

# Assignment 3

## **Team 11**

Ankur Vora  
Krutika Dedhia  
Kinjal Gada

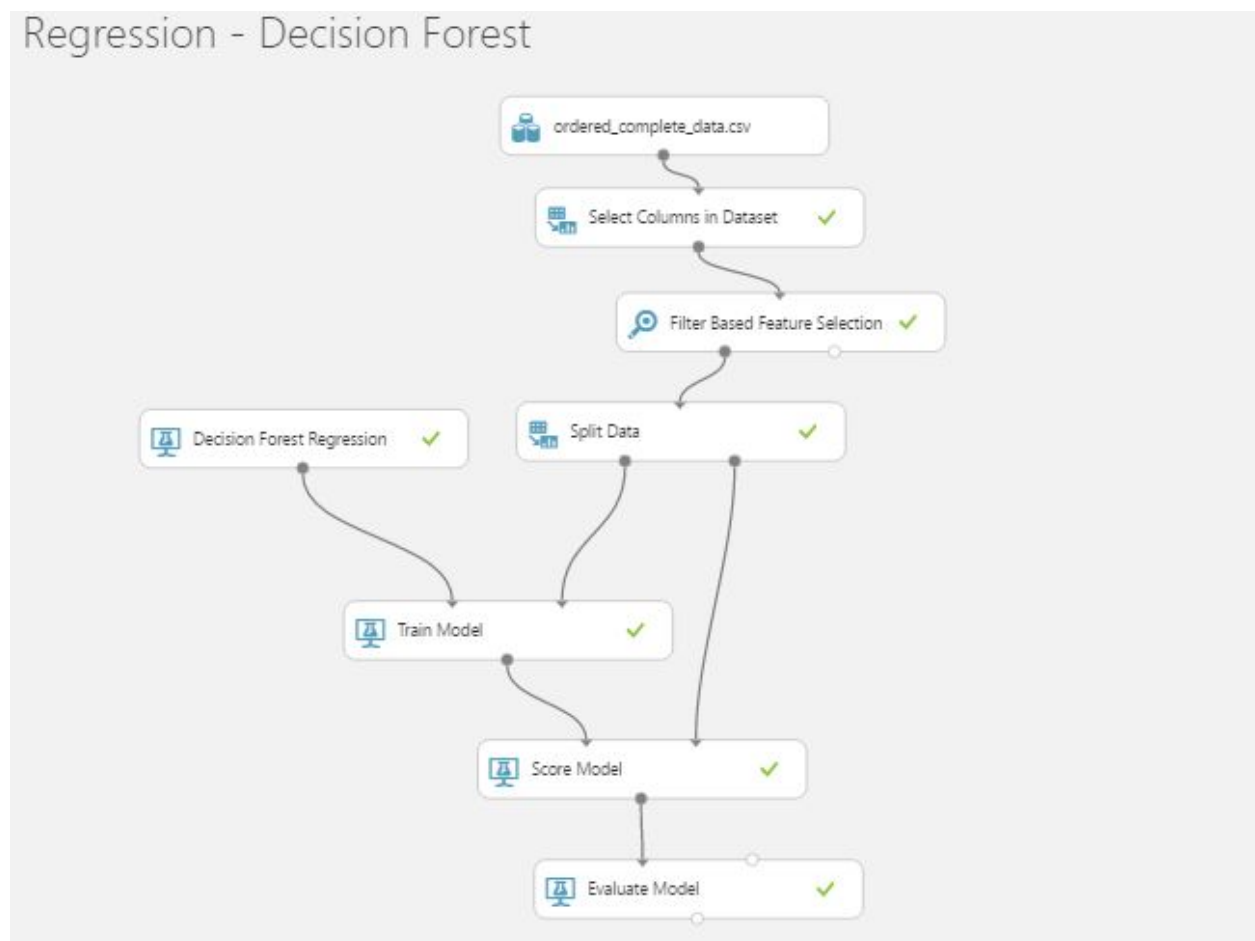
# Introduction

Using data from midterm, deploy the models as a web service.

## Regression

Used components in Azure to build trained regression models - decision forest, linear regression and neural network. Scored and evaluated the model and based on the accuracy selected Random forest as the best for our data. Deployed each model as a web service to call the API to predict the value for the normalized consumption

### Decision Forest



Regression - Decision Forest &gt; Filter Based Feature Selection &gt; Filtered dataset

rows  
621817columns  
12

	norm_consumption	Dew_PointF	TemperatureF	date	Humidity	WindDirDegrees	VisibilityMPH	Sea_Level_PressureIn
view as								
0.227	35.6	35.6	20130101	100	150	5.6	29.42	
0.227	33.8	35.6	20130101	93	150	4.3	29.39	
0.236	33.8	33.8	20130101	100	140	6.2	29.39	
0.227	33.8	35.6	20130101	93	160	5	29.36	
0.227	35.6	35.6	20130101	100	160	6.2	29.33	
0.227	37.4	37.4	20130101	100	180	6.2	29.33	
0.227	37.4	37.4	20130101	100	180	6.2	29.33	
0.245	37.4	37.4	20130101	100	180	6.2	29.3	
0.209	37.4	37.4	20130101	100	180	5	29.3	
0.236	37.4	37.4	20130101	100	180	5	29.3	







Regression - Decision Forest &gt; Filter Based Feature Selection &gt; Filtered dataset

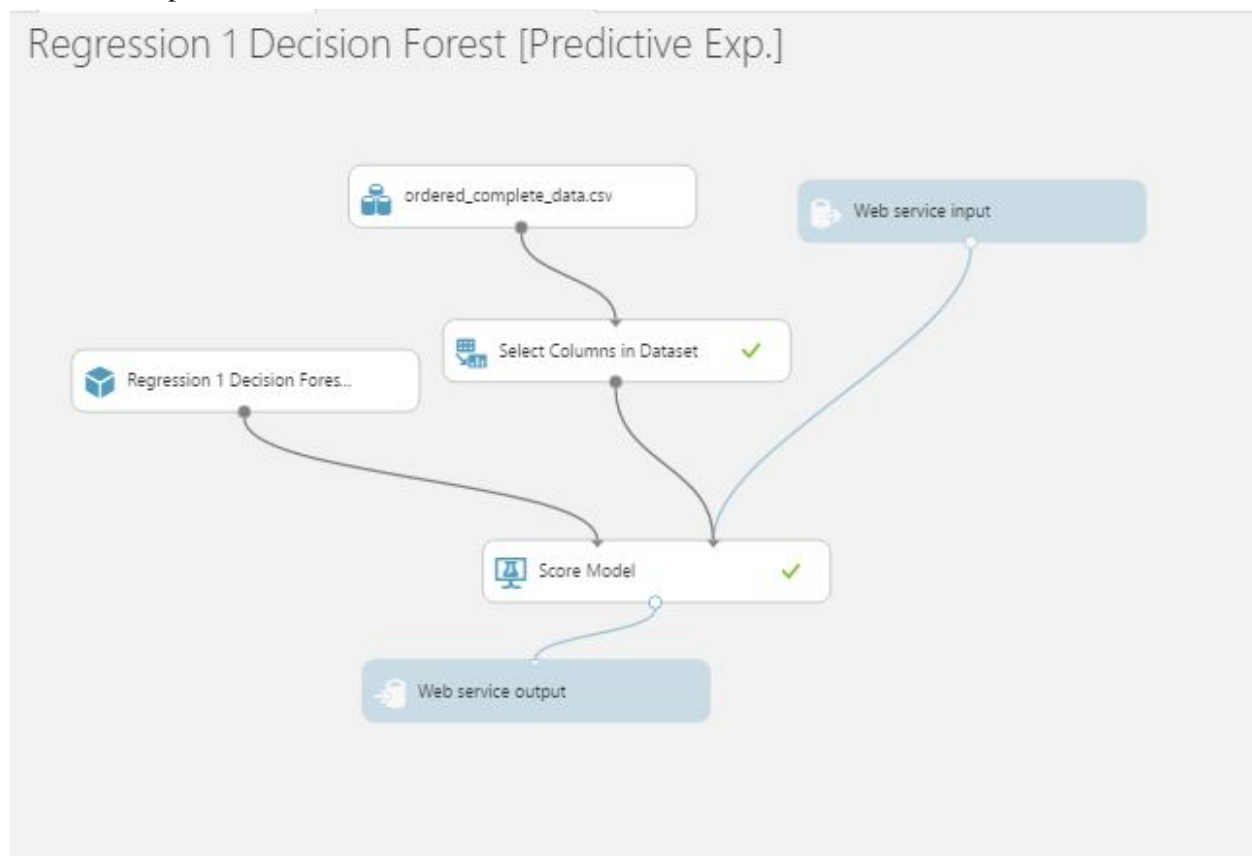
rows  
621817columns  
12

peratureF	date	Humidity	WindDirDegrees	VisibilityMPH	Sea_Level_PressureIn	isHoliday	Wind_SpeedMPH	hour	building
	20130101	100	150	5.6	29.42	1	11.5	0	Building 1
	20130101	93	150	4.3	29.39	1	13.8	1	Building 1
	20130101	100	140	6.2	29.39	1	12.7	2	Building 1
	20130101	93	160	5	29.36	1	10.4	3	Building 1
	20130101	100	160	6.2	29.33	1	10.4	4	Building 1
	20130101	100	180	6.2	29.33	1	11.5	5	Building 1
	20130101	100	180	6.2	29.33	1	12.7	6	Building 1
	20130101	100	180	6.2	29.3	1	12.7	7	Building 1
	20130101	100	180	5	29.3	1	11.5	8	Building 1
	20130101	100	180	5	29.3	1	12.7	9	Building 1

**Evaluation:**

Regression - Decision Forest &gt; Evaluate Model &gt; Evaluation results

rows	columns						
1	6						
		Negative Log Likelihood	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
view as							
		-290806.069217	0.009274	0.01767	0.601174	0.323426	0.676574

**Predictive Experiment:**

## Web Service:

regression 1 decision forest [predictive exp.]

