

Introdução ao Azure Machine Learning

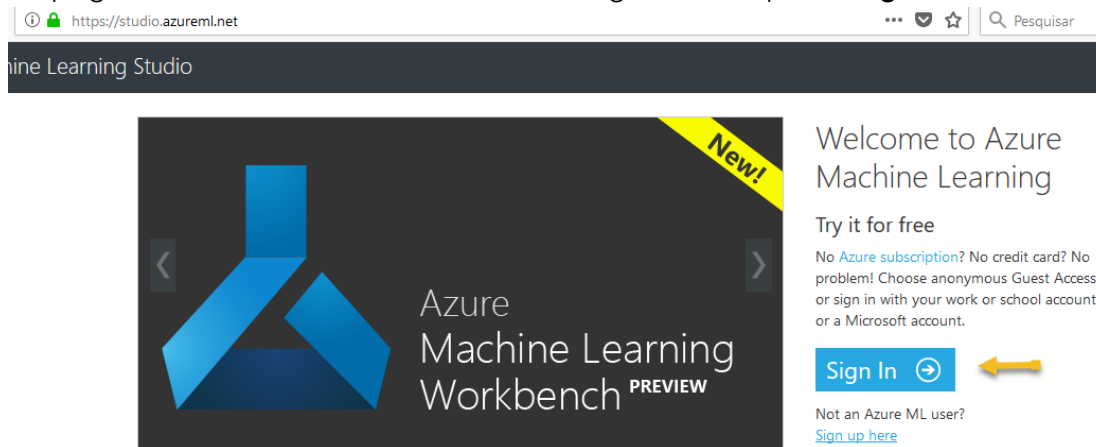
Módulo 4 – Prática Final

Introdução

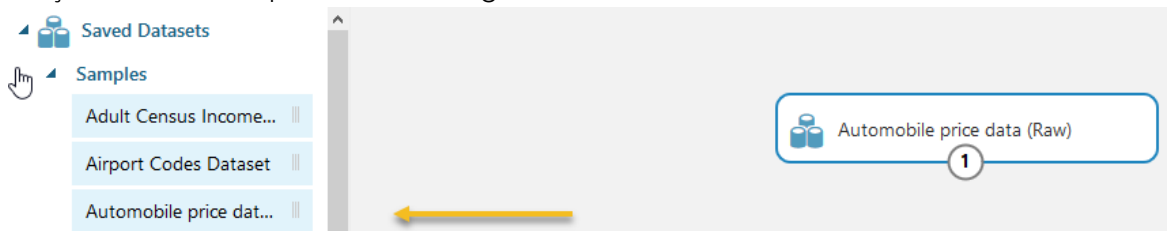
Neste laboratório você irá realizar diversas práticas para fixação dos conteúdos abordados.

Lab1 – Criando Experimento no Azure Machine Learning

- Na página do Microsoft Azure Machine Learning Studio, clique em **Sign In**, conforme tela abaixo:

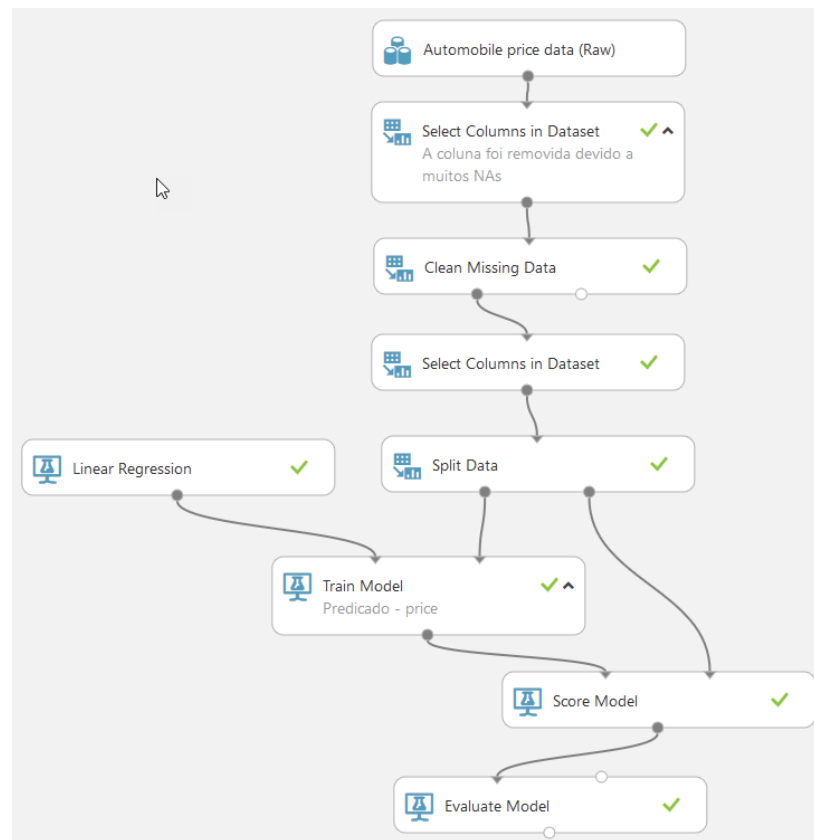


- Cri um novo experimento e arraste o exemplo **Automobile price data (Raw)** para a área de edição, conforme apresentado a seguir.



