

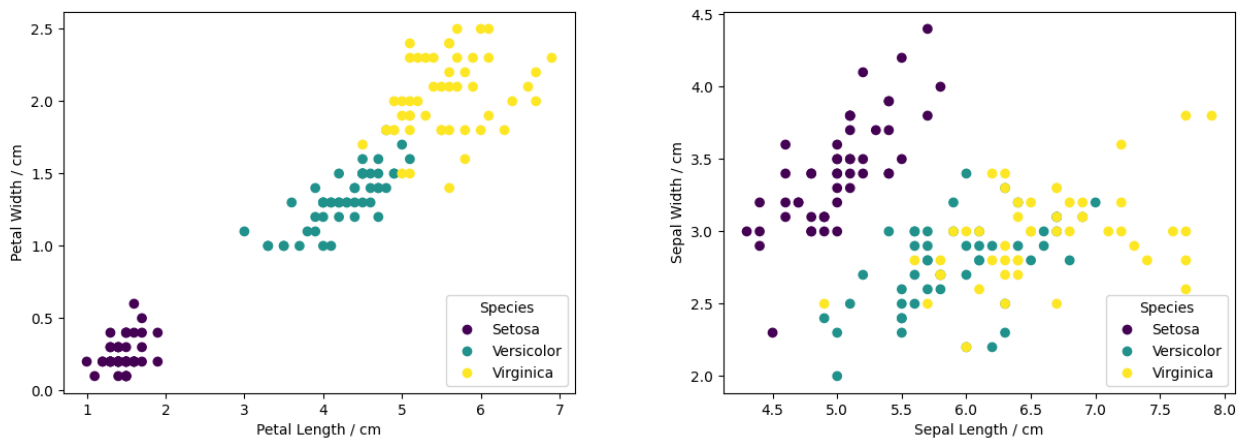
# Iris Classifier

Classify a dataset of three different iris flowers based on their sepal and petal dimensions.

## Learning Objectives

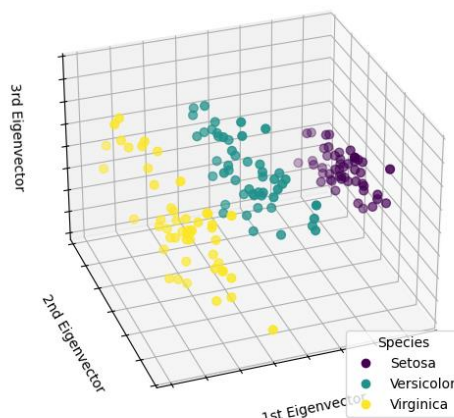
- Visualising Data
- K-Nearest Neighbour Classification
- Support Vector Machine Classification
- Decision Tree Classification

## Visualising the Data



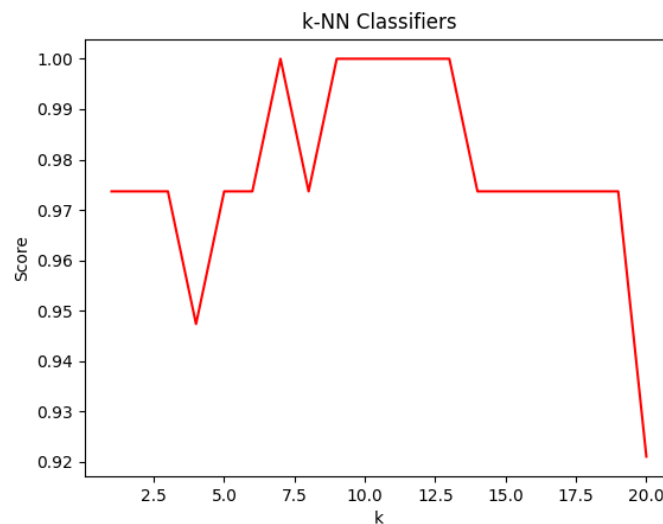
Principal Component Analysis (PCA) was applied to the iris dataset, creating three new features which are a linear combination of the four original features. This transformation was chosen to maximise the variance between the three species. The species can be identified using on the first feature.