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) [preview](#)

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

qPL3pQetjdtpe9gzLas5SbasW0B6sw9YXivvY0iYDOob3+Tgdqk9K9adadAkg1/E1+gsukMLUKVSc0FY4pw==

Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	<a href="#">Test</a> <a href="#">Test preview</a>	Excel 2013 or later    Excel 2010 or earlier workbook	12/2/2016 7:34:30 PM
BATCH EXECUTION	<a href="#">Test preview</a>	Excel 2013 or later workbook	12/2/2016 7:34:30 PM

## Output:

input1

date

20160303

hour

5

building

Building7

isHoliday

1

TemperatureF

78

Dew\_PointF

33.5

Humidity

23

Sea\_Level\_PressureIn

30

VisibilityMPH

4.5

Wind\_SpeedMPH

89

WindDirDegrees

180

Test Request-Response

output1

date

20160303

hour

5

building

Building7

isHoliday

1

TemperatureF

78

Dew\_PointF

33.5

Humidity

23

Sea\_Level\_PressureIn

30

VisibilityMPH

4.5

Wind\_SpeedMPH

89

WindDirDegrees

180

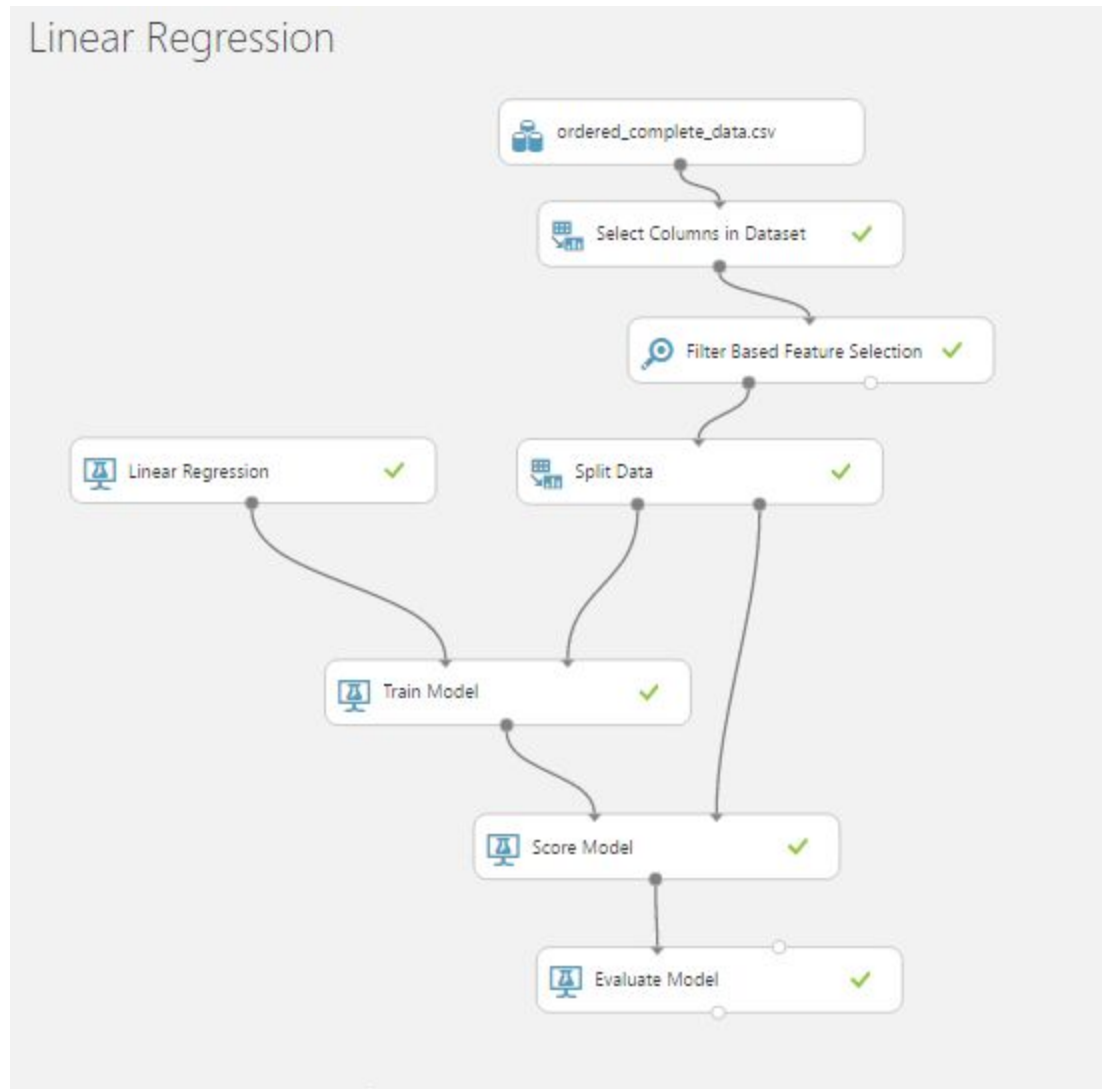
Scored Label Mean

0.0028741180237868

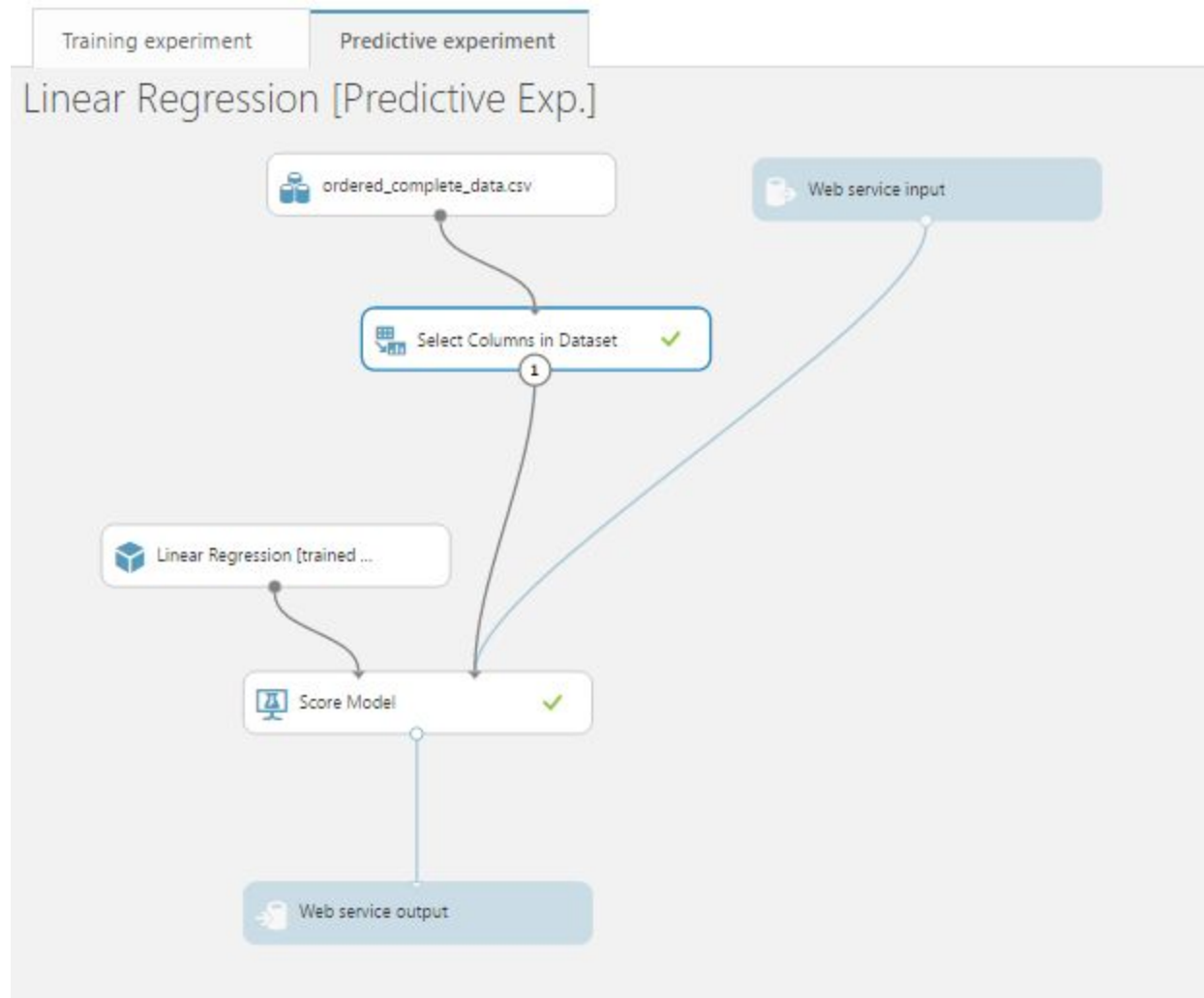
Scored Label Standard Deviation

0.00401704537174725

## Linear Regression:



## Predictive Experiment



## Web Service:

linear regression [predictive exp.]

DASHBOARD CONFIGURATION

General [New Web Services Experience](#) [preview](#)

Published experiment:

[View snapshot](#) [View latest](#)

Description:

No description provided for this web service.

API key:

INXs9qSZTEv02LeHfVR+Py6aRqRbihaRqlojikiHu8gKZohNZ7NTqHfP1PD30S578CC+Nyp5Z53VLg==

Default Endpoint:

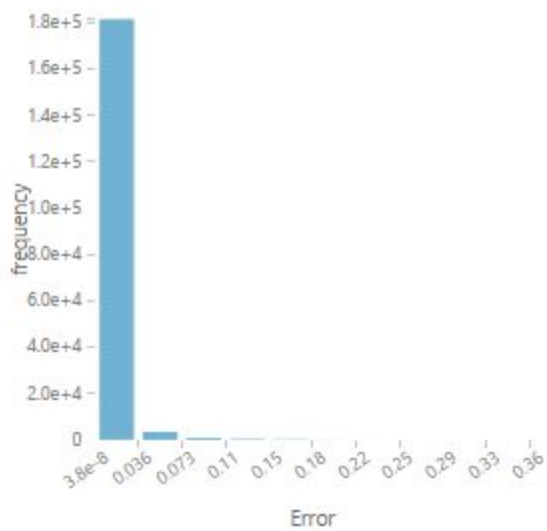
API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	<a href="#">Test</a> <a href="#">Test preview</a>	<a href="#">Excel 2013 or later</a> <a href="#">Excel 2010 or earlier workbook</a>	12/2/2016 11:25:49 PM
BATCH EXECUTION	<a href="#">Test</a> <a href="#">Test preview</a>	<a href="#">Excel 2013 or later workbook</a>	12/2/2016 11:25:49 PM

Linear Regression &gt; Evaluate Model &gt; Evaluation results

## Metrics

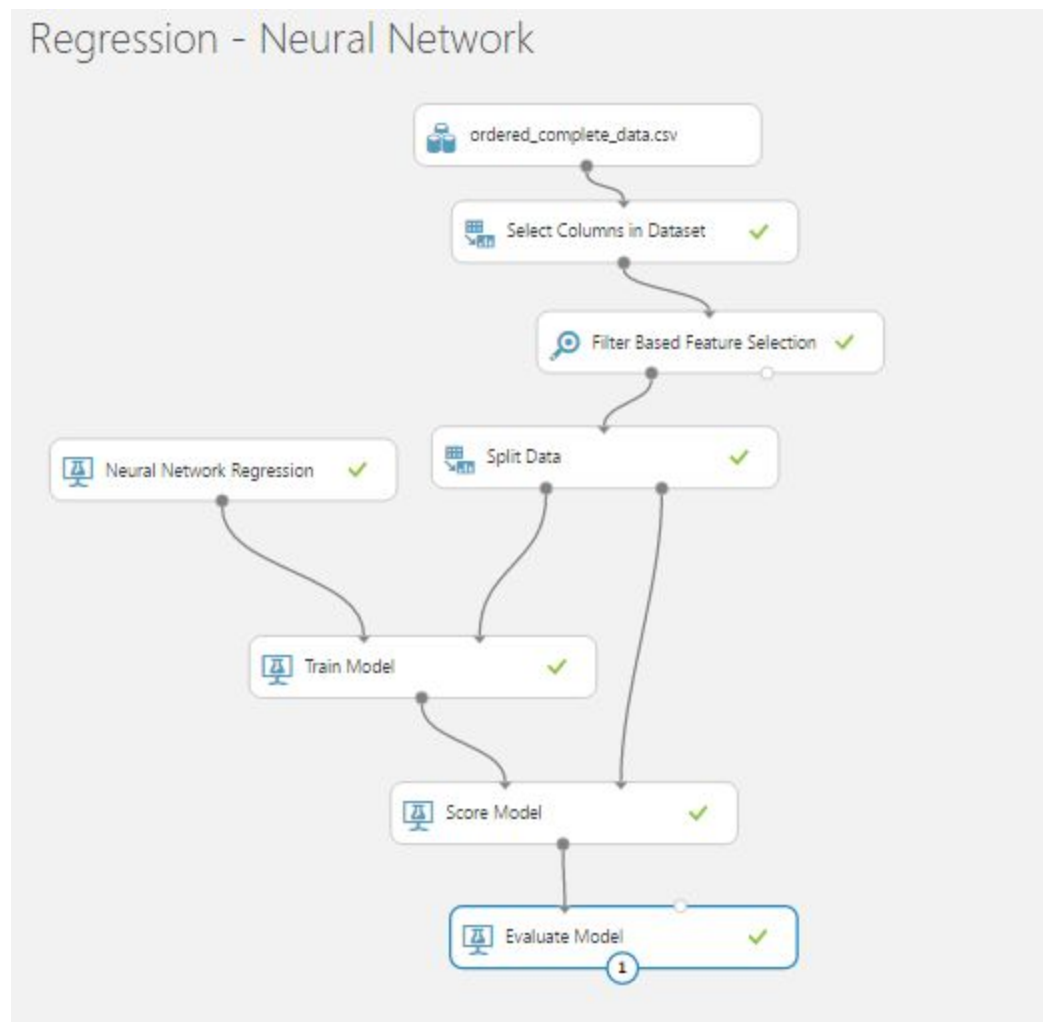
Mean Absolute Error	0.009568
Root Mean Squared Error	0.01669
Relative Absolute Error	0.620235
Relative Squared Error	0.288558
Coefficient of Determination	0.711442

