



Experiment No. 7

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

Batch: B2/B4**Roll No.: 16010420117****Experiment No.:7****Title: Case study:**

Resources needed:**Describe the following points with respect to the business under consideration,****1. Problem faced by the business**

The system is designed such that it helps to estimate the price of a Car based upon the different features present, so it helps the business to give the correct amount of pricing to their Car Models, so that the customers are attracted and also are satisfied with price range assigned based upon it's features. So, assigning Price range accordingly is challenging without the use of ML algorithms.

2. Approach/ Methodology followed by the business

The methodology which we followed was that we took the dataset of Automobile Price Data. Then we split the dataset into train and test. Then we used Linear Regression and trained the machine learning model with that dataset and in the end we tested that model.

3. Skillsets , infrastructure and other impact on the business during**implementation Skillset:** Cloud computing**Infrastructure:** Microsoft Azure Cloud services

There were no as such impacts on the business, as any of the business service lines were not being used.

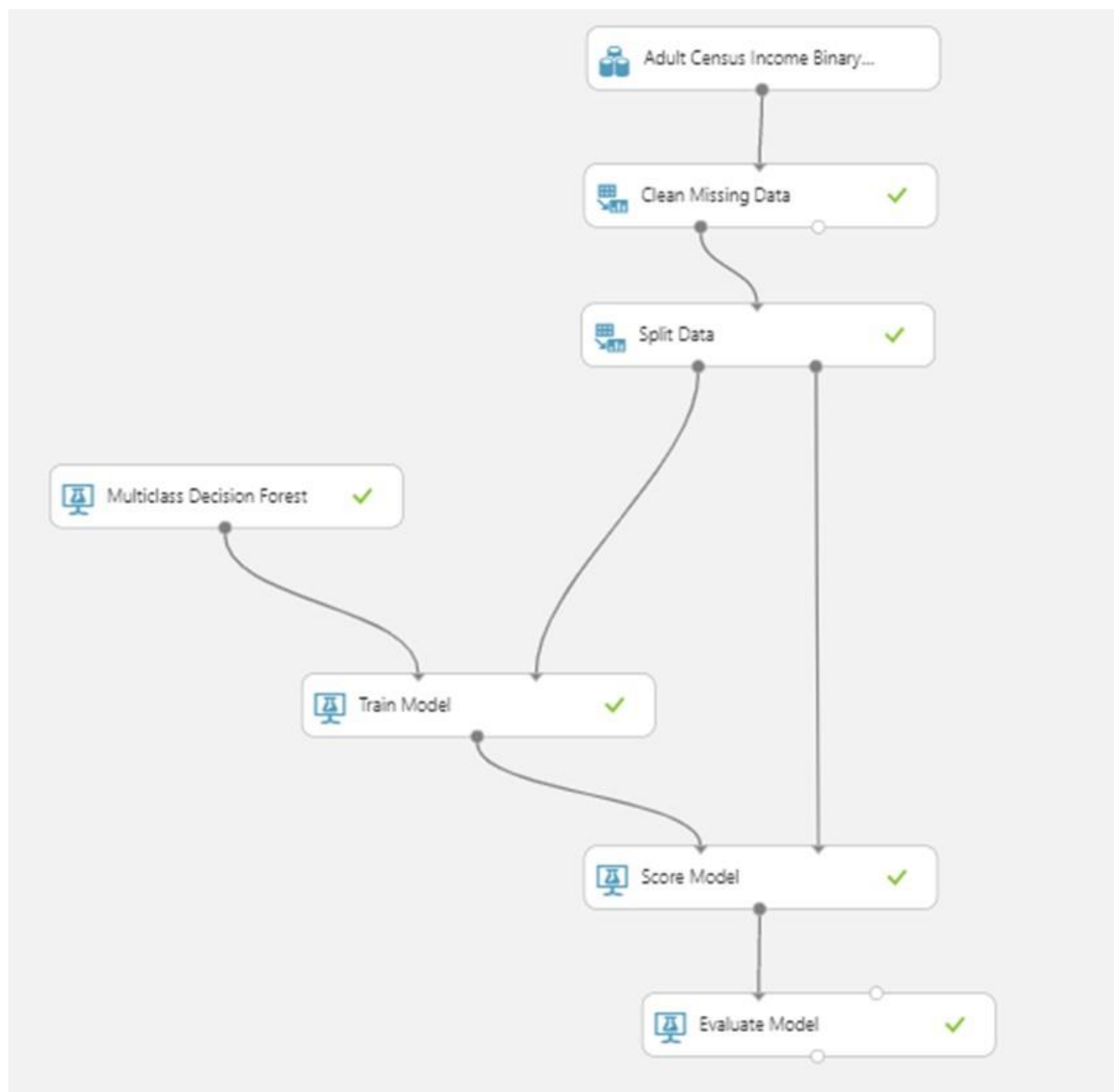
4. Similar approaches followed by other businesses**Step 1.** Sign-in using Microsoft account on studio.azureml.net**Step 2.** Creating workspace for our Machine Learning project.**Step 3.** Select New option on bottom right:**Step 4.** Click on Blank experiment and write name and summary of experiment**Step 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

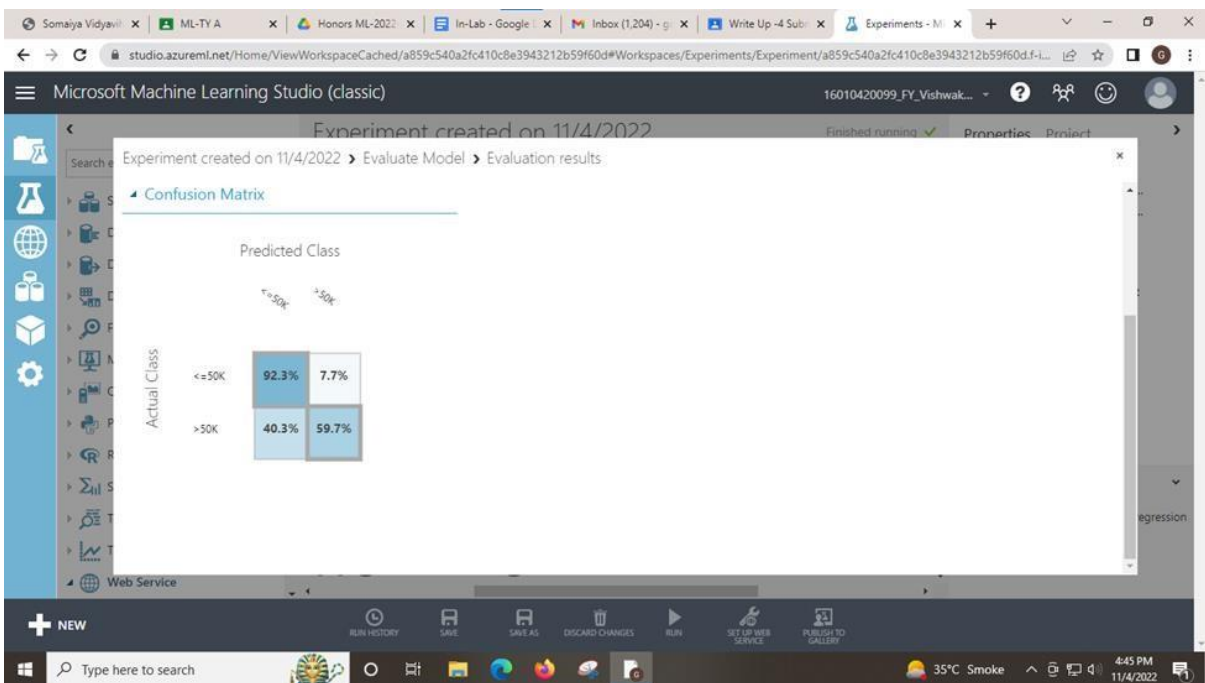
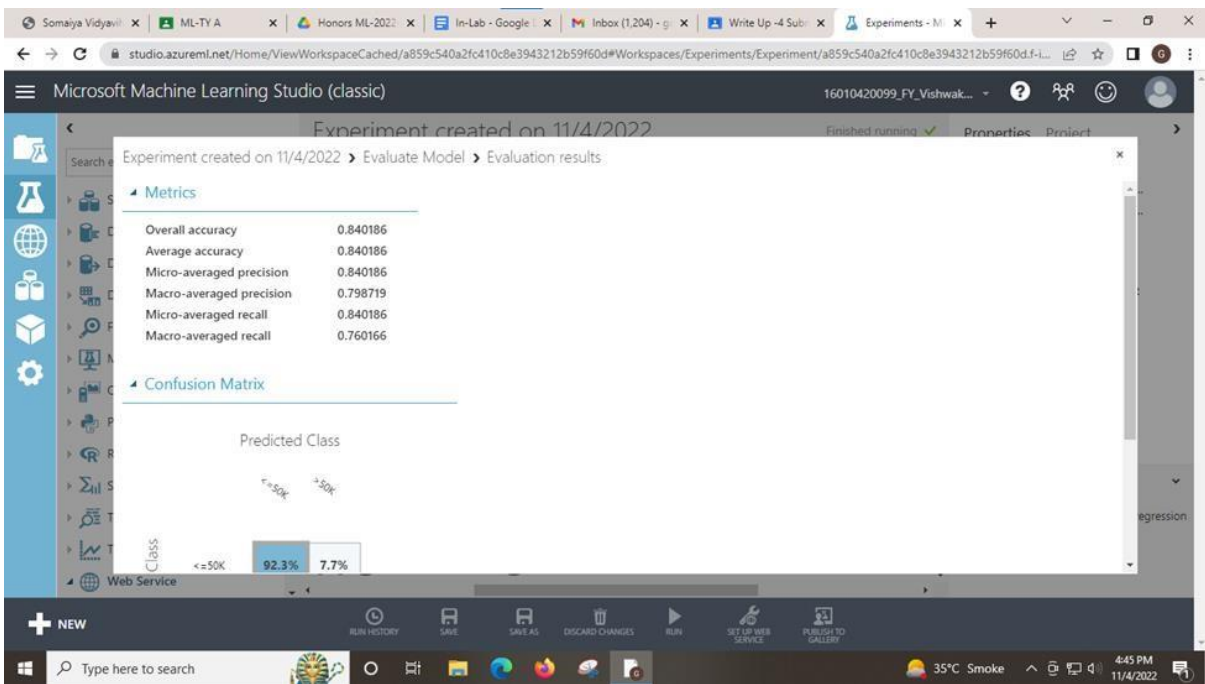
- Step 10.** Again choose 'select columns in dataset'
- Step 11.** Now, launch column selector and include all the columns based on which prediction is to be done: make, body-style, wheel-base, engine-size, horsepower, peak-rpm, highway-mpg, price
- Step 12.** Now, select 'split data' from list and drag 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:

Dataset used: Adult Income Census

Classes: Two classes on the “income” column labeled as- ‘<50K’ and ‘>50K’ Screenshots:





Post Lab Question- Answers (If Any):

Q. Differentiate between linear and nonlinear regression and write a note on converting non-linear model into linear model

Ans: Linear regression always uses a linear equation, $Y = a + bx$, where x is the explanatory variable and Y is the dependent variable. In multiple linear regression, multiple equations are added together but the parameters are still linear. If the model equation does not follow the $Y = a + bx$ form then the relationship between the dependent and independent variables will not be linear. There are many different forms of non-linear models. A random forest regression is considered a non-linear model. Random forest models are ensemble learning methods for regression which grow a forest of regression trees and then average the outcomes. This cannot be expressed as an equation. In regression trees, the splitting decision is based on minimizing the Residual Sum of Squares (RSS). The variable which has the greatest possible reduction in RSS is chosen as the root node. The tree splitting takes a top-down greedy approach, meaning the algorithm makes the best split at the current step rather than saving a split for better results on future nodes.

CO: Apply concepts of different types of Learning and Neural Network

Conclusion: Successfully Learnt and Implemented the model in Azure ML Studio.

KJSCE/IT/TY/SEM-VI/ML(H)/2022-23

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

Signature of faculty in-charge with date**References:****Books/ Journals/ Websites:**

