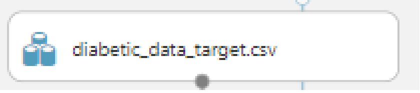
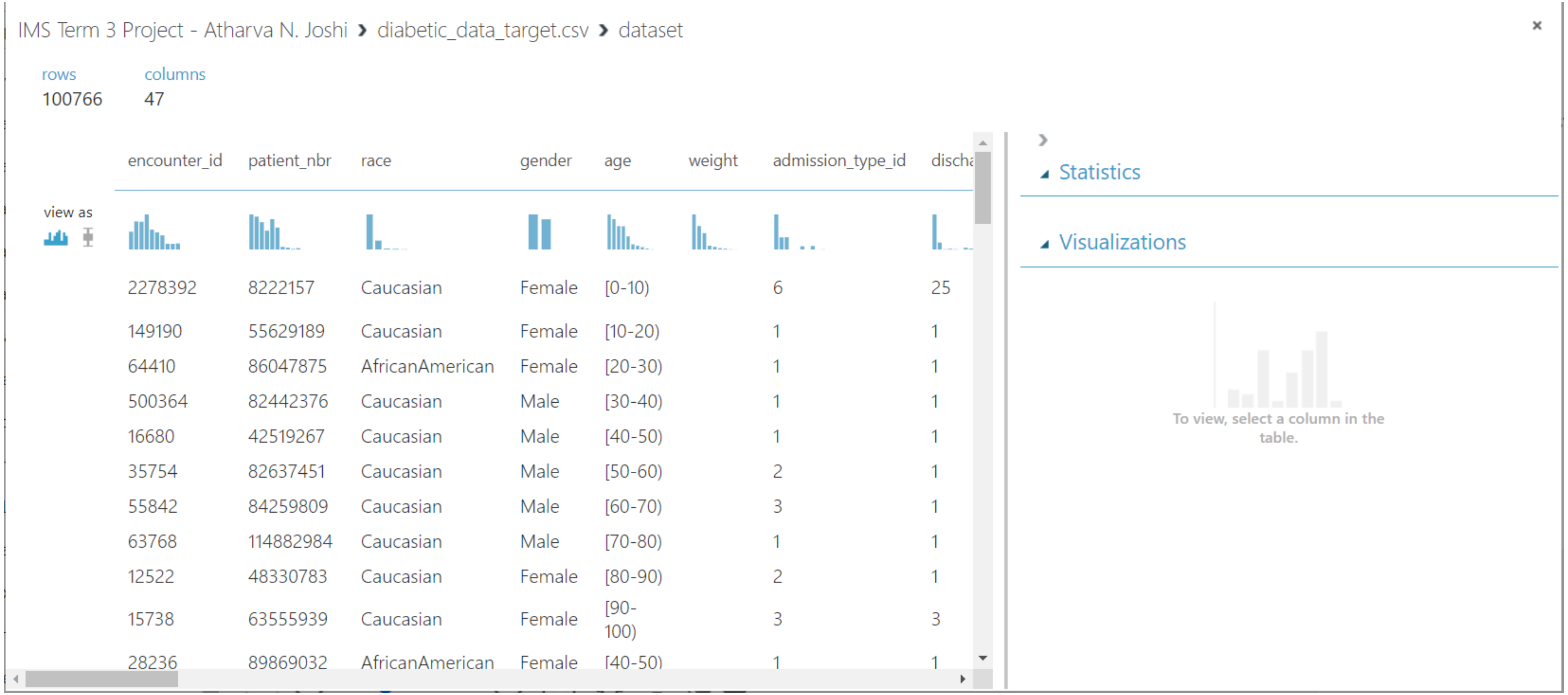
IMS Term 3 Project

- Atharva N. Joshi

First we load the dataset and try to do EDA. Azure gives us all necessary information without any extra commands



diabetic_data_target.csv



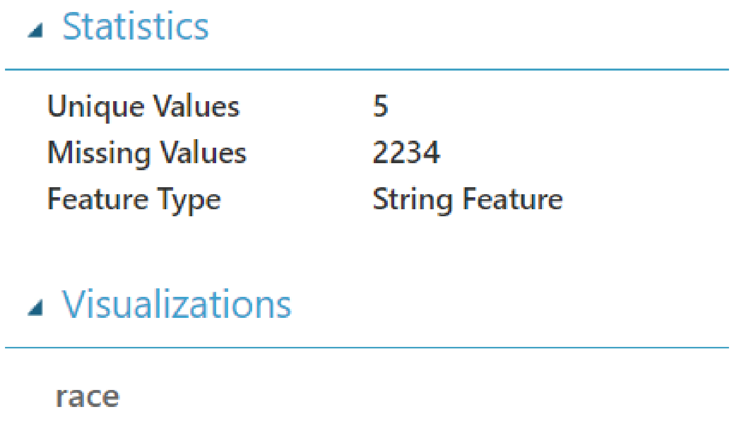
IMS Term 3 Project - Atharva N. Joshi › diabetic_data_target.csv › dataset

rows 100766

columns 47

encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discha
2278392	8222157	Caucasian	Female	[0-10)		6	25
149190	55629189	Caucasian	Female	[10-20)		1	1
64410	86047875	AfricanAmerican	Female	[20-30)		1	1
500364	82442376	Caucasian	Male	[30-40)		1	1
16680	42519267	Caucasian	Male	[40-50)		1	1
35754	82637451	Caucasian	Male	[50-60)		2	1
55842	84259809	Caucasian	Male	[60-70)		3	1
63768	114882984	Caucasian	Male	[70-80)		1	1
12522	48330783	Caucasian	Female	[80-90)		2	1
15738	63555939	Caucasian	Female	[90-100)		3	3
28236	89869032	AfricanAmerican	Female	[40-50)		1	1

We see that some columns have many null values. We remove those columns except race which has comparatively less null values and hence, will be imputed

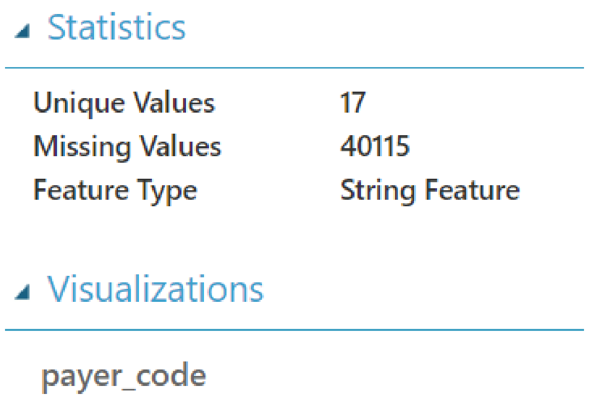


Statistics

Unique Values	5
Missing Values	2234
Feature Type	String Feature

Visualizations

race

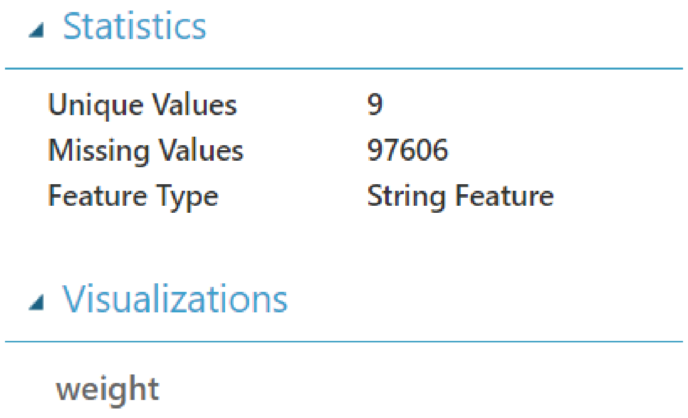


Statistics

Unique Values	17
Missing Values	40115
Feature Type	String Feature

Visualizations

payer_code

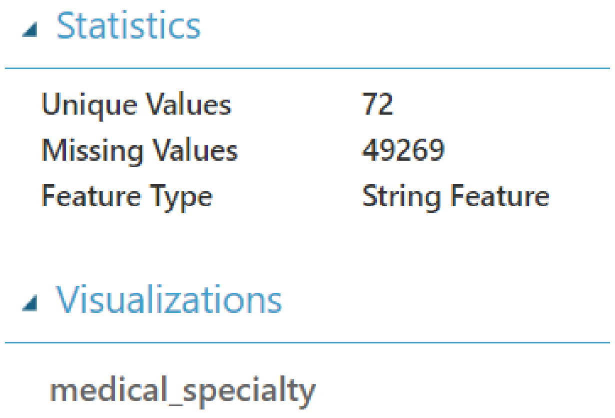


Statistics

Unique Values	9
Missing Values	97606
Feature Type	String Feature

Visualizations

weight

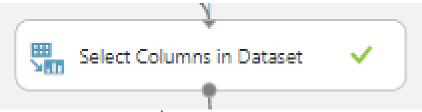


Statistics

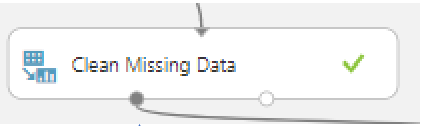
Unique Values	72
Missing Values	49269
Feature Type	String Feature

Visualizations

medical_specialty

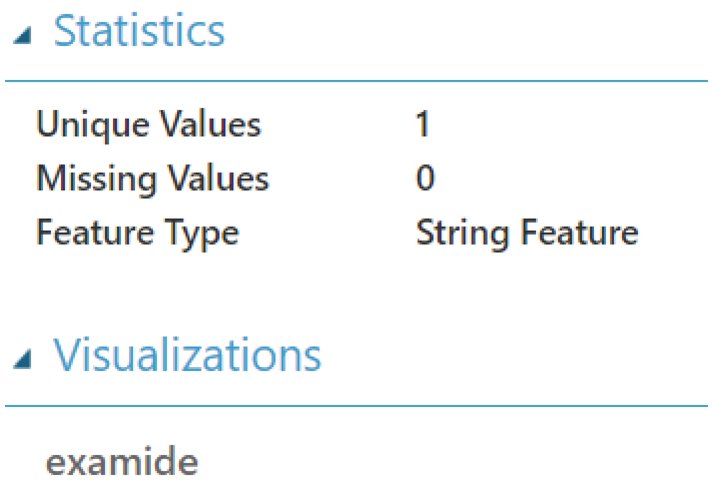


To include/exclude certain columns



To impute null values

We can also see that the below two columns have only one value. So we can exclude those columns as well.