## Error Histogram





## Neural Network:

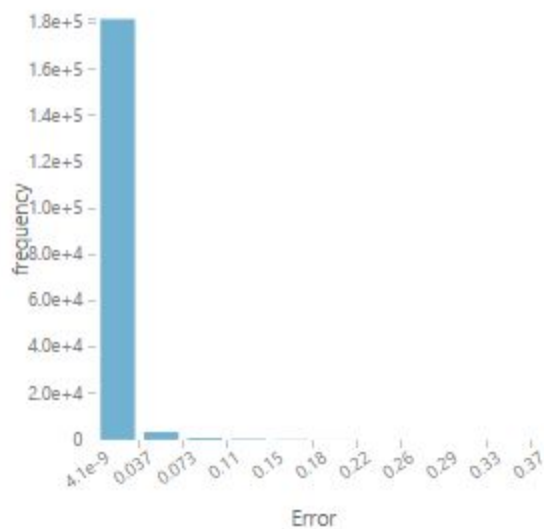


## Regression - Neural Network &gt; Evaluate Model &gt; Evaluation results

## Metrics

Mean Absolute Error	0.009801
Root Mean Squared Error	0.016674
Relative Absolute Error	0.635345
Relative Squared Error	0.287993
Coefficient of Determination	0.712007

## Error Histogram



## Web Service:

regression - neural network [predictive exp.]

DASHBOARD CONFIGURATION

General New Web Services Experience [preview](#)

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

pHJldMuWZCge54AjpOT4YMuFID6FDpPncQ8LVXGBUj3FuYvThR7/BmZm33SZKCurh2AeXTOY-g\*\*

Default Endpoint

API HELP PAGE

TEST

[Test](#) [Test preview](#)

APPS

[Excel 2013 or later](#) [Excel 2010 or earlier workbook](#)

LAST UPDATED

12/2/2016 11:30:52 PM

REQUEST/RESPONSE

BATCH EXECUTION

[Test](#) [Test preview](#)[Excel 2013 or later workbook](#)

12/2/2016 11:30:52 PM

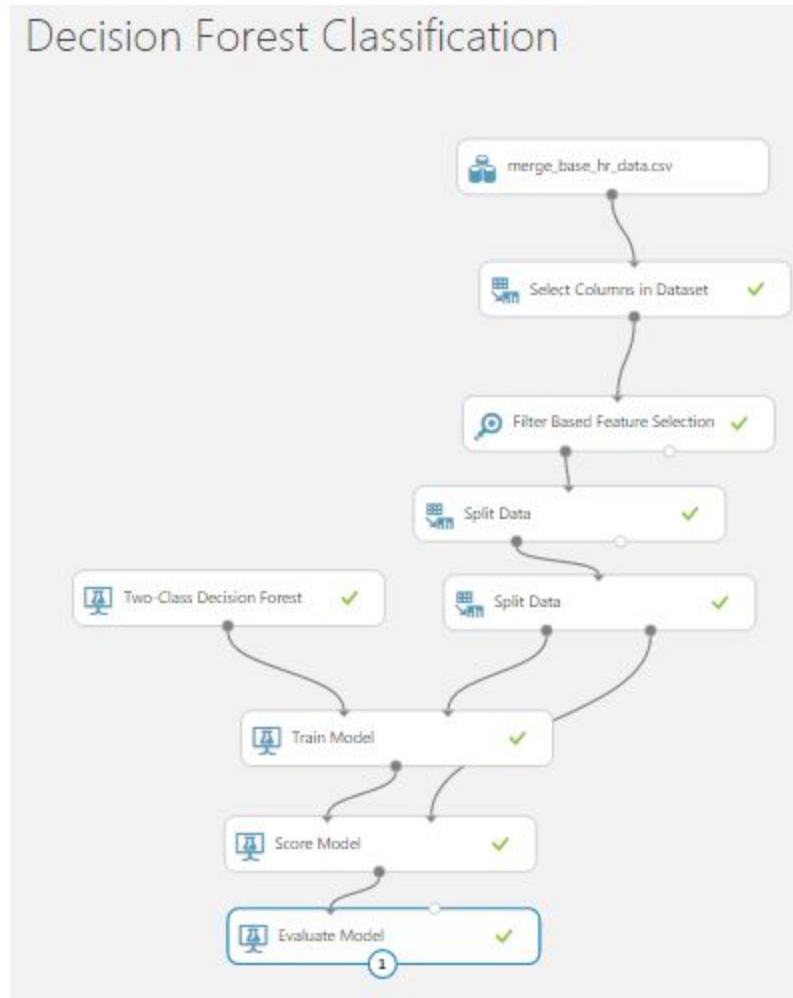
Best Model - Neural Network. The coefficient of determination value is high amongst all 3.

## Classification

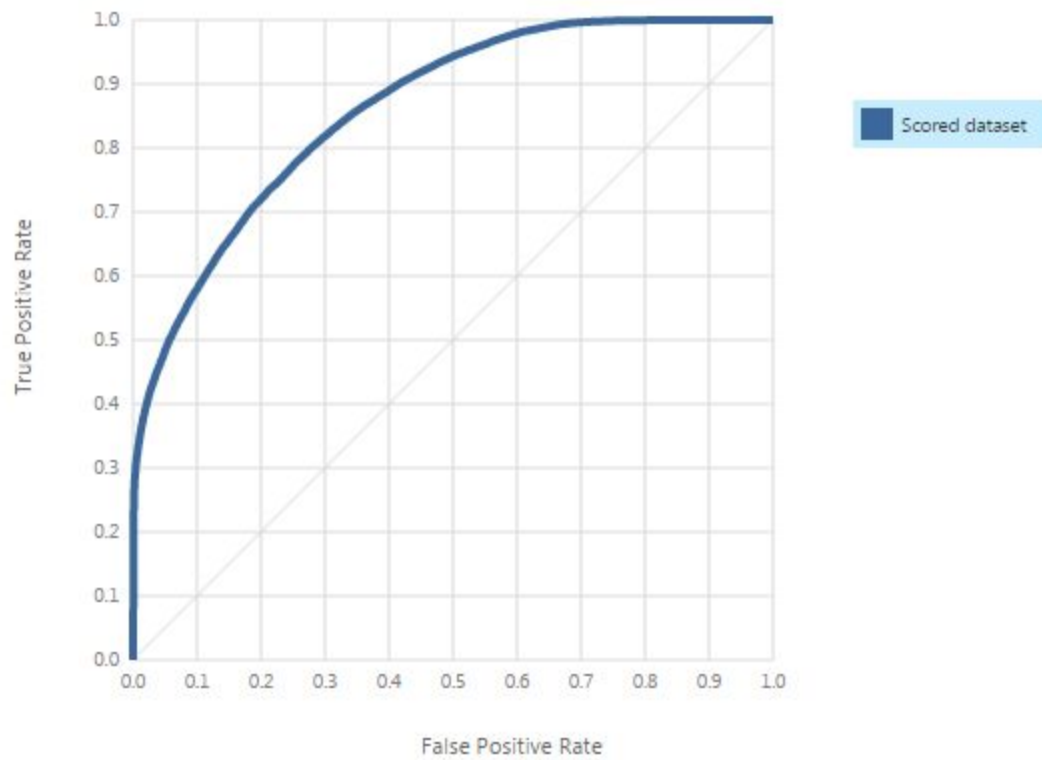
Used the classification techniques - random forest, logistic regression and neural network to train models and score and evaluate data. The comparison for each evaluation is visualised and the ROC curve is displayed. Based on all the evaluations depending on the AUC value Random Forest gave the best results for AUC. Deployed each model as a web service to take input from the user and predict the value for the normalised power consumption.

## Decision Forest

### Decision Forest Classification

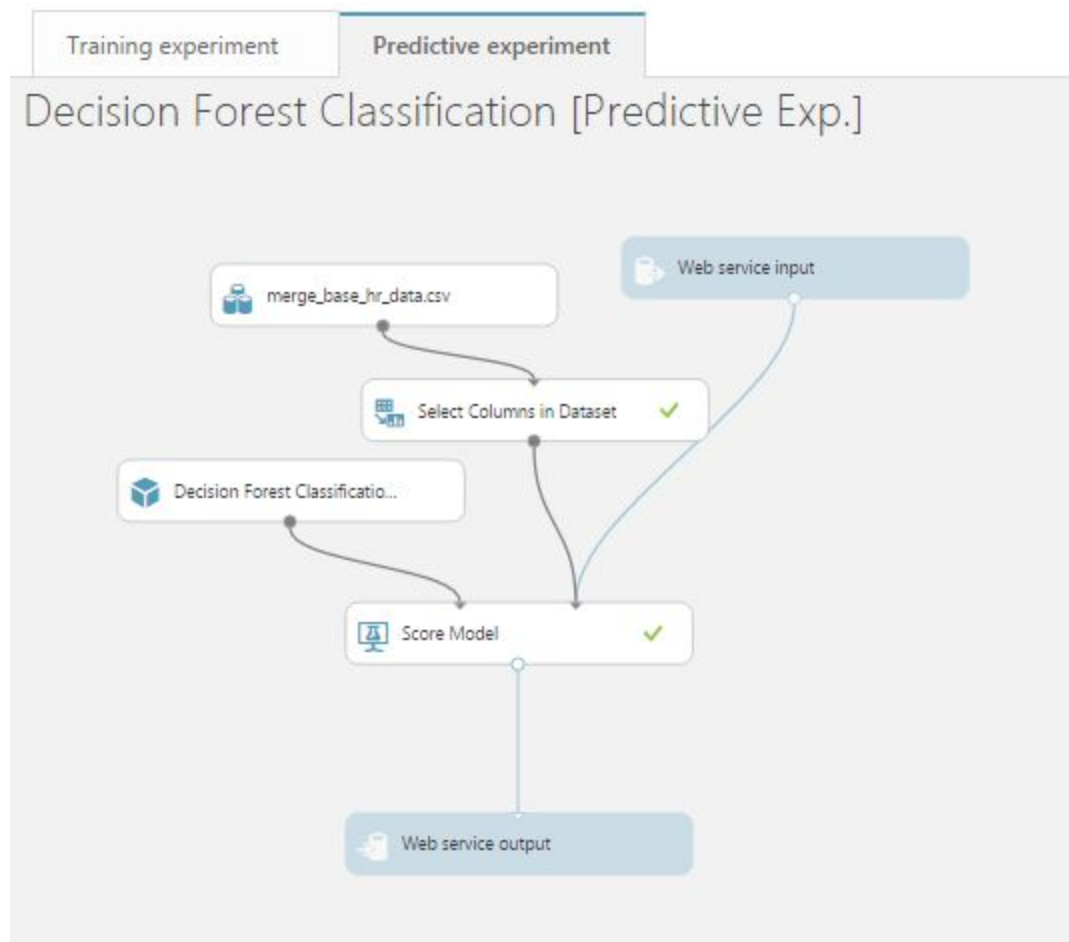


ROC PRECISION/RECALL LIFT



True Positive	False Negative	Accuracy	Precision	Threshold	<input type="text" value="0.5"/>	AUC
<b>62485</b>	<b>10689</b>	<b>0.773</b>	<b>0.780</b>	<b>0.5</b>		<b>0.861</b>
False Positive	True Negative	Recall	F1 Score			
<b>17587</b>	<b>33602</b>	<b>0.854</b>	<b>0.815</b>			
Positive Label	Negative Label					
<b>Low</b>	<b>High</b>					