- Arraste o módulo **Select Columns in Dataset** e adicione todas as colunas e exceto a coluna *normalized-losses*.
- Clique no módulo e digite o texto "**A coluna foi removida devido a muitos NAs**".
- Arraste o módulo **Clean Missing Data** e escolha **Remove entire Row**, irá excluir a linha se alguma coluna estiver faltando.
- Arraste o módulo **Select Columns in Dataset** e adicione somente as colunas: *make*, *body-style*, *wheel-base*, *engine-size*, *horsepower*, *peak-rpm*, *highway-mpg*, *price*.
- Adicione o módulo **Split Data** para dividir em 75% dos dados para treinamento do modelo e 25% para ser usado no módulo **Score Model**.
- Conecte o módulo **Split Data** no **Train Model** e como também o módulo **Linear Regression**.
- Em seguida, no 2o. dataset do Split Data conecte ao **Score Model**.
- Arraste o **Evaluate Model** e conecte ao **Score Model**.

- Verifique se após a sequência de passos que você criou está conforme a figura abaixo.



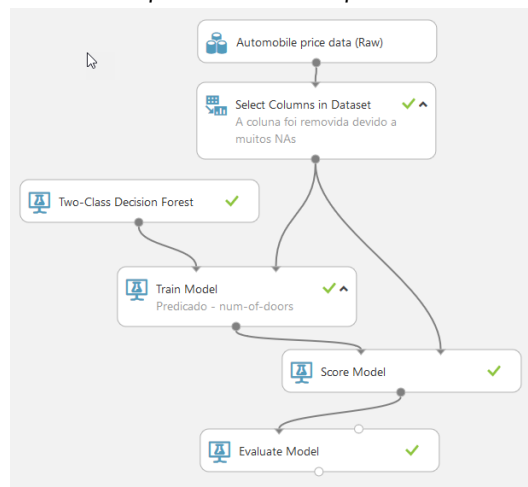
- Salve e execute o experimento.

Lab2 – Criando Experimento no Azure Machine Learning usando Algoritmo de Classificação para Aprendizagem de Máquina Supervisionada

- No experimento criado no Laboratório 1, clique em SAVE AS e salve com o nome de LabClassificacao.
- Exclua os seguintes módulos:
 - Clean Missing Data
 - Select Columns in DataSet
 - Split Data
 - Linear Regression
- Após o passo anterior deve ficar dessa forma o experimento.



- Clique no módulo **Select Columns in DataSet** e adicione a coluna *num-of-doors*.
- Clique em Train Model e mude a coluna *price* para *num-of-doors*.
- Arraste o módulo **Two-Class Decision Forest** e configure para **1** a opção *Number of decision trees*.
- Verifique se o experimento criado por você se apresenta conforme exibido abaixo.



- Salve e execute o experimento.

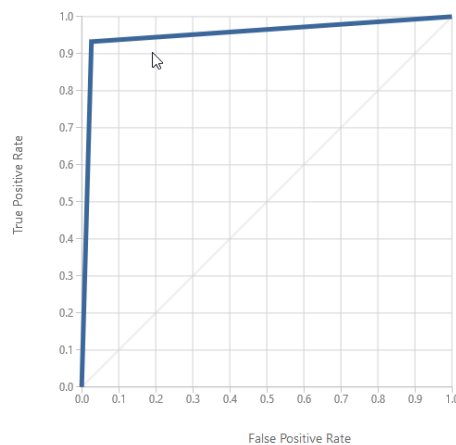
- Clique com o botão direito no **Score Model**, escolha **Scored dataset** e em seguida **Visualize**.

rows	columns													
205	27													
curb-weight	engine-type	num-of-cylinders	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price	Scored Labels	Scored Probabilities
2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	13495	two	1
2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	16500	two	1
2823	ohcv	six	152	mpfi	2.68	3.47	9	154	5000	19	26	16500	two	1
2337	ohc	four	109	mpfi	3.19	3.4	10	102	5500	24	30	13950	four	0
2824	ohc	five	136	mpfi	3.19	3.4	8	115	5500	18	22	17450	four	0
2507	ohc	five	136	mpfi	3.19	3.4	8.5	110	5500	19	25	15250	two	1