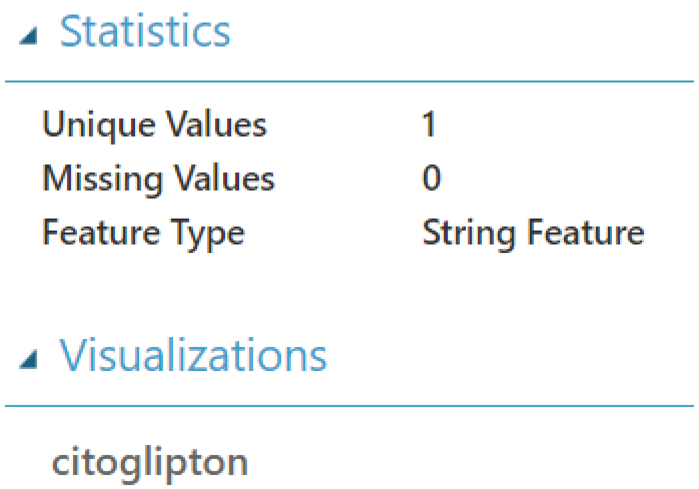


◢ Statistics

Unique Values	1
Missing Values	0
Feature Type	String Feature

◢ Visualizations

examide



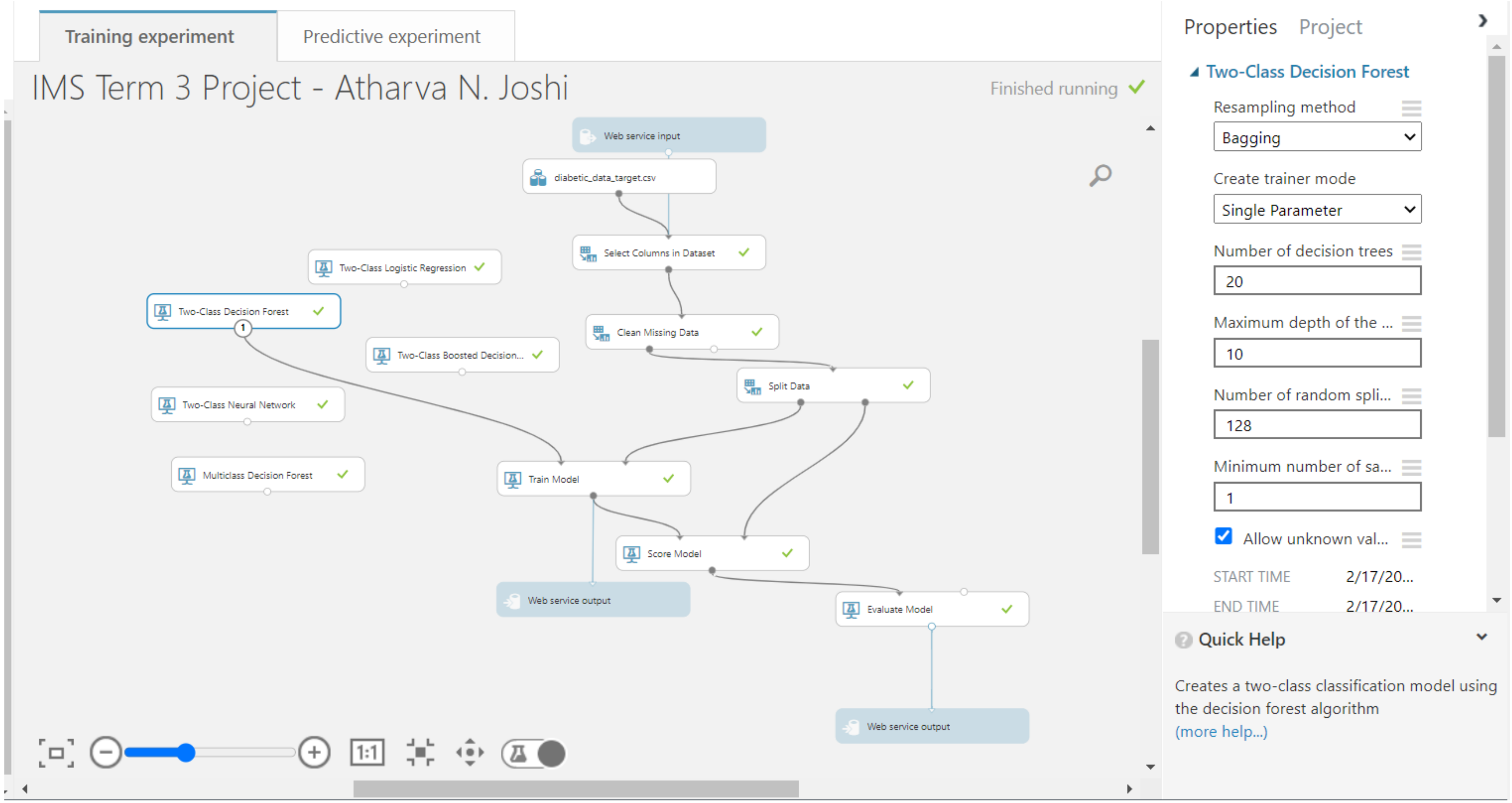
◢ Statistics

Unique Values	1
Missing Values	0
Feature Type	String Feature

◢ Visualizations

citoglipton

Two Class Decision Forest



IMS Term 3 Project - Atharva N. Joshi

Finished running ✓

Two-Class Decision Forest

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision trees: 20

Maximum depth of the ...: 10

Number of random spli...: 128

Minimum number of sa...: 1

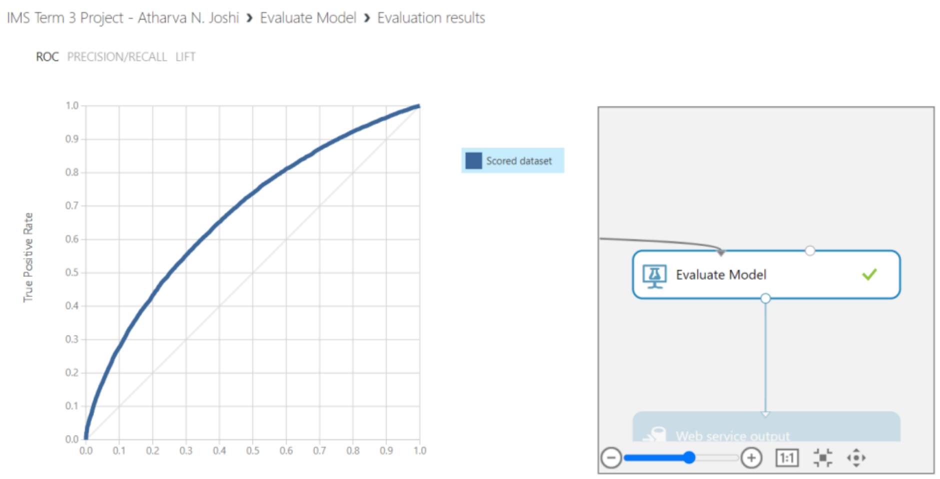
☑ Allow unknown val...

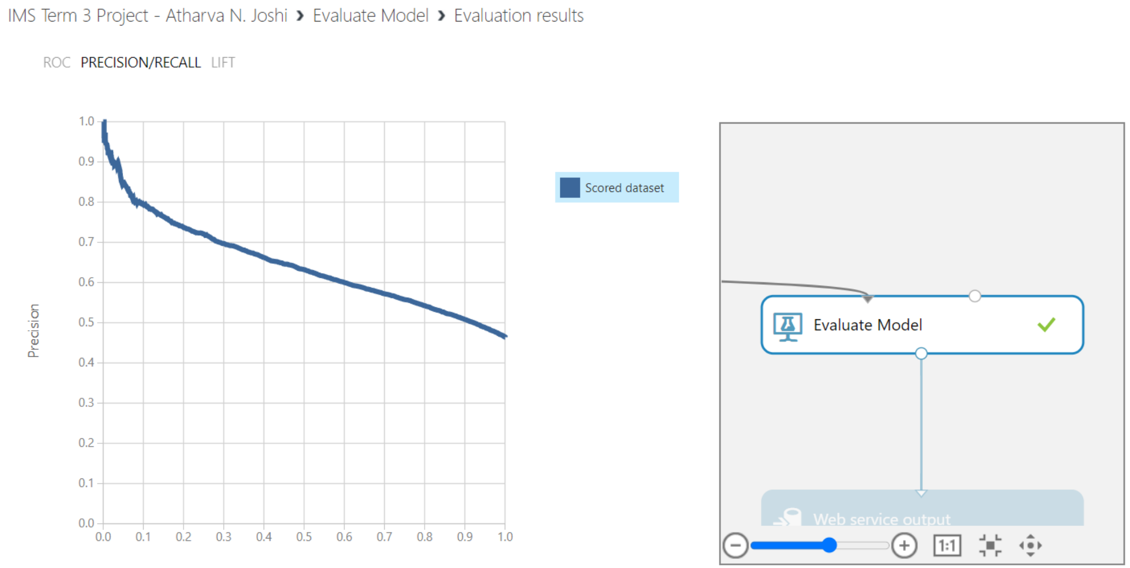
START TIME 2/17/20...