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	25316	525	0.208	0.611	0.511	0.980	0.346	0.514	0.990	0.003
(0.800,0.900]	8268	1575	0.287	0.665	0.617	0.941	0.459	0.554	0.959	0.015
(0.700,0.800]	10480	3721	0.401	0.719	0.716	0.883	0.602	0.609	0.886	0.055
(0.600,0.700]	11534	6411	0.545	0.760	0.789	0.820	0.760	0.689	0.761	0.141
(0.500,0.600]	6888	5360	0.644	0.773	0.815	0.780	0.854	0.759	0.656	0.226
(0.400,0.500]	5816	6920	0.746	0.764	0.823	0.736	0.933	0.846	0.521	0.347
(0.300,0.400]	2787	4987	0.809	0.746	0.818	0.707	0.972	0.912	0.424	0.440
(0.200,0.300]	1393	3974	0.852	0.725	0.809	0.684	0.991	0.962	0.346	0.516
(0.100,0.200]	467	2874	0.879	0.706	0.800	0.667	0.997	0.985	0.290	0.572
(0.000,0.100]	225	14842	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.861



### decision forest classification [predictive exp.]

DASHBOARD    CONFIGURATION

General    [New Web Services Experience](#) [preview](#)

Published experiment

[View snapshot](#)    [View latest](#)

Description

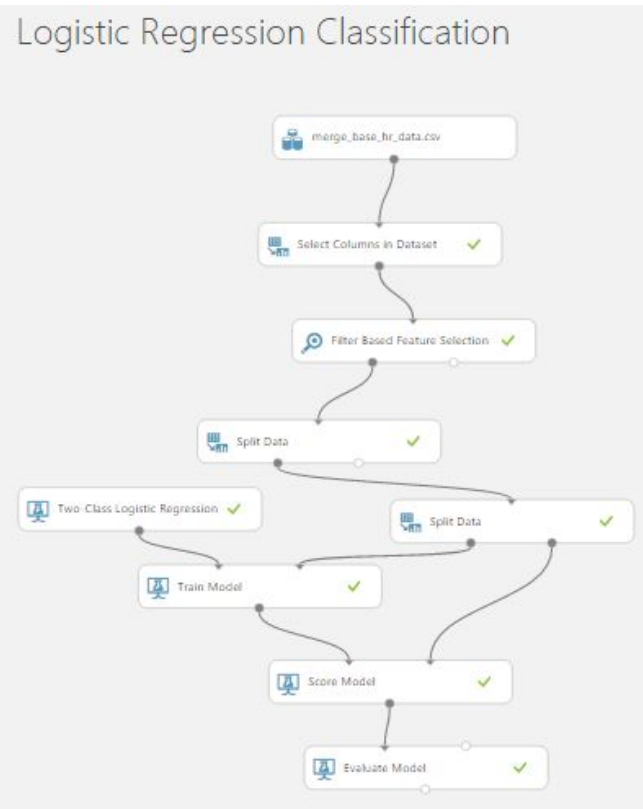
No description provided for this web service.

API key

`tUhiZK8w@r0FYXjy+hqbu6W1pH1HqNxP2n048y4ZWyMI+GWL0AOJC5bIGfxBX+qrr5zmhFj0I0godMHhkzLow==`

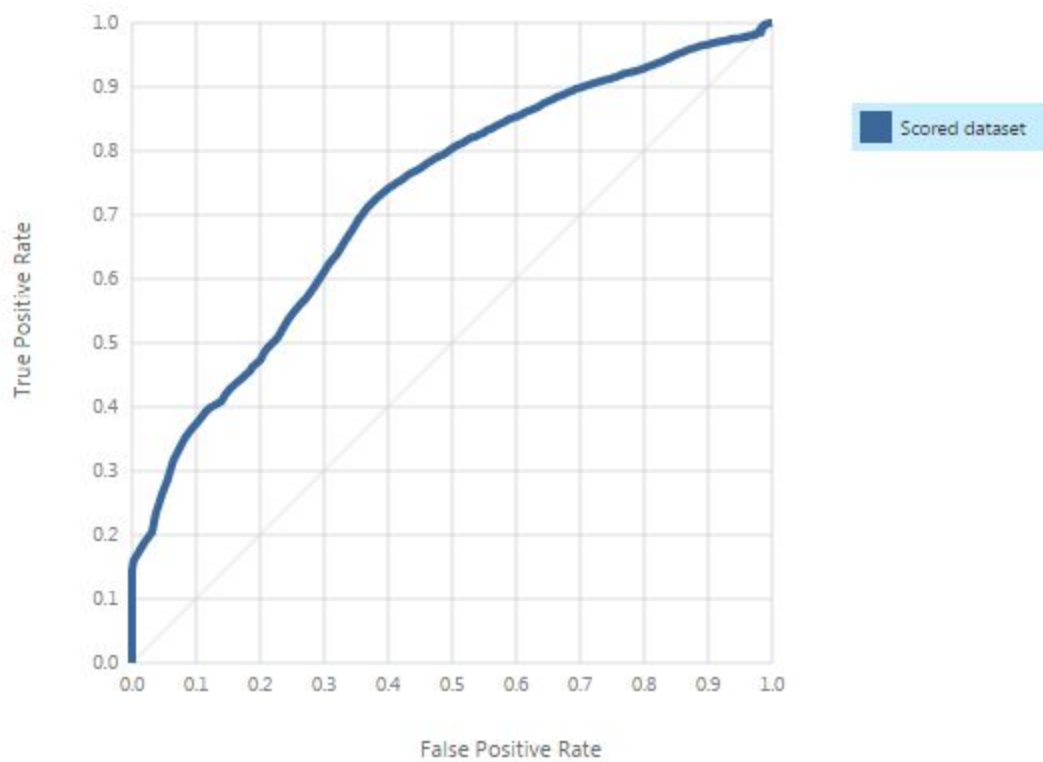
Default Endpoint

API HELP PAGE	TEST	APPS	LAST UPDATED
REQUEST/RESPONSE	<a href="#">Test</a> <a href="#">Test preview</a>	Excel 2013 or later     Excel 2010 or earlier workbook	12/2/2016 7:48:11 PM
BATCH EXECUTION	<a href="#">Test preview</a>	Excel 2013 or later workbook	12/2/2016 7:48:11 PM



Linear Regression

ROC PRECISION/RECALL LIFT



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
69888	3286	0.619	0.613	0.5	0.724
False Positive	True Negative	Recall	F1 Score		
44155	7034	0.955	0.747		
Positive Label	Negative Label				
Low	High				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.800,0.900]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.700,0.800]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.600,0.700]	54874	21154	0.611	0.683	0.736	0.722	0.750	0.621	0.587	0.201
(0.500,0.600]	15014	23001	0.917	0.619	0.747	0.613	0.955	0.682	0.137	0.590
(0.400,0.500]	1550	4438	0.965	0.595	0.740	0.595	0.976	0.599	0.051	0.674
(0.300,0.400]	320	986	0.976	0.590	0.738	0.591	0.981	0.532	0.031	0.693
(0.200,0.300]	247	438	0.981	0.588	0.738	0.590	0.984	0.501	0.023	0.701
(0.100,0.200]	240	267	0.985	0.588	0.738	0.590	0.987	0.493	0.018	0.706
(0.000,0.100]	929	905	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.724

