



[AZURE ML PERFORMED ON BIKE SHARING DATA]



By Sreetam Dev
[sreetamkumardev@gmail.com]

Table of Contents

1.Setting up the resource group, workspace for the bike share data.	2
2.Reviewing and creating the workspace for the data.	2
3.Setting up the Azure ML studio.	3
4.Setting up the compute instances, compute clusters, inference clusters before proceeding to data exploration, modelling and deployment.	3
5.Setting up the dataset information, data storage requirements and creating the BIKE-SHARE-DATA dataset.....	4
6.Creating Automated ML environment for finding the best model and validation metrics for the dataset . Here, the metric for validation is chosen to normalized root mean squared error. And also the exit strategy for the modelling can be fine-tuned to save computation resource usage and time.....	4
7.Examining the end results in terms of metrics, of the processed Auto ML environment.	5
8.The best metric returned, was performed by the following model. Now, this differs from the result which we achieved on the Jupyter notebook as the modelling performed over here has not undergone feature selection.	5
9.Now, we are exploring the designer module of the Azure, where can create the data pipeline based on our earlier findings from the Jupyter notebook data script for the “Bike Share Data”. We need to drag and drop methods defined on the left hand side into the workspace for each implementation that we need to execute. For e.g. Here we have “Cleaning Missing Data method” for handling missing instances.....	6
10.Visualization on the data file is performed to explore if the changes have been implemented or not.	6
11. Columns are being selected to handle instances of features that are important, as discovered from the Jupyter data script for our predictive modelling	7
12.Data normalization is performed, to reduce weightage of any influencing feature that might bias the predictive modelling.	7
13.Linear Regression along with Train method are drawn in order to train and test the data file.	8
14.Fine tuning the hyper parameters of the Decision Forest Regression model to achieve better prediction results.	8
15.Altering the previous Normalizing method to MinMax for data scaling.....	9
16.Score Model is being fined tuned in terms of setting.	9
17.The similar process is repeated with Decision Forest Regression to compare both the outcomes in terms of metrics to fine tune our modelling process. At the end nodes we are using Score Model and Evaluate model methods to draw outputs for both modelling techniques.	10
18.The model is being processed to fetch results in terms of prediction and metrics to evaluate the model.	10
19.Since Decision Forest Regression model performed better, the inference pipeline is being created as the model we be prepared for deployment.....	11
20.The inference pipeline has been created and the web service input node is being allocated to the pipeline so that user can provide input after deployment and model will fetch the predictive results for the input instance provided.	11
21.The endpoints are being defined for the web api for the prepared data model.	12
22.A separate data output instance is being designed for the model stored within the notebook to be displayed by the web api.	12
23.The endpoint and key of the default practise script of Azure ML is being replaced with our Endpoint’s.	13
24.Once the environment is ready, the output python script is connected and submitted. Once processed it is deployed.	13

1.Setting up the resource group, workspace for the bike share data.

Microsoft Azure

Search resources, services, and docs (G+/)

azure.sreetamdev@gma...
DEFAULT DIRECTORY

All services > New > Machine Learning >

Machine Learning

Create a machine learning workspace

Basics Networking Advanced Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *

Resource group *
[Create new](#)

Workspace details

Specify the name, region, and edition for the workspace.

Workspace name *

Region *

For your convenience, these resources are added automatically to the workspace, if regionally available: Azure Storage, Azure Application Insights, Azure Key Vault

Click to stop screen recording

[Review + create](#) [< Previous](#) [Next : Networking](#)

2.Reviewing and creating the workspace for the data.

Microsoft Azure

Search resources, services, and docs (G+/)

azure.sreetamdev@gma...
DEFAULT DIRECTORY

All services > New > Machine Learning >

Machine Learning

Create a machine learning workspace

Validation passed

Basics Networking Advanced Tags **Review + create**

Basics

Subscription	Free Trial
Resource group	(New) Prediction_Bike_Sharing_Resource
Region	Australia East
Workspace name	Prediction_Bike_Sharing_Workspace