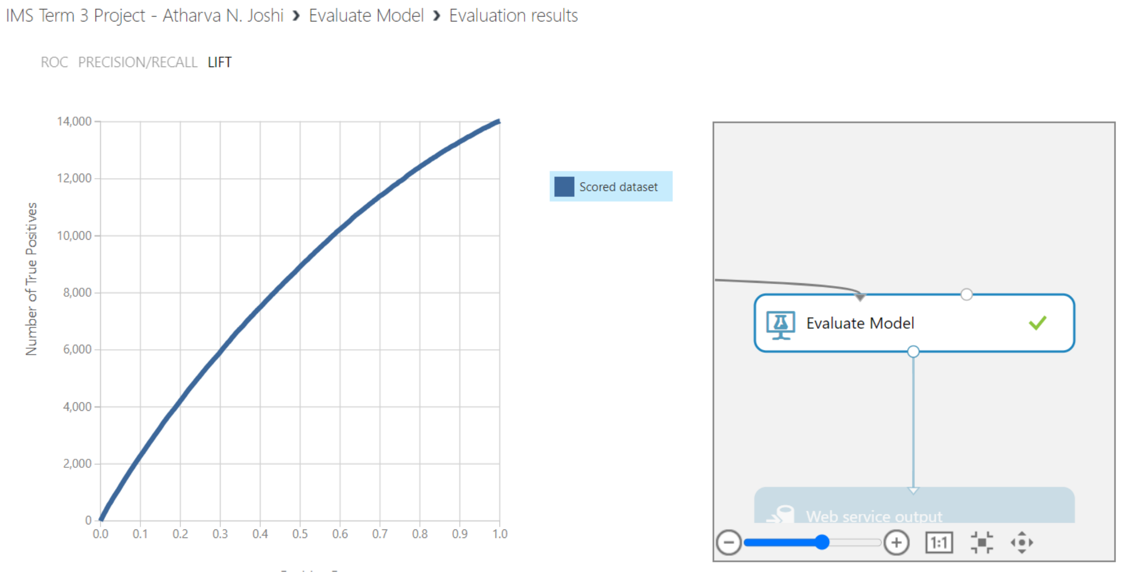
END TIME 2/17/20...

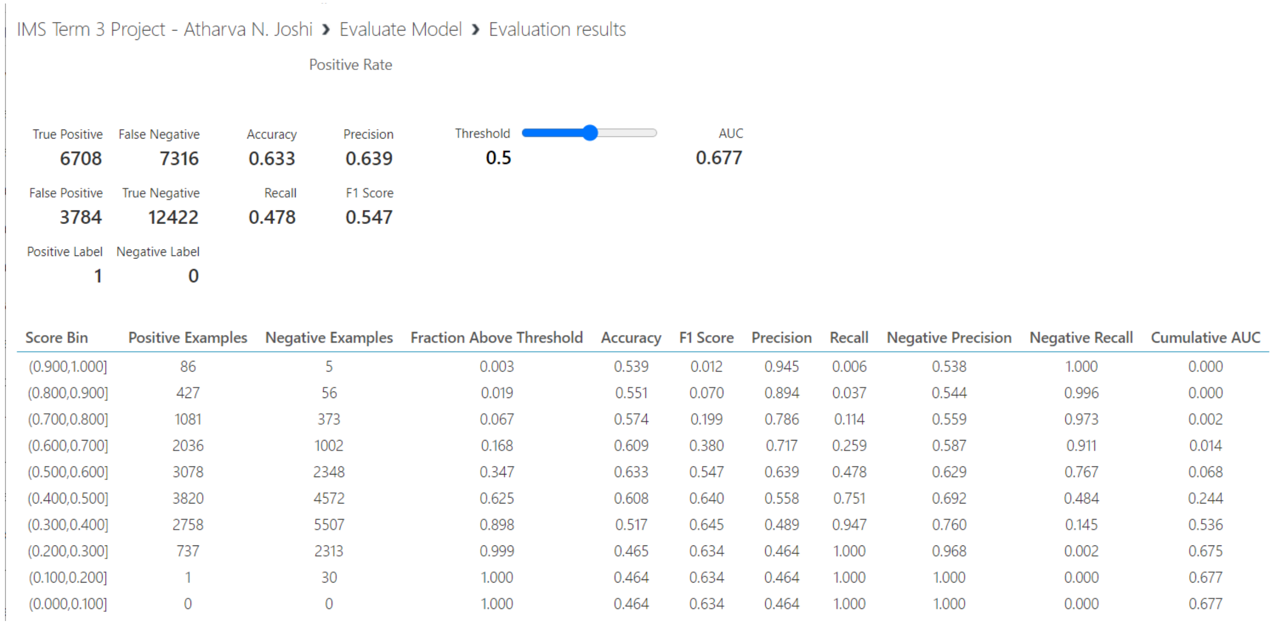
Creates a two-class classification model using the decision forest algorithm
(more help...)

Result – Best among all Algorithms


IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results


IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results


IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results


IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results

Positive Rate

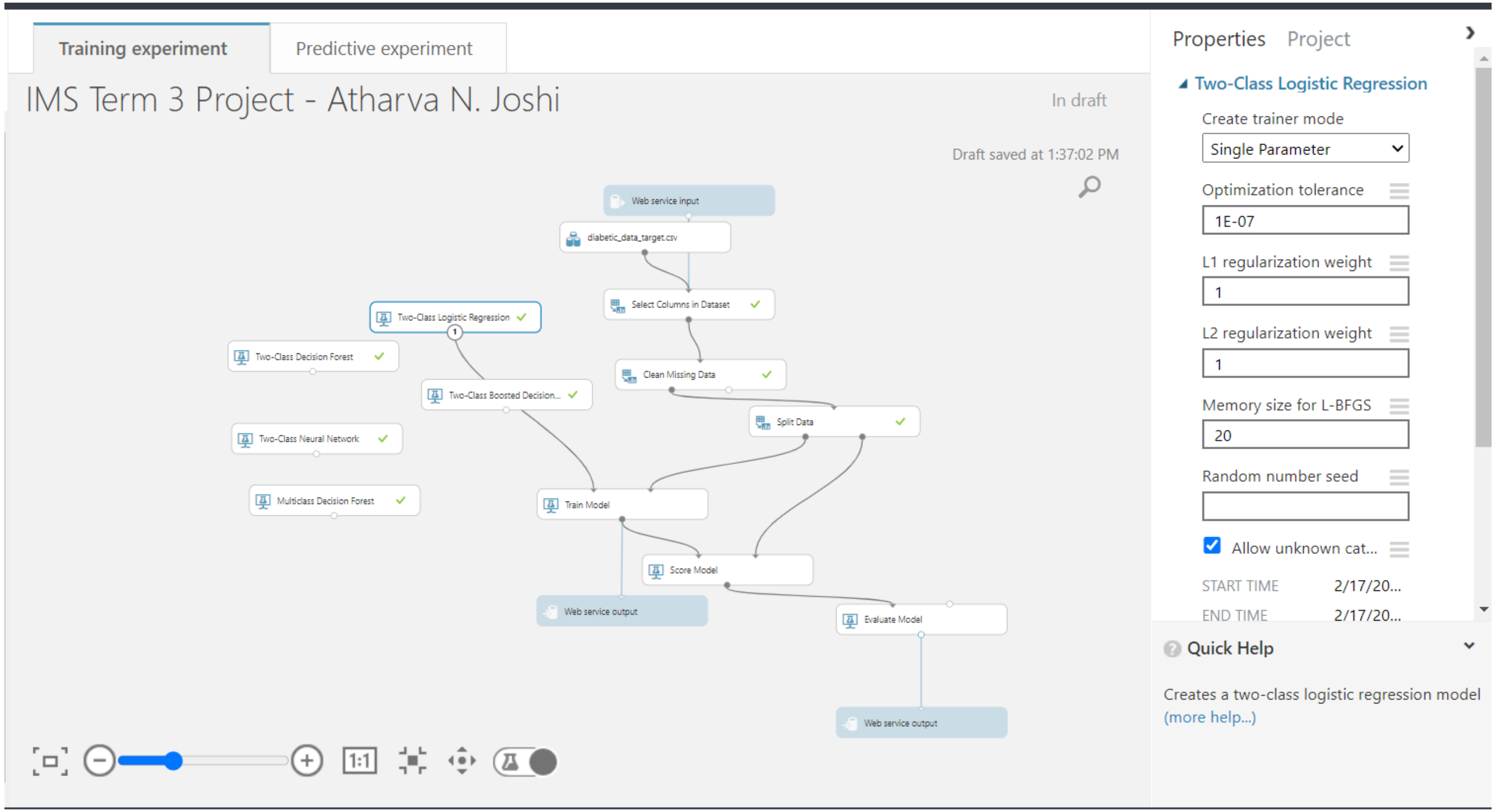
True Positive 6708
False Negative 7316
Accuracy 0.633
Precision 0.639
Threshold 0.5
AUC 0.677

False Positive 3784
True Negative 12422
Recall 0.478
F1 Score 0.547

Positive Label 1
Negative Label 0

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	86	5	0.003	0.539	0.012	0.945	0.006	0.538	1.000	0.000
(0.800,0.900]	427	56	0.019	0.551	0.070	0.894	0.037	0.544	0.996	0.000
(0.700,0.800]	1081	373	0.067	0.574	0.199	0.786	0.114	0.559	0.973	0.002
(0.600,0.700]	2036	1002	0.168	0.609	0.380	0.717	0.259	0.587	0.911	0.014
(0.500,0.600]	3078	2348	0.347	0.633	0.547	0.639	0.478	0.629	0.767	0.068
(0.400,0.500]	3820	4572	0.625	0.608	0.640	0.558	0.751	0.692	0.484	0.244
(0.300,0.400]	2758	5507	0.898	0.517	0.645	0.489	0.947	0.760	0.145	0.536
(0.200,0.300]	737	2313	0.999	0.465	0.634	0.464	1.000	0.968	0.002	0.675
(0.100,0.200]	1	30	1.000	0.464	0.634	0.464	1.000	1.000	0.000	0.677
(0.000,0.100]	0	0	1.000	0.464	0.634	0.464	1.000	1.000	0.000	0.677

Two Class Logistic Regression



IMS Term 3 Project - Atharva N. Joshi

In draft

Draft saved at 1:37:02 PM

Properties Project

Two-Class Logistic Regression

Create trainer mode: Single Parameter

Optimization tolerance: 1E-07

L1 regularization weight: 1

L2 regularization weight: 1

Memory size for L-BFGS: 20

Random number seed:

☑ Allow unknown cat...