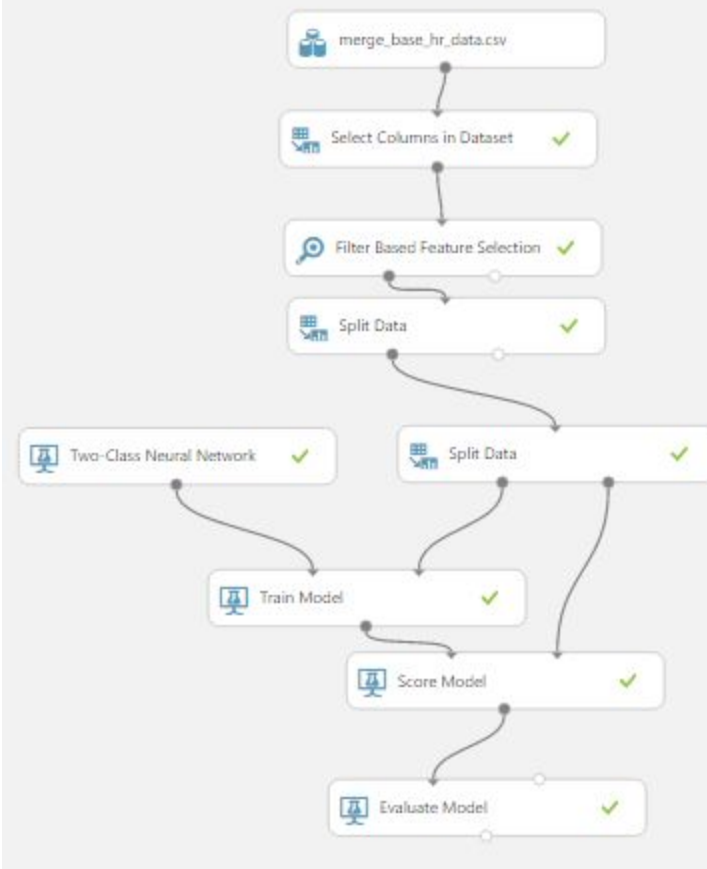


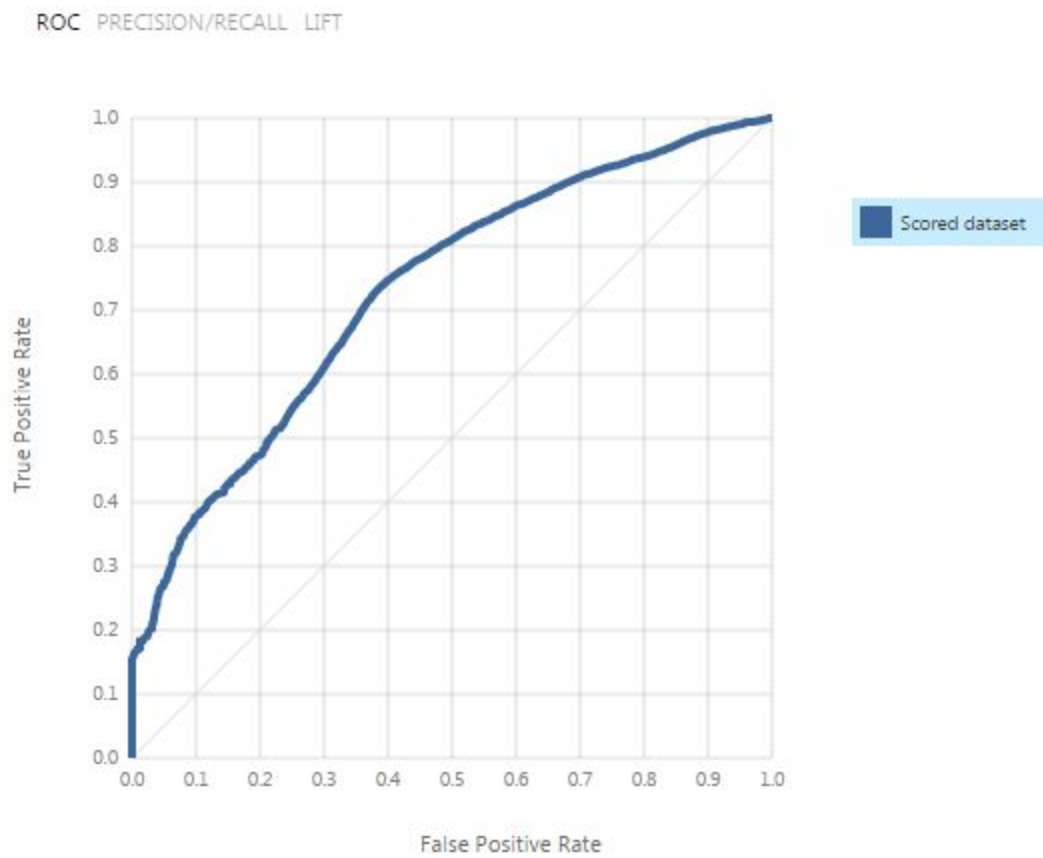
## Neural Network

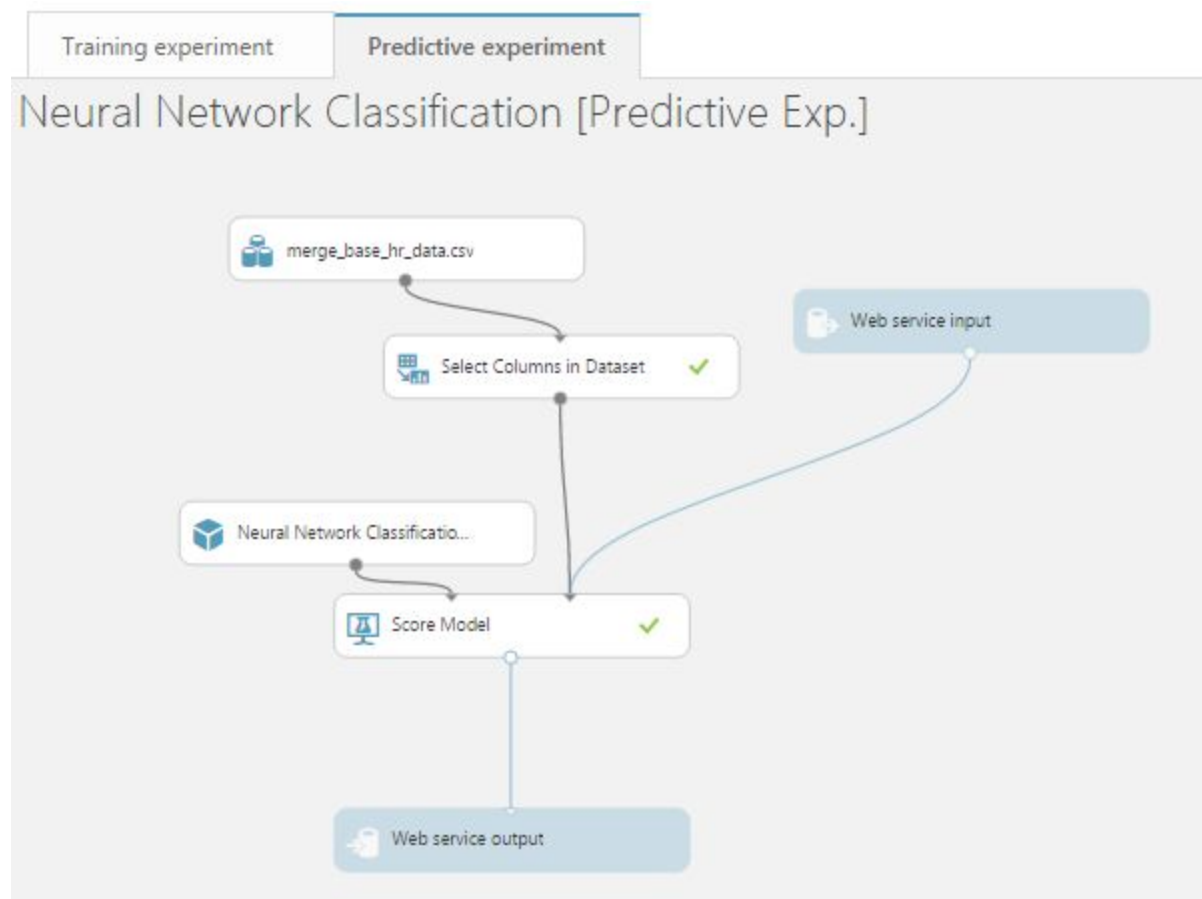
True Positive	False Negative	Accuracy	Precision	Threshold		AUC
<b>65992</b>	<b>7182</b>	<b>0.660</b>	<b>0.653</b>	<b>0.5</b>		<b>0.729</b>
False Positive	True Negative	Recall	F1 Score			
<b>35112</b>	<b>16077</b>	<b>0.902</b>	<b>0.757</b>			
Positive Label	Negative Label					
<b>Low</b>	<b>High</b>					

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.800,0.900]	25866	4269	0.242	0.585	0.501	0.858	0.353	0.498	0.917	0.021
(0.700,0.800]	12804	8017	0.410	0.624	0.623	0.759	0.528	0.530	0.760	0.090
(0.600,0.700]	16996	9175	0.620	0.687	0.741	0.722	0.761	0.629	0.581	0.207
(0.500,0.600]	10326	13651	0.813	0.660	0.757	0.653	0.902	0.691	0.314	0.429
(0.400,0.500]	4924	9766	0.931	0.621	0.751	0.612	0.969	0.736	0.123	0.608
(0.300,0.400]	2180	6005	0.997	0.590	0.742	0.590	0.999	0.797	0.006	0.723
(0.200,0.300]	71	99	0.998	0.590	0.742	0.589	1.000	0.967	0.004	0.725
(0.100,0.200]	4	141	0.999	0.589	0.741	0.589	1.000	0.957	0.001	0.728
(0.000,0.100]	3	66	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.729

## Neural Network Classification

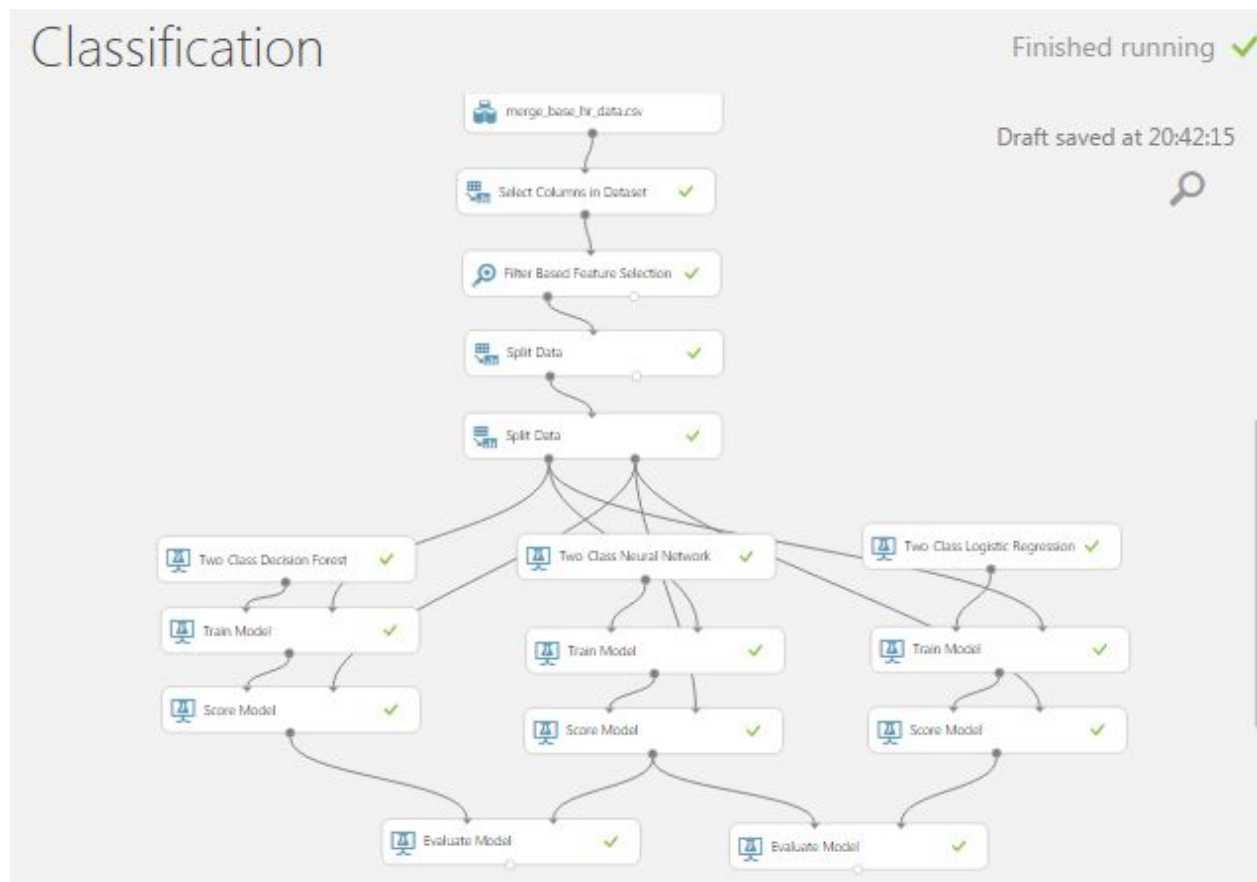




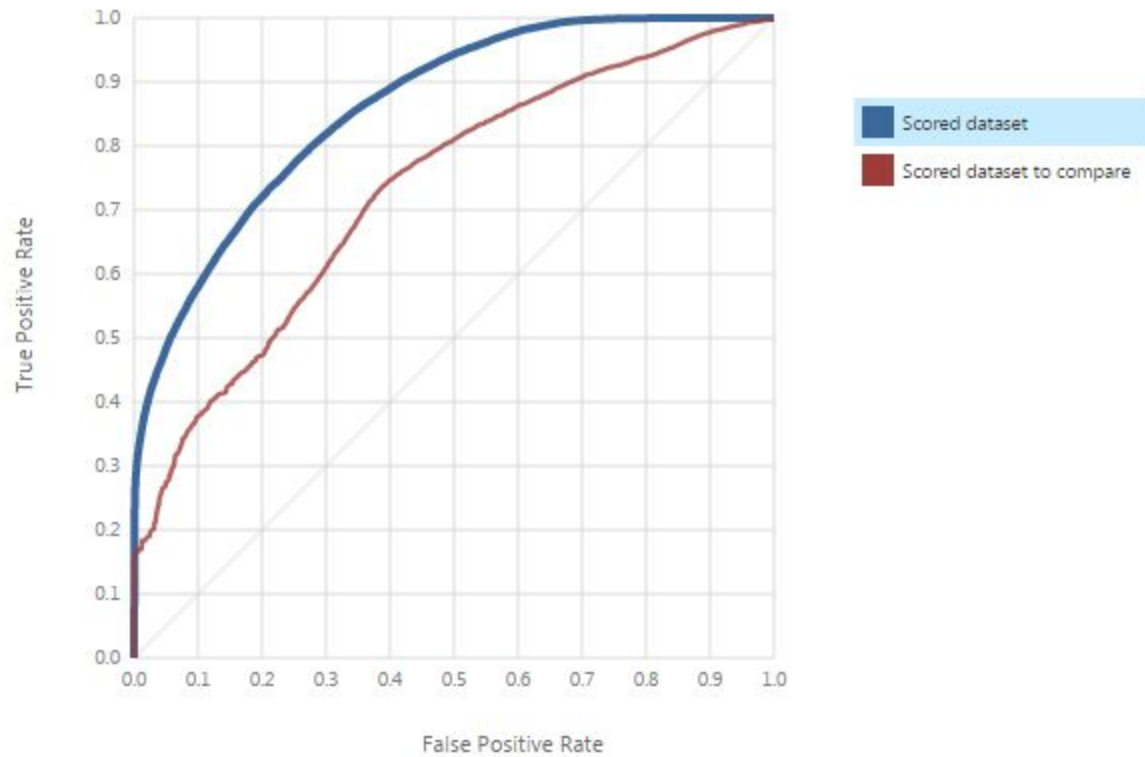


Comparison

Here is the combination of decision forest, neural network and linear regression classification model with their results for the same dataset.



