**Project Plan**

1. Investigation of the problem and available solutions… References
2. Writing the Python code to implement the solution
3. Documentation and explanations, including comments and README

The ***Iris* flower data set** or **Fisher's *Iris* data set** is a multivariate data set introduced by the British statistician and [biologist](https://en.wikipedia.org/wiki/Biologist) Ronald Fisher in his 1936 paper The use of multiple measurements in taxonomic problems as an example of [linear discriminant analysis](https://en.wikipedia.org/wiki/Linear_discriminant_analysis)[1]

The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor)[2].

Four features were measured from each sample: the length and the width of the [sepals](https://en.wikipedia.org/wiki/Sepal) and [petals](https://en.wikipedia.org/wiki/Petal), in centimetres. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

Based on Fisher's linear discriminant model, this data set became a typical test case for many statistical classification techniques in machine learning such as support vector machines[3].

The use of this data set in cluster analysis however is not common, since the data set only contains two clusters with rather obvious separation. One of the clusters contains Iris setosa, while the other cluster contains both Iris virginica and Iris versicolor and is not separable without the species information Fisher used. This makes the data set a good example to explain the difference between supervised and unsupervised techniques in data mining: Fisher's linear discriminant model can only be obtained when the object species are known

[1] https://en.wikipedia.org/wiki/Iris\_flower\_data\_set

[2] R. A. Fisher (1936). "The use of multiple measurements in taxonomic problems". [Annals of Eugenics](https://en.wikipedia.org/wiki/Annals_of_Eugenics). 7 (2): 179–188. [doi](https://en.wikipedia.org/wiki/Digital_object_identifier):[10.1111/j.1469-1809.1936.tb02137.x](https://doi.org/10.1111%2Fj.1469-1809.1936.tb02137.x).

[3] ["UCI Machine Learning Repository: Iris Data Set"](https://archive.ics.uci.edu/ml/datasets/iris). archive.ics.uci.edu. Retrieved 2017-12-01.