Networking

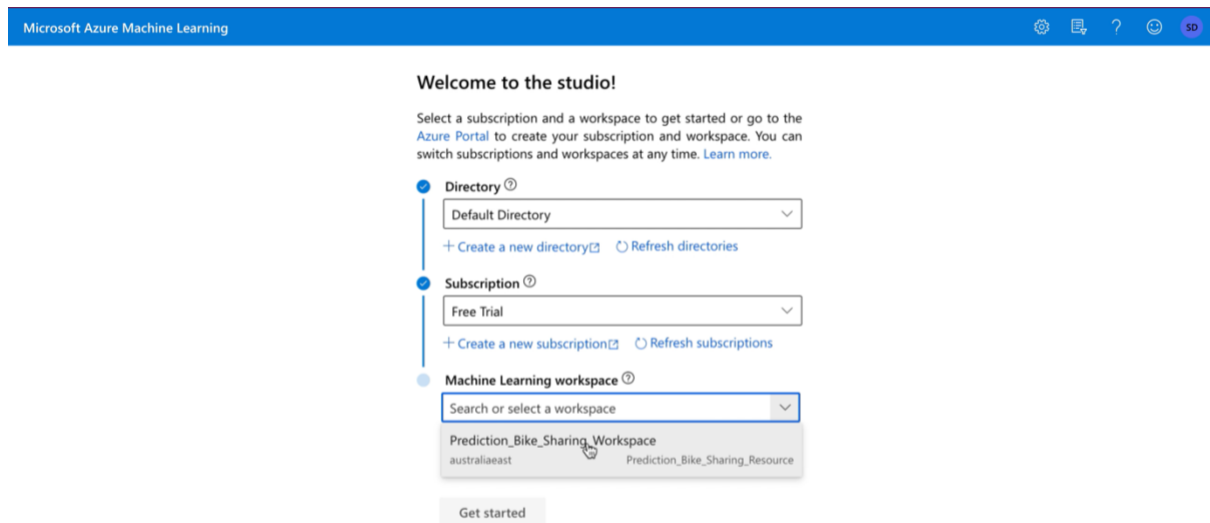
Connectivity method	Public endpoint (all networks)
---------------------	--------------------------------

Advanced

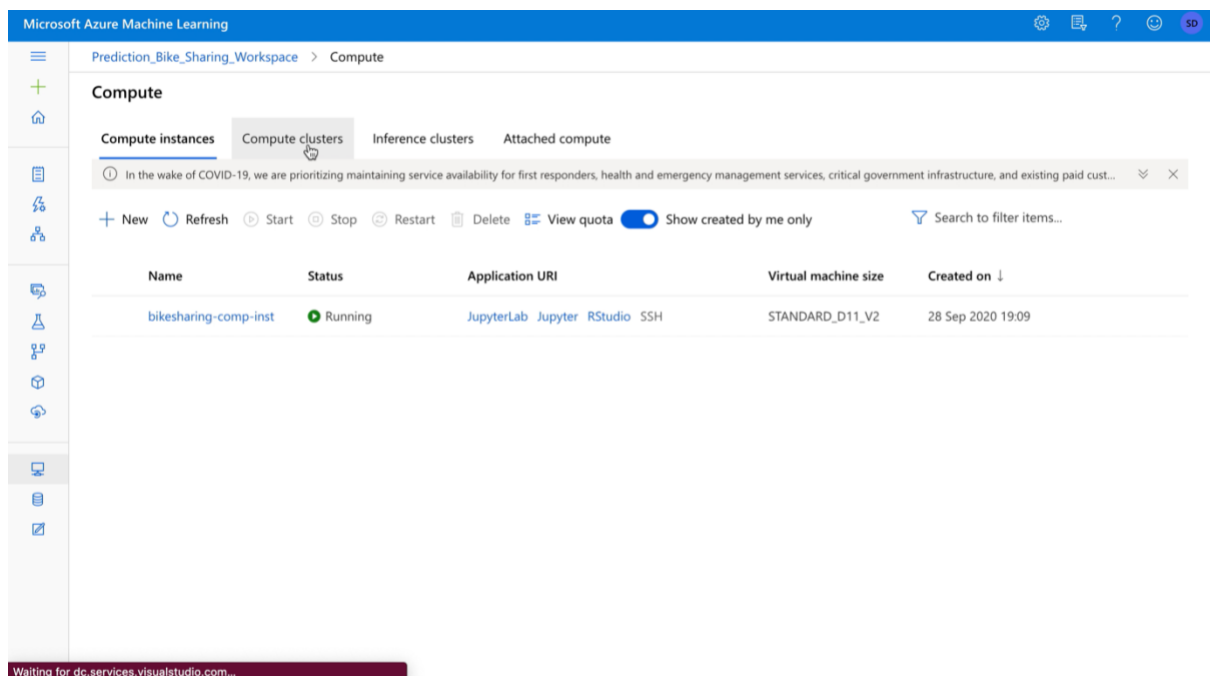
Encryption type	Microsoft-managed keys
Enable HBI Flag	Disabled

[Create](#) [< Previous](#) [Next >](#) [Download a template for automation](#)

3.Setting up the Azure ML studio.



4.Setting up the compute instances, compute clusters, inference clusters before proceeding to data exploration, modelling and deployment.



5. Setting up the dataset information, data storage requirements and creating the BIKE-SHARE-DATA dataset.

The screenshot shows the 'Create dataset from local files' wizard in the Microsoft Azure Machine Learning interface. The 'Basic info' step is active, showing the following fields:

- Name:** BIKE-SHARE-DATA
- Dataset version:** 1
- Dataset type:** Tabular
- Description:** Two-year historical log corresponding to years 2011 and 2012 from Capital Bikeshare system, Washington D.C., USA

Navigation buttons at the bottom include 'Back', 'Next', and 'Cancel'.