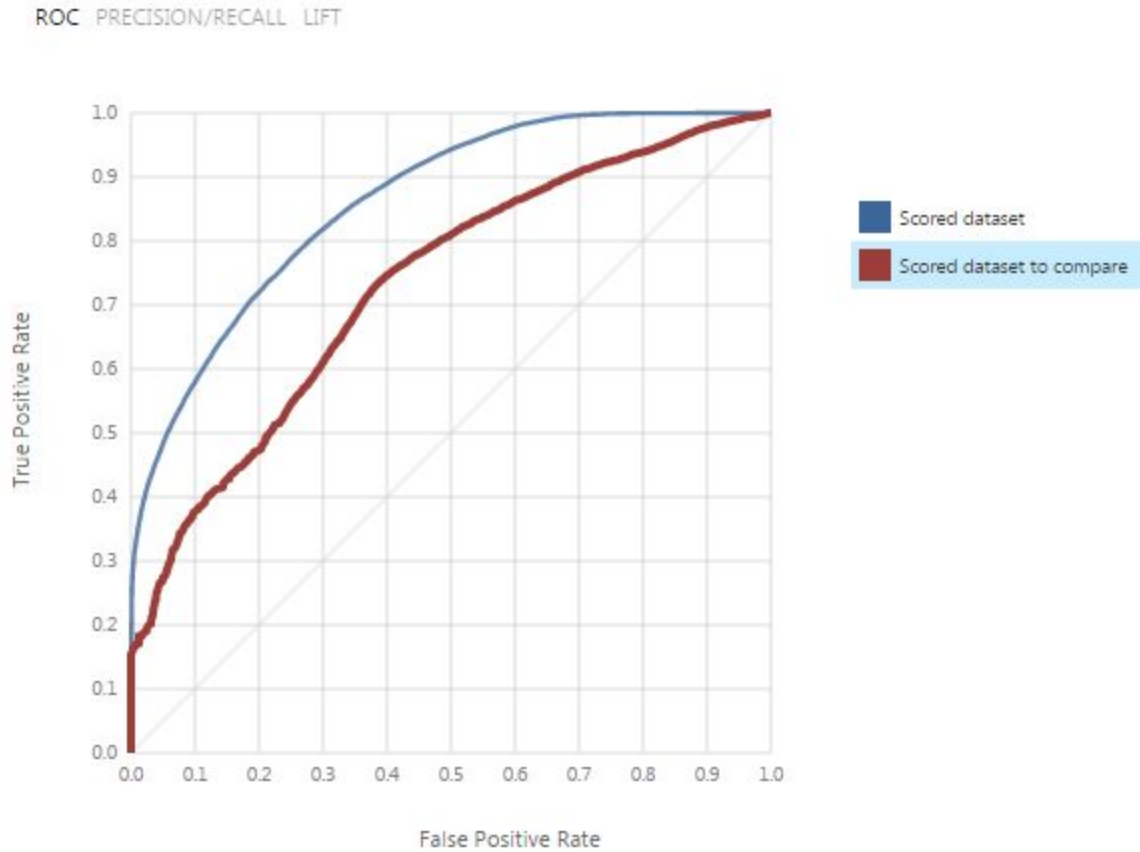
Following is the comparison between model developed by decision forest and neural network. This Following shows AUC for decision forest for AUC is 0.86 and in the graph it is shown in blue.



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
62485	10689	0.773	0.780	0.5	0.861
False Positive	True Negative	Recall	F1 Score		
17587	33602	0.854	0.815		
Positive Label	Negative Label				
Low	High				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	25316	525	0.208	0.611	0.511	0.980	0.346	0.514	0.990	0.003
(0.800,0.900]	8268	1575	0.287	0.665	0.617	0.941	0.459	0.554	0.959	0.015
(0.700,0.800]	10480	3721	0.401	0.719	0.716	0.888	0.602	0.609	0.886	0.055
(0.600,0.700]	11534	6411	0.545	0.760	0.789	0.820	0.760	0.689	0.761	0.141
(0.500,0.600]	6888	5360	0.644	0.773	0.815	0.780	0.854	0.759	0.656	0.226
(0.400,0.500]	5816	6920	0.746	0.764	0.823	0.736	0.933	0.846	0.521	0.347
(0.300,0.400]	2787	4987	0.809	0.746	0.818	0.707	0.972	0.912	0.424	0.440
(0.200,0.300]	1393	3974	0.852	0.725	0.809	0.684	0.991	0.962	0.346	0.516
(0.100,0.200]	467	2874	0.879	0.706	0.800	0.667	0.997	0.985	0.290	0.572
(0.000,0.100]	225	14842	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.861

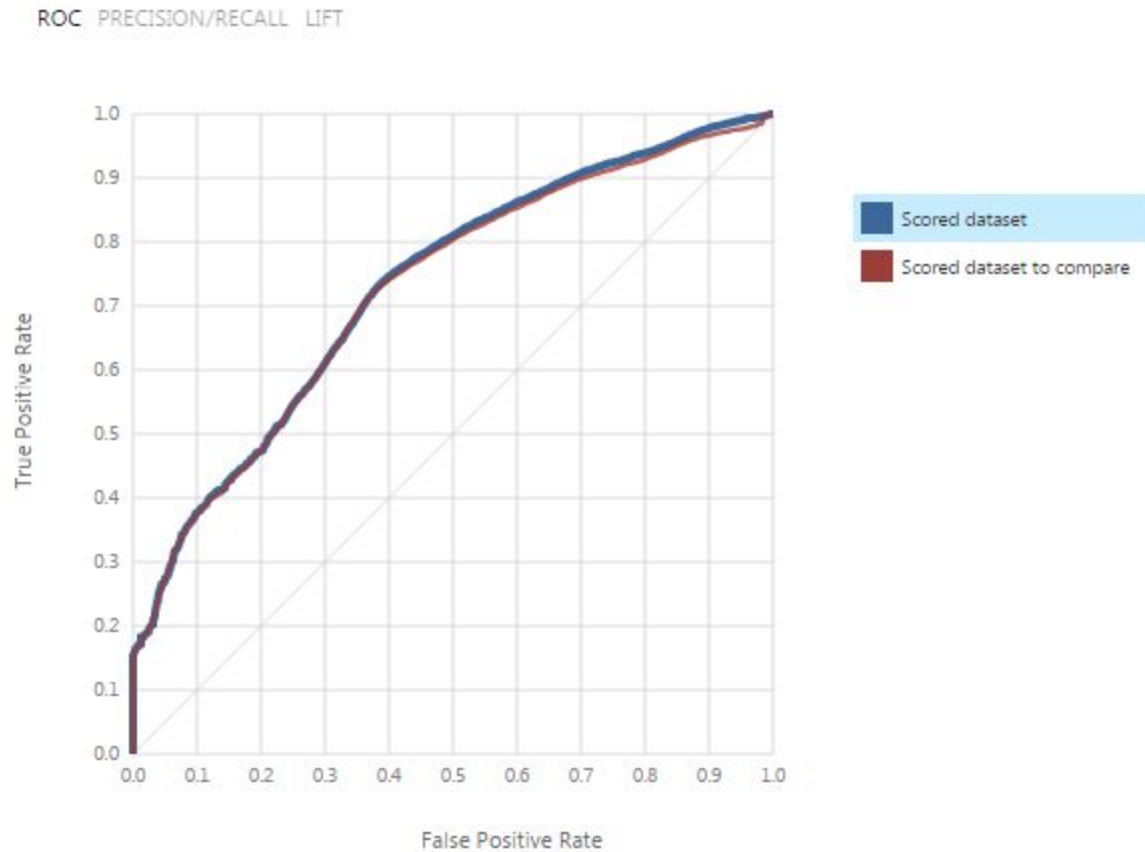
Following shows AUC for Nneural network in red



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
65992	7182	0.660	0.653	0.5	0.729
False Positive	True Negative	Recall	F1 Score		
35112	16077	0.902	0.757		
Positive Label	Negative Label				
Low	High				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.800,0.900]	25866	4269	0.242	0.585	0.501	0.858	0.353	0.498	0.917	0.021
(0.700,0.800]	12804	8017	0.410	0.624	0.623	0.759	0.528	0.530	0.760	0.090
(0.600,0.700]	16996	9175	0.620	0.687	0.741	0.722	0.761	0.629	0.581	0.207
(0.500,0.600]	10326	13651	0.813	0.660	0.757	0.653	0.902	0.691	0.314	0.429
(0.400,0.500]	4924	9766	0.981	0.621	0.751	0.612	0.969	0.736	0.123	0.608
(0.300,0.400]	2180	6005	0.997	0.590	0.742	0.590	0.999	0.797	0.006	0.723
(0.200,0.300]	71	99	0.998	0.590	0.742	0.589	1.000	0.967	0.004	0.725
(0.100,0.200]	4	141	0.999	0.589	0.741	0.589	1.000	0.957	0.001	0.728
(0.000,0.100]	3	66	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.729

Following is the comparison between neural network and linear regression. Blue is highlighted for neural network and other demonstrates linear regression