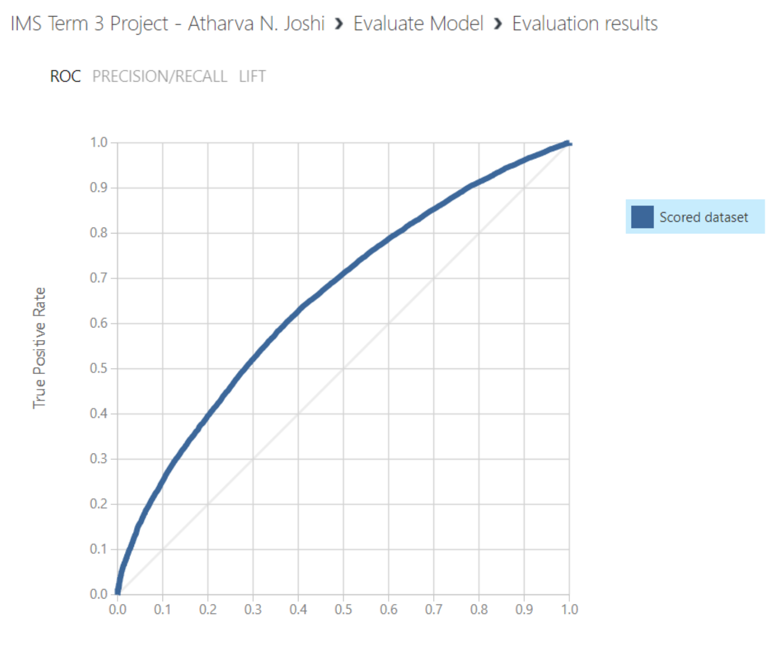
START TIME 2/17/20...

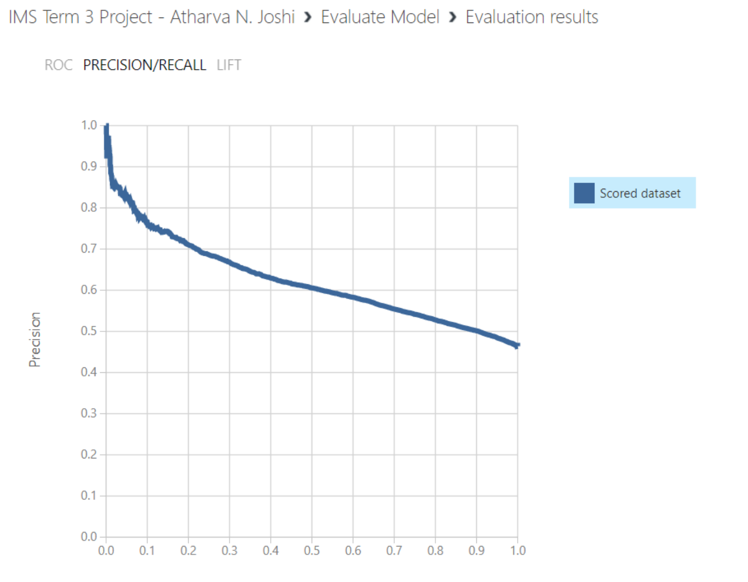
END TIME 2/17/20...

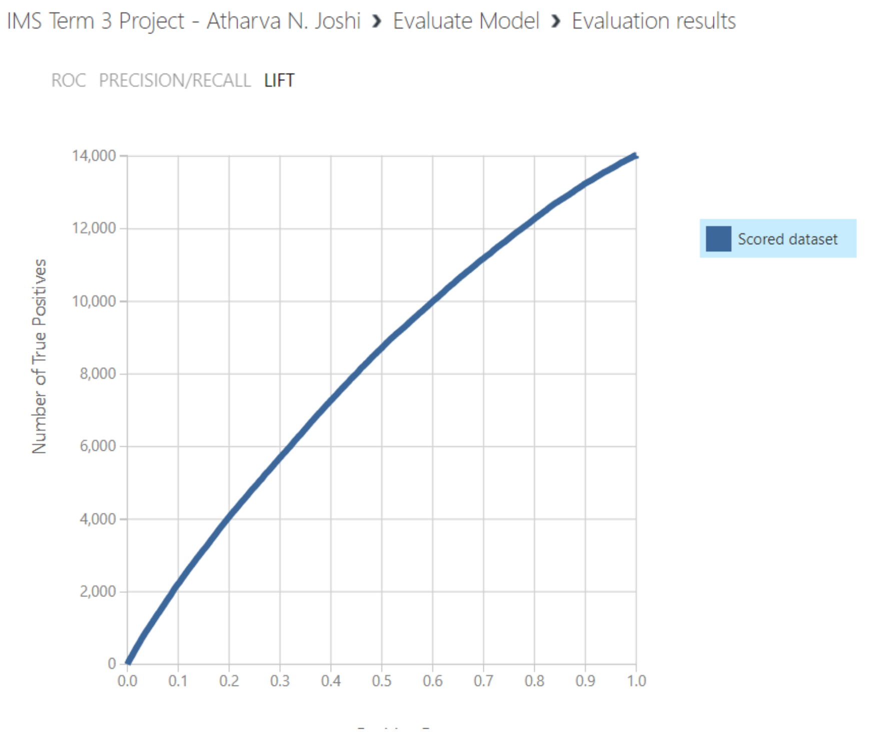
Quick Help

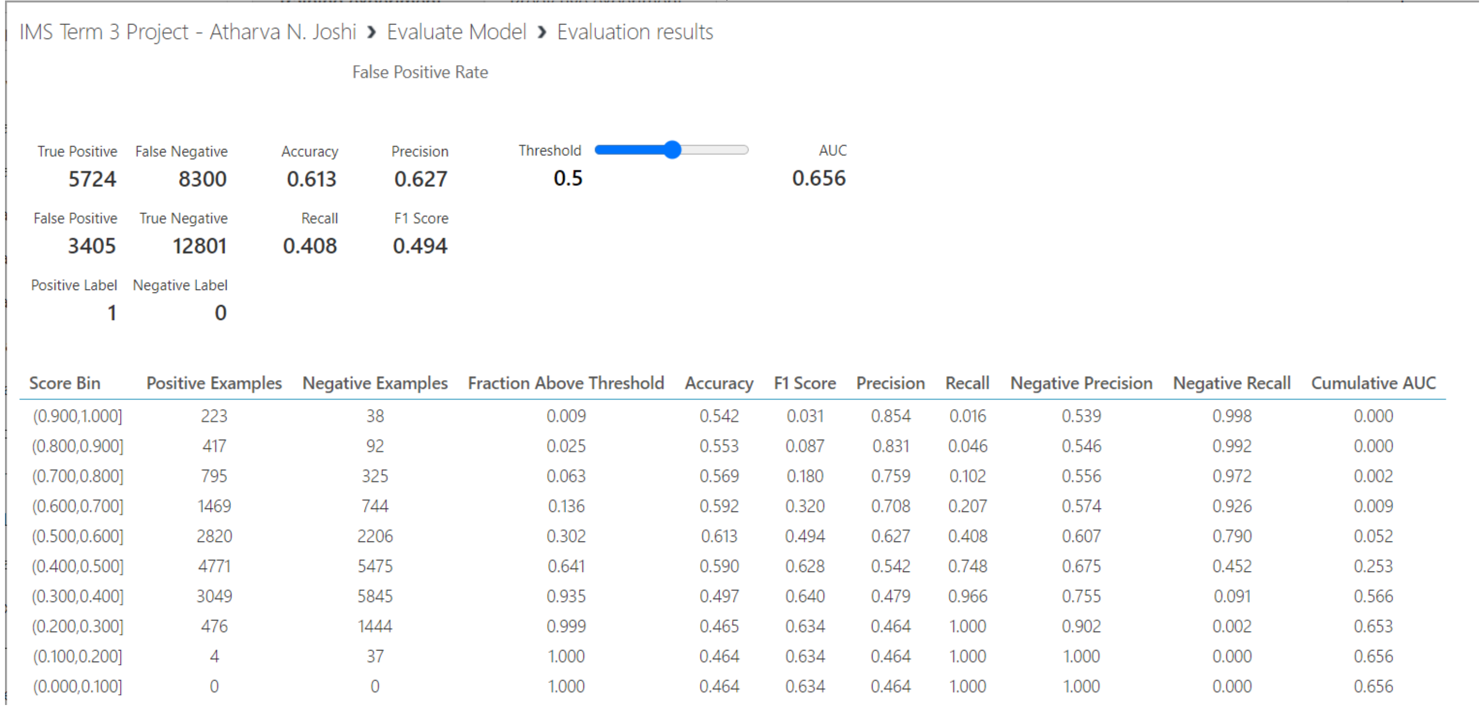
Creates a two-class logistic regression model (more help...)