6. Creating Automated ML environment for finding the best model and validation metrics for the dataset. Here, the metric for validation is chosen to be normalized root mean squared error. And also the exit strategy for the modelling can be fine-tuned to save computation resource usage and time.

The screenshot shows the 'Additional configurations' dialog for an Automated ML run. The 'Primary metric' is set to 'Normalized root mean squared error'. The 'Explain best model' checkbox is checked. The 'Blocked algorithms' list includes:

- ElasticNet
- DecisionTree
- KNN
- ExtremeRandomTrees

A dropdown menu is open, showing a list of algorithms that Automated ML will not use during training:

- GradientBoosting
- LassoLars
- SGD
- RandomForest
- LightGBM
- XGBoostRegressor
- FastLinearRegressor
- OnlineGradientDescentRegressor

Navigation buttons at the bottom include 'Back', 'Next', 'Save', and 'Cancel'.

7.Examining the end results in terms of metrics, of the processed Auto ML environment.

The screenshot shows the Microsoft Azure Machine Learning interface. The left sidebar contains navigation options: New, Home, Author, Notebooks, Automated ML (selected), Designer, and Assets. The main area displays the 'Run 5' status as 'Completed'. Below this, there are tabs for Details, Model, Explanations (preview), Metrics (selected), Outputs + logs, Images, Child runs, and Snapshot. The Metrics tab shows a list of metrics on the left and their values on the right. The metrics listed are: explained_variance, mean_absolute_error, mean_absolute_percentage_error, median_absolute_error, normalized_mean_absolute_error, normalized_median_absolute_error, normalized_root_mean_squared_error, predicted_true, r2_score, and residuals. The values for the first five metrics are: explained_variance (0.995590395407404), mean_absolute_error (81.26004478015443), mean_absolute_percentage_error (6.5123558567201325), median_absolute_error (51.49793075908722), and normalized_mean_absolute_error (0.009348831659014544).

Metric	Value
explained_variance	0.995590395407404
mean_absolute_error	81.26004478015443
mean_absolute_percentage_error	6.5123558567201325
median_absolute_error	51.49793075908722
normalized_mean_absolute_error	0.009348831659014544

8.The best metric returned, was performed by the following model. Now, this differs from the result which we achieved on the Jupyter notebook as the modelling performed over here has not undergone feature selection.

The screenshot shows the Microsoft Azure Machine Learning interface. The left sidebar contains navigation options: New, Home, Author, Notebooks, Automated ML (selected), Designer, and Assets. The main area displays the 'Run 5' status as 'Completed'. Below this, there are tabs for Details, Model (selected), Explanations (preview), Metrics, Outputs + logs, Images, Child runs, and Snapshot. The Model tab shows a 'Model summary' section with the following information: Algorithm name (MaxAbsScaler, LightGBM), Normalized root mean squared error (0.01455), Sampling (100.00 %), Registered models (No registration yet), and Deploy status (No deployment yet).

Model summary

Algorithm name
MaxAbsScaler, LightGBM

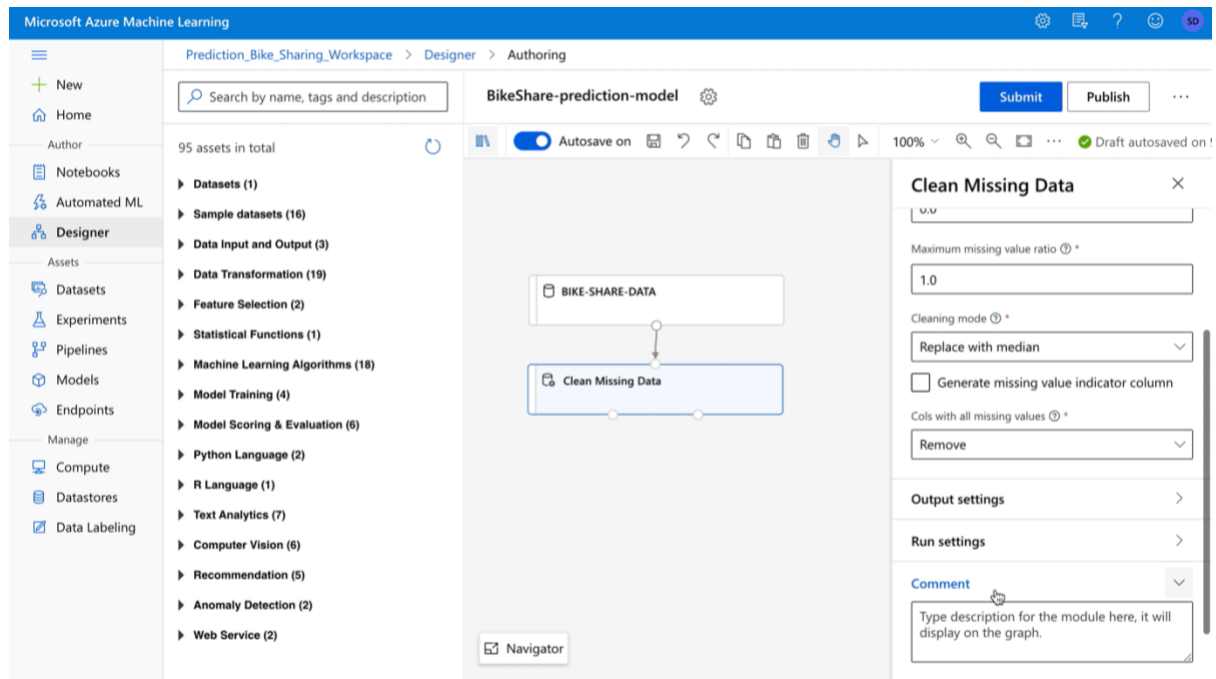
Normalized root mean squared error
0.01455 [View all other metrics](#)

