# Iris dataset

Using Orange software for supervised learning of Iris dataset to categorise it into it species. Species is the target and the predictors are:

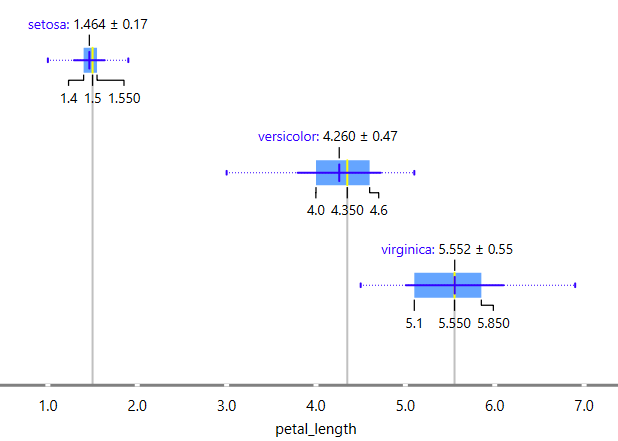
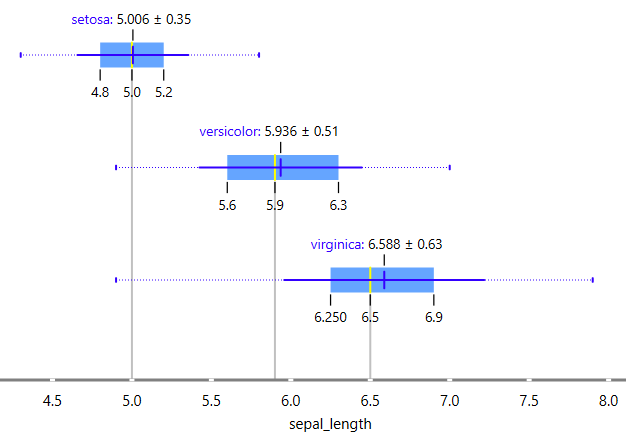
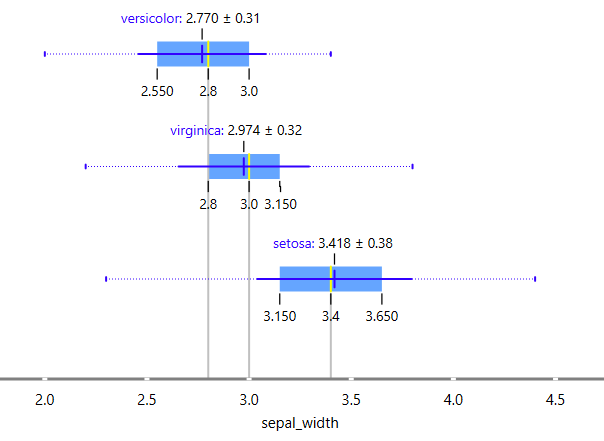
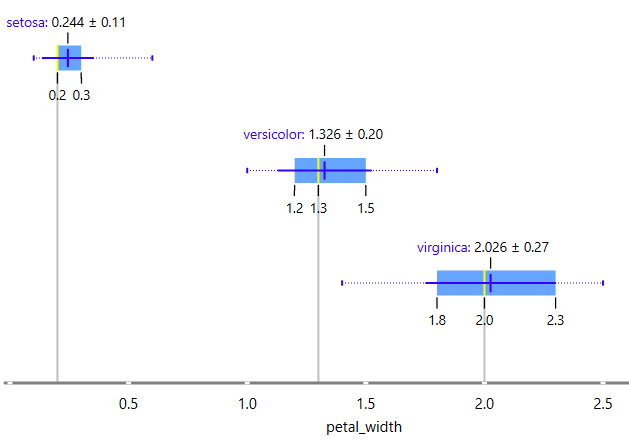
## Overview of the Machine Learning Process Flow. A diagram of a company AI-generated content may be incorrect.

## Exploratory Data Analysis:

A screenshot of a computer

AI-generated content may be incorrect.

Box plot



A screenshot of a computer

AI-generated content may be incorrect.Ranking of the predictors

The box plot and ranking indicate that sepal\_width has a minimal impact on species classification.

## Preprocess Data

Data splitting

A screenshot of a computer

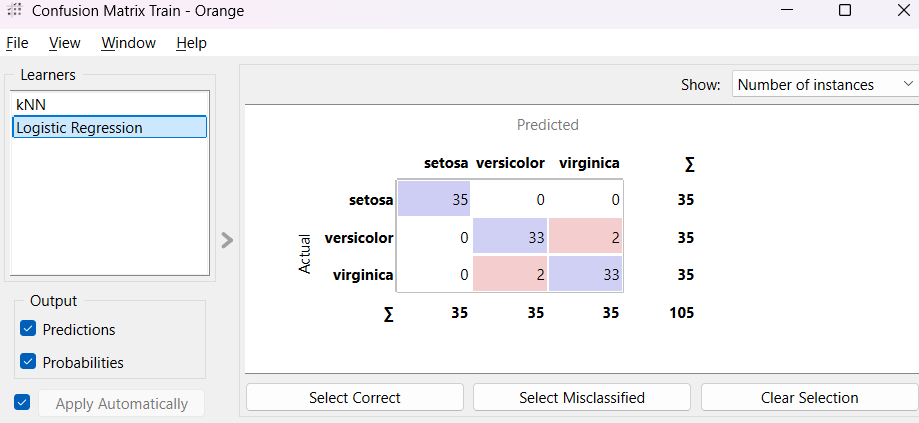
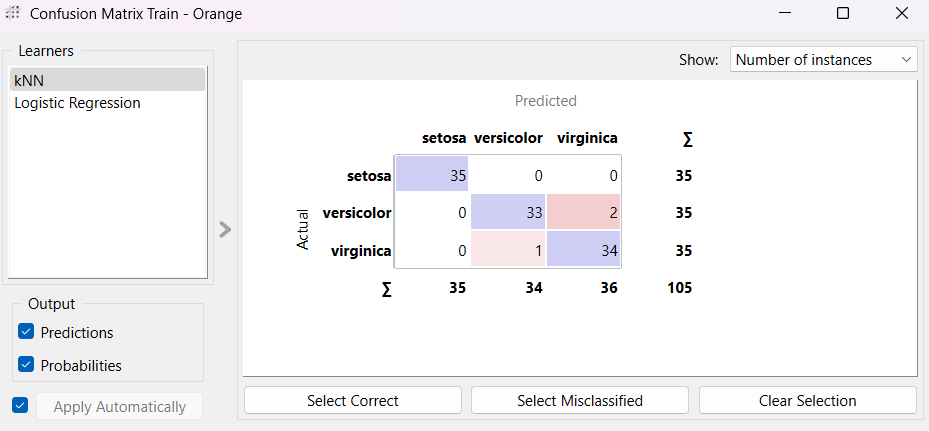
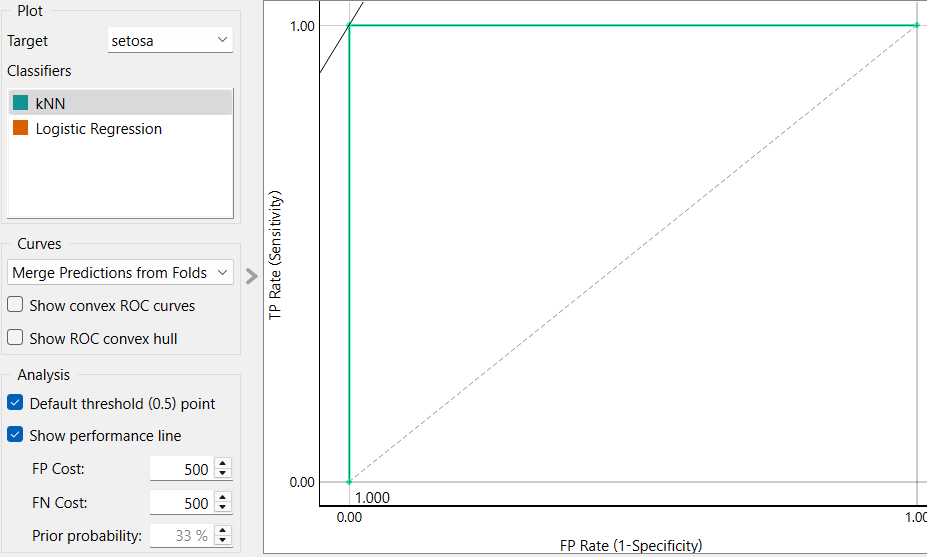
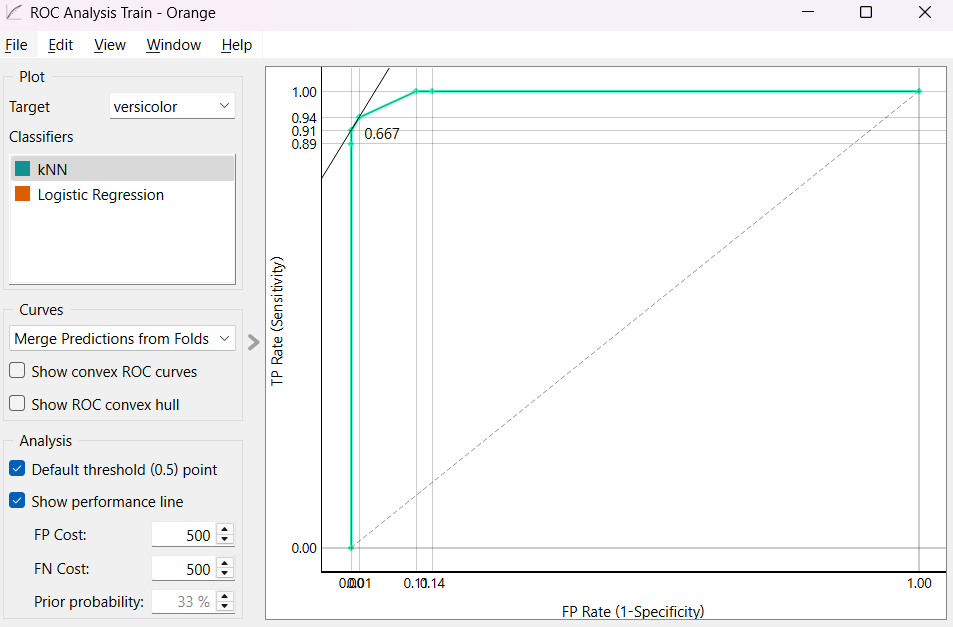
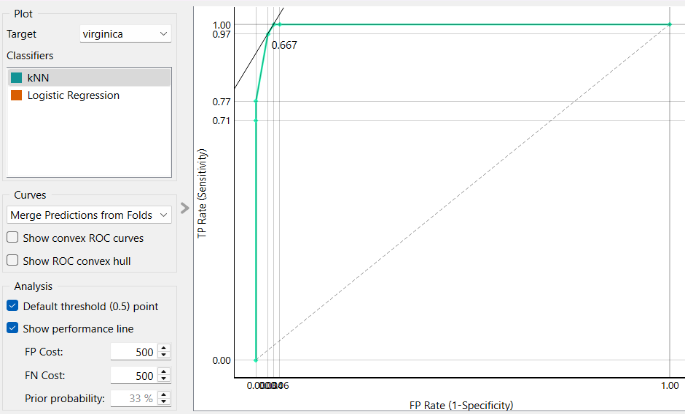
AI-generated content may be incorrect.The data is split into train and test set with 70:30 ratio.

## A diagram of a data processing process AI-generated content may be incorrect.Model selection

I have chosen to use Logistic Regression and KNN models for species classification. The training set is used to train the model, and the test set is used to validate it.

## Evaluate the model's performance

ROC analysis of the training dataset using the kNN model.



Based on the confusion matrix of the model, the kNN algorithm demonstrates superior performance.

## Validating with Test data

Confusion matrix for the test data

A screenshot of a computer

AI-generated content may be incorrect.

The confusion matrix shows that the model effectively classifies the species.