Result





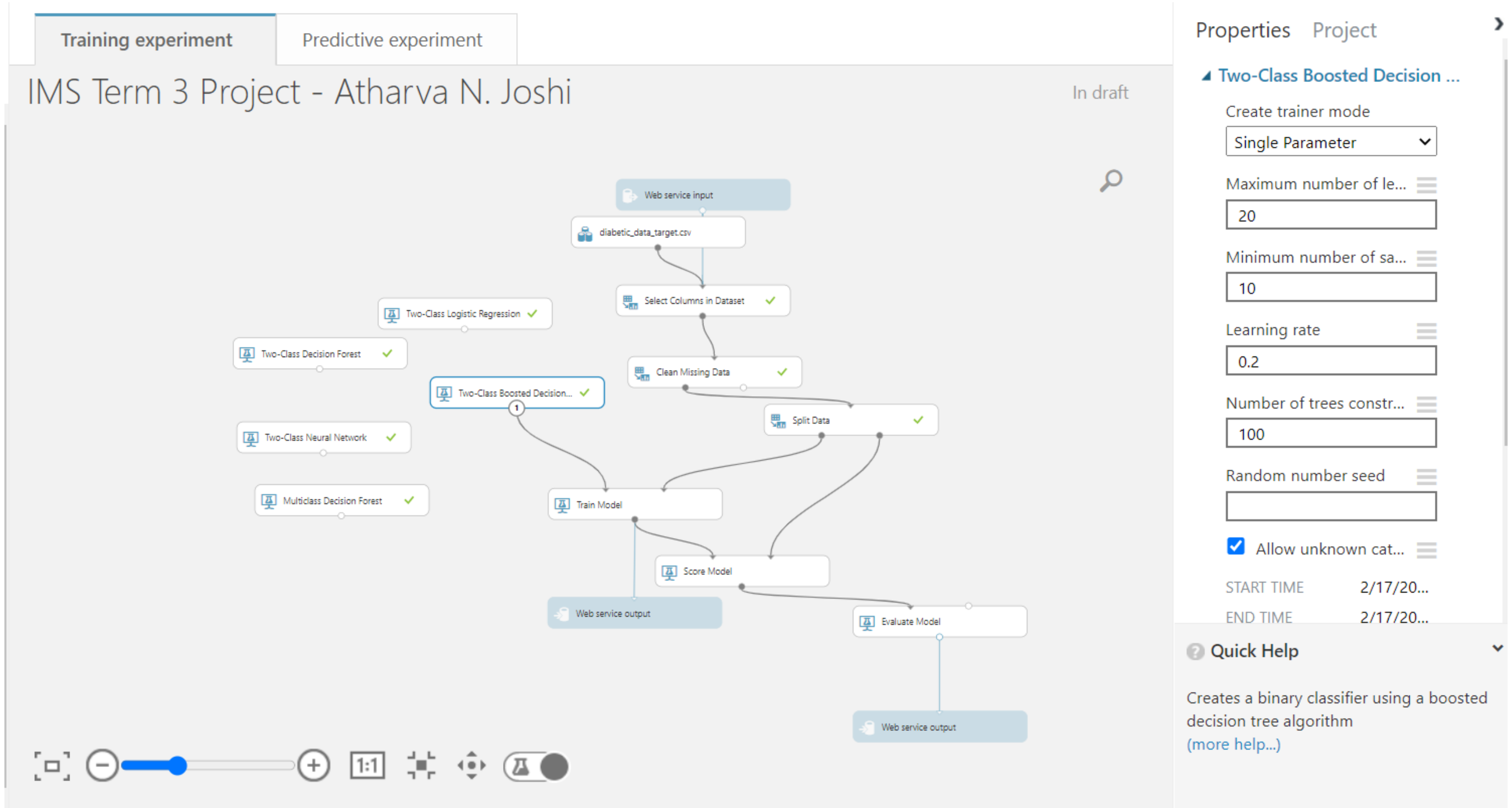



IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results
False Positive Rate

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
5724	8300	0.613	0.627	0.5	0.656
False Positive	True Negative	Recall	F1 Score		
3405	12801	0.408	0.494		
Positive Label	Negative Label				
1	0				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	223	38	0.009	0.542	0.031	0.854	0.016	0.539	0.998	0.000
(0.800,0.900]	417	92	0.025	0.553	0.087	0.831	0.046	0.546	0.992	0.000
(0.700,0.800]	795	325	0.063	0.569	0.180	0.759	0.102	0.556	0.972	0.002
(0.600,0.700]	1469	744	0.136	0.592	0.320	0.708	0.207	0.574	0.926	0.009
(0.500,0.600]	2820	2206	0.302	0.613	0.494	0.627	0.408	0.607	0.790	0.052
(0.400,0.500]	4771	5475	0.641	0.590	0.628	0.542	0.748	0.675	0.452	0.253
(0.300,0.400]	3049	5845	0.935	0.497	0.640	0.479	0.966	0.755	0.091	0.566
(0.200,0.300]	476	1444	0.999	0.465	0.634	0.464	1.000	0.902	0.002	0.653
(0.100,0.200]	4	37	1.000	0.464	0.634	0.464	1.000	1.000	0.000	0.656
(0.000,0.100]	0	0	1.000	0.464	0.634	0.464	1.000	1.000	0.000	0.656

Two Class Boosted Decision Forest



IMS Term 3 Project - Atharva N. Joshi

In draft

Two-Class Boosted Decision ...

Create trainer mode: Single Parameter

Maximum number of le...: 20

Minimum number of sa...: 10

Learning rate: 0.2

Number of trees constr...: 100

Random number seed:

☑ Allow unknown cat...

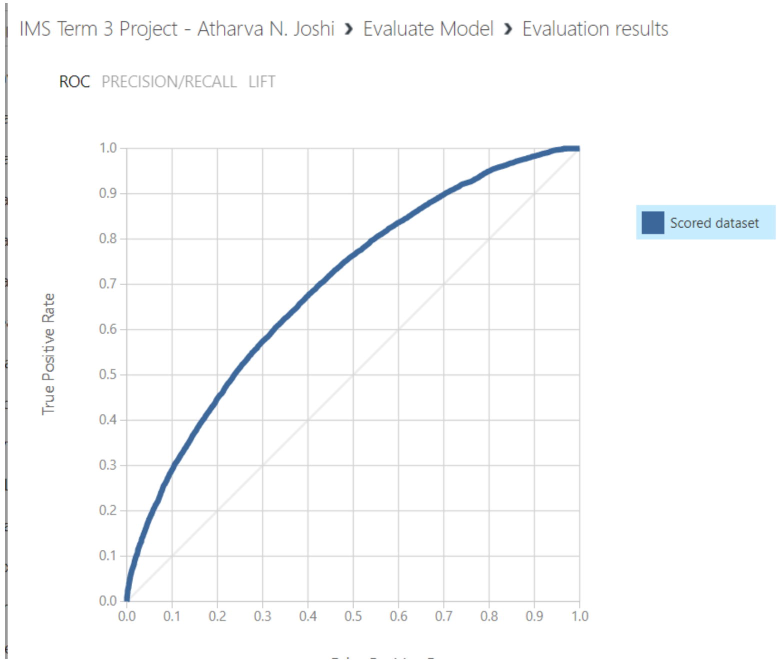
START TIME 2/17/20...

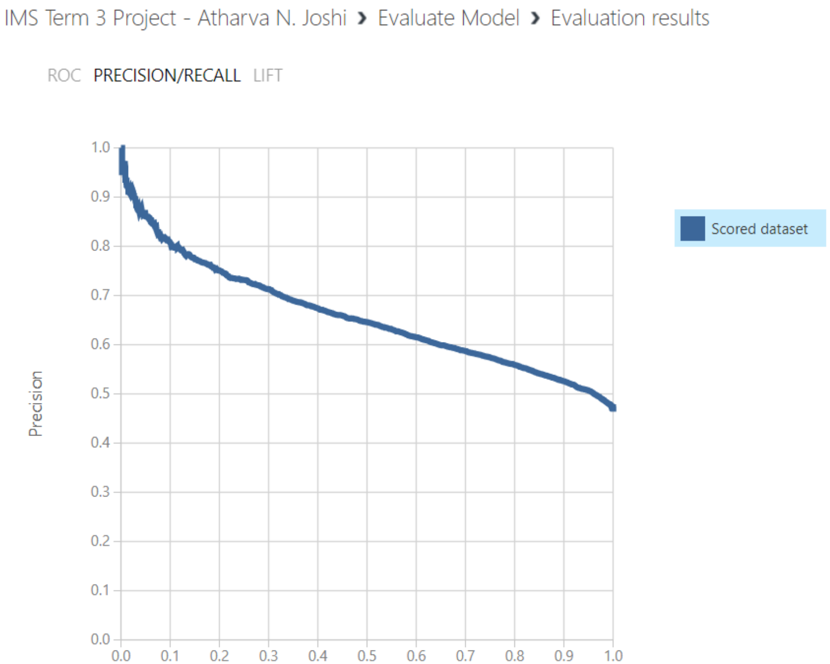
END TIME 2/17/20...

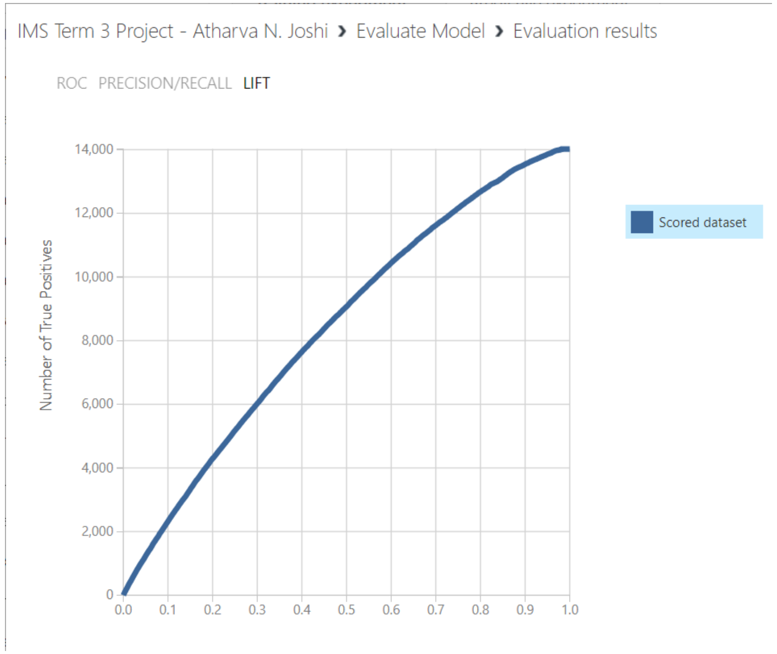
Creates a binary classifier using a boosted decision tree algorithm