Sampling
100.00 %

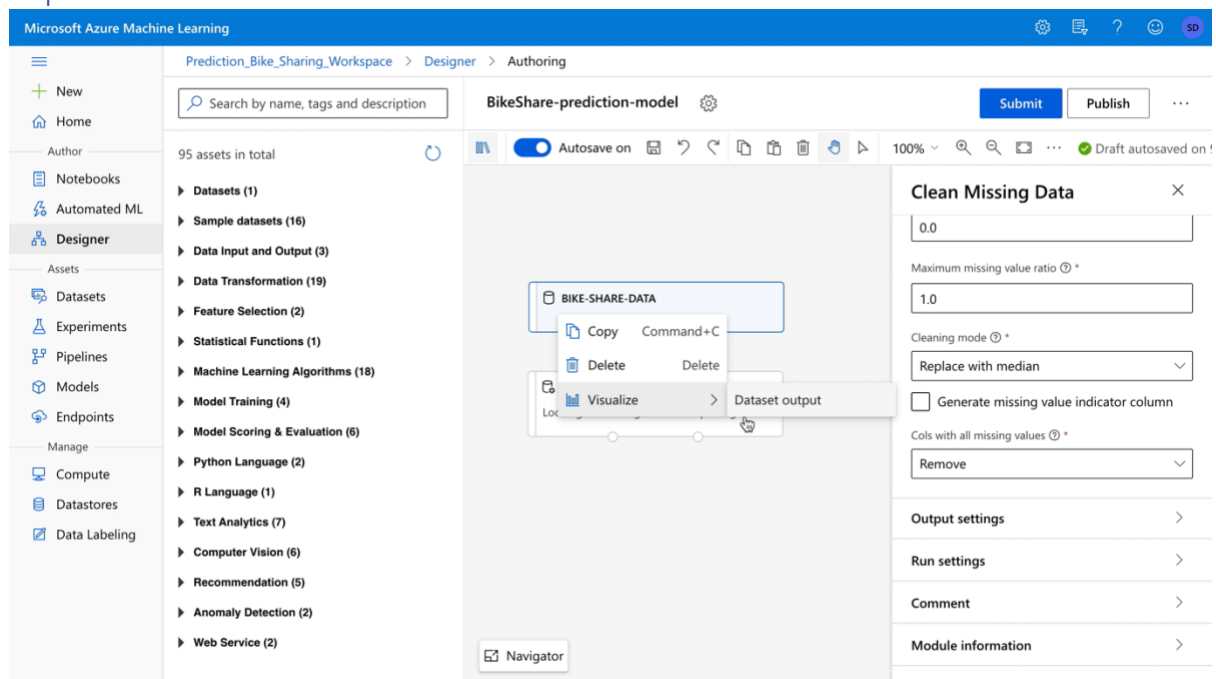
Registered models
No registration yet

Deploy status
No deployment yet

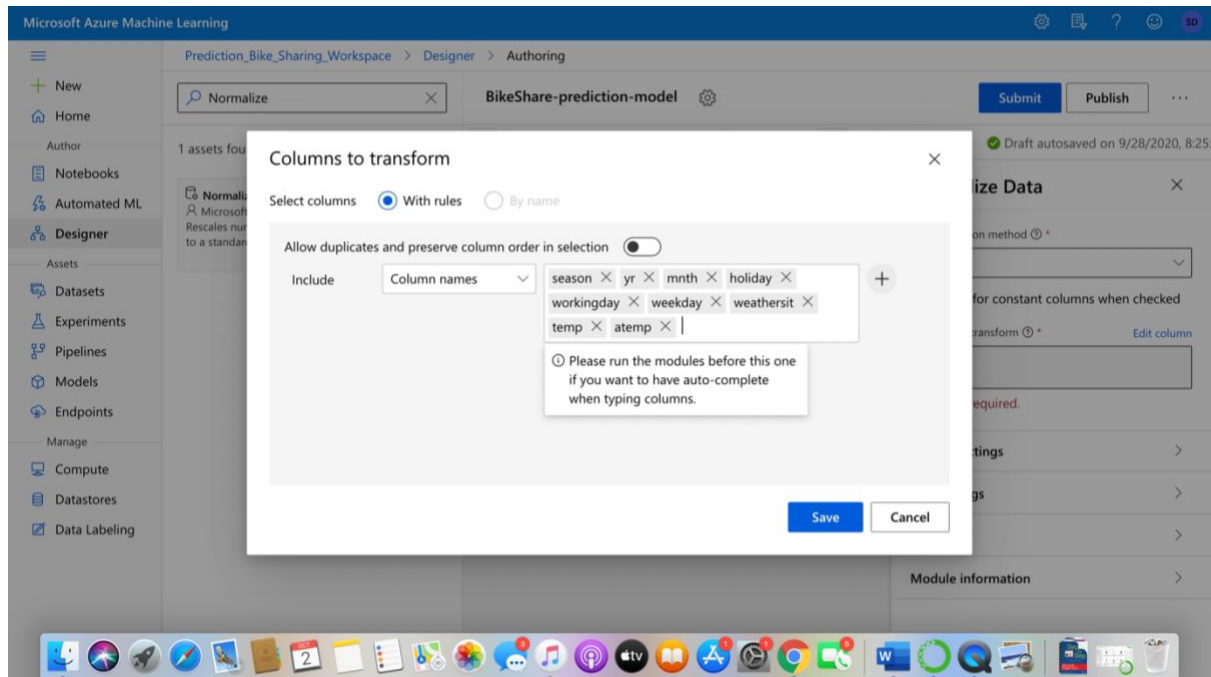
9. Now, we are exploring the designer module of the Azure, where can create the data pipeline based on our earlier findings from the Jupyter notebook data script for the “Bike Share Data”. We need to drag and drop methods defined on the left hand side into the workspace for each implementation that we need to execute. For e.g. Here we have “Cleaning Missing Data method” for handling missing instances.



