# Iris Flower Dataset

To practice using the *k*-nearest neighbor algorithm in python, we'll be using a very famous dataset, the *Iris flower data set* from statistician and biologist Ronald Fisher in 1936. This dataset consists of measurements taken from 150 iris flowers from three different iris species, Iris setosa, Iris virginica, and Iris versicolor. Four features were measured for each flower sample: sepal length, sepal width, petal length, and petal width.
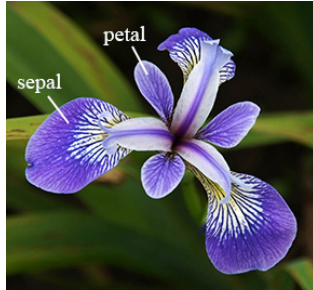
The *petals* are the colorful leaves of a flower, which surround the reproductive parts of the flower.

The *sepals* are outside of the petals, and they protect the flower while it's in bud. When the flower blooms, the sepals support the petals. In most flowers, the sepals are green, but for these irises, the sepals are purple.

When Fisher first introduced the iris data set, he showed how a machine learning method called *linear discriminant analysis* can be used for classification of these iris species. Since then, the Iris dataset has become one of the most commonly used datasets for practicing, testing, and demonstrating machine learning classification algorithms. As you continue to study machine learning, you're likely to come across the iris dataset again and again!