

**Московский авиационный институт
(Национальный исследовательский университет)**

Факультет: «Информационные технологии и прикладная математика»

Кафедра: 806 «Вычислительная математика и программирование»

Дисциплина: «Численные методы»

**Лабораторная работа 1
Azure ML**

Студент: Девяткина Д.В.

Группа: 80-304

Дата:

Оценка:

Постановка задачи

Ваша задача познакомиться с платформой Azure Machine Learning, реализовывая полный цикл разработки решения задачи машинного обучения, используя три различных алгоритма, реализованные на этой платформе.

Задача 1

Предсказывание цены автомобиля на основании его производителя, типа распределительного вала, размера двигателя и количества лошадиных сил.

Был использован алгоритм *линейной регрессии*, а конкретно *множественной линейной регрессии* (так как цена в этом эксперименте ищется в зависимости от 4 предикторов), как наиболее простой способ построения прогнозирования. В ЛР 0 после построения графика, можно было увидеть, что зависимость цены от количества лошадиных сил можно было считать линейной.

Из датасета выделила 70 % для обучающей выборки.

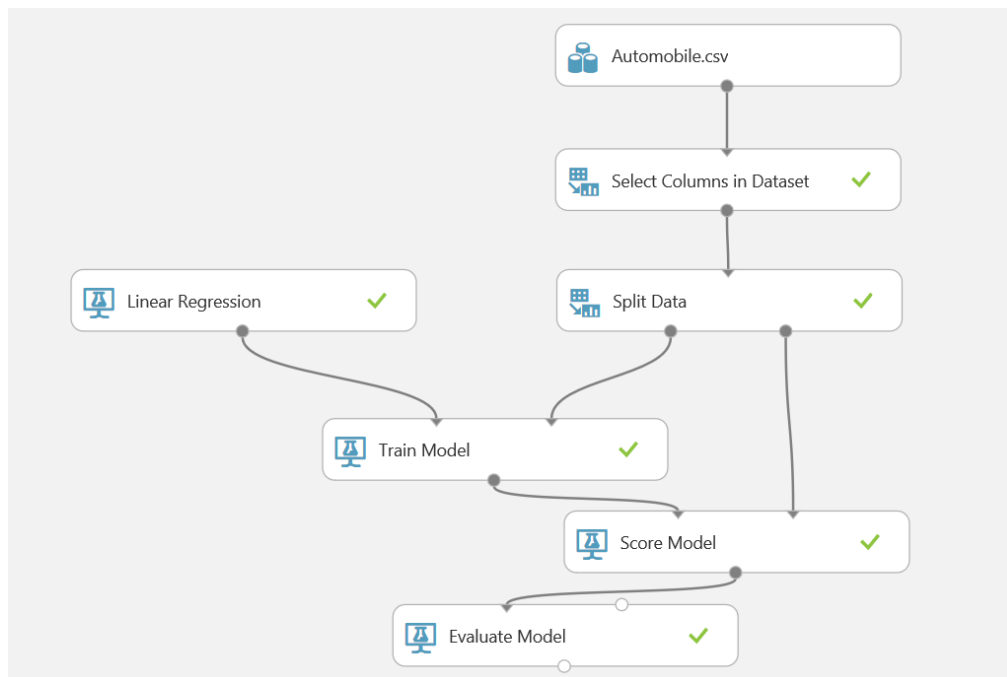


Рис 1. Модель эксперимента

linreg price prediction > Score Model > Scored dataset


rows	columns
60	6
	makeengine_typeengine_sizehorsepowerpriceScored Labels
view as	
	mazdaohc916867958005.685989
	toyotohc987069387136.07893
	dodgeohc906855725958.625465
	dodgeohc9810285588393.089612
	bmwohc1641212456526888.658492
	toyotohc146116998913637.15678
	volvohcv1731342148517919.579779
	toyotohc926287786208.014826
	subaruohef1088292338489.182739
	volvohc1451062247016539.293917
	toyotohc1461161119913637.15678
	jaguardohc2581763555032246.060532

Рис 2. Полученные оценки

Здесь мы можем посмотреть на цену, которая была у автомобиля изначально и сравнить её с полученной в ходе расчета.

Metrics

Mean Absolute Error	1729.520744
Root Mean Squared Error	2334.415066
Relative Absolute Error	0.267213
Relative Squared Error	0.079516
Coefficient of Determination	0.920484

Рис 3. Метрики

Средняя абсолютная ошибка, среднеквадратичная ошибка, относительные ошибки и коэффициент смешанной корреляции (детерминированности).

▲ Error Histogram

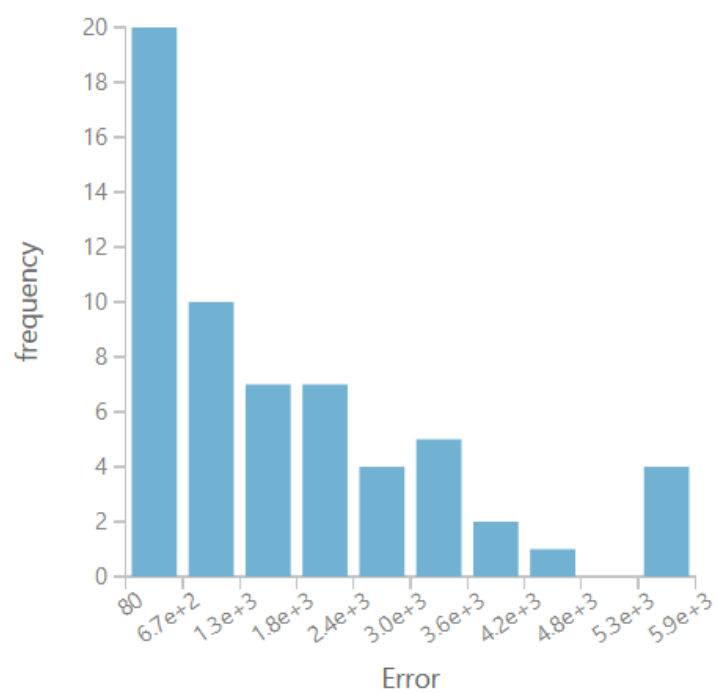


Рис 4. Ошибки и их частоты

Задача 2

Мультиклассовая классификация типа кузова (хэтчбек, кабриолет, седан и тд) по всем параметрам датасета.

Для классификации я выбрала *дерево решений*, он прост для представления и понимания, сами деревья строятся в Azure ML Studio, является надежным, не требует подготовки данных и работает со всеми типами данных.

Выделила 80% процентов датасета для обучения.

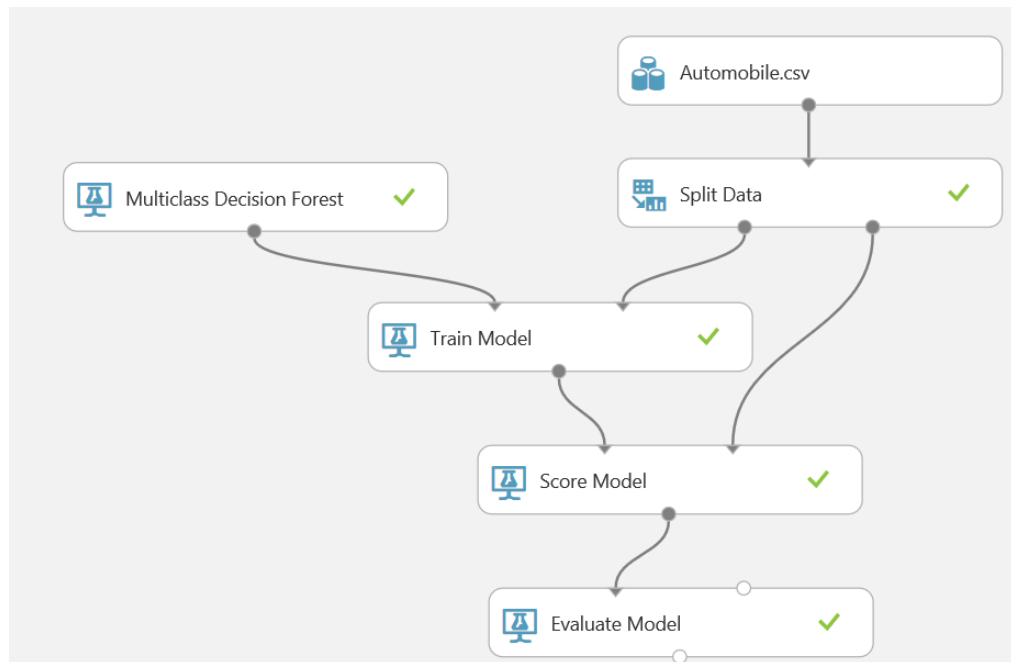


Рис 5. Модель эксперимента

desforest body prediction ▶ Train Model ▶ Trained model

trees constructed
8

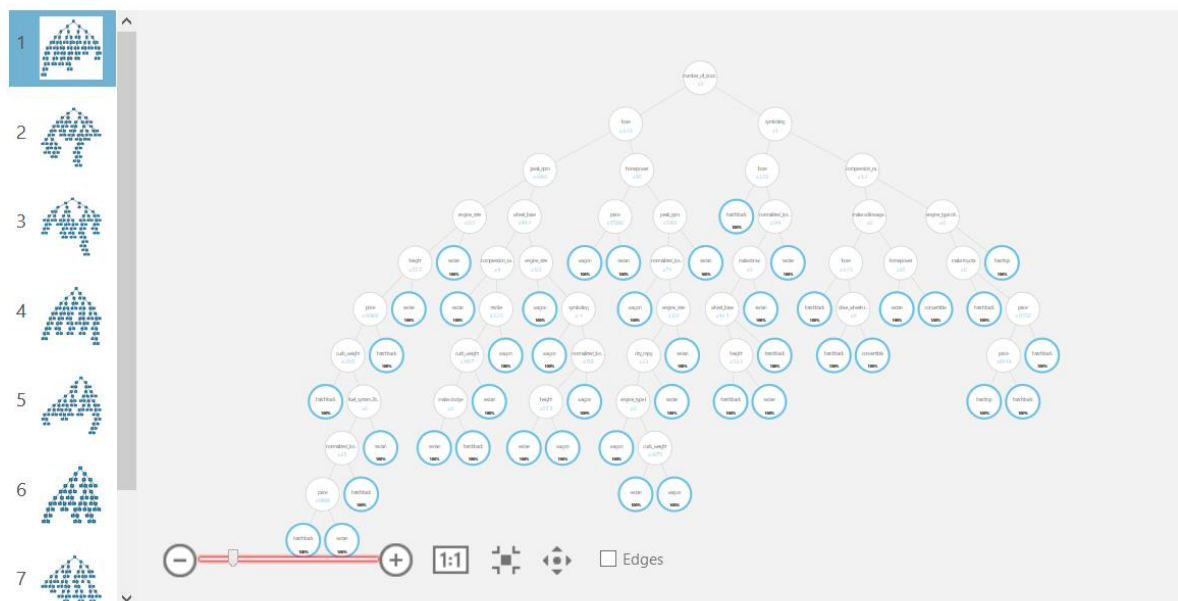


Рис 6. Построенные деревья решений





Scored Probabilities for Class "convertible"	Scored Probabilities for Class "hardtop"	Scored Probabilities for Class "hatchback"	Scored Probabilities for Class "sedan"	Scored Probabilities for Class "wagon"	Scored Labels
					
0	0	1	0	0	hatchback
0	0	1	0	0	hatchback
0	0	0	0.875	0.125	sedan
0	0	0	1	0	sedan
0.25	0.125	0.375	0.25	0	hatchback
0	0	1	0	0	hatchback
0	0.125	0.75	0.125	0	hatchback
0	0	0.25	0.75	0	sedan
0	0	0	0.75	0.25	sedan

Рис 7. Результат Scored Model

В результате мы получаем оцененные вероятности каждого класса для всех позиций и класс, который наиболее вероятен.

desforest body prediction ▶ Evaluate Model ▶ Evaluation results

Confusion Matrix

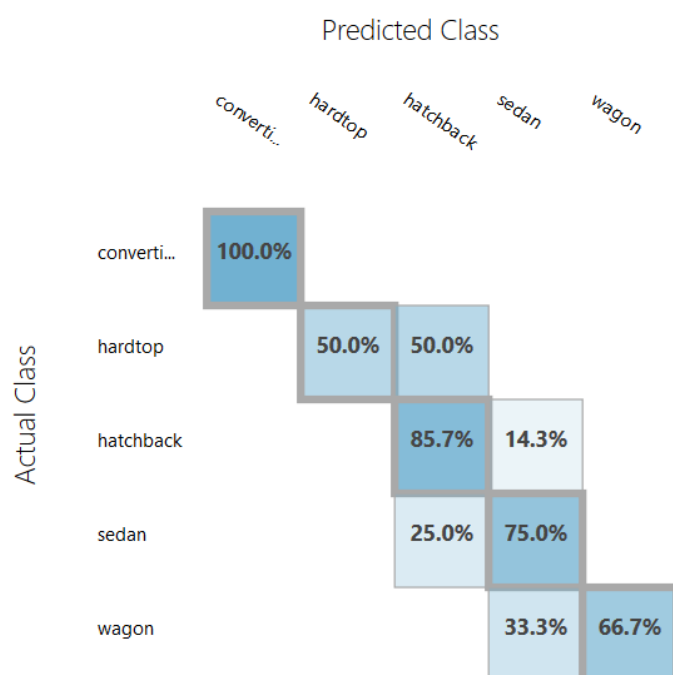


Рис 8. Матрица ошибок

Metrics

Overall accuracy	0.775
Average accuracy	0.91
Micro-averaged precision	0.775
Macro-averaged precision	0.9
Micro-averaged recall	0.775
Macro-averaged recall	0.754762

Рис 9. Метрики

Общая точность, средняя точность, микро- и макро- усредненная точность и полнота.

Задача 3

Классификация по количеству дверей (2 или 4) по всем параметрам.

Был выбран *метод опорных векторов* для бинарной классификации, он довольно широко используется и относится к обучению с учителем.

Из датасета выделила 70 % для обучающей выборки.