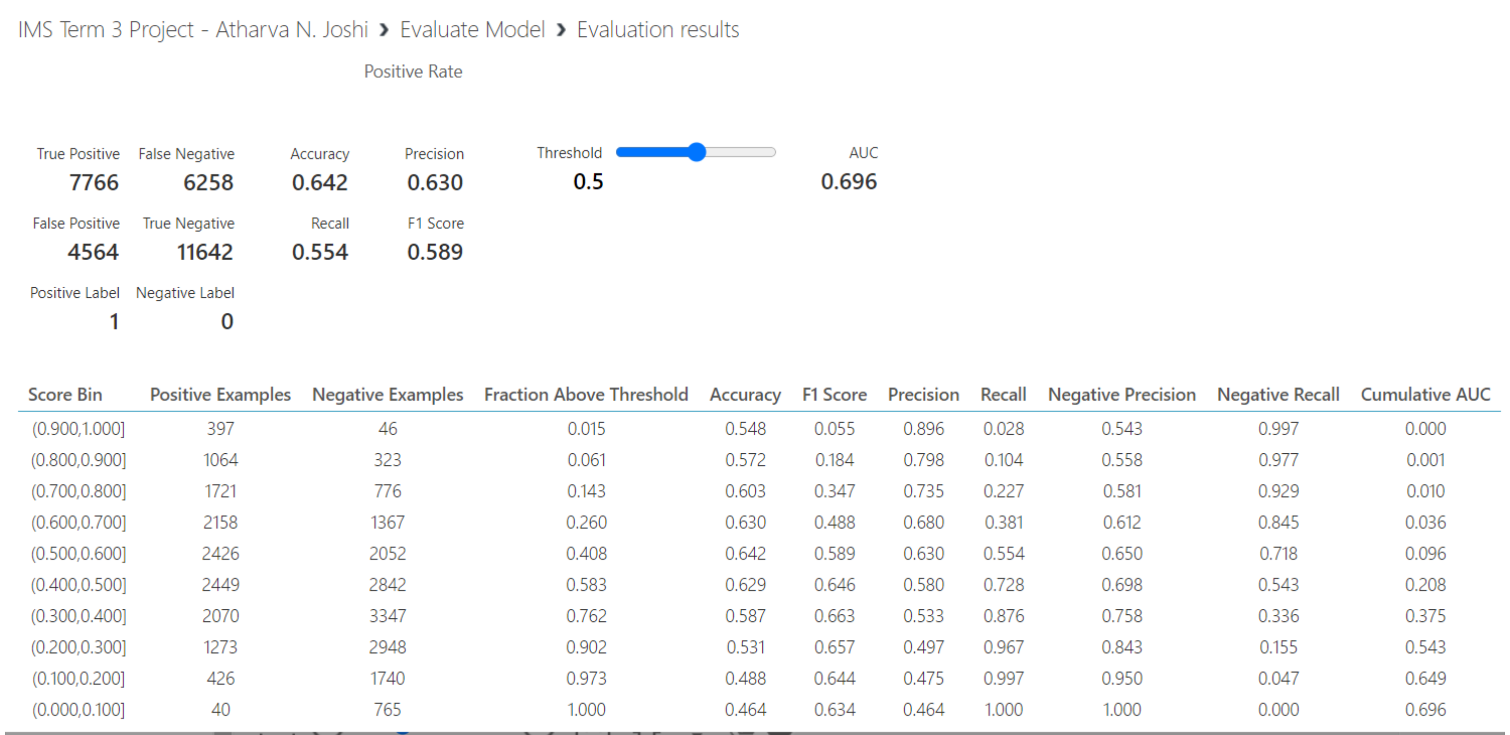
(more help...)

Result









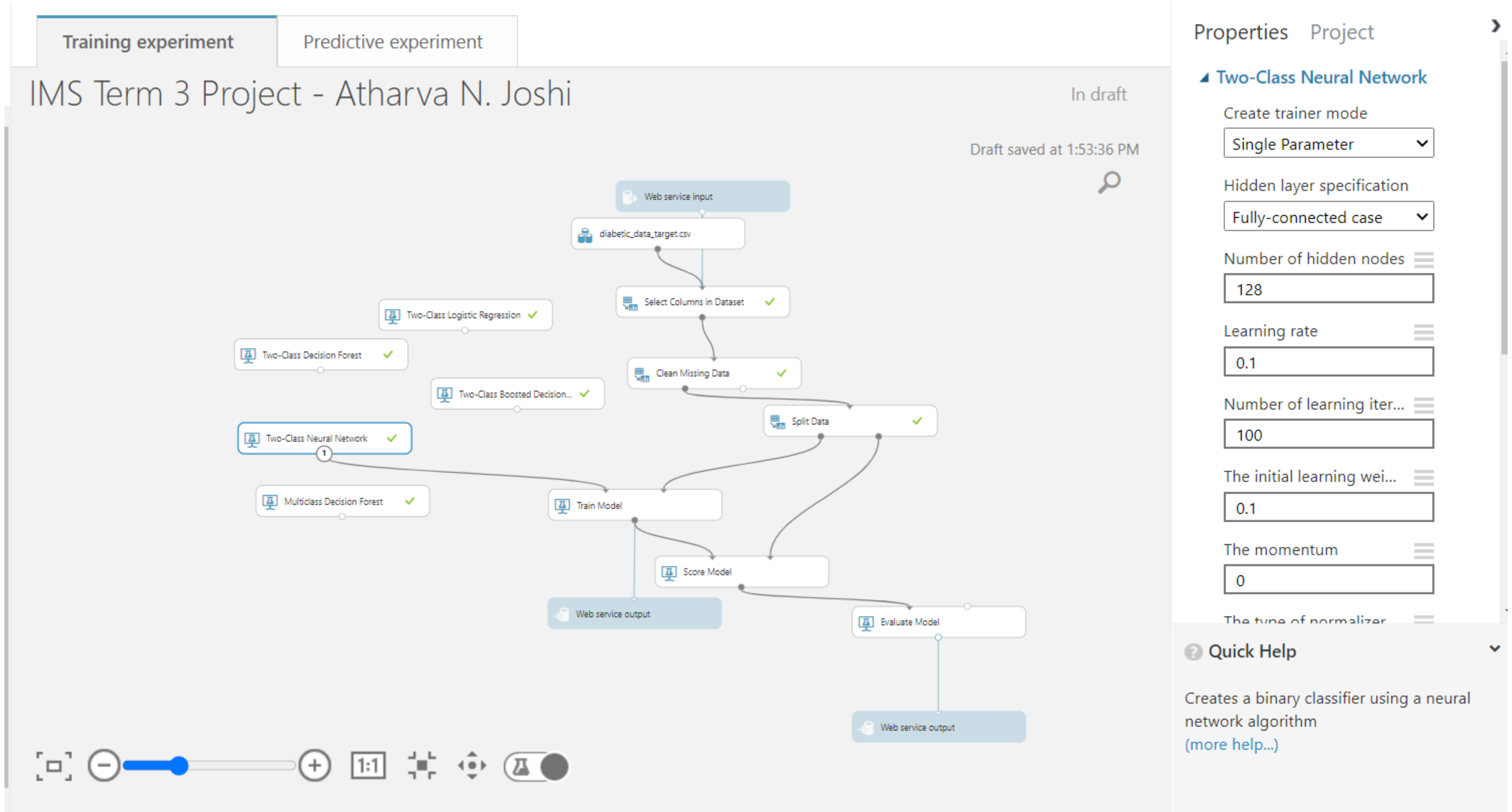
IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results

Positive Rate

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
7766	6258	0.642	0.630	0.5	0.696
False Positive	**True Negative**	**Recall**	**F1 Score**		
4564	11642	0.554	0.589		
Positive Label	**Negative Label**				
1	0				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	397	46	0.015	0.548	0.055	0.896	0.028	0.543	0.997	0.000
(0.800,0.900]	1064	323	0.061	0.572	0.184	0.798	0.104	0.558	0.977	0.001
(0.700,0.800]	1721	776	0.143	0.603	0.347	0.735	0.227	0.581	0.929	0.010
(0.600,0.700]	2158	1367	0.260	0.630	0.488	0.680	0.381	0.612	0.845	0.036
(0.500,0.600]	2426	2052	0.408	0.642	0.589	0.630	0.554	0.650	0.718	0.096
(0.400,0.500]	2449	2842	0.583	0.629	0.646	0.580	0.728	0.698	0.543	0.208
(0.300,0.400]	2070	3347	0.762	0.587	0.663	0.533	0.876	0.758	0.336	0.375
(0.200,0.300]	1273	2948	0.902	0.531	0.657	0.497	0.967	0.843	0.155	0.543
(0.100,0.200]	426	1740	0.973	0.488	0.644	0.475	0.997	0.950	0.047	0.649
(0.000,0.100]	40	765	1.000	0.464	0.634	0.464	1.000	1.000	0.000	0.696