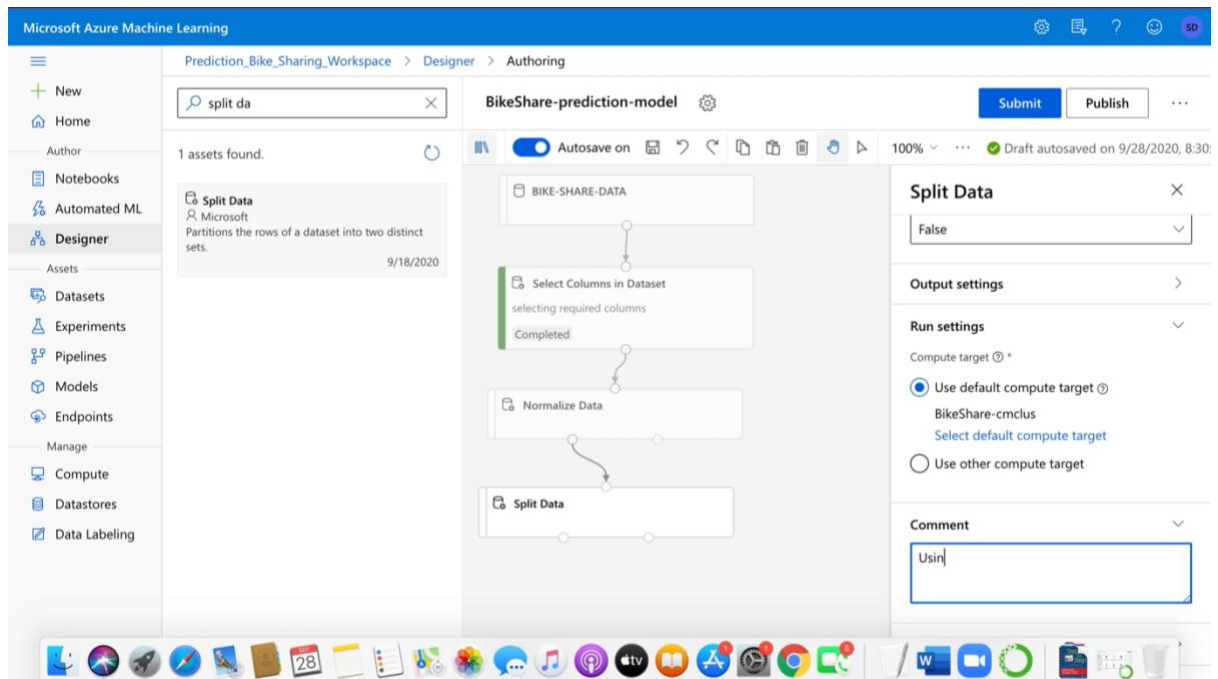
10. Visualization on the data file is performed to explore if the changes have been implemented or not.



11. Columns are being selected to handle instances of features that are important, as discovered from the Jupyter data script for our predictive modelling .



12.Data normalization is performed, to reduce weightage of any influencing feature that might bias the predictive modelling.