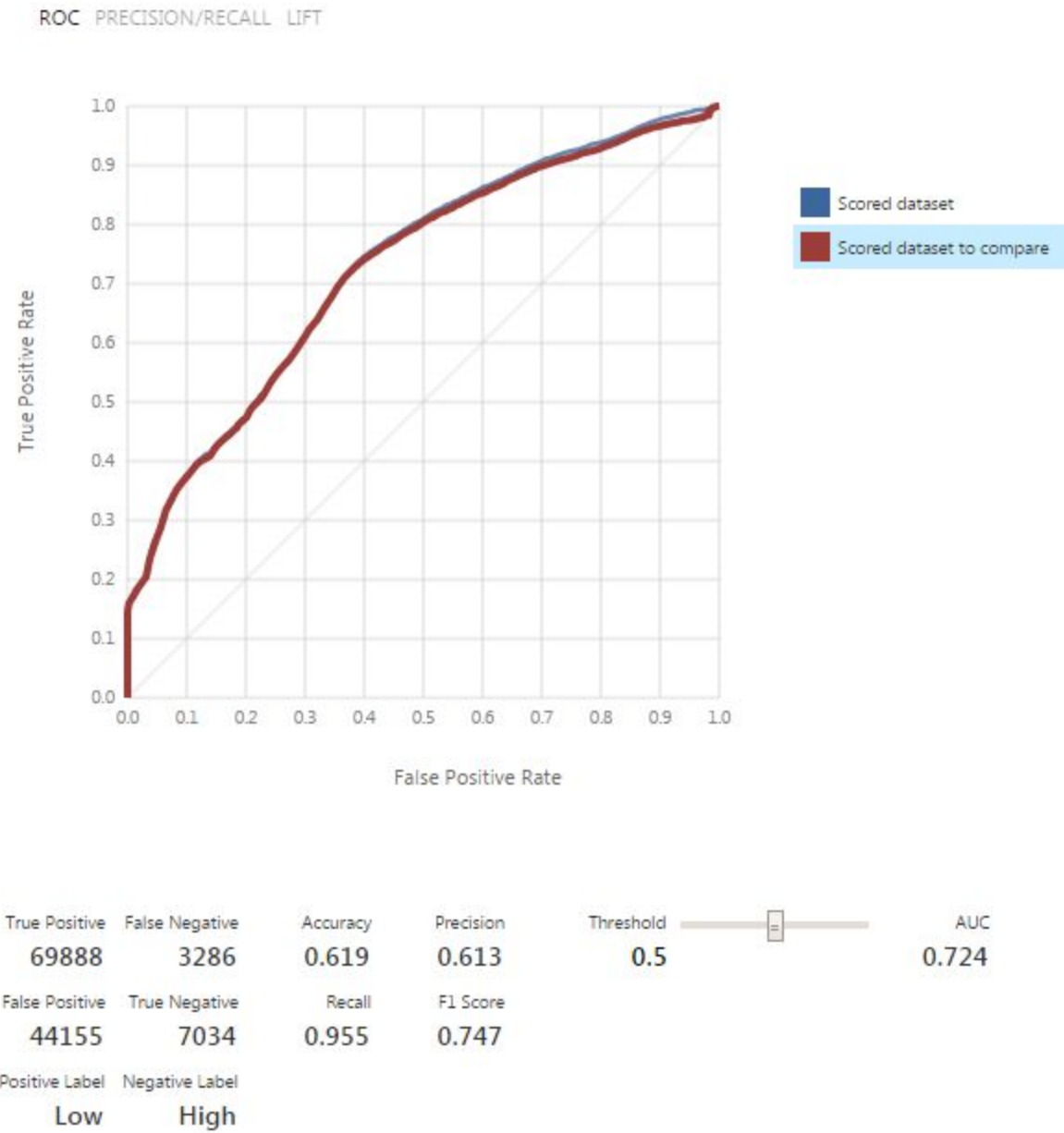


True Positive	False Negative	Accuracy	Precision	Threshold	AUC
65992	7182	0.660	0.653	0.5	0.729
False Positive	True Negative	Recall	F1 Score		
35112	16077	0.902	0.757		
Positive Label	Negative Label				
Low	High				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.800,0.900]	25866	4269	0.242	0.585	0.501	0.858	0.353	0.498	0.917	0.021
(0.700,0.800]	12804	8017	0.410	0.624	0.623	0.759	0.528	0.530	0.760	0.090
(0.600,0.700]	16996	9175	0.620	0.687	0.741	0.722	0.761	0.629	0.581	0.207
(0.500,0.600]	10316	13651	0.813	0.660	0.757	0.653	0.902	0.691	0.314	0.429
(0.400,0.500]	4924	9766	0.931	0.621	0.751	0.612	0.969	0.736	0.123	0.608
(0.300,0.400]	2180	6005	0.997	0.590	0.742	0.590	0.999	0.797	0.006	0.723
(0.200,0.300]	71	99	0.998	0.590	0.742	0.589	1.000	0.967	0.004	0.725
(0.100,0.200]	4	141	0.999	0.589	0.741	0.589	1.000	0.957	0.001	0.728
(0.000,0.100]	3	66	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.729



Here linear regression shown below:



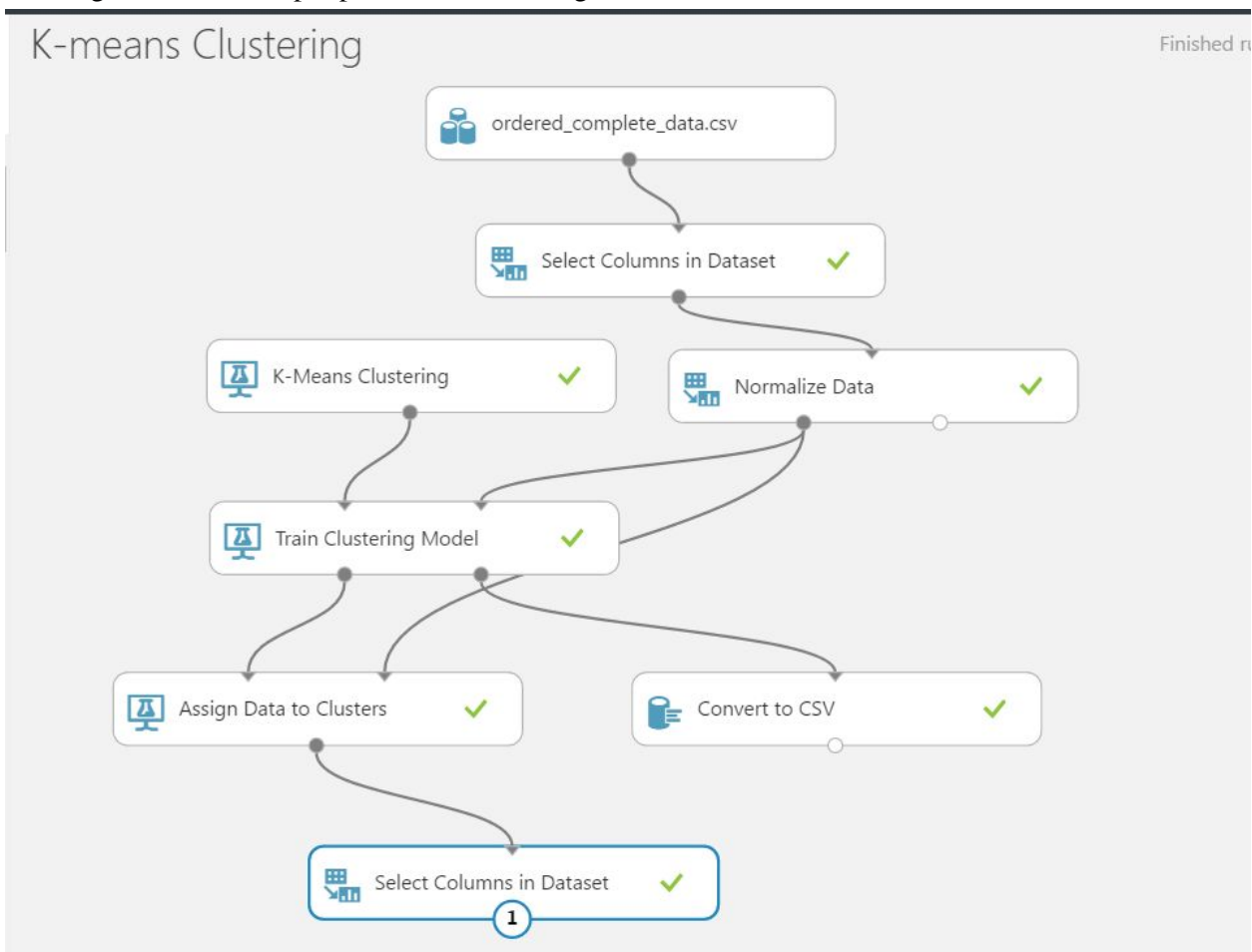
Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.800,0.900]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.700,0.800]	0	0	0.000	0.412	0.000	1.000	0.000	0.412	1.000	0.000
(0.600,0.700]	54874	21154	0.611	0.683	0.736	0.722	0.750	0.621	0.587	0.201
(0.500,0.600]	15014	23001	0.917	0.619	0.747	0.613	0.955	0.682	0.137	0.590
(0.400,0.500]	1550	4438	0.965	0.595	0.740	0.595	0.976	0.599	0.051	0.674
(0.300,0.400]	320	986	0.976	0.590	0.738	0.591	0.981	0.532	0.031	0.693
(0.200,0.300]	247	438	0.981	0.588	0.738	0.590	0.984	0.501	0.023	0.701
(0.100,0.200]	240	267	0.985	0.588	0.738	0.590	0.987	0.493	0.018	0.706
(0.000,0.100]	929	905	1.000	0.588	0.741	0.588	1.000	1.000	0.000	0.724

**Best Model - Random Forest . AUC value is high amongst all 3**

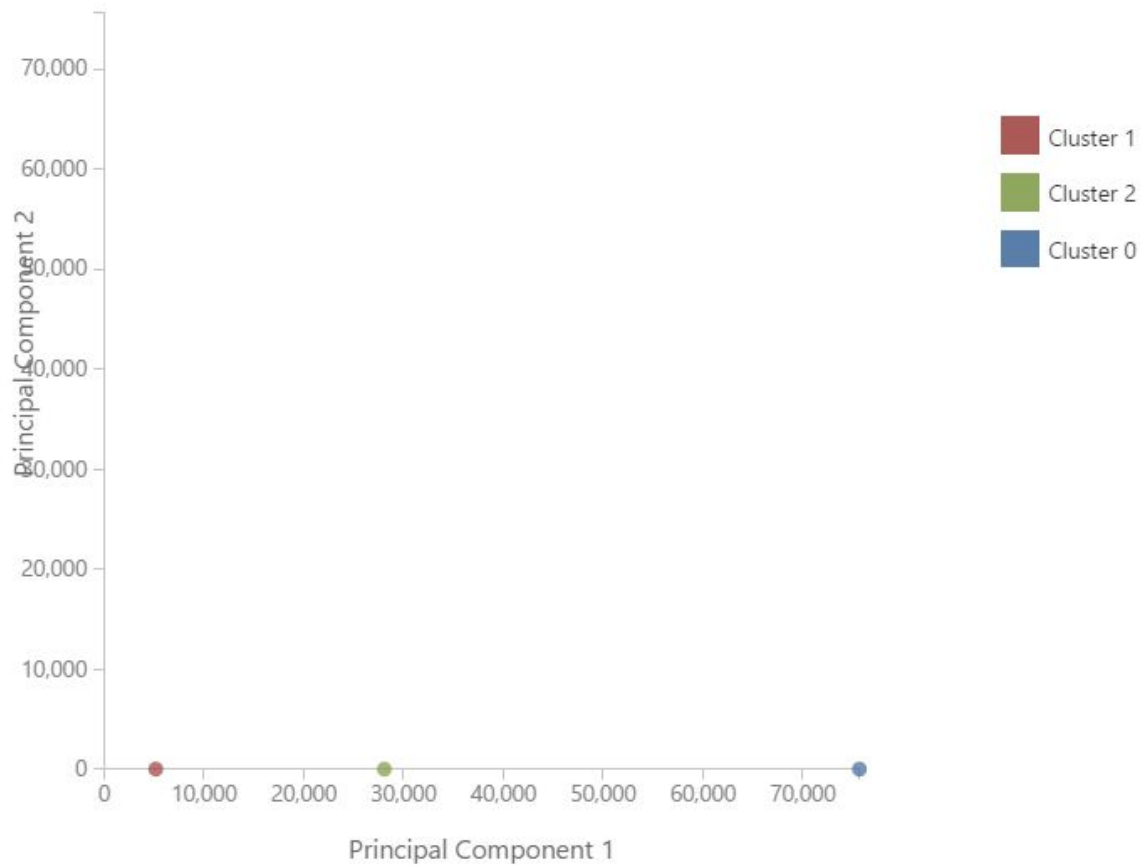
Clustering :