Two Class Neural Network



IMS Term 3 Project - Atharva N. Joshi

In draft

Draft saved at 1:53:36 PM

Two-Class Neural Network

Create trainer mode: Single Parameter

Hidden layer specification: Fully-connected case

Number of hidden nodes: 128

Learning rate: 0.1

Number of learning iter...: 100

The initial learning wei...: 0.1

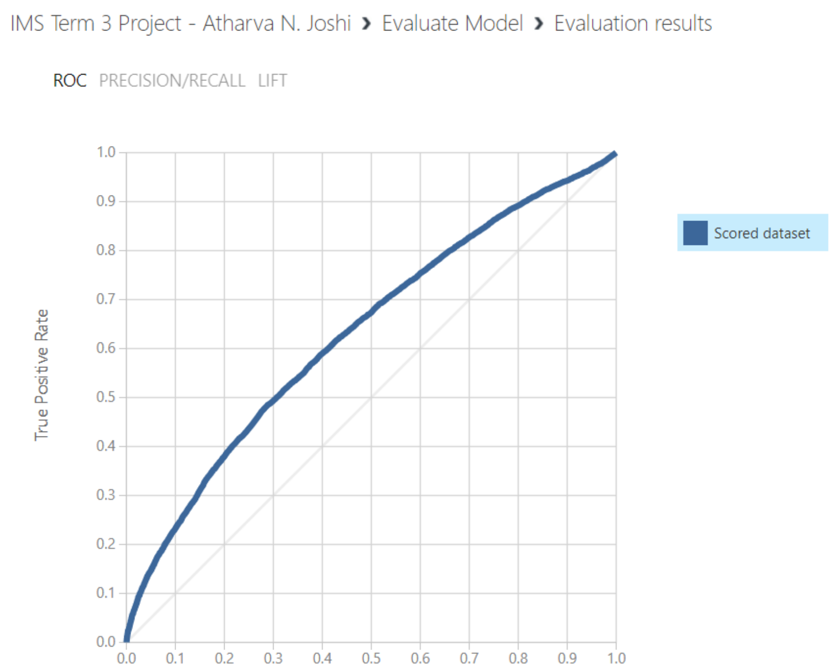
The momentum: 0

The type of normalizer

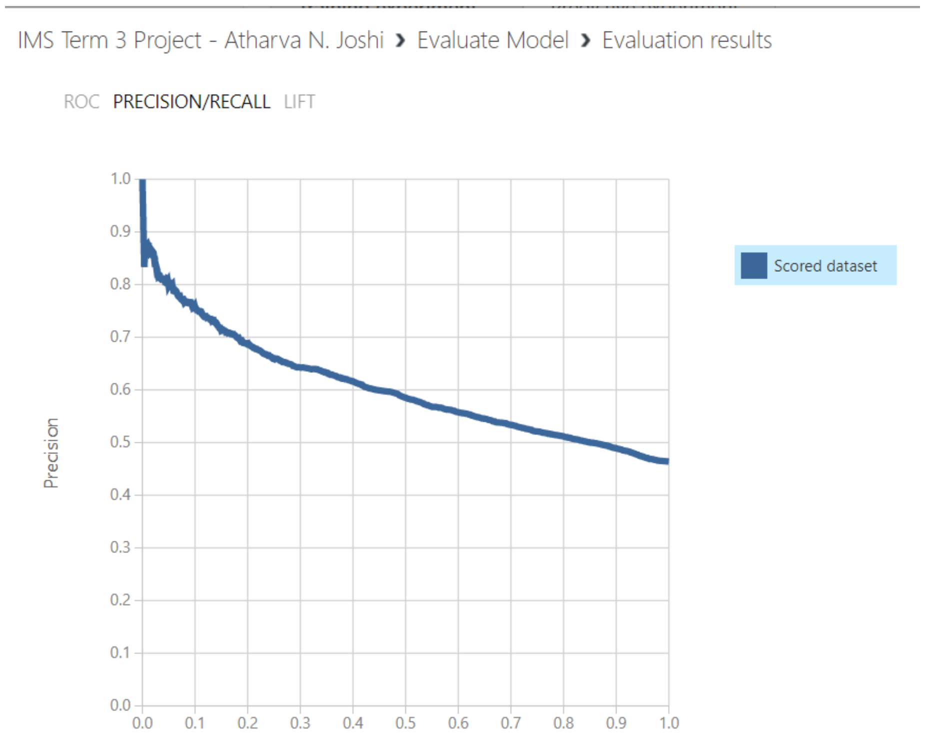
Quick Help

Creates a binary classifier using a neural network algorithm
(more help...)

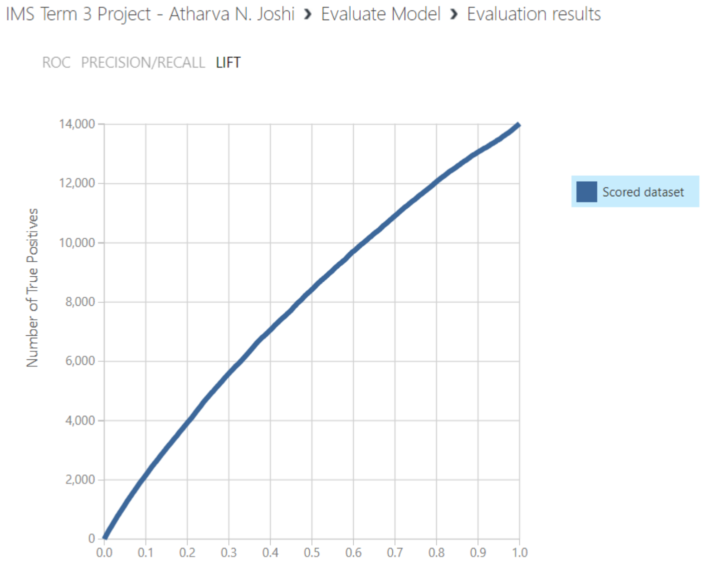
Result



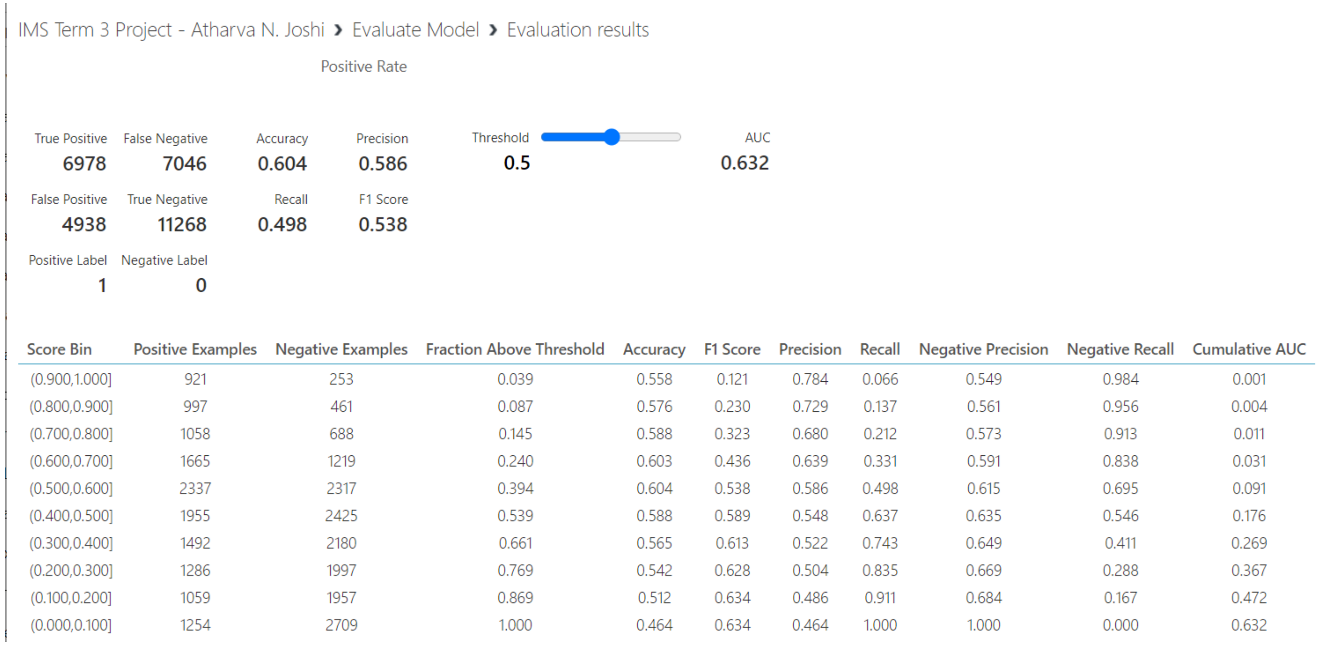
IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results



IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results



IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results



IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results

Positive Rate

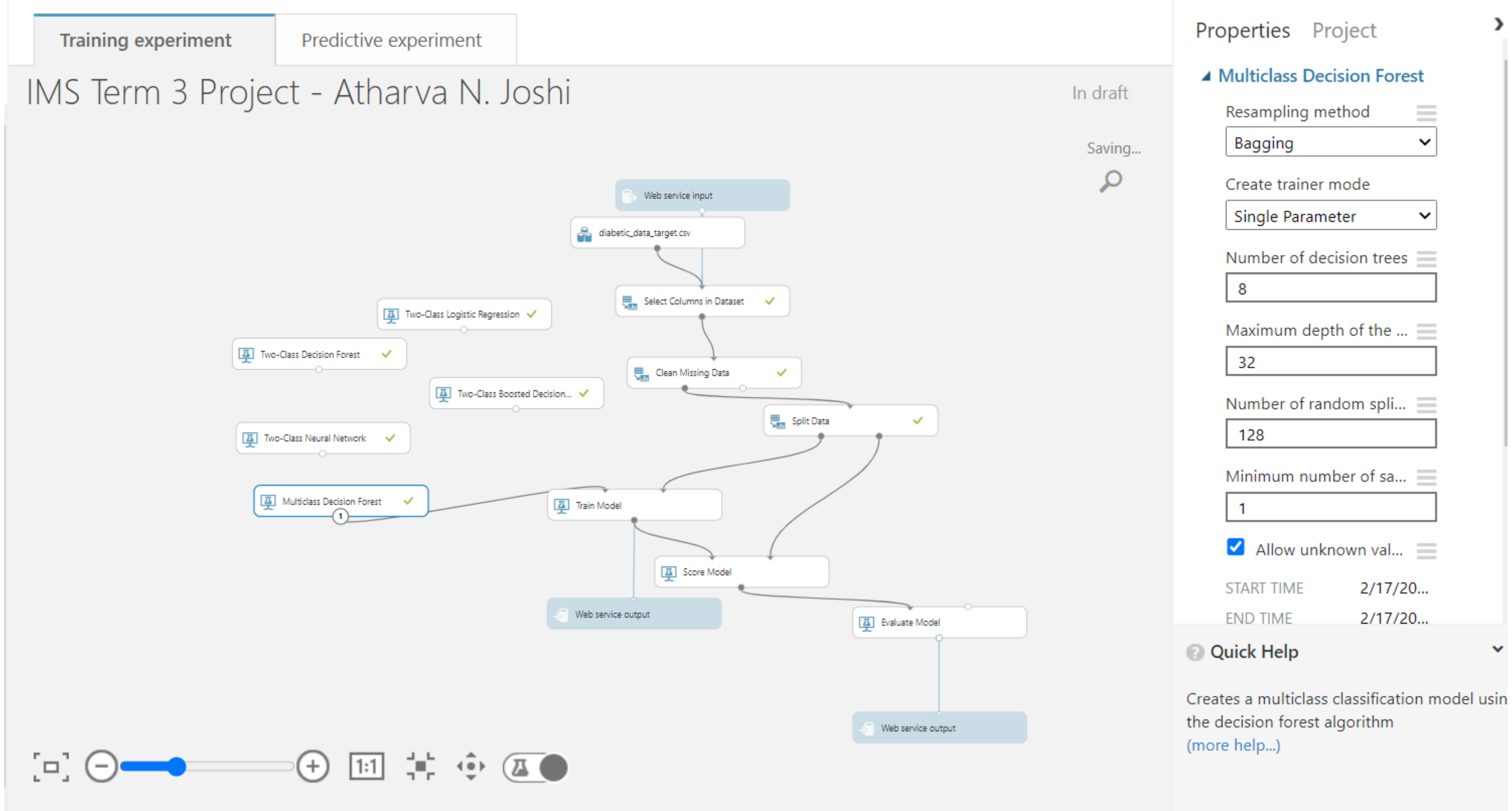
True Positive	False Negative	Accuracy	Precision	Threshold	AUC
6978	7046	0.604	0.586	0.5	0.632