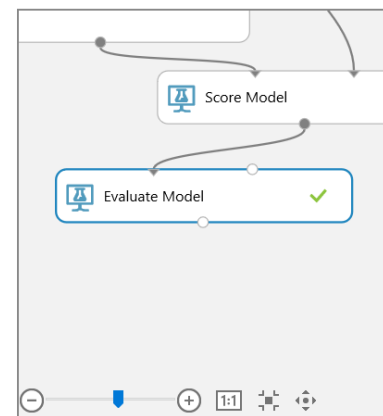
- No conjunto de dados observados temos apenas duas classes 0 e 1 na coluna **Scored Probabilities**. Vamos agora fechar e no experimento clique com o botão direito no **Evaluate Model** e escolha **Evaluation results** e em seguida **Visualize**.

LabClassificacao > Evaluate Model > Evaluation results

ROC PRECISION/RECALL LIFT



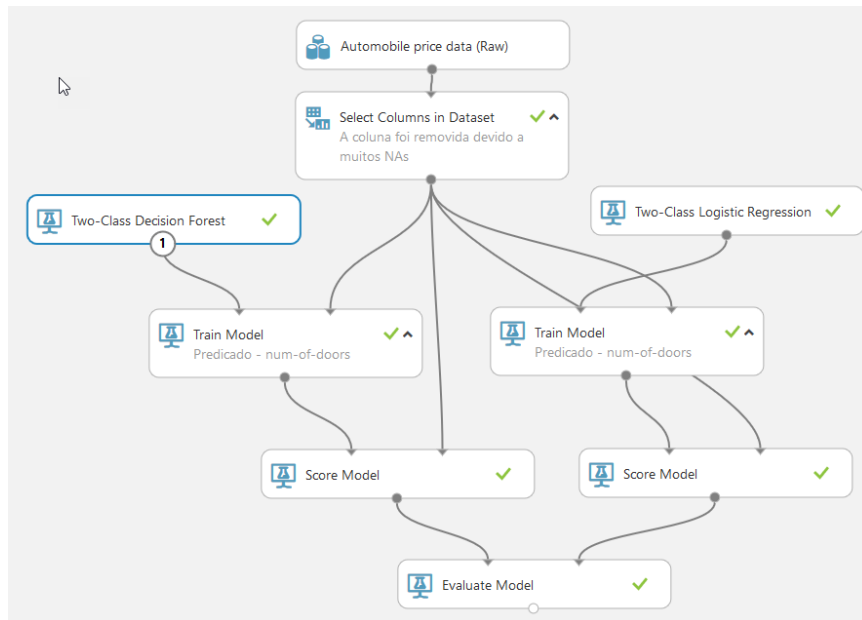
Scored dataset



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
83	6	0.956	0.965	0.5	0.941
False Positive	True Negative	Recall	F1 Score		
3	111	0.933	0.949		

- Observe o valor da acurácia, bem como os falsos negativos e positivos.

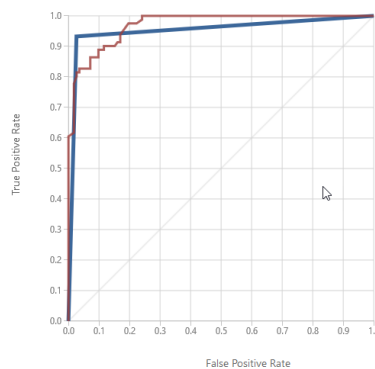
- Vamos modificar o experimento adicionando o módulo **Two-Class Logistic Regression**, bem como fazendo as modificações apresentadas a seguir.



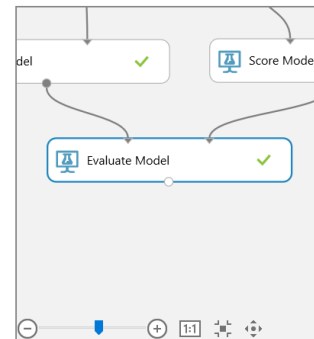
- Clique com o botão direito no **Evaluate Model** e escolha **Evaluation results** e em seguida Visualize.

LabClassificacao2 > Evaluate Model > Evaluation results

ROC PRECISION/RECALL LIFT



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
83	6	0.956	0.965	0.5	0.941
False Positive	True Negative	Recall	F1 Score		
3	111	0.933	0.949		



- Clique em **Scored dataset to compare**.
- Na sua opinião este novo modelo é melhor do que o anterior? Justifique a sua resposta, enviando para o e-mail: marcondesalexandre@gmail.com. Estarei retornando sobre análise enviada.