

Experiment No. 9

Title: Case study: Big data platform / analytics as business need)



Batch: A1**Roll No.: 1714014****Experiment No.:9****Case study: Bike Buyers Prediction**

Resources needed: Internet Connection, Microsoft Azure Account

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

The system is so designed that it helps to predicts whether a person is interested in purchasing a bike, so it helps the sales team to improvise their business by targeting the audience who they are sure will be most likely be looking to buy a bike so that their company can gain good profits and increase their total revenue.

2. Approach/ Methodology followed by the business**Answer:**

The methodology is such that we first collected a dataset of people which had attributes like age, income, gender, no of cars possessed and some other attributes. Then we select the attributes which are needed for the model and then we normalize those attributes. Then we split the dataset into training dataset and testing dataset. Then we trained the machine learning model using Two Class Logistic Regression with that dataset and then we tested that model. Then we evaluated the model for finding out its accuracy and precision.

3. Skillsets, infrastructure and other impact on the business during implementation**Answer:**

Skillset: Cloud computing

Infrastructure: Microsoft Azure Cloud services

There were no as such impacts on the business during implementation.

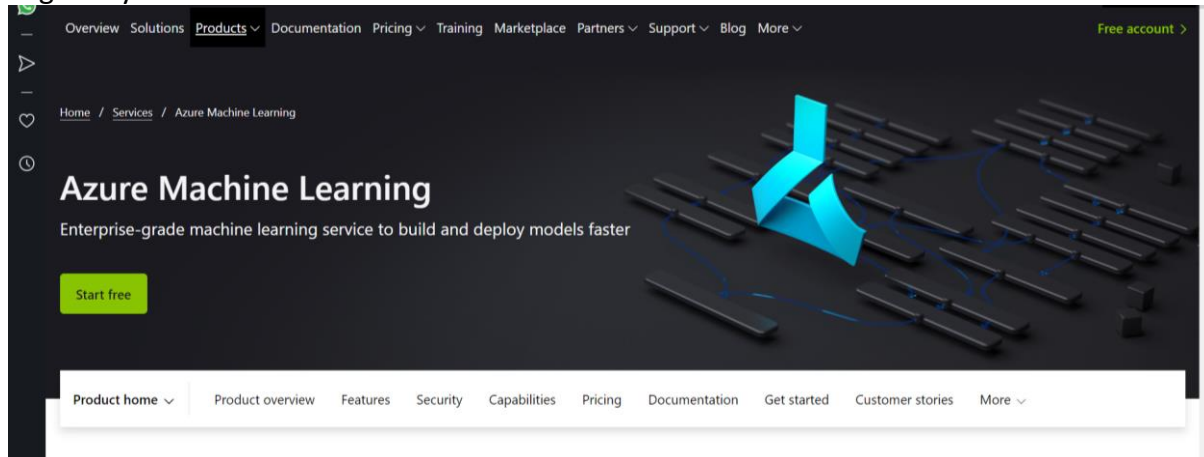
4. Similar approaches followed by other businesses**Answer:**

Other businesses are still not using ML in order to predict whether a person is interested in purchasing a bike or not and they rather use traditional methods and thus some companies face problems like loss in business revenues.

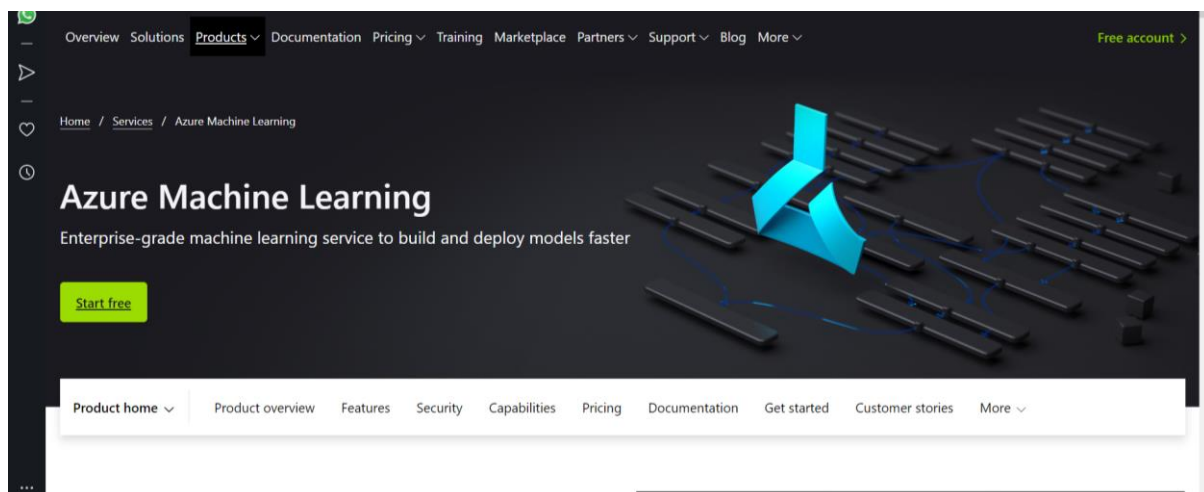
Steps:

Go to <https://azure.microsoft.com/en-in/services/machine-learning/>

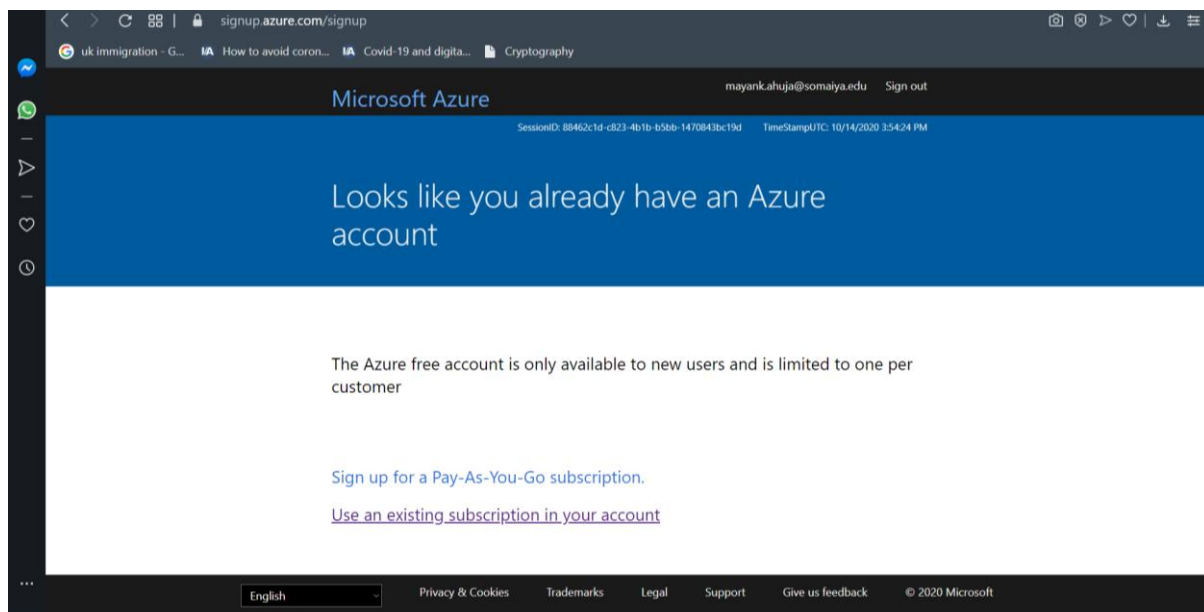
Login to your account.