K-means Clustering

Used to determine the outliers. Used to identify data points located away from the centroid. Clusters Building based on the input parameters and assigns it to a cluster. No of clusters used is 3



K-means Clustering ▶ Train Clustering Model ▶ Results dataset

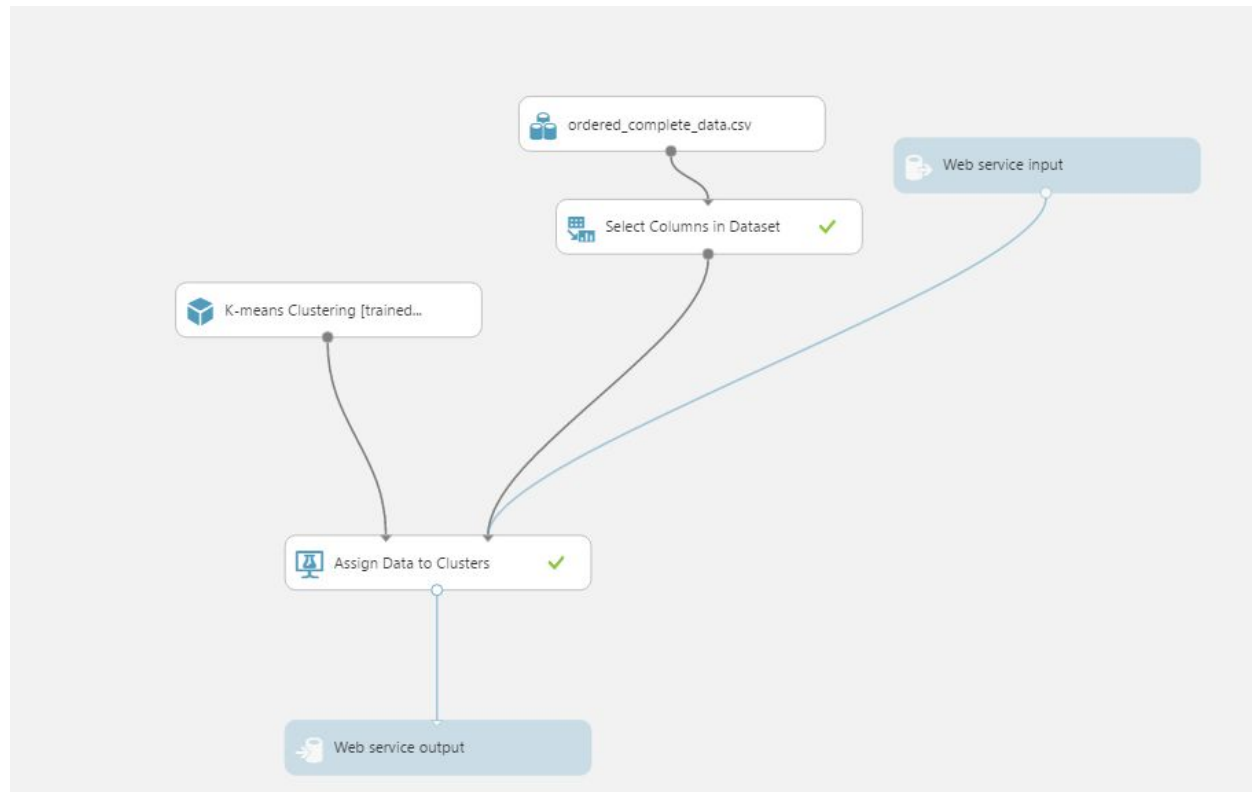


K-means Clustering &gt; Select Columns in Dataset &gt; Results dataset

rows  
621817

columns  
2

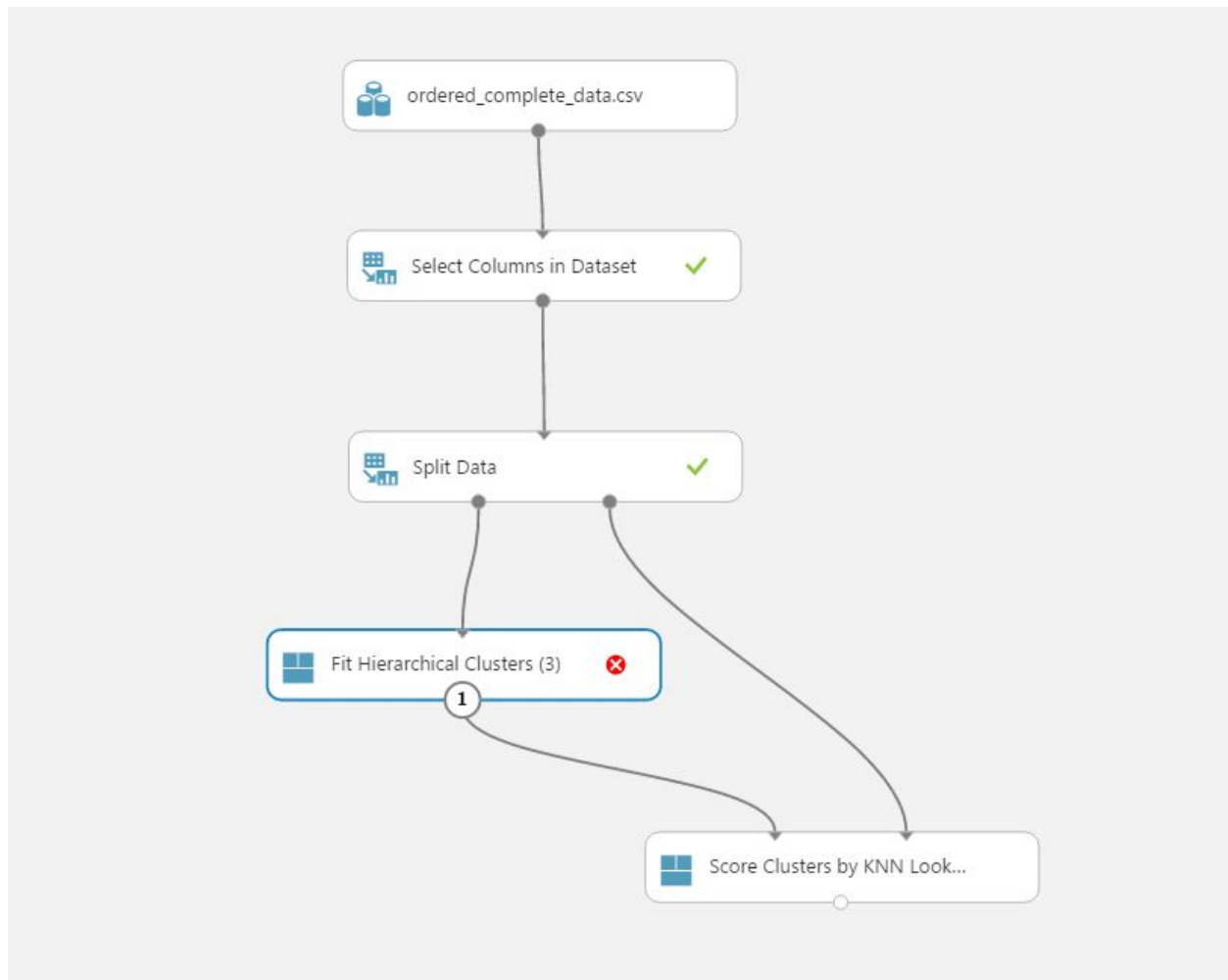
view as	BuildingID	Assignments
		
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1
	5198	1

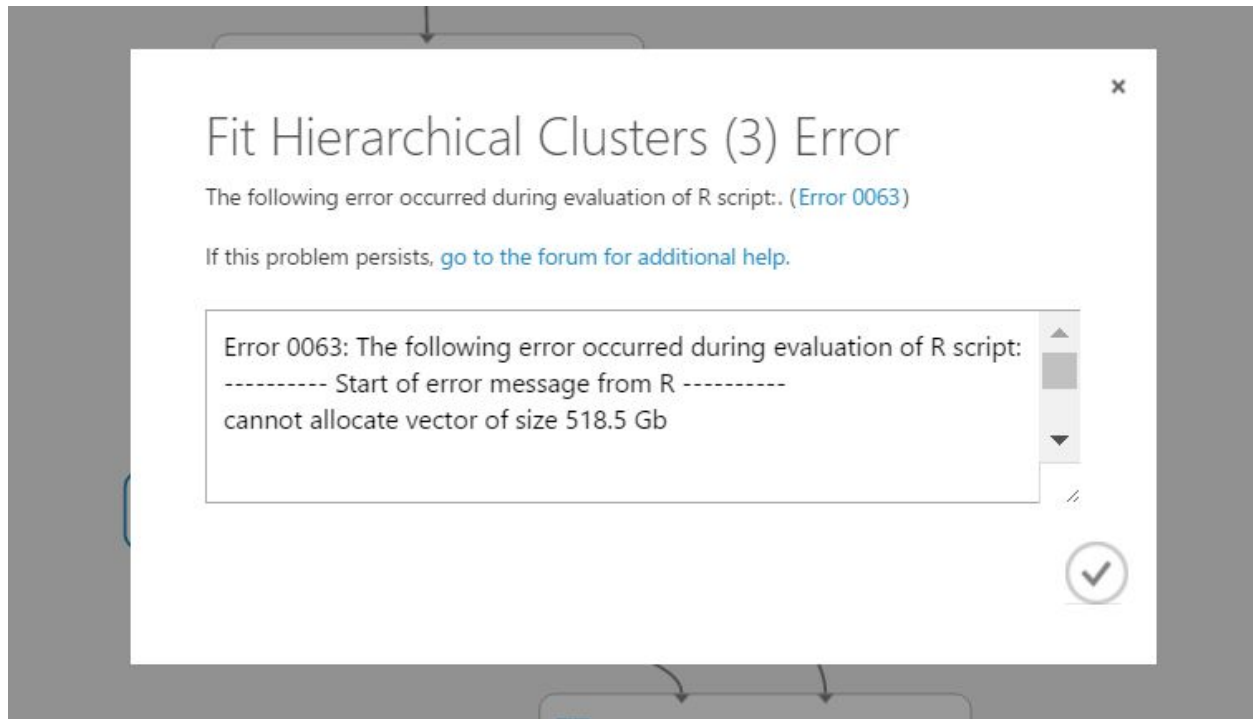


Quickstart	Dashboard	Batch Request Log	Configure	Consume	Test	Swagger API
BuildingID	<input type="text" value="83427"/>			BuildingID	83427	
meternumb	<input type="text" value="1"/>			meternumb	1	
norm_consumption	<input type="text" value="0.25"/>			norm_consumption	0.25	
TemperatureF	<input type="text" value="38"/>			TemperatureF	38	
Dew_PointF	<input type="text" value="33"/>			Dew_PointF	33	
Humidity	<input type="text" value="90"/>			Humidity	90	
Sea_Level_PressureIn	<input type="text" value="30"/>			Sea_Level_PressureIn	30	
VisibilityMPH	<input type="text" value="5.6"/>			VisibilityMPH	5.6	
				Assignments	0	
				DistancesToClusterCenter no.0	7045.88292666954	

[Test Request-Response](#)

Hierarchical Clustering :





localhost:8080/AdsAzureWebServices/

Apps gada\_kin\_001643864 My.EquityApartments JobApplications - Go Welcome! | LinkedIn Welcome, Kinjal - Bla myNEU - Northeastern Northeastern Univers chu

# Azure Web Services Demo

Please complete the following form:

**Date**

20140506

**Select Hour**

17

**Algorithm type**

☐ KNN

☒ Random Forest

☐ Neural Network