KJSCE/IT/LY-B.Tech/SEM-VII/ML/2022-23



**Experiment No. 6**

**Title: Develop a classification model using Microsoft Azure Machine Learning Studio**

KJSCE/IT/LY-B.Tech/SEM-VII/ML/2022-23

**Batch: B1** **Roll No.1914078** **Experiment No.:6**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Title: Case study: Iris Flower Classification**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Resources needed:**



Microsoft azure studio

**Describe the following points with respect to the business under consideration,**

1. **Problem faced by the business:**

The Iris flower data set is a multivariate data set introduced by the British statistician and biologist Ronald Fisher. The use of multiple measurements in taxonomic problems makes it difficult to manually classify the flowers..

1. **Approach/ Methodology followed by the business:**

Existing 4 numerical attributes for identification of the classes are sepal length, sepal width, and petal length and petal width in centimeters. The classes are IRIS Setosa, IRIS versicolor. In this application 50 data of each class are used for training and are then tested..

1. **Skill Sets, infrastructure and other impact on the business during implementation:**

A software engineer that has experience in machine learning will prove to be very helpful.

1. **Similar approaches followed by other businesses**

A classifier model will work best.

KJSCE/IT/LY-B.Tech/SEM-VII/ML/2022-23

**Step 1.** Sign-in using Microsoft account on studio.azureml.net **Step 2.** Creating workspace for our Machine Learning project.

**Step 3.** Select New option on bottom right:

**Step 4.** Click on Blank experiment and write name and summary of experiment **Step 5.** Select From Saved Datasets-> Samples-> dataset of your choice

**Step 6.** Now, search ‘Select columns in dataset’ from items and drag it

**Step 7.** Now, click on launch column selector-> with rules->exclude column normalized-losses as that column contains many rows/records with empty values.

**Step 8.** Search and select ‘Clean Missing Data’ from items list

**Step 9.** Now, select cleaning mode -> Remove entire row as it will remove the entire row wherever missing value is found

**Step 10.** Again choose ‘select columns in dataset’

**Step 11.** Now, launch column selector and include all the columns based on which

prediction is to be done: make, body-style, wheel-base, engine-size, horsepower, peak-rpm, highway-mpg, price

**Step 12.** Now, select ‘split data’ from list and drag it

**Step 13.** For Split data, enter the fraction of data which is needed for training while rest will be used for testing

**Step 14.** Now, Select ‘Linear Regression’ as the algorithm to be used and ‘Train Model’ from list

**Step 15.** For training model, click on launch column selector, include price column as Price is what is to be predicted

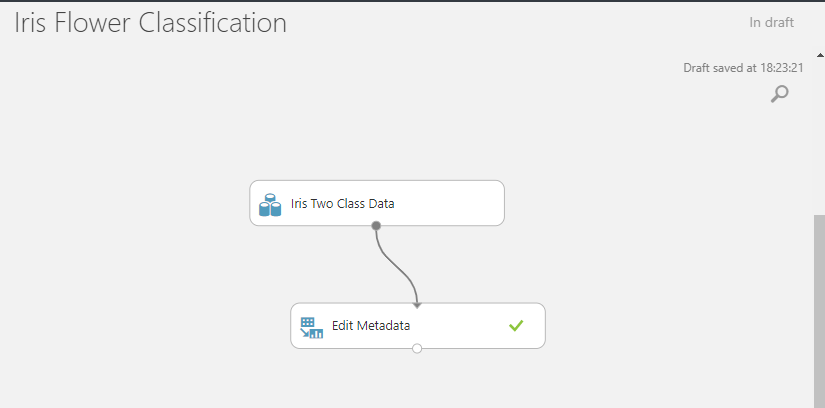
**Step 16.** Add Score Model from list drag it and make connections **Step 17.** Now, Add Evaluate Model from list and make connections **Step 18.** Now, Click on Run

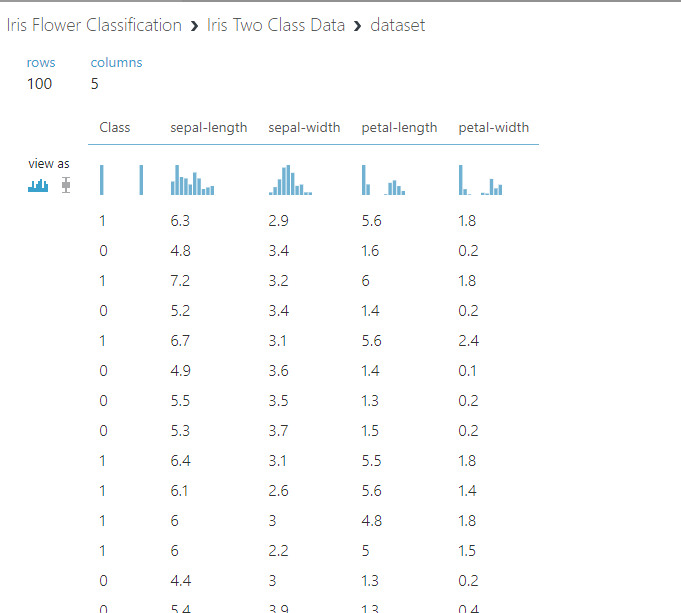
**Step 19.** To check prediction results, right click on Score Model, select visualize

**Step 20.** To check Evaluation results, right click on Evaluation Model, select visualize

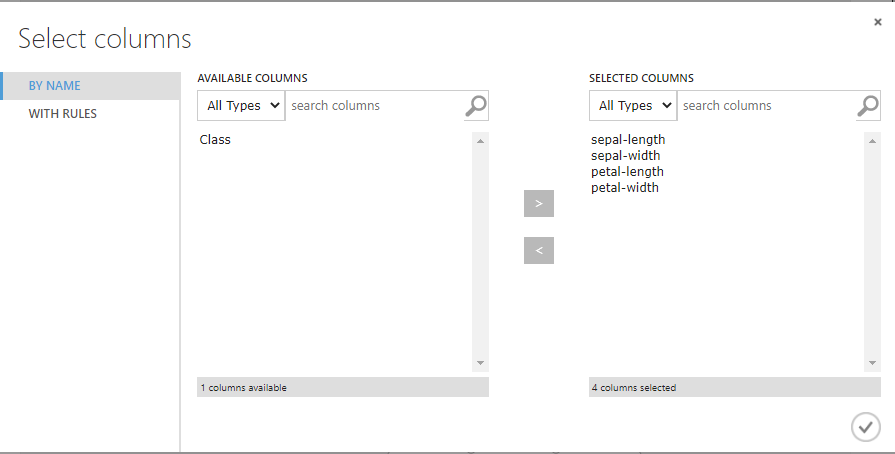
**Results:**

1. Uploading and choosing dataset

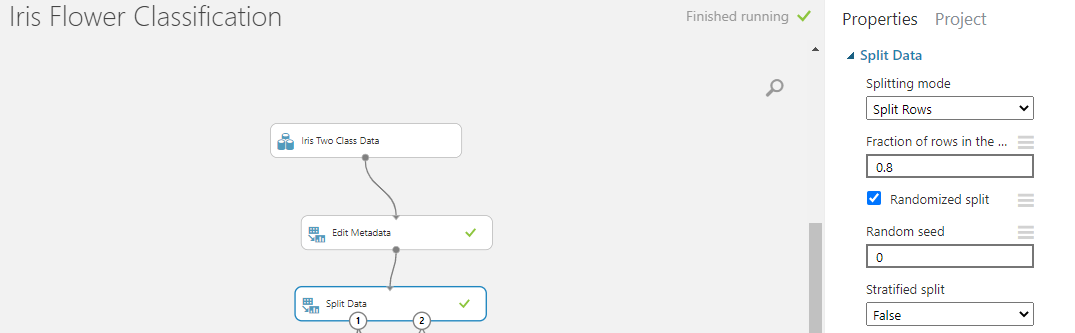




1. Selecting columns in dataset



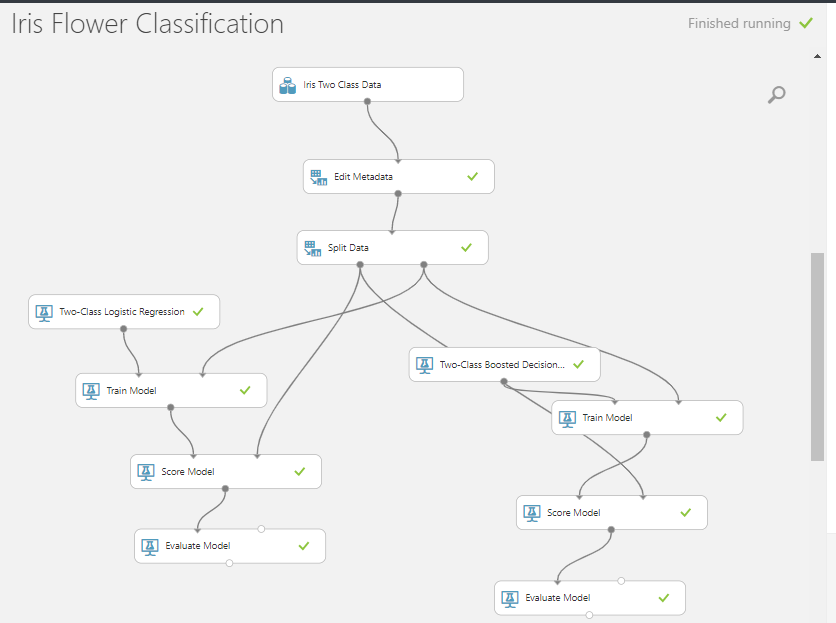
1. Splitting dataset



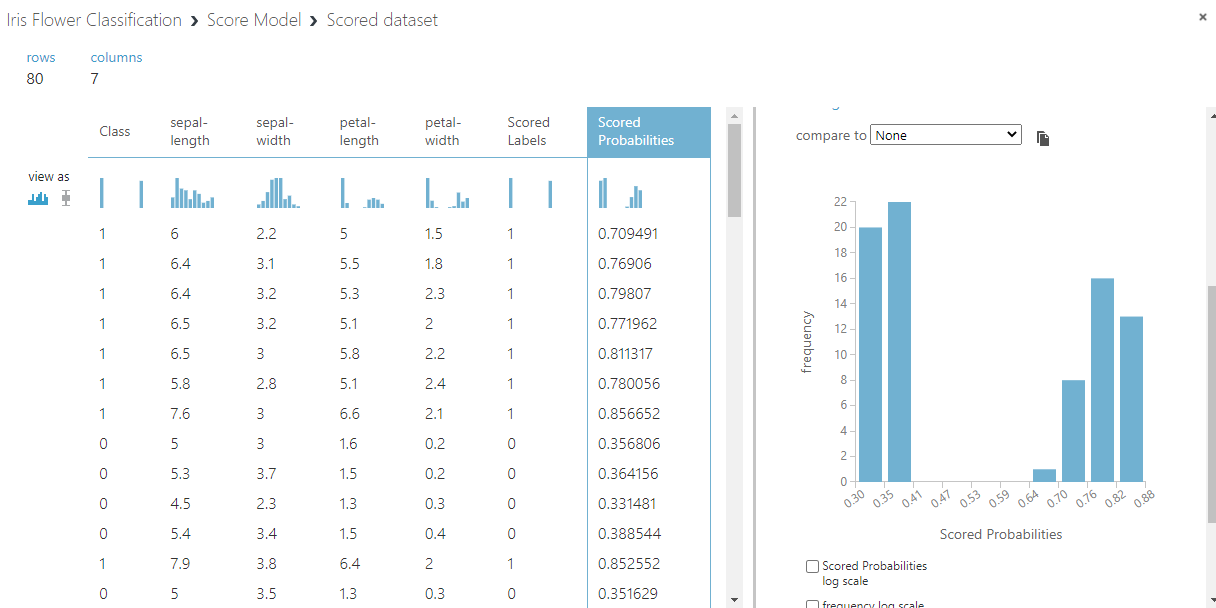
1. Training classification models(Since we have two classes 0-1, we will be using the following models) and Evaluating the model.

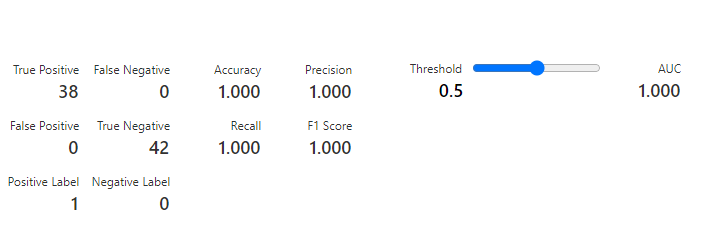
**Two-Class Logistic Regression:** This module is used for predicting binary outcomes with logistic regression. This algorithm predicts the two target classes by fitting data to a logistic function. The logistic function estimates the probability of occurrence of an event (in this case, iris class 0 or 1), and converts it into the target class of the dependent variable, Class.

**Two-Class Boosted Decision Tree:** This module creates a binary classifier using the boosted decision tree algorithm. This algorithm is based on the ensemble learning model, in which every tree builds upon the previous tree by correcting its error. For the data used in this guide, every single tree will make predictions on the target class of the dependent variable, Class. The final predictions are based on the entire ensemble of trees taken together.

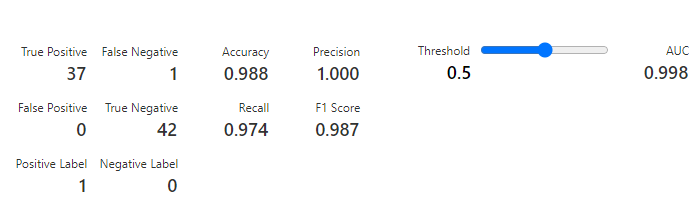


1. Logistic Regression Evaluation





1. Boosted Decision Evaluation



Of the above models, logistic regression has given accuracy of 1 and decision boosted gave an accuracy of 98.8.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Outcomes:**

CO2 - Apply concepts of different types of Learning and Neural Networks



**Conclusion: (Conclusion to be based on the objectives and outcomes achieved)**

We were able to perform an experiment on Azure ML Studio using various classification Machine Learning algorithms.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Grade: AA / AB / BB / BC / CC / CD /DD**

**Signature of faculty in-charge with date**

**References:**

**Books/ Journals/ Website**

****