MACHINELOGISTICPROJECT

Inthisproject,IhaveusedLogisticRegression MachineLearningmodelfortheirisdatasetfromKagglewebsite PROBLEMSTATEMENT:

TheIrisflowerdatasetorFisher'sIrisdata

setisamultivariatedatasetintroducedbytheBritishstatisticianand biologistRonaldFisherinhis1936paperTheuseofmultiple measurementsintaxonomicproblemsasanexampleoflinear discriminantanalysis.ItissometimescalledAnderson'sIrisdataset becauseEdgarAndersoncollectedthedatatoquantifythe morphologicvariationofIrisflowersofthreerelatedspecies.Twoof thethreespecieswerecollectedintheGaspéPeninsula"allfromthe samepasture,andpickedonthesamedayandmeasuredatthe sametimebythesamepersonwiththesameapparatus".

MLMETHODOLOGY:

Logisticregressionisastatisticalmodelthatinits

basicformusesalogisticfunctiontomodelabinarydependent variable,althoughmanymorecomplexextensionsexist.In regressionanalysis,logisticregression(orlogitregression)is estimatingtheparametersofalogisticmodel(aformofbinary regression).

DATASETDESCRIPTION:

Somerelevantcolumnsinthedataset

· Sepallength(cm)

·Sepalwidth(cm)

·petallength(cm)

·petalwidth(cm)

·Flower\_type

PRE\_PROCESSING:

Pre\_processingreferstothe

transformationsappliedtoourdatabeforefeedingittothealgorithm.

1)%matplotlibinline importmatplotlib.pyplotasplt importnumpyasnp importpandasaspd

fromsklearn.model\_selectionimporttrain\_test\_split fromsklearnimportdatasets fromsklearn.metricsimport mean\_squared\_error,accuracy\_score,classification\_report fromsklearn.preprocessingimportStandardScaler fromsklearn.linear\_modelimportLogisticRegression

2)iris=datasets.load\_iris()

3)print(iris.data.shape)

p

r

i

n

t

(

i

r

i

s

.

t

a

r

g

e

t

.

s

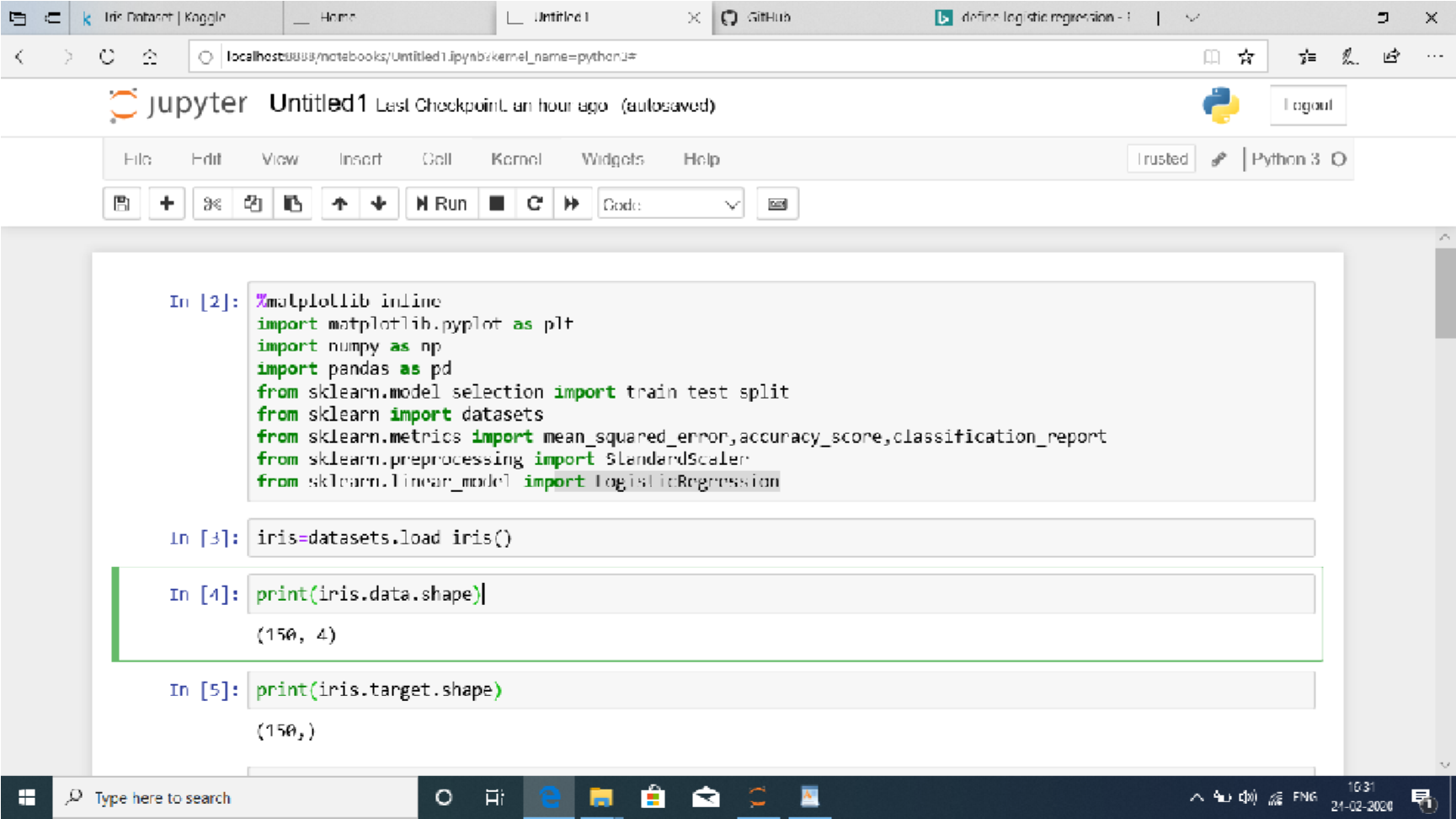
h

a

p

e

)



p

r

i

n

t

(

i

r

i

s

.

f

e

a

t

u

r

e

\_

n

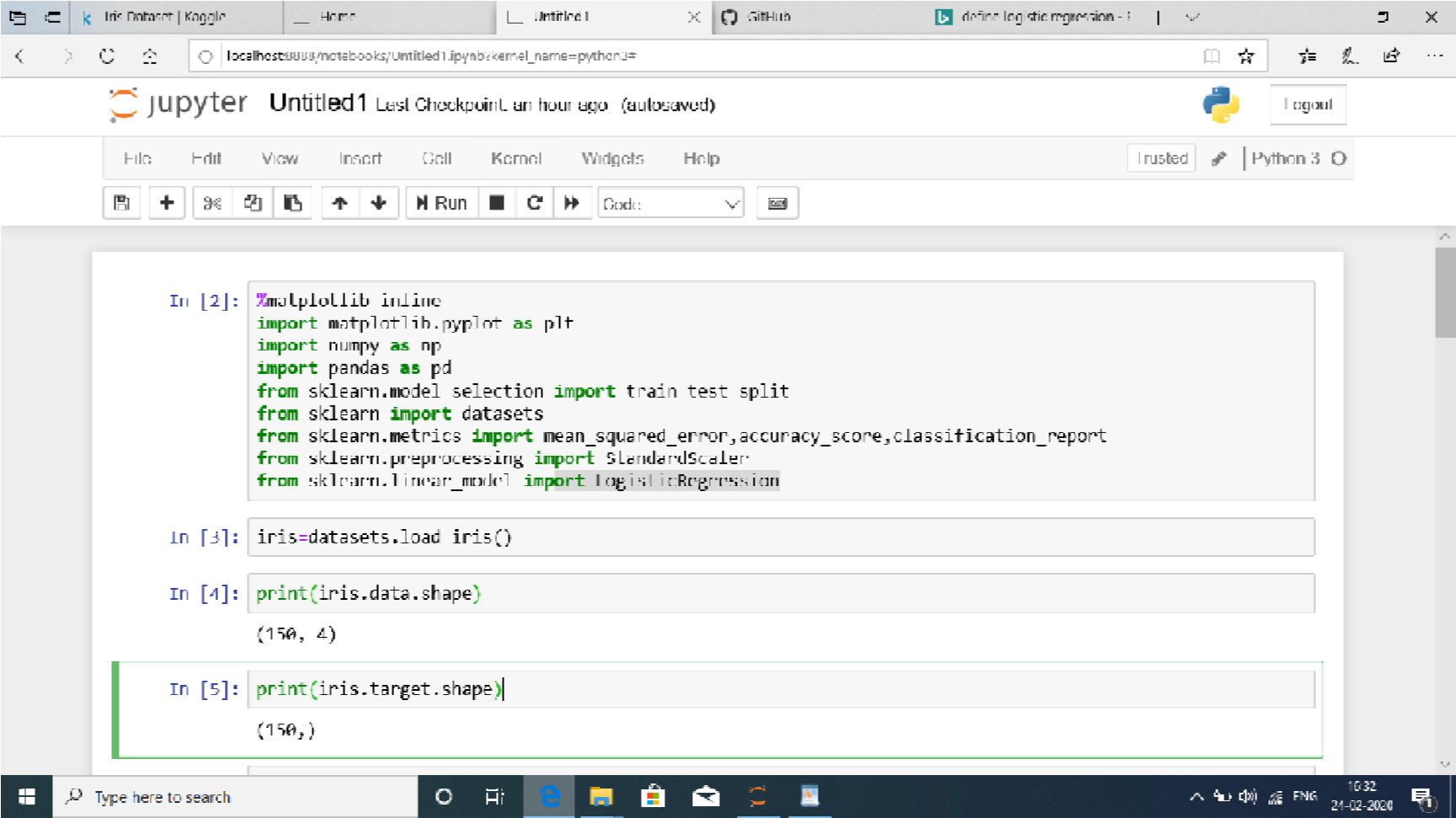
a

m

e

s

)



print(iris.target)

p

r

i

n

t

(

'

T

a

r

g

e

t

:

1

:

{

}

,

2

:

{

}

,

3

:

{

}

.

f

o

r

m

a

t

(

n

p

.

s

u

m

(

i

r

i

s

.

t

a

r

g

e

t

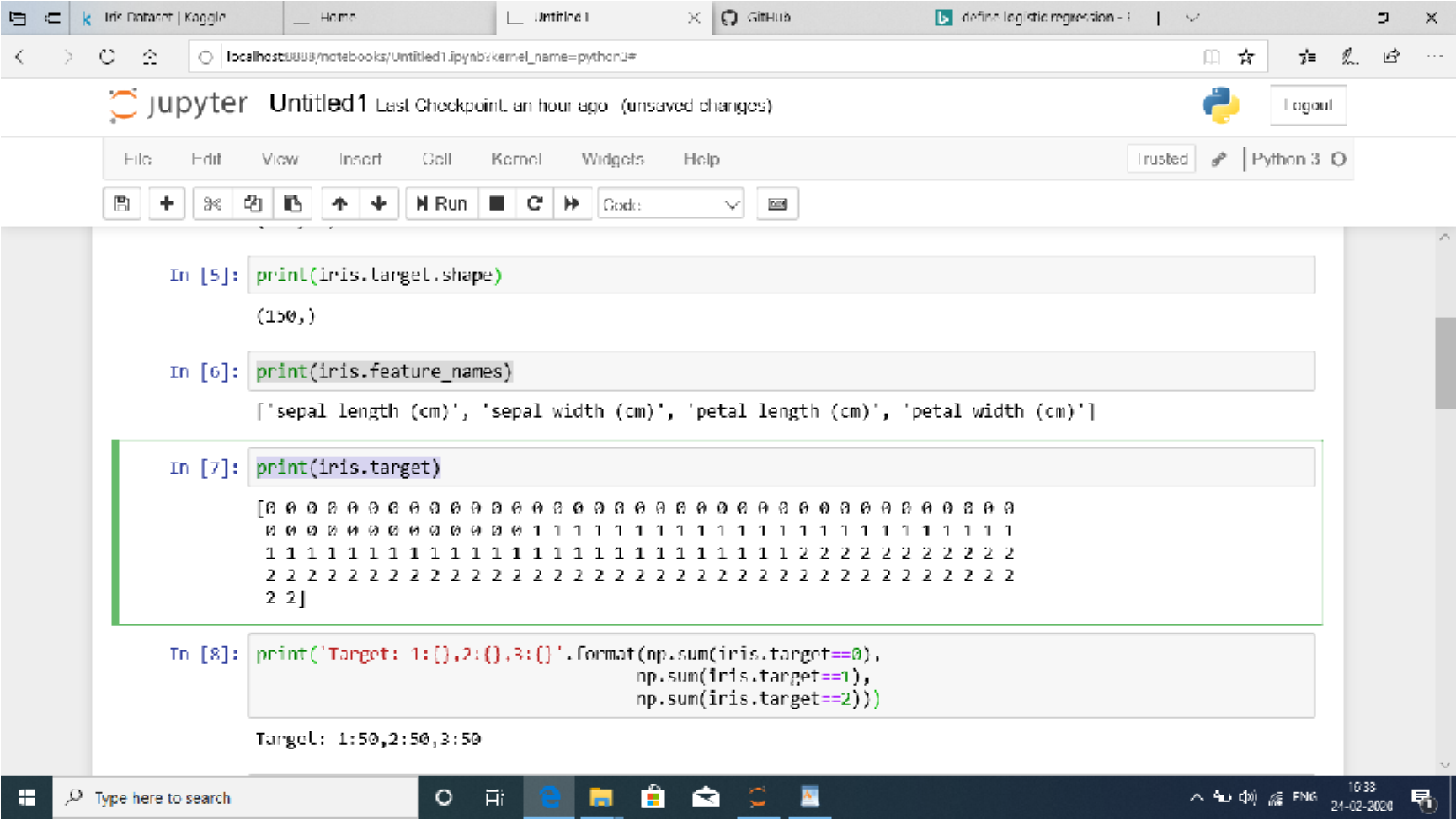
=

=

0

)

,



np.sum(iris.target==1),

n

p

.

s

u

m

(

i

r

i

s

.

t

a

r

g

e

t

=

=

2

)

)

)

d

f

=

p

d

.

D

a

t

a

F

r

a

m

e

(

i

r

i

s

.