First three PCA dimensions

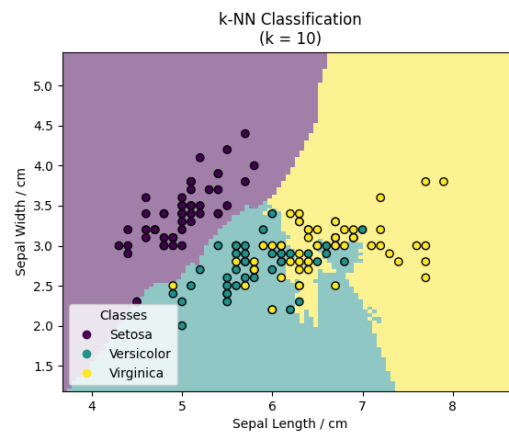
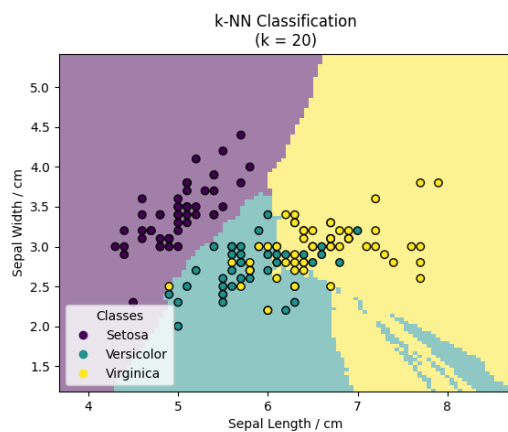
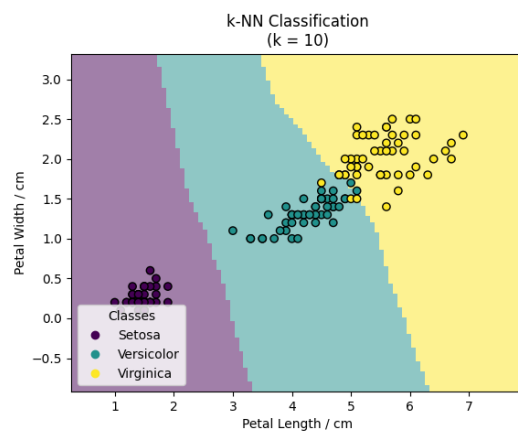
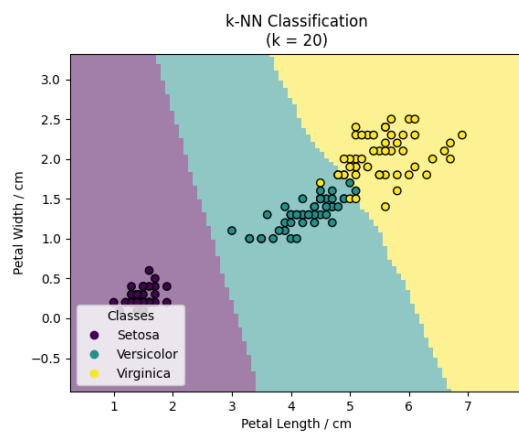


## k-Nearest Neighbour Classification

k-Nearest Neighbour (kNN) models were fitted to a random training subset of the iris dataset and tested on the remaining data. The value of  $k$  was varied from 1 to 20.



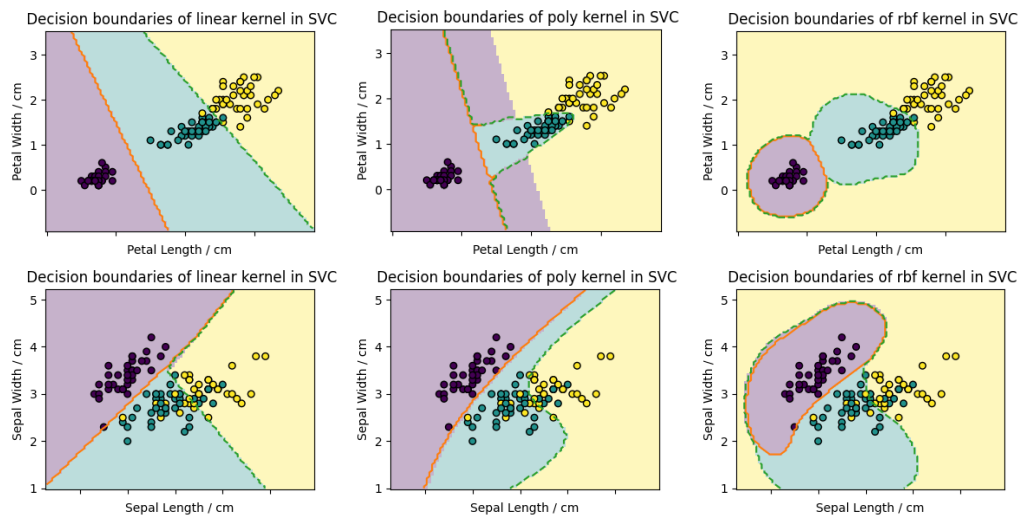
If the chosen value of  $k$  was too high, the classifier would start selecting nearby points belonging to a different species.



# Support Vector Machine Classification

Support Vector Machine (SVM) models were fitted to a random training subset of the iris dataset and tested on the remaining data. Four different SVM kernels were compared: linear, RBF, polynomial, and sigmoid. Each kernel was paired with the same tuning parameters:  $C = 1$ ,  $\gamma = 2$ .

Kernel	Score
Linear	1.0000
RBF	0.9737
Polynomial	0.9474



It is clear from the boundary plots that the sigmoid kernel is unable to describe the distribution of the three species in the feature space. The other three kernels performed well; the linear kernel achieved the highest score.