False Positive	True Negative	Recall	F1 Score
4938	11268	0.498	0.538

Positive Label	Negative Label
1	0

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	921	253	0.039	0.558	0.121	0.784	0.066	0.549	0.984	0.001
(0.800,0.900]	997	461	0.087	0.576	0.230	0.729	0.137	0.561	0.956	0.004
(0.700,0.800]	1058	688	0.145	0.588	0.323	0.680	0.212	0.573	0.913	0.011
(0.600,0.700]	1665	1219	0.240	0.603	0.436	0.639	0.331	0.591	0.838	0.031
(0.500,0.600]	2337	2317	0.394	0.604	0.538	0.586	0.498	0.615	0.695	0.091
(0.400,0.500]	1955	2425	0.539	0.588	0.589	0.548	0.637	0.635	0.546	0.176
(0.300,0.400]	1492	2180	0.661	0.565	0.613	0.522	0.743	0.649	0.411	0.269
(0.200,0.300]	1286	1997	0.769	0.542	0.628	0.504	0.835	0.669	0.288	0.367
(0.100,0.200]	1059	1957	0.869	0.512	0.634	0.486	0.911	0.684	0.167	0.472
(0.000,0.100]	1254	2709	1.000	0.464	0.634	0.464	1.000	1.000	0.000	0.632

Multi Class Decision Forest



IMS Term 3 Project - Atharva N. Joshi

In draft

Saving...

Multiclass Decision Forest

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision trees: 8

Maximum depth of the ...: 32

Number of random spli...: 128

Minimum number of sa...: 1

☑ Allow unknown val...

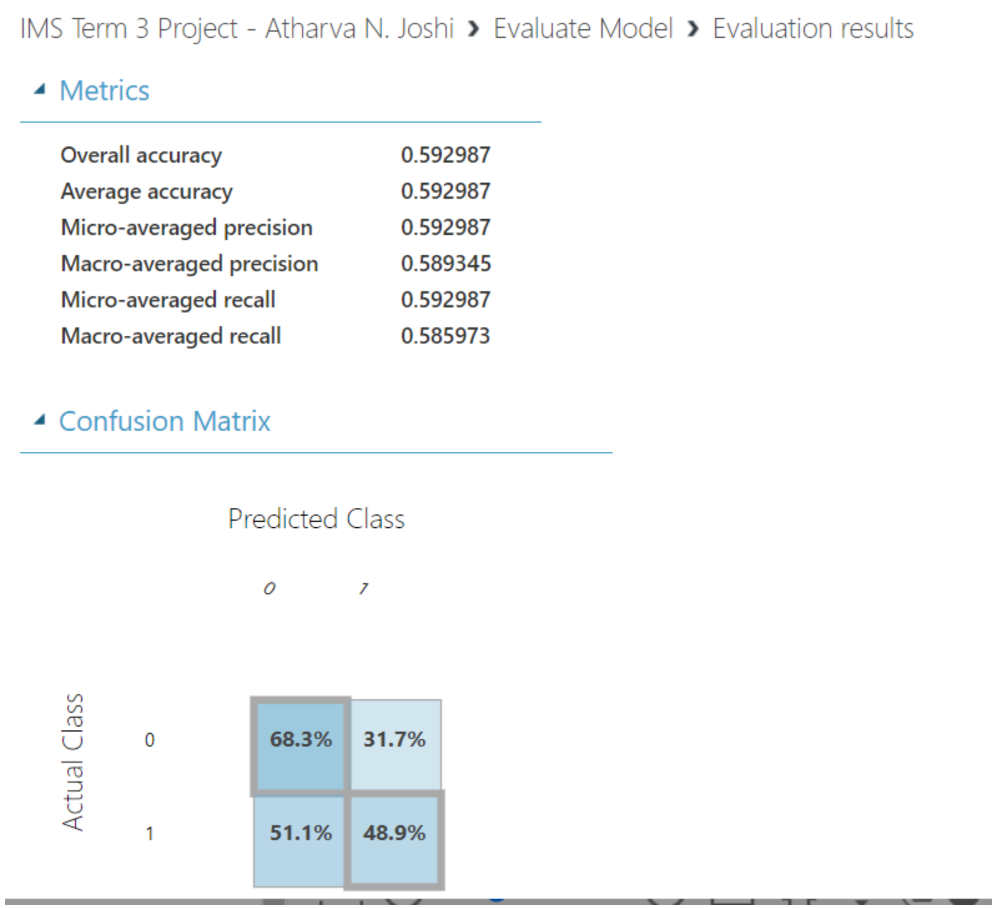
START TIME 2/17/20...

END TIME 2/17/20...

Quick Help

Creates a multiclass classification model usin the decision forest algorithm (more help...)

Result



IMS Term 3 Project - Atharva N. Joshi › Evaluate Model › Evaluation results

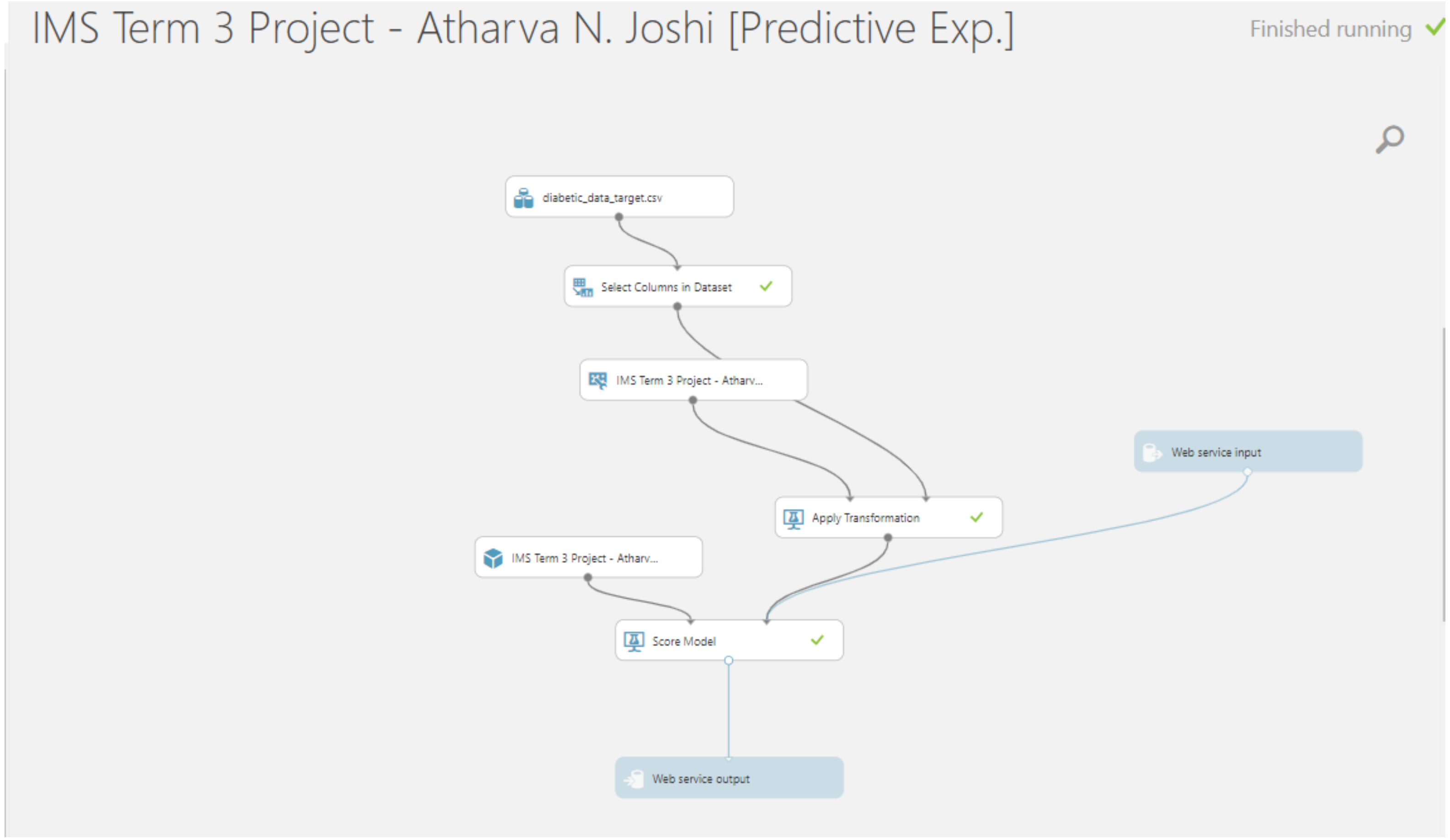
Metrics

Overall accuracy	0.592987
Average accuracy	0.592987
Micro-averaged precision	0.592987
Macro-averaged precision	0.589345
Micro-averaged recall	0.592987
Macro-averaged recall	0.585973

Confusion Matrix

Actual Class	Predicted Class 0	Predicted Class 1
0	68.3%	31.7%
1	51.1%	48.9%

Deployment over Azure


IMS Term 3 Project - Atharva N. Joshi [Predictive Exp.]
Finished running ✓
diabetic_data_target.csv
Select Columns in Dataset ✓
IMS Term 3 Project - Atharv...
Web service input
Apply Transformation ✓
IMS Term 3 Project - Atharv...
Score Model ✓
Web service output

Deployment Link

https://gallery.azure.ai/Experiment/IMS-Term-3-Project-Atharva-N-Joshi-Predictive-Exp