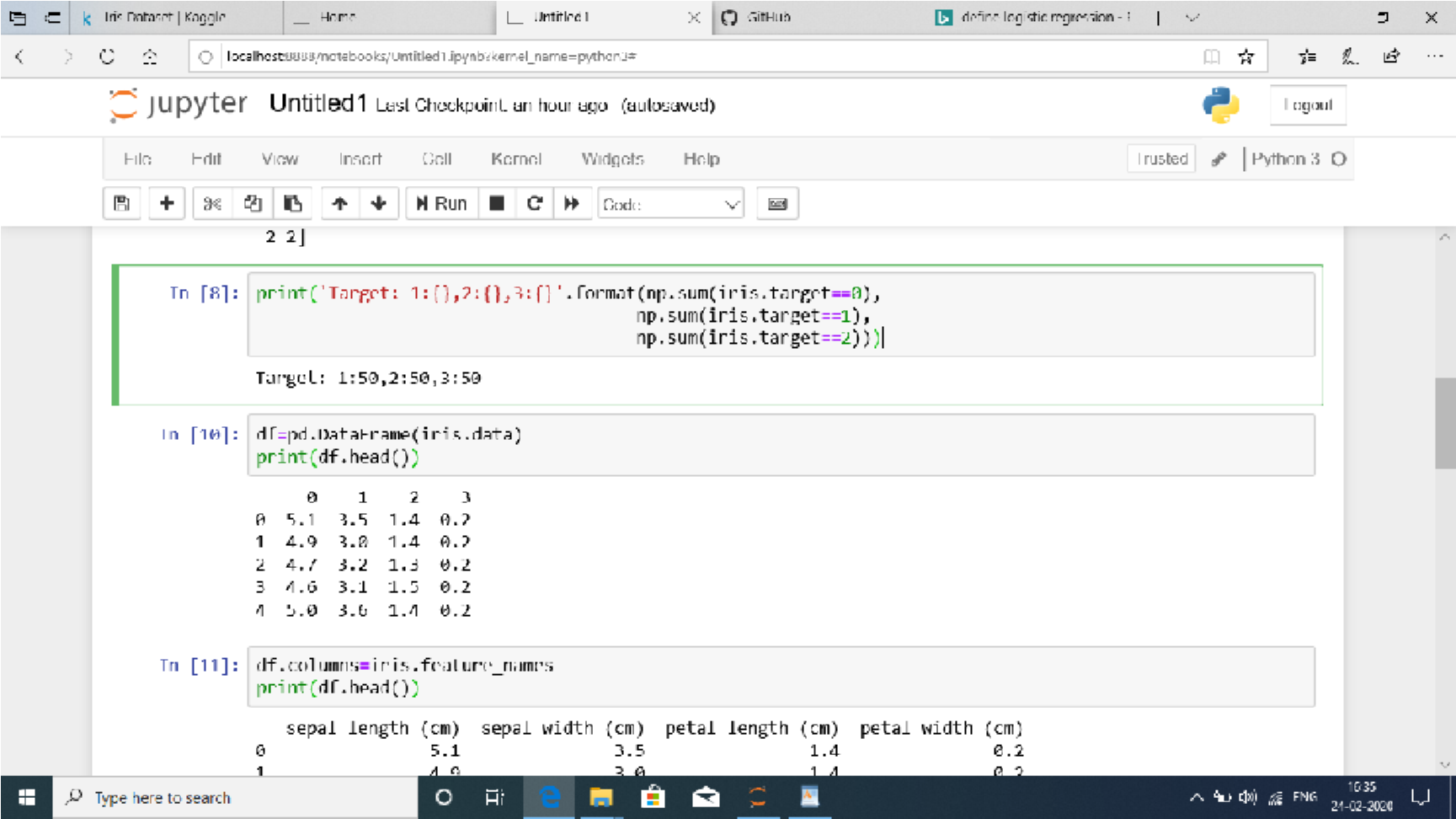
d

a

t

a

)



print(df.head())

d

f

.

c

o

l

u

m

n

s

=

i

r

i

s

.

f

e

a

t

u

r

e

\_

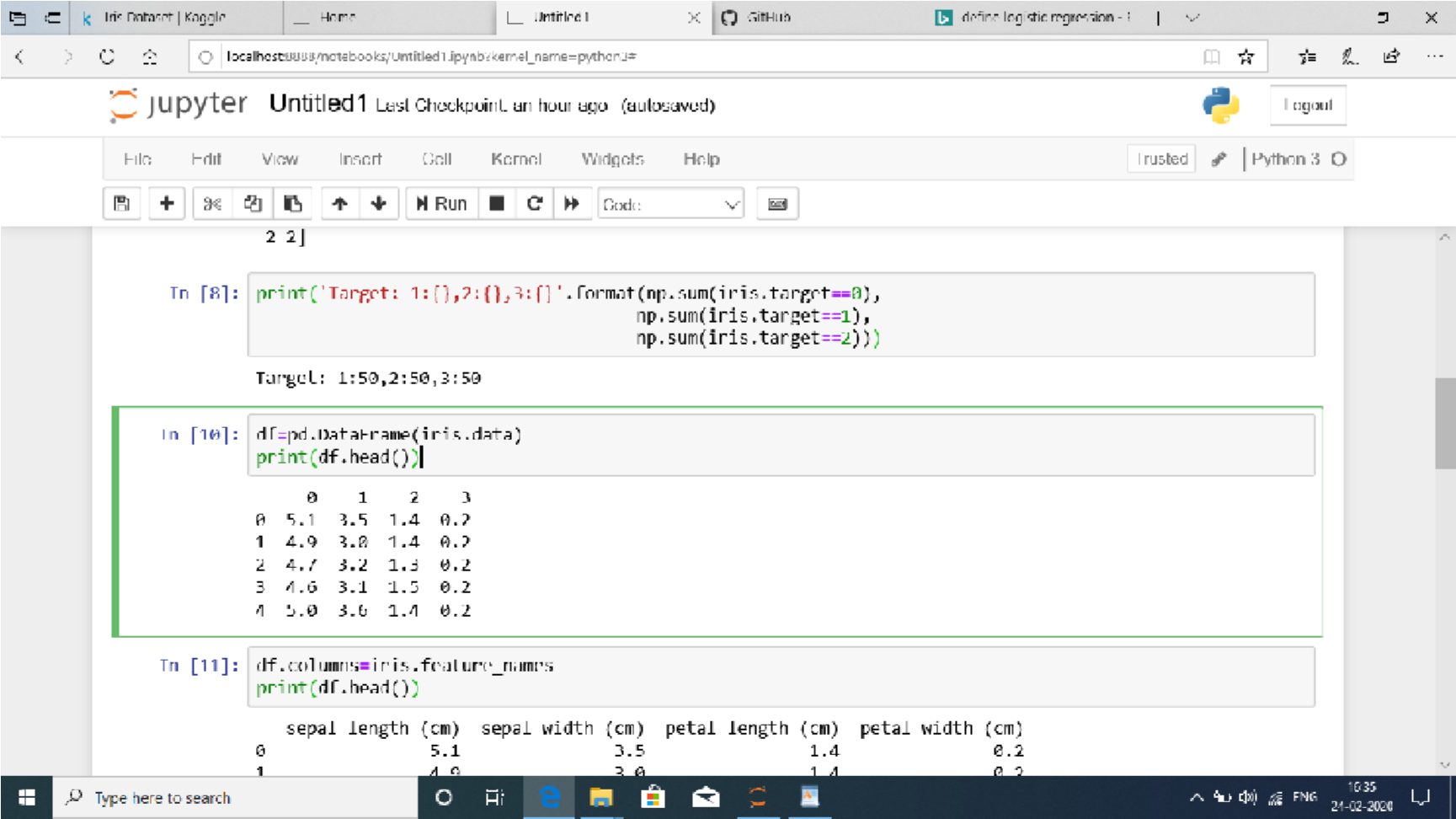
n

a

m

e

s



print(df.head()) df['flowe\_type']=iris.target

p

r

i

n

t

(

d

f

.