13. Linear Regression along with Train method are drawn in order to train and test the data file.

The screenshot displays the Microsoft Azure Machine Learning Designer interface. The workspace is titled "BikeShare-prediction-model". On the left, the "Assets" pane shows a search for "Linear" resulting in one asset: "Linear Regression" (Microsoft, 9/18/2020). The central canvas shows a pipeline with three steps: "Linear Regression", "Split Data", and "Train Model". The "Linear Regression" step is selected, and its configuration panel on the right is visible. The configuration includes:

- Solution method: Online Gradient Descent
- Create trainer mode: SingleParameter
- Learning rate: 0.00001
- Number of training epochs: 10
- L2 regularization weight: 0.001
- Normalize features: ☒
- Decrease learning rate: ☐
- Random number seed: (empty field)

14. Fine tuning the hyper parameters of the Decision Forest Regression model to achieve better prediction results.

The screenshot displays the Microsoft Azure Machine Learning Designer interface. The workspace is titled "BikeShare-prediction-model". On the left, the "Assets" pane shows a search for "Decision Forest Regression" resulting in six assets: "Decision Forest Regression", "Fast Forest Quantile Regression", "Linear Regression", "Neural Network Regression", and "Poisson Regression". The central canvas shows a pipeline with four steps: "Select Columns in Dataset", "Normalize Data", "Split Data", and "Train Model". The "Select Columns in Dataset" step is selected, and its configuration panel on the right is visible. The configuration includes:

- Create trainer mode: SingleParameter
- Number of decision trees: 8
- Maximum depth of the decision trees: 30
- Minimum number of samples per leaf node: 1
- Resampling method: Bagging Resampling

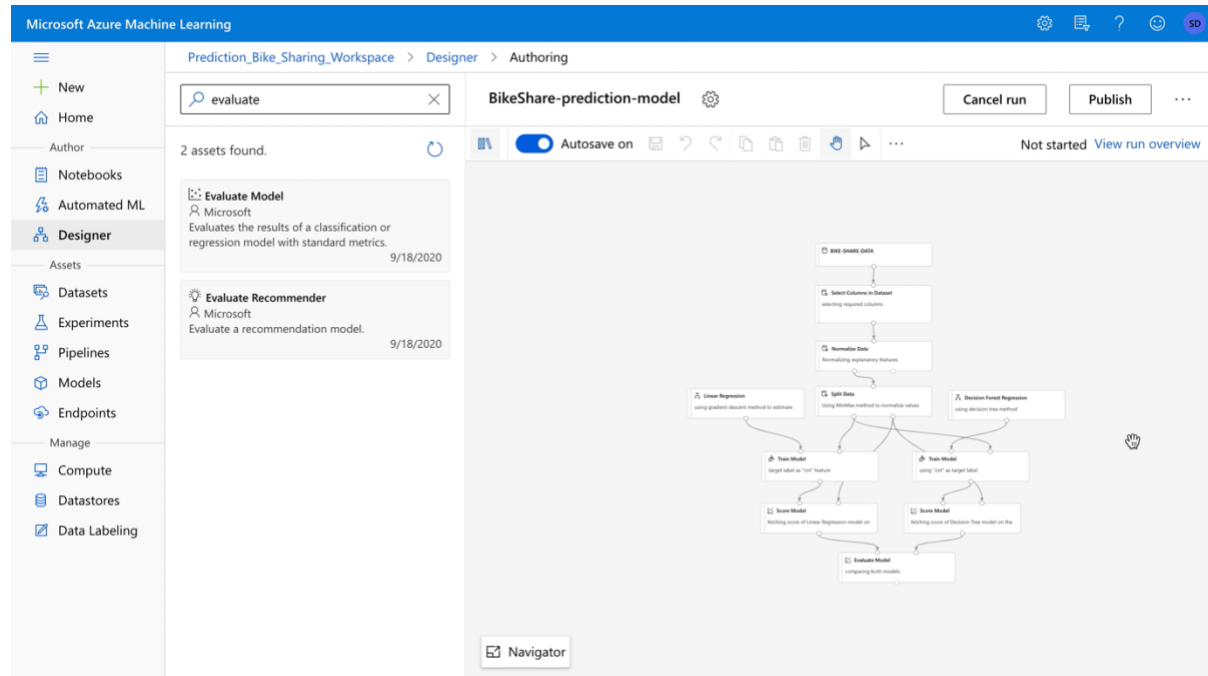
15. Altering the previous Normalizing method to MinMax for data scaling.

The screenshot shows the Microsoft Azure Machine Learning Designer interface. The workspace is titled 'BikeShare-prediction-model'. The left sidebar shows the 'Designer' tab selected. The main workspace contains a flowchart with the following modules: 'Select Columns in Dataset' (Completed), 'Normalize Data', 'Split Data' (Using MinMax method to normalize), and 'Train Model'. The 'Normalize Data' module is selected, and its configuration panel is open on the right. The 'Transformation method' is set to 'MinMax'. The 'Columns to transform' list includes: season, yr, mnth, holiday, workingday, weekday, weathersit, temp, atemp, hum, windspeed. The 'Output settings' panel shows the 'Comment' field with the text 'Normalizing explanatory fe'. The 'Module information' panel is also visible.

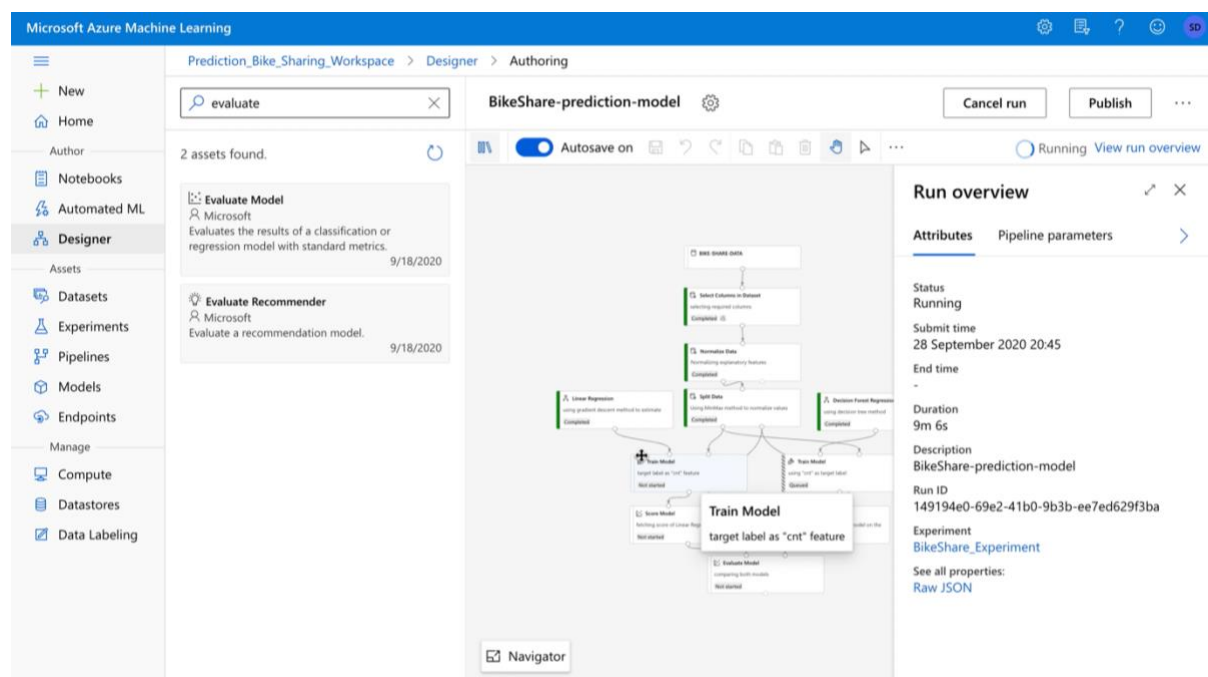
16. Score Model is being fined tuned in terms of setting.

The screenshot shows the Microsoft Azure Machine Learning Designer interface. The workspace is titled 'BikeShare-prediction-model'. The left sidebar shows the 'Designer' tab selected. The main workspace contains a flowchart with the following modules: 'Normalize Data' (Normalizing explanatory features), 'Split Data' (Using MinMax method), 'Train Model' (target label as "cnt" feature), and 'Score Model'. The 'Score Model' module is selected, and its configuration panel is open on the right. The 'Append score columns to output' checkbox is checked. The 'Output settings' panel is also visible.

17. The similar process is repeated with Decision Forest Regression to compare both the outcomes in terms of metrics to fine tune our modelling process. At the end nodes we are using Score Model and Evaluate model methods to draw outputs for both modelling techniques.



18. The model is being processed to fetch results in terms of prediction and metrics to evaluate the model.



19. Since Decision Forest Regression model performed better, the inference pipeline is being created as the model we be prepared for deployment.

Microsoft Azure Machine Learning

Prediction_Bike_Sharing_Workspace > Designer > Authoring

evaluate

BikeShare-prediction-model

Submit Create inference pipeline Publish

Run finished View run overview

2 assets found.

Evaluate Model
Microsoft
Evaluates the results of a classification or regression model with standard metrics.
9/18/2020

Evaluate Recommender
Microsoft
Evaluate a recommendation model.
9/18/2020

Run overview

Attributes Pipeline parameters

Status Completed

Submit time 28 September 2020 21:45

End time 28 September 2020 21:52

Duration 7m 34s

Description BikeShare-prediction-model

Run ID 97778698-f523-47b1-89c3-6ce6c314b168

Experiment BikeShare_Experiment

See all properties: Raw JSON

20. The inference pipeline has been created and the web service input node is being allocated to the pipeline so that user can provide input after deployment and model will fetch the predictive results for the input instance provided.

Microsoft Azure Machine Learning

Prediction_Bike_Sharing_Workspace > Designer > Authoring

Training pipeline Real-time inference pipeline

evaluate

BikeShare-prediction-model-real time inference

Submit Deploy

Draft autosaved on 9/28/2020, 9:55:38 PM

2 assets found.

Evaluate Model
Microsoft
Evaluates the results of a classification or regression model with standard metrics.
9/18/2020

Evaluate Recommender
Microsoft
Evaluate a recommendation model.
9/18/2020

Run overview

Attributes Pipeline parameters

Status Completed

Submit time 28 September 2020 21:45

End time 28 September 2020 21:52

Duration 7m 34s

Description BikeShare-prediction-model

Run ID 97778698-f523-47b1-89c3-6ce6c314b168

Experiment BikeShare_Experiment

See all properties: Raw JSON

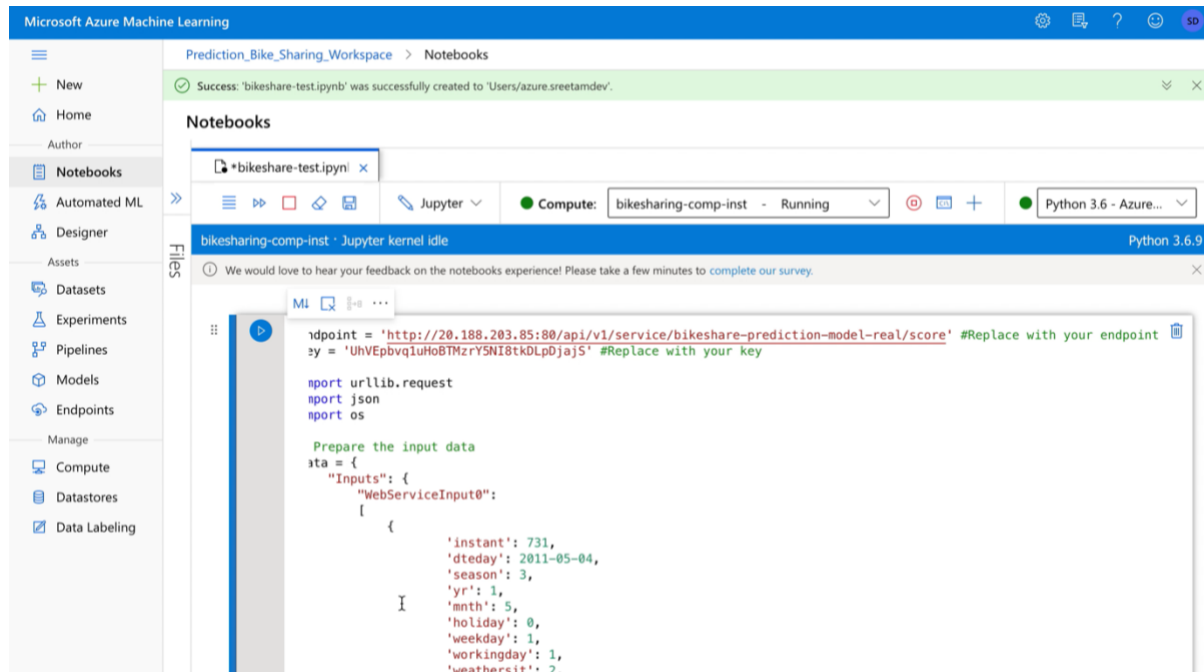
21.The endpoints are being defined for the web api for the prepared data model.

The screenshot shows the Microsoft Azure Machine Learning interface. The left sidebar contains navigation options: New, Home, Author, Notebooks, Automated ML, Designer, Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Manage, Compute, Datastores, and Data Labeling. The main pane displays the 'bikeshare-prediction-model-real' endpoint configuration. The 'Consume' tab is active, showing 'Basic consumption info' with a REST endpoint URL and authentication types (Using key selected). Below this, the 'Consumption option' section shows 'Consumption types' with tabs for C#, Python, and R. The Python tab is selected, displaying a code snippet for installing the Microsoft.AspNet.WebApi.Client NuGet package.

22.A separate data output instance is being designed for the model stored within the notebook to be displayed by the web api.

The screenshot shows the Microsoft Azure Machine Learning interface with a Jupyter notebook open. The left sidebar is the same as in the previous screenshot. The main pane shows a notebook titled 'bikeshare-test.ipynb'. A green notification bar at the top indicates 'Success: 'bikeshare-test.ipynb' was successfully created to 'Users/azure.sreetamdev''. The notebook content shows a REST API call using the requests library, with a JSON body containing various parameters like 'instant', 'dteday', 'season', 'yr', 'mnth', 'holiday', 'weekday', 'workingday', 'weathersit', 'temp', 'atemp', 'hum', 'windspeed', 'casual', and 'registered'. The code is written in Python 3.6.9.

23.The endpoint and key of the default practise script of Azure ML is being replaced with our Endpoint's.



24.Once the environment is ready, the output python script is connected and submitted. Once processed it is deployed.

