



# Predictive Model Using IBM Cognos Analytics

Decision Tree

## Sample Dataset Used

The dataset used is the iris dataset. The dataset was used in R.A. Fisher's classic 1936 paper, The Use of Multiple Measurements in Taxonomic Problems, and can also be found in the UCI Machine Learning Repository.

It includes three iris species with 50 samples each and some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other.

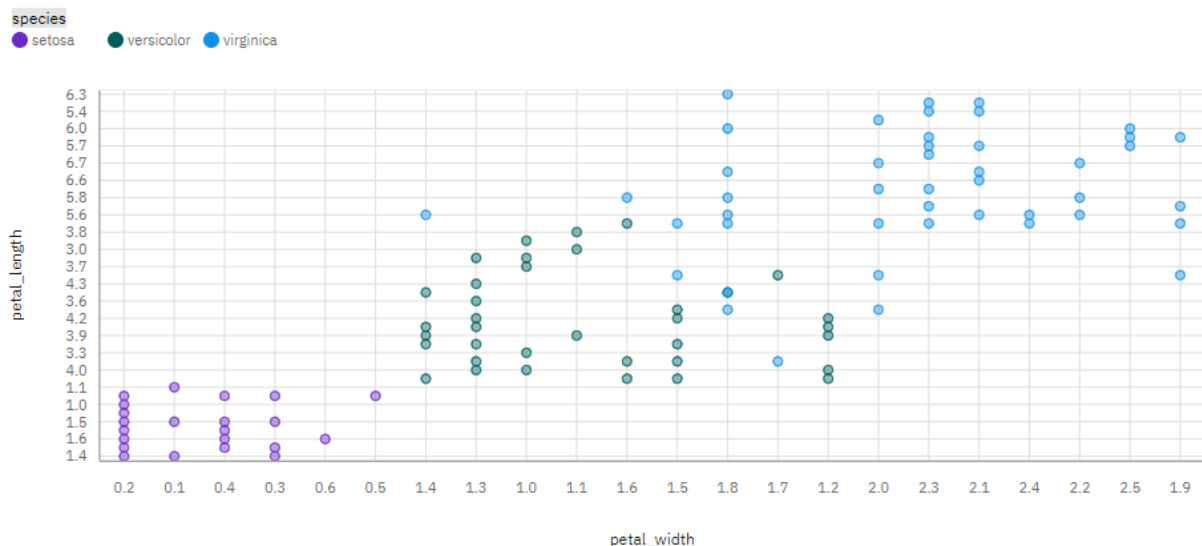
The columns in this dataset are:

- Id
- Sepal Length
- Sepal Width
- Petal Length
- Petal Width
- Species

## Exploration

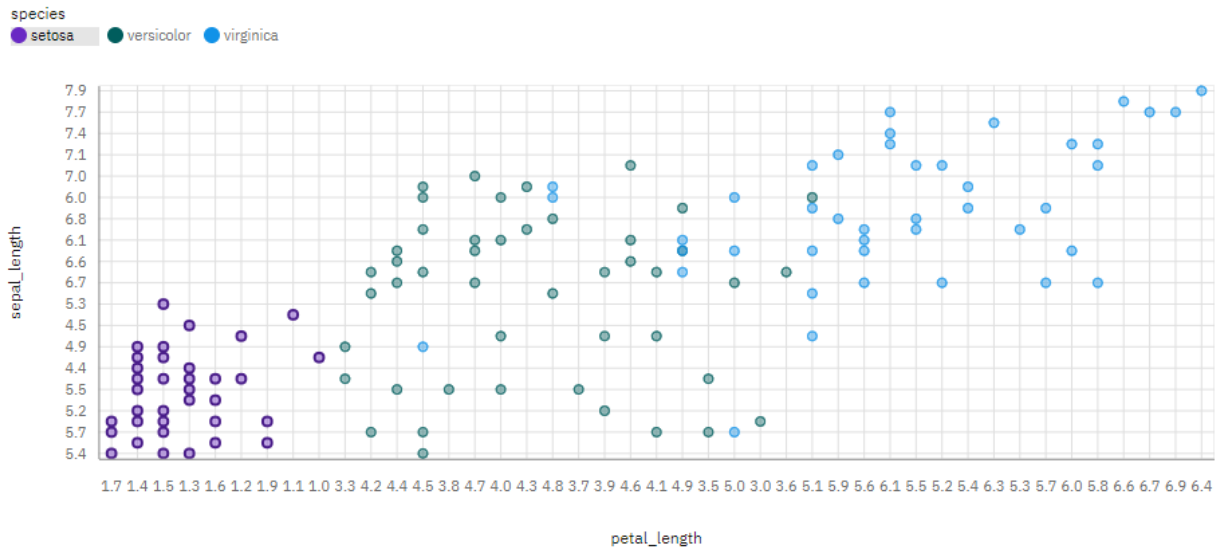
Exploratory analysis was done on the dataset, and a scatter plot was used to visualize the correlation between the Petal Length and Petal Width columns, grouped by color for the various species. Below, it can be denoted that the Setosa species does not exceed a petal length of 1.1 cm and a width of 0.5. also, the other two species can be read from the visual

petal\_width by petal\_length with points for petal\_length



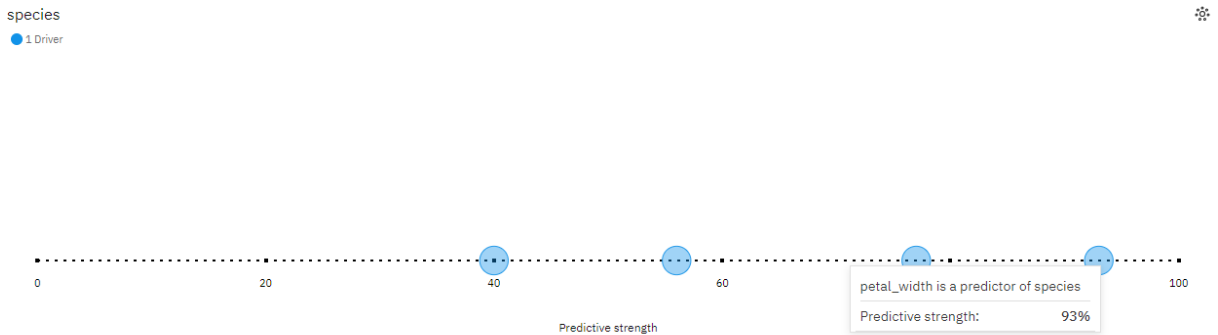
The correlation between Petal Length and Sepal Length was also visualized and grouped by color for the various species.

petal\_length by sepal\_length colored by species



## Predictive model

To create a predictive model using the decision tree in Cognos Analytics. Firstly, I used the driver analysis tool to visualize the predictive strength of various variables with the target variable “species.” It was seen that petal width is the highest predictor of species, with a predictive strength of 93%, shown below:



The petal width was used as the driver for the model, with species as the target variable. Since the species column is a categorical variable, it denotes the various flower species by color as seen below, with the petal width variable as the splitting variable.



Looking at the rules of the model, which is so straightforward, it shows that flowers with petal width 1.8, 2.5, 1.9, 2.1, 2.2, 2.0, 2.4, and 2.3 all falls under Virginica specie, which is a representation of 98% of the sample data considered in the model.

species

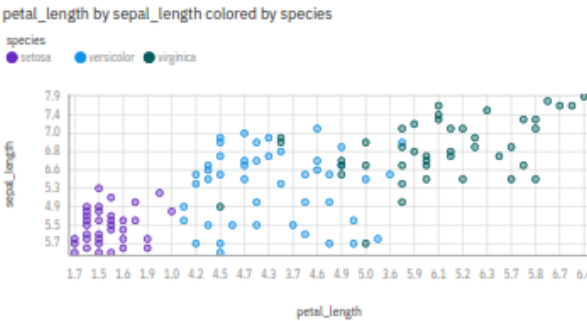
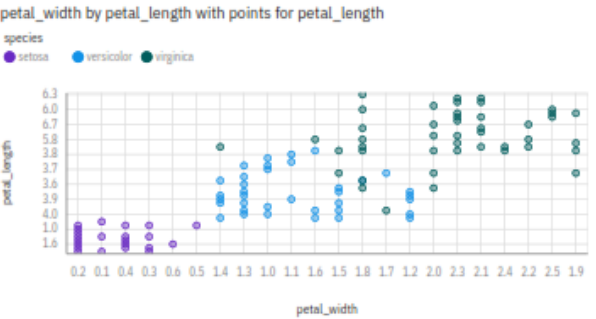
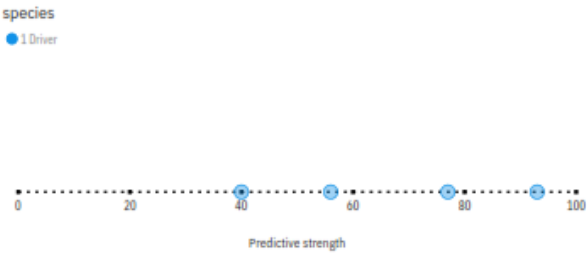
Tree sunburst Tree diagram Rules

Target category species

All setosa versicolor virginica

↓	Target category	Rules	Records
98%		petal_width = 1.8, 2.5, 1.9, 2.1, 2.2, 2.0, 2.4, 2.3	46 (31%)
96%		petal_width = 0.2, 0.4, 0.3, 0.1, Other	52 (35%)
92%		petal_width = 1.4, 1.5, 1.3, 1.6, 1.0, 1.1, 1.2	52 (35%)

In conclusion, this is just a summary of a rough analysis showing the use of IBM Cognos Analytics and the decision tree feature. To round up, a dashboard created from the exploratory visuals is provided.



species

Target category

All

species

● setosa ● versicolor ● virginica

Target category	Rules	Records
98%	petal_width = 1.8, 2.5, 1.9, 2.1, 2.2, 2.0, 2.4, 2.3	46 (31%)
96%	petal_width = 0.2, 0.4, 0.3, 0.1, Other	52 (35%)
92%	petal_width = 1.4, 1.5, 1.3, 1.6, 1.0, 1.1, 1.2	52 (35%)