Text

Description automatically generated

A screenshot of a computer

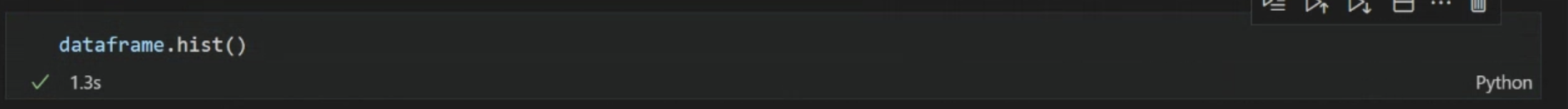
Description automatically generated with medium confidence

Graphical user interface

Description automatically generated

Text

Description automatically generated



Graphical user interface, text, application, chat or text message

Description automatically generated

Chart, bar chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text

Description automatically generated

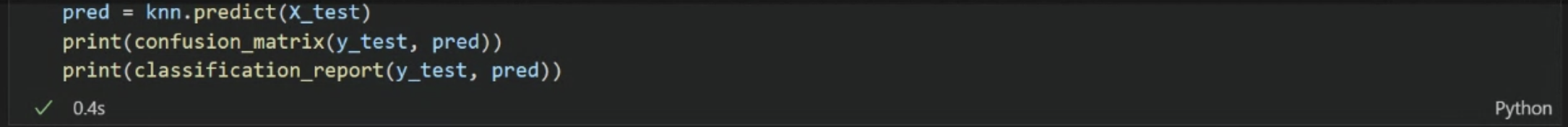
Empezamos a probar con un valor para n\_neighbors; es un proceso de prueba/error has ta dar con el valor apropiado.

Los valores de train y test deben ser cercanos

Sacar un 0,5 es el peor resultado, ya que es un 50% de acierto.

0,1 ó 0,9 son valores considerados un buen resultado de Accuracy.





Como tenemos un clasificador con cinco clases tenemos una matriz de 5x5

Text

Description automatically generated

Cada línea del informe corresponde a una clase

Text

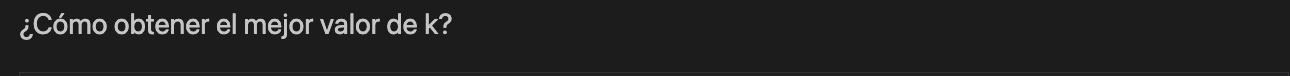
Description automatically generated

El ideal sería tener el dataset balanceado, si hay 5 clases cada una debería tener un 20% de los datos.

No tener este balanceo afecta al accuracy

Macro recoge las muestras sin tener en cuenta los pesos

Weighted tiene en cuenta el peso de cada clase en el total del dataset



Text

Description automatically generated

Probamos varios modelos con k diferente y nos guardamos el score; el modelo no lo guardamos.

Scatter chart

Description automatically generated with medium confidence

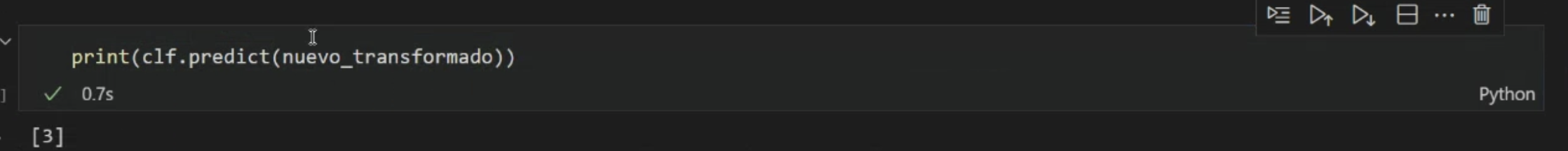
Chart, histogram

Description automatically generated

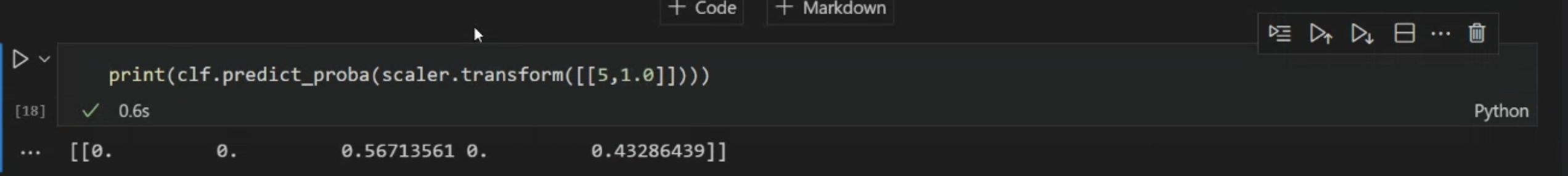
El score mas alto va asociado al k=7

Graphical user interface, application

Description automatically generated



Si llega una nueva review con cinco palabras y un sentiment value de 1, la predicción es que nos va a tener 3 estrellas.



Queremos saber la probabilidad que se le ha dado a cada clase, con esos nuevos datos.