h

e

a

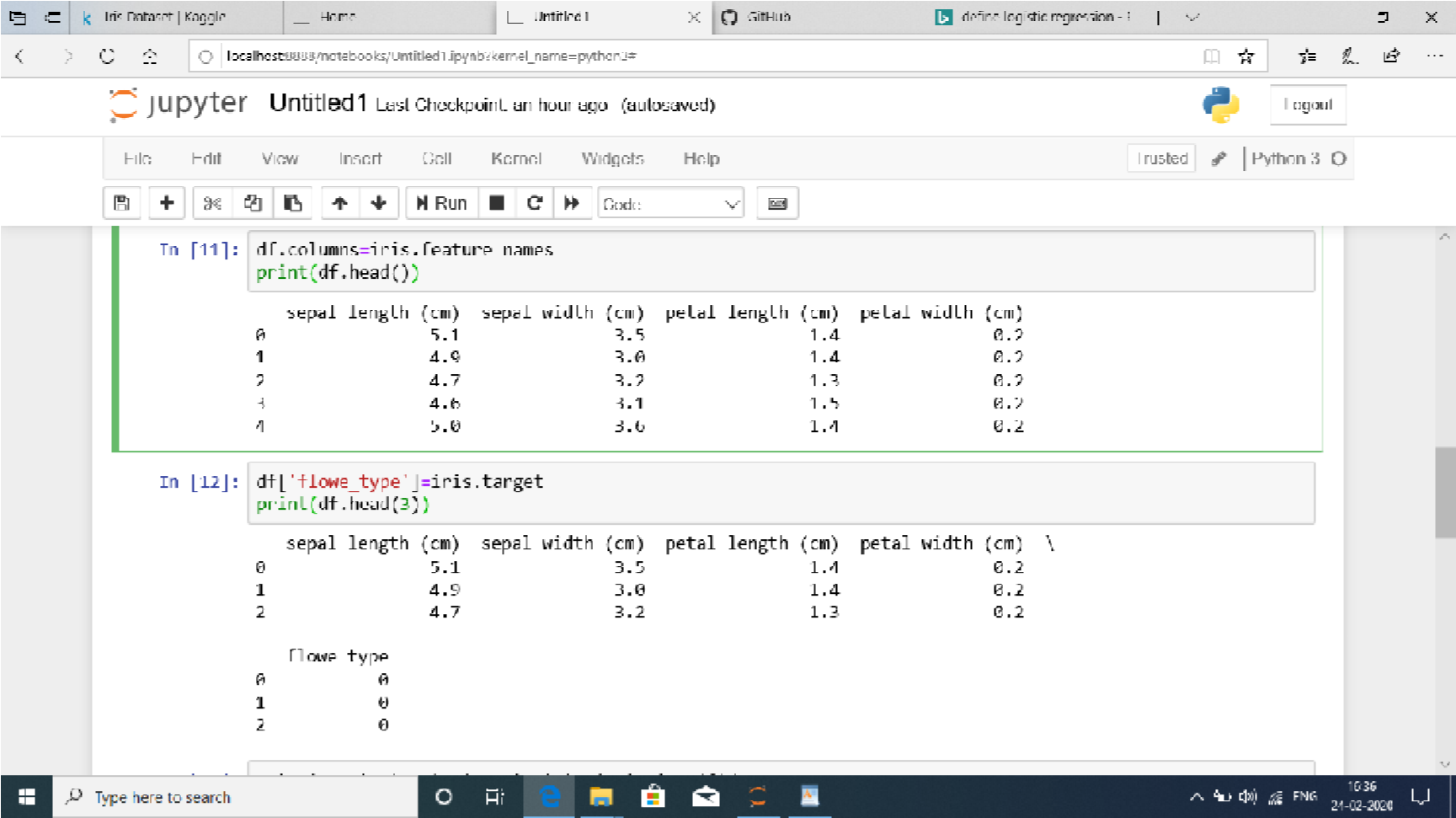
d

(

3

)

