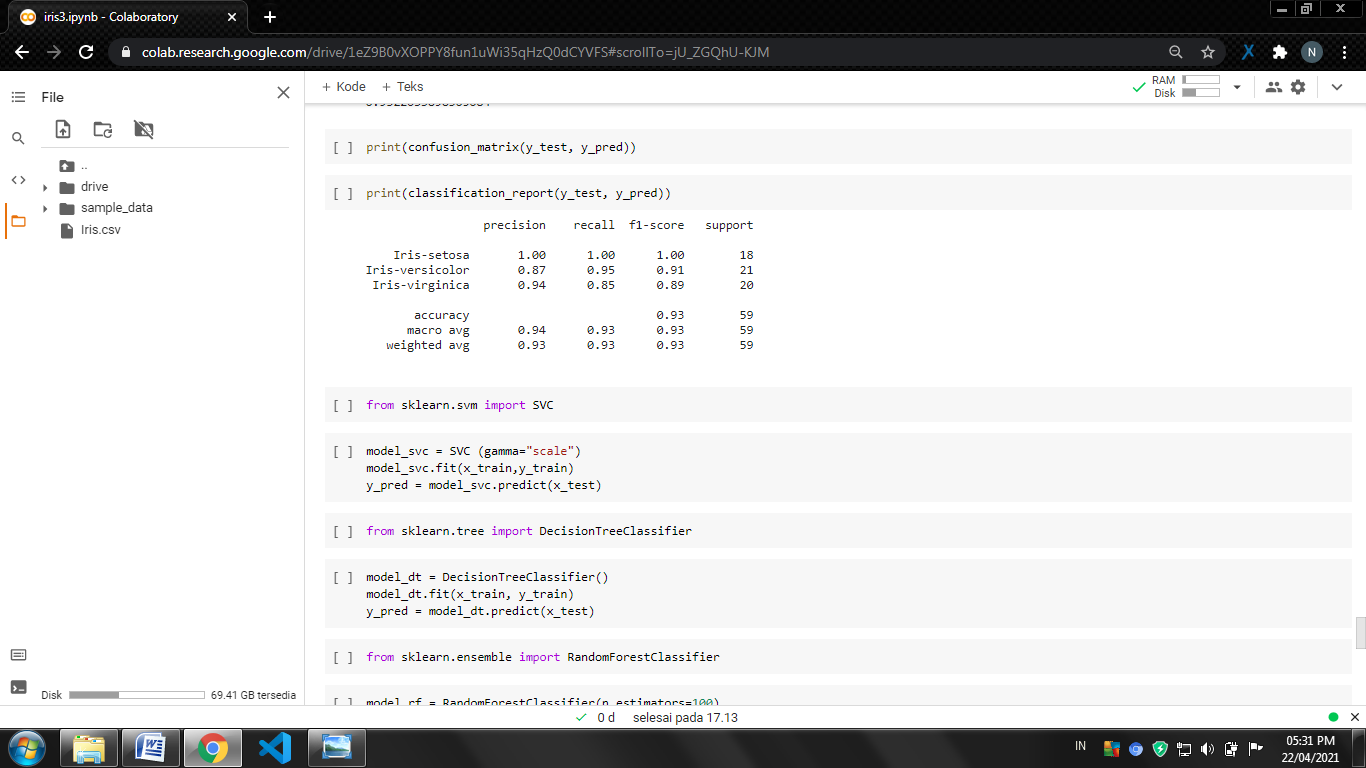
1. Classification Report



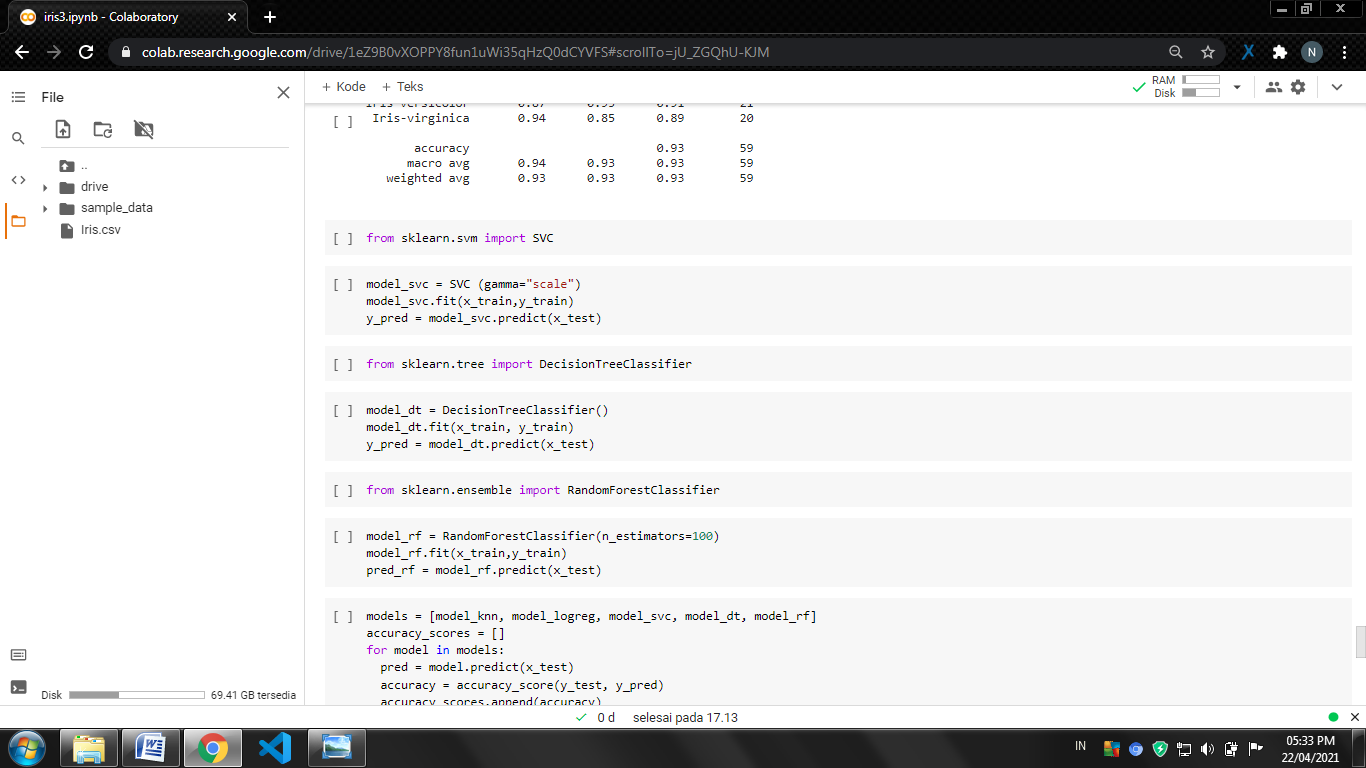
1. Support Vector Classifier



1. Decision Tree Classifier



1. Random Forest Classifier



1. Accuracy comparision for various models