## Decision Tree Classification

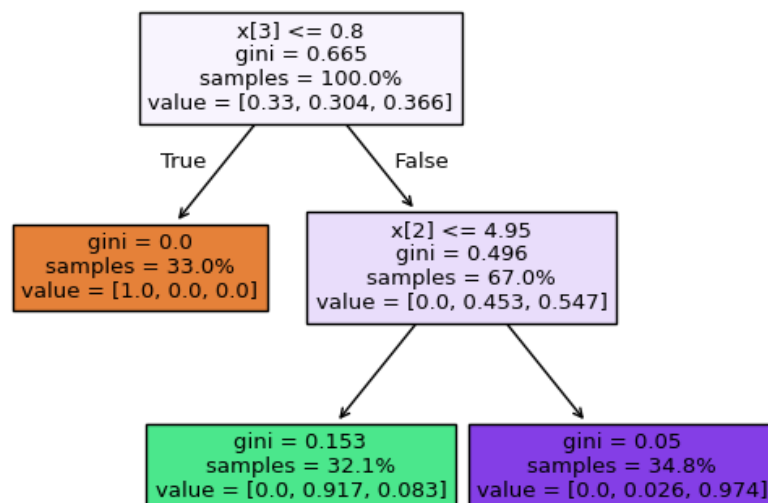
A Decision Tree model was fitted to a random training subset of the iris dataset and tested on the remaining data. Next, Decision Tree was pruned to a maximum depth of 3 nodes. The performance of these trees was compared.

Decision Tree	Score
Full	0.9737
Maximum Depth = 3	0.8947

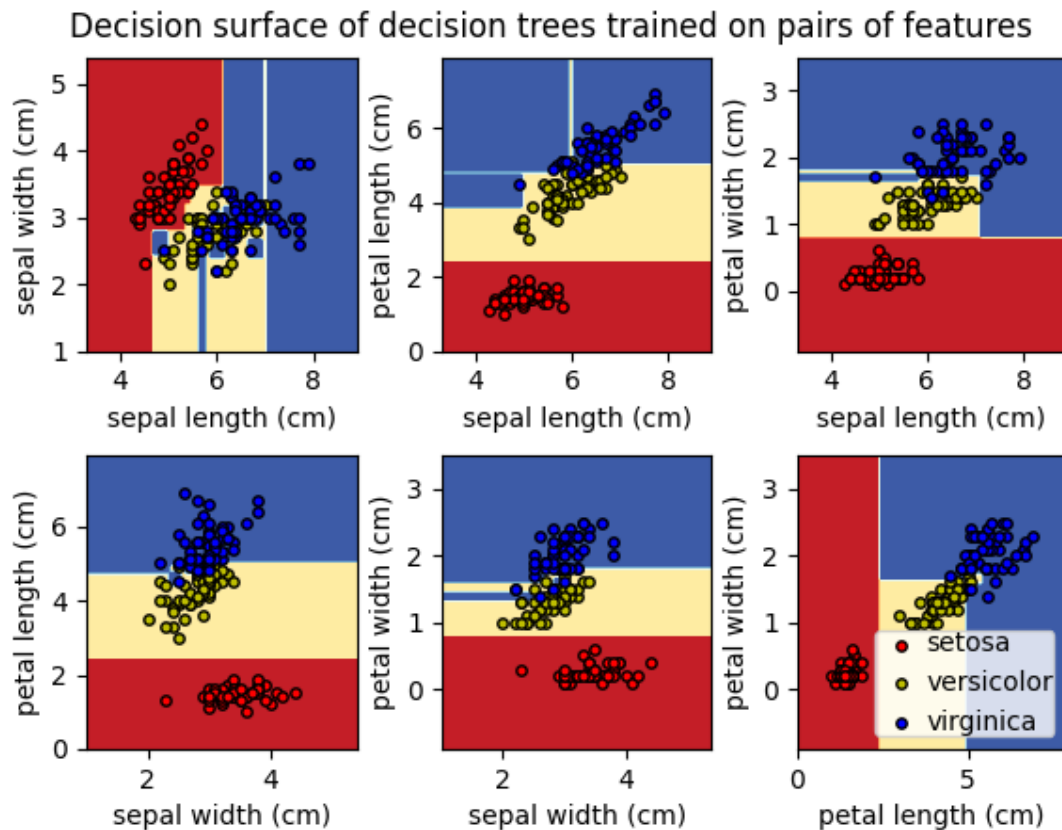
Full Decision Tree



Pruned Decision Tree



Decision Trees were trained on each pair of features, and the resulting boundaries were plotted.



## Conclusion

All three Machine Learning models (kNN, SVM, and Decision Tree) were able to classify the iris species with high accuracy. By visualising the data, it is clear that the species can easily be identified by considering the petal dimensions. The highest performing models were kNN ( $k=10$ ) and SVM (linear kernel). As expected, the SVM with a sigmoid kernel was unable to classify the three species. The full decision tree achieved a very high score, but pruning the tree resulted in a noticeably poorer performance.