)



Building,trainingandevaluationofallMLmodels x\_train,x\_test,y\_train,y\_test=train\_test\_split( iris.data,iris.target,test\_size=0.25,random\_state=42) scaler=StandardScaler() x\_train\_std=scaler.fit\_transform(x\_train) x\_test\_std=scaler.fit\_transform(x\_test) print(x\_train\_std.shape) print(x\_test\_std.shape)

p

r

i

n

t

(

y

\_

t

r

a

i

n

.

s

h

a

p

e

)

p

r

i

n

t

(

y

\_

t

e

s

t

.

s

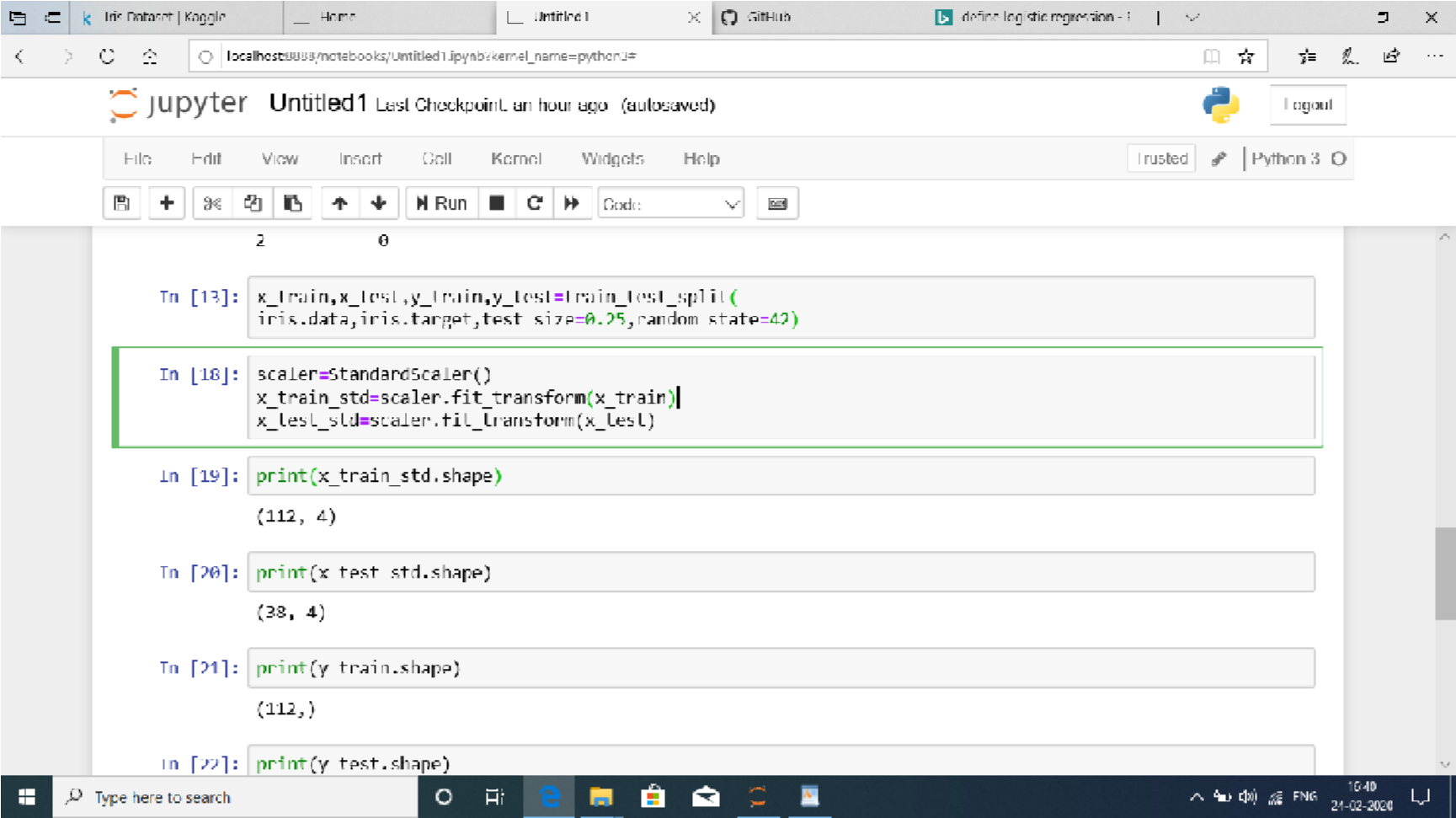
h

a

p

e

)



plt.scatter(x\_train[:,:1],x\_train[:,1:2]) plt.xlabel('SepalLength') plt.ylabel('Sepalwidth') plt.title('IrisDataset-Trainingset') plt.show()

C

O

N

C

L

U

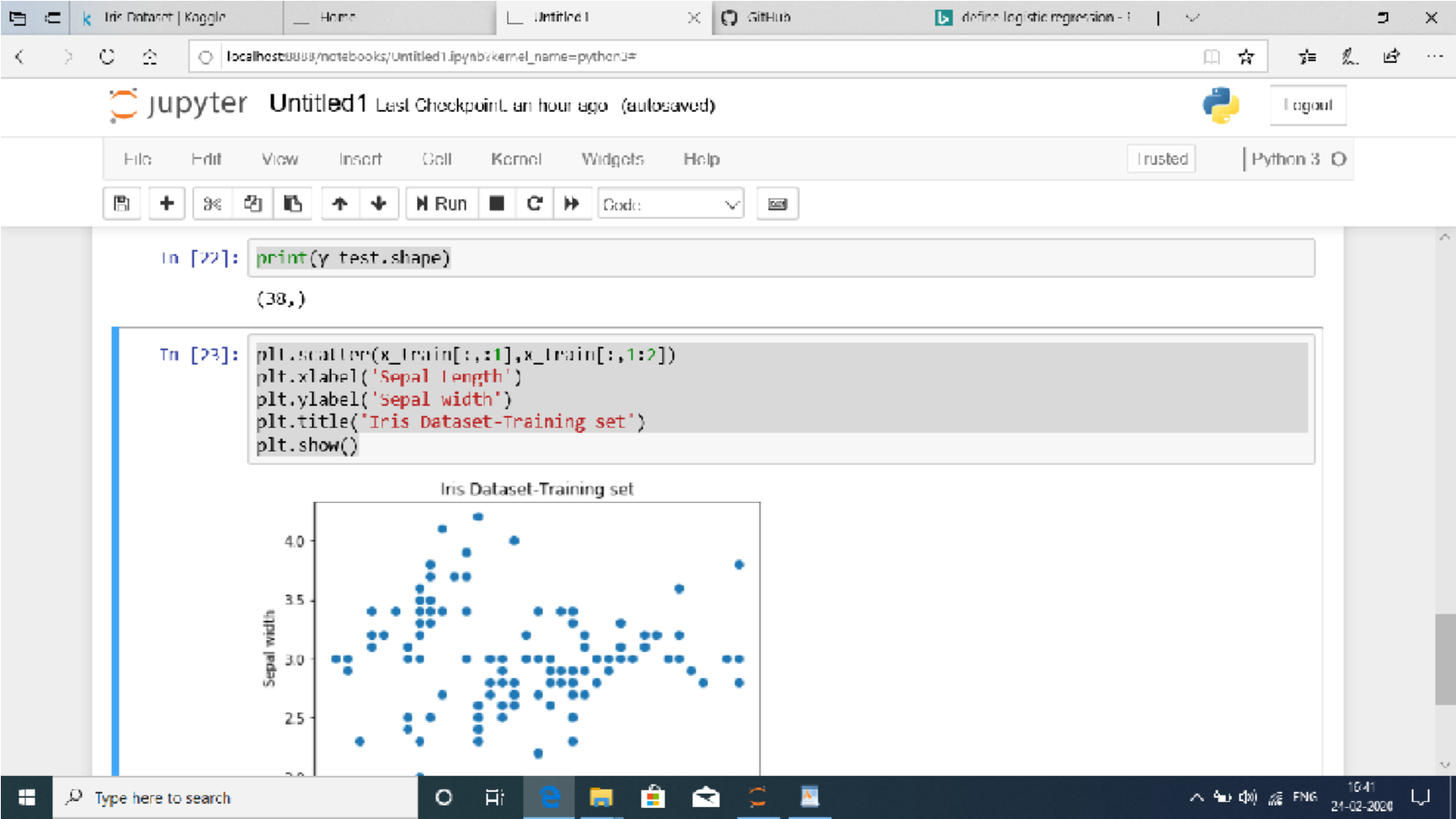
S

I

O

N

:



Modelweightsofdatasetsarecalculated

byusingLogisticRegressionandcalculatethetrainingandtesting dataset.