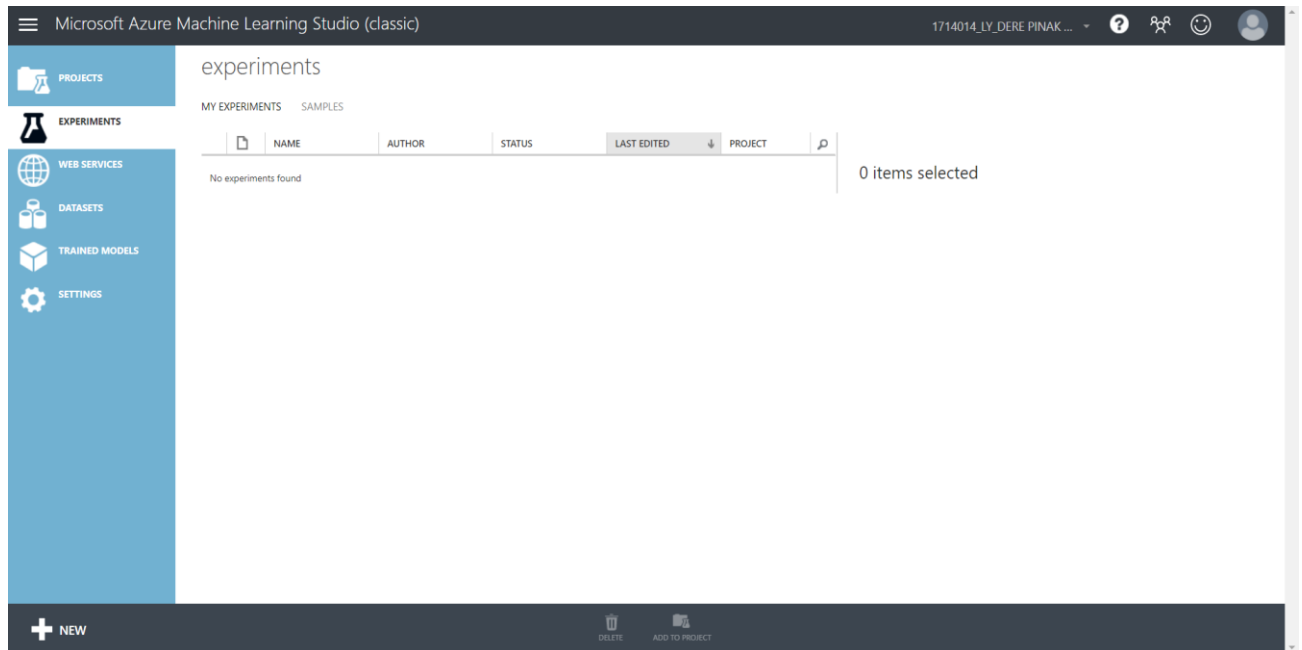
Click on start free



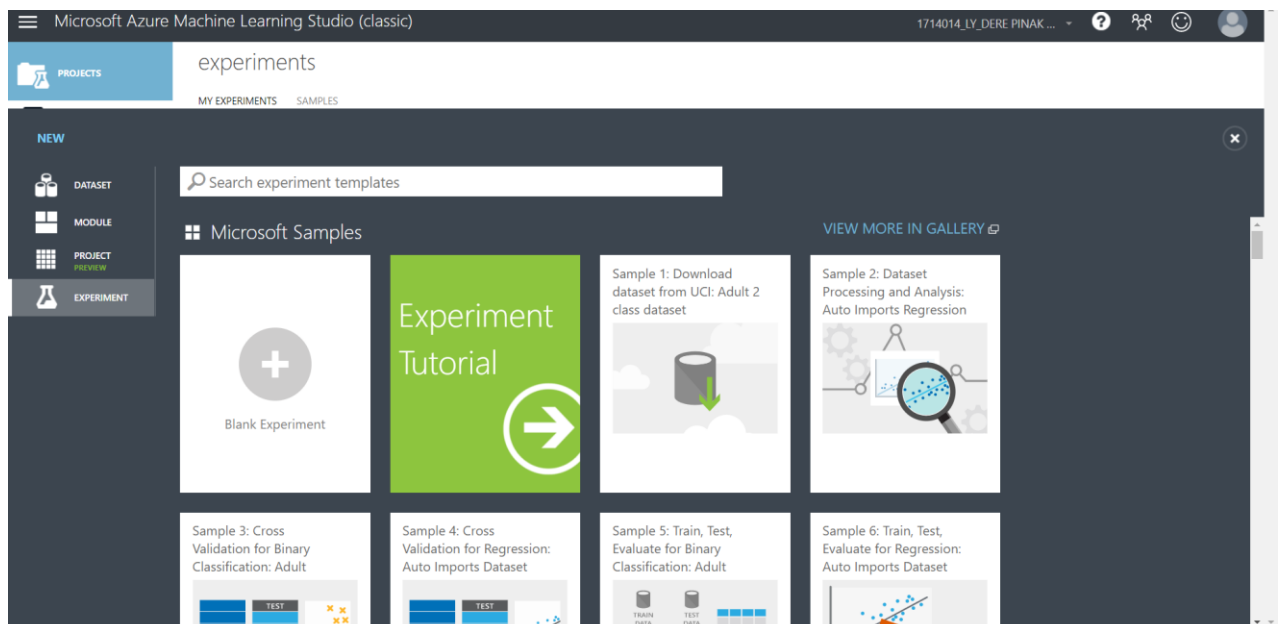
Use existing subscription on your account.



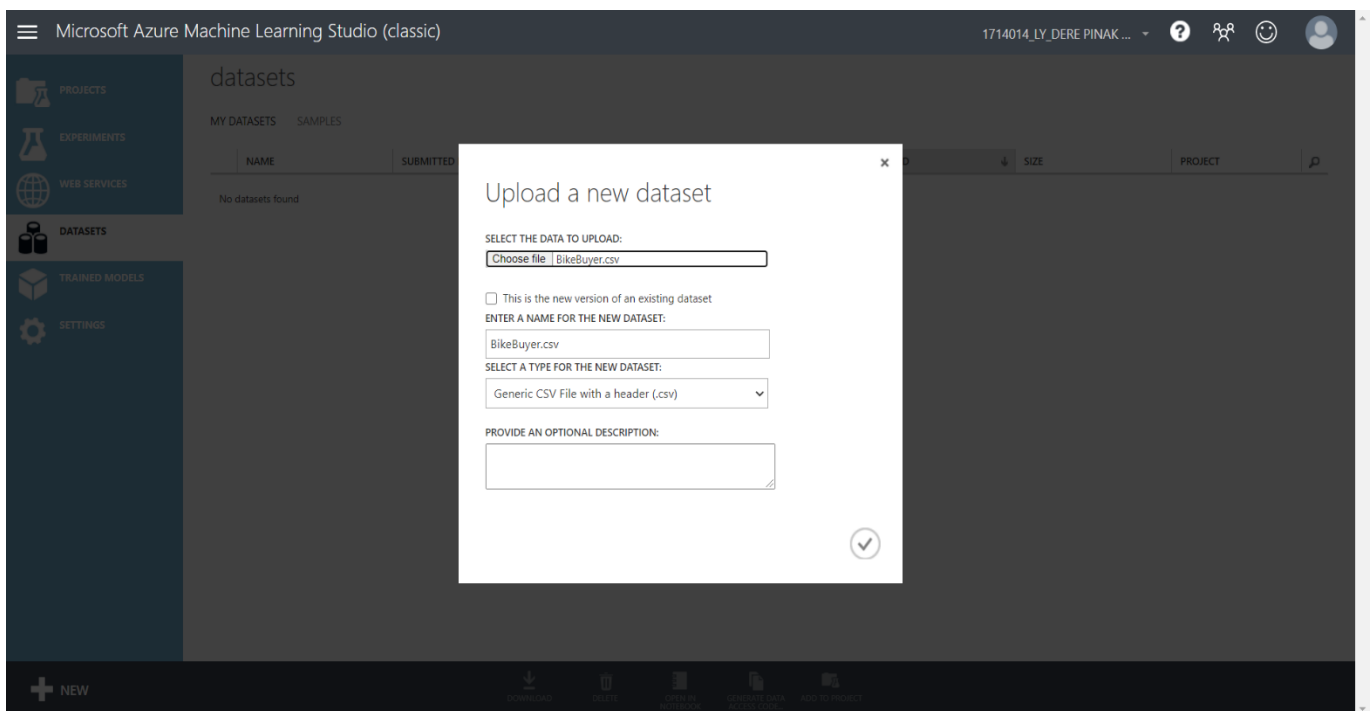
Open Azure and Login.



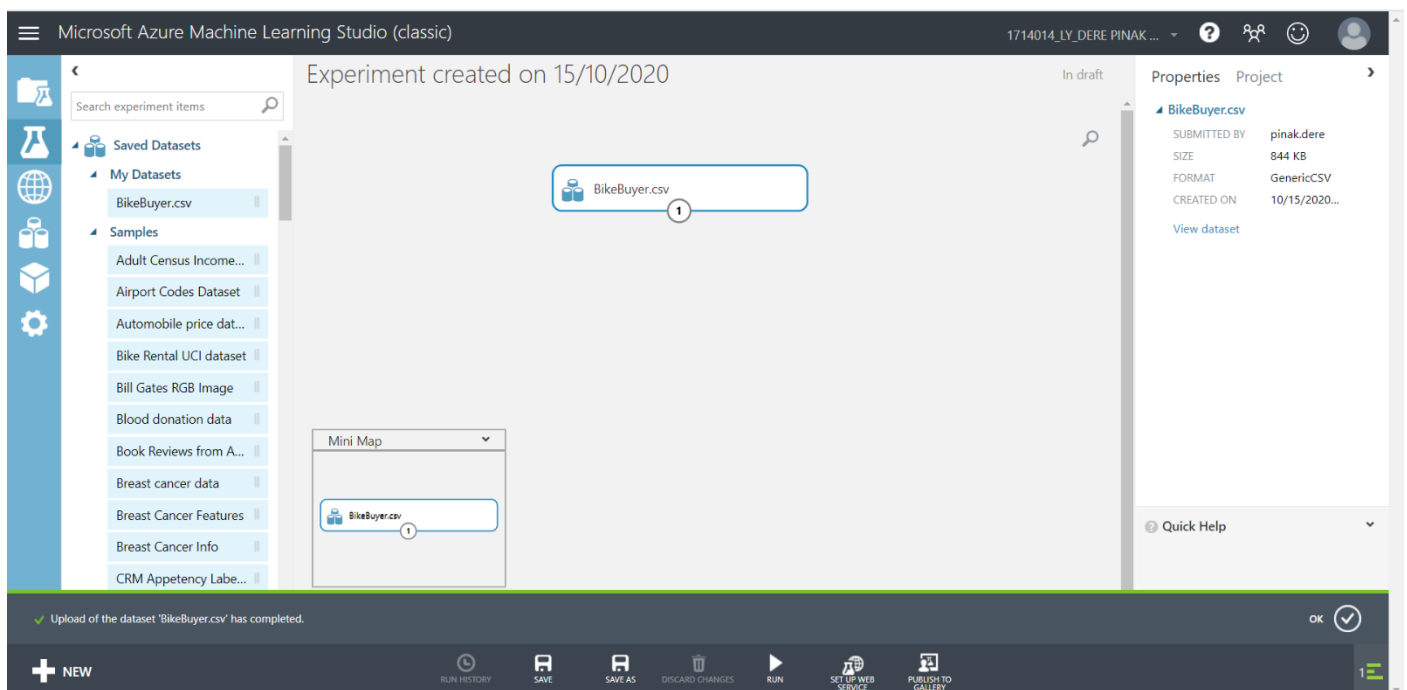
Go to experiments section and create a blank experiment.



Upload the dataset to import from the local files.



Add the BikeBuyers.csv file to the layout of the experiment.



Add Select Columns in Dataset to select certain columns for prediction.

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. The top bar indicates the experiment was created on 15/10/2020 and is currently in draft status. The left sidebar contains a search bar and a list of operations categorized under Data Transformation, Manipulation, Feature Selection, and Statistical Functions. The main workspace displays a workflow diagram with two steps: 'BikeBuyer.csv' (dataset (GenericCSV)) and 'Select Columns in Dataset'. A 'Mini Map' view at the bottom left shows the overall workflow structure. The right sidebar contains the 'Properties' and 'Project' tabs, with the 'Experiment Properties' section showing the status as 'InDraft'. The bottom status bar indicates that the upload of the dataset 'BikeBuyer.csv' has completed.

Select the attributes which we want to use for the model development.

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface with the 'Select columns' dialog box open. The dialog box has two tabs: 'BY NAME' and 'WITH RULES'. The 'BY NAME' tab is active, showing a list of available columns. The 'SELECTED COLUMNS' list on the right contains 12 columns: Marital Status, Yearly Income, Cars, Occupation, Commute Distance, Region, Children, Age, Education, Home Owner, Gender, and BikeBuyer. The bottom status bar indicates that the upload of the dataset 'BikeBuyer.csv' has completed.

Normalize the dataset.

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface. The top bar shows the user's name '1714014_LY_DERE PINAK ...' and the experiment title 'Experiment created on 15/10/2020'. The left sidebar contains a search bar with 'norma' and a list of data transformation steps: 'Data Transformation', 'Scale and Reduce', and 'Normalize Data'. The main workspace shows a workflow diagram with three steps: 'BikeBuyer.csv' (dataset (GenericCSV)), 'Select Columns in Dataset', and 'Normalize Data'. A 'Mini Map' in the bottom left shows the workflow overview. The right sidebar shows the 'Properties' panel for the 'Normalize Data' step, with the 'Transformation method' set to 'ZScore' and 'Use 0 for constant col...' checked. The 'Columns to transform' section shows 'Selected columns: Column type: Numeric, All'. A 'Quick Help' section at the bottom right explains that the step 'Rescales numeric data to constrain dataset values to a standard range'. The bottom status bar indicates 'Upload of the dataset "BikeBuyer.csv" has completed.' and includes buttons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

The screenshot displays the Microsoft Azure Machine Learning Studio (classic) interface, showing the 'Select columns' dialog box. The top bar shows the user's name '1714014_LY_DERE PINAK ...' and the experiment title 'Experiment created on 15/10/2020'. The left sidebar contains a search bar with 'norma' and a list of data transformation steps: 'Data Transformation', 'Scale and Reduce', and 'Normalize Data'. The main workspace shows the 'Select columns' dialog box with the 'WITH RULES' tab selected. The 'Begin With' section shows 'ALL COLUMNS' and 'NO COLUMNS' buttons. The 'Include' section shows a dropdown menu with 'column type' and 'Numeric' selected. The right sidebar shows the 'Properties' panel for the 'Normalize Data' step, with the 'Transformation method' set to 'ZScore' and 'Use 0 for constant col...' checked. The 'Columns to transform' section shows 'Selected columns: Column type: Numeric, All'. A 'Quick Help' section at the bottom right explains that the step 'Rescales numeric data to constrain dataset values to a standard range'. The bottom status bar indicates 'Upload of the dataset "BikeBuyer.csv" has completed.' and includes buttons for 'NEW', 'RUN HISTORY', 'SAVE', 'SAVE AS', 'DISCARD CHANGES', 'RUN', 'SET UP WEB SERVICE', and 'PUBLISH TO GALLERY'.

Add Split Data function to split the data such that 80% will be used for model training and 20% will be used for testing the model.

Microsoft Azure Machine Learning Studio (classic) interface showing a workflow for 'Bike Buyers Prediction'. The workflow includes the following steps:

- BikeBuyers.csv
- Select Columns in Dataset
- Normalize Data
- Split Data
- Train Model
- Score Model
- Evaluate Model

The 'Split Data' node is highlighted, and the 'Properties' pane on the right shows its configuration:

- Splitting mode:** Split Rows
- Fraction of rows in the ...:** 0.8
- Randomized split:** ☒
- Random seed:** 0
- Stratified split:** False
- START TIME:** 10/15/2...
- END TIME:** 10/15/2...
- ELAPSED TIME:** 0:00:00.0...
- STATUS CODE:** Finished
- STATUS DETAILS:** Task output was

Quick Help: Split the rows of a dataset into two distinct sets (more help...)

Add Train Model function for training the model.

Microsoft Azure Machine Learning Studio (classic) interface showing a workflow for 'Experiment created on 15/10/2020'. The workflow includes the following steps:

- BikeBuyer.csv
- Select Columns in Dataset
- Normalize Data
- Split Data
- Train Model

The 'Train Model' node is highlighted, and the 'Properties' pane on the right shows its configuration:

- Label column:** Selected columns: Launch the selector tool to make a selection
- Launch column selector:** Launch column selector

Quick Help: Train a previously created classification or regression model

Mini Map: A small overview of the workflow is shown in the bottom left corner.

Upload of the dataset 'BikeBuyer.csv' has completed.

Add Two-Class Logistic Regression to the layout

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface. The main workspace displays a workflow diagram with the following steps: **BikeBuyer.csv** (dataset) → **Select Columns in Dataset** (operation) → **Normalize Data** (operation) → **Split Data** (operation) → **Two-Class Logistic Regression** (model) → **Train Model** (operation). A tooltip for the **Select Columns in Dataset** operation reads: "Select Columns in Dataset. Selects columns to include or exclude from a dataset in an operation. Formerly known as Project Columns. L1 regularization weights." The right-hand sidebar shows the **Properties** panel for the **Two-Class Logistic Regression** model, with settings for **Create trainer mode** (Single Parameter), **Optimization tolerance** (1E-07), **L1 regularization weight** (1), **Memory size for L-BFGS** (20), **Random number seed**, and a checked **Allow unknown categ...** option. The bottom toolbar includes icons for **NEW**, **RUN HISTORY**, **SAVE**, **SAVE AS**, **DISCARD CHANGES**, **RUN**, **SET UP WEB SERVICE**, and **PUBLISH TO GALLERY**.

Select the attribute on which prediction will be performed.

The screenshot shows the Microsoft Azure Machine Learning Studio (classic) interface with the **Bike Buyers Prediction** experiment. A dialog box titled "Select a single column" is open, allowing the user to choose a column for prediction. The dialog has two tabs: **BY NAME** and **WITH RULES**. Under the **WITH RULES** tab, there is a table with columns **Include** and **column names**. The **Include** column has a dropdown menu set to **Include**, and the **column names** column has a dropdown menu set to **column names**. The **BikeBuyer** column is selected. The right-hand sidebar shows the **Properties** panel for the **Train Model** operation, with settings for **Label column** (Selected columns: Launch the selector tool to make a selection) and a **Launch column selector** button. The bottom toolbar includes icons for **NEW**, **RUN HISTORY**, **SAVE**, **SAVE AS**, **DISCARD CHANGES**, **RUN**, **SET UP WEB SERVICE**, and **PUBLISH TO GALLERY**.

Execute the training of the model

Microsoft Azure Machine Learning Studio (classic)

1714014_LY_DERE PINAK ...

Bike Buyers Prediction

Finished running ✓

Search experiment items

My Datasets

- BikeBuyer.csv

Samples

- Adult Census Income...
- Airport Codes Dataset
- Automobile price dat...
- Bike Rental UCI dataset
- Bill Gates RGB Image
- Blood donation data
- Book Reviews from A...
- Breast cancer data
- Breast Cancer Features
- Breast Cancer Info
- CRM Appetency Labe...
- CRM Churn Labels Sh...
- CRM Dataset Shared

Two-Class Logistic Regression ✓

Split Data ✓

Train Model ✓

Properties Project

Experiment Properties

START TIME 10/15/2...

END TIME 10/15/2...

STATUS CODE Finished

STATUS DETAILS None

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

Add Score Model to the layout so that the trained model can be tested.

Microsoft Azure Machine Learning Studio (classic)

1714014_LY_DERE PINAK ...

Bike Buyers Prediction

In draft

Draft saved at 01:51:15

Search experiment items

Machine Learning

- Score
- Score Model

Two-Class Logistic Regression ✓

Split Data ✓

Train Model ✓

Score Model 1

Properties Project

Score Model

Append score column...

Quick Help

Score a trained classification or regression model
(more help...)

Add Evaluate Model to layout to evaluate the accuracy and precision of the model with respect to the results obtained.

Microsoft Azure Machine Learning Studio (classic) interface showing the 'Bike Buyers Prediction' workflow. The workflow includes steps: BikeBuyers.csv, Select Columns In Dataset, Normalize Data, Split Data, Train Model, Score Model, and Evaluate Model. The 'Evaluate Model' step is highlighted with a green checkmark. The right sidebar shows the 'Properties' pane with a 'Summary' section stating 'Prediction of Bike buyers using Logistic Regression.' and a 'Description' section explaining the prediction goal. The bottom toolbar includes buttons for NEW, RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

Check the weightage of the attributes for prediction. So, here the income and number of cars possessed are the two most important attributes for prediction of the outcome.

Bike Buyers Prediction > Train Model > Trained model

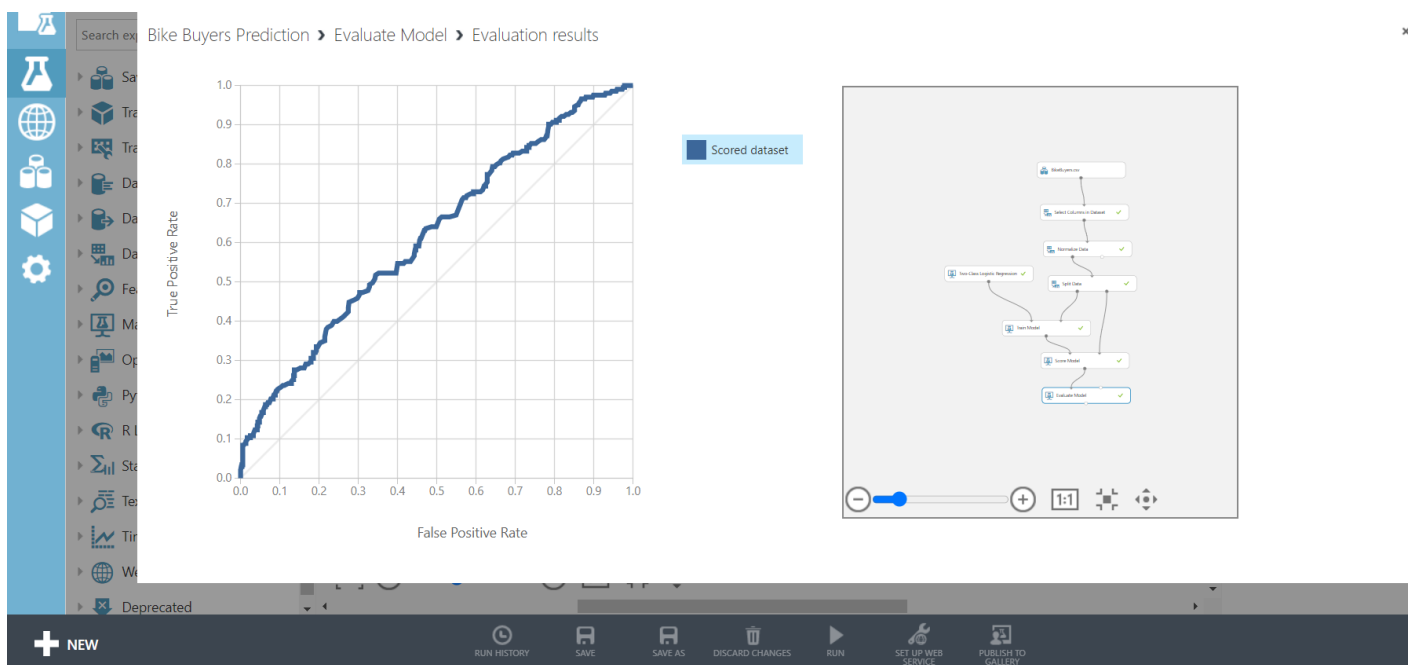
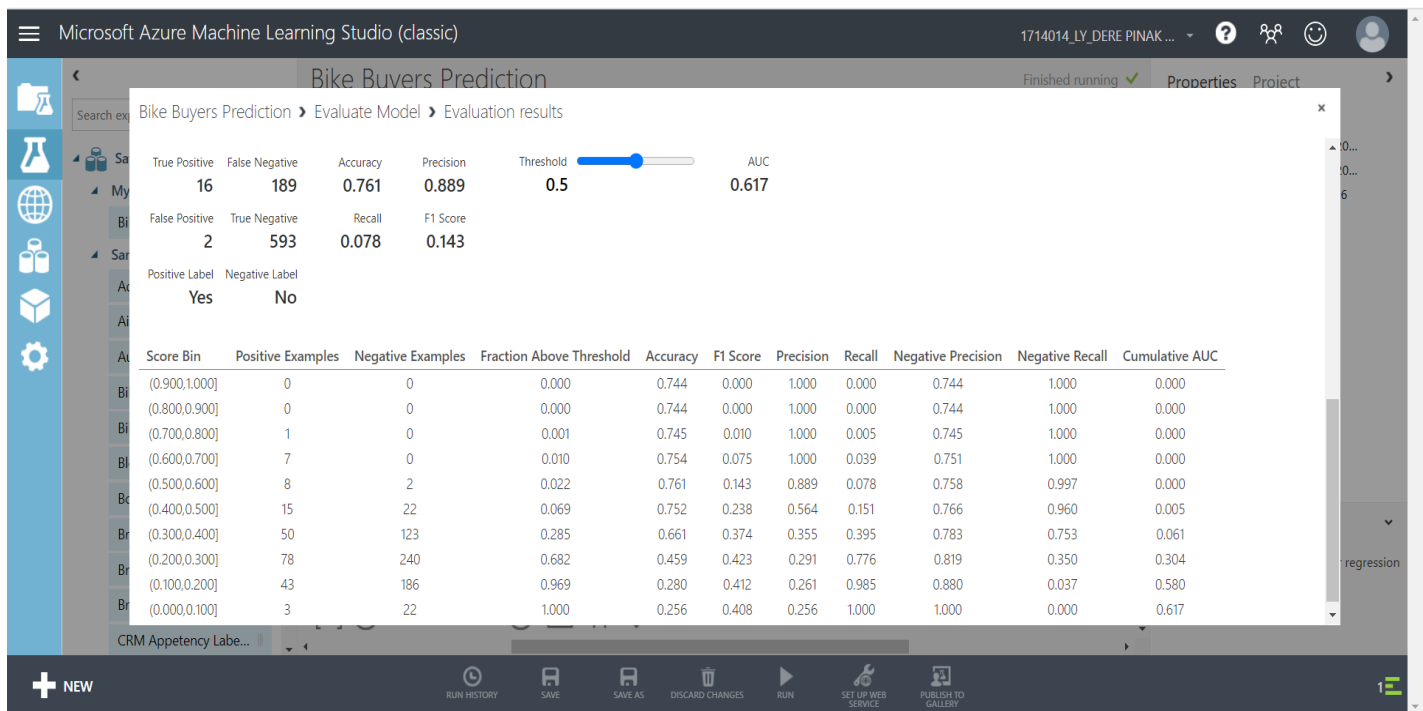
Feature Weights

Feature	Weight
Yearly Income	1.38128
Cars	-0.89267
Region_Pacific_2	0.829824
Bias	-0.793642
Children	-0.685978
Marital Status_Married_0	-0.410857
Education_Graduate	-0.410424
Degree_1	-0.410424
Commute Distance_5-10 Miles_4	-0.385682
Commute Distance_10+ Miles_1	-0.359797
Commute Distance_0-1 Miles_0	0.28029
Occupation_Professional_3	0.271654
Occupation_Manual_2	-0.238731
Home Owner_No_0	-0.220494

Deprecated

Bottom toolbar of the Microsoft Azure Machine Learning Studio (classic) interface. It includes buttons for RUN HISTORY, SAVE, SAVE AS, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

Visualize the Evaluate Model component to check out the results.



Questions:

Discuss the tangible and intangible benefits the business has observed after the implementation.

Ans:

- Increased the chances of high business revenue.
- Manpower got saved due to automation.
- This model can be used now anywhere in the enterprise.
- Lower expenditure.

Outcomes: Realize adequate perspectives of big data analytics in various applications.

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

Thus, I was able to train a model which predicts whether a customer is interested in purchasing a bike or not based on their age, income, gender, education, number of cars possessed and some other attributes using Azure ML Studio and then was able to publish it on GitHub:

<https://github.com/Pinak1998/BikeBuyersPrediction>

Link for the Azure Experiment:

<https://gallery.azure.ai/Experiment/Bike-Buyers-Prediction-44>

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

Signature of faculty in-charge with date

References:

Books/ Journals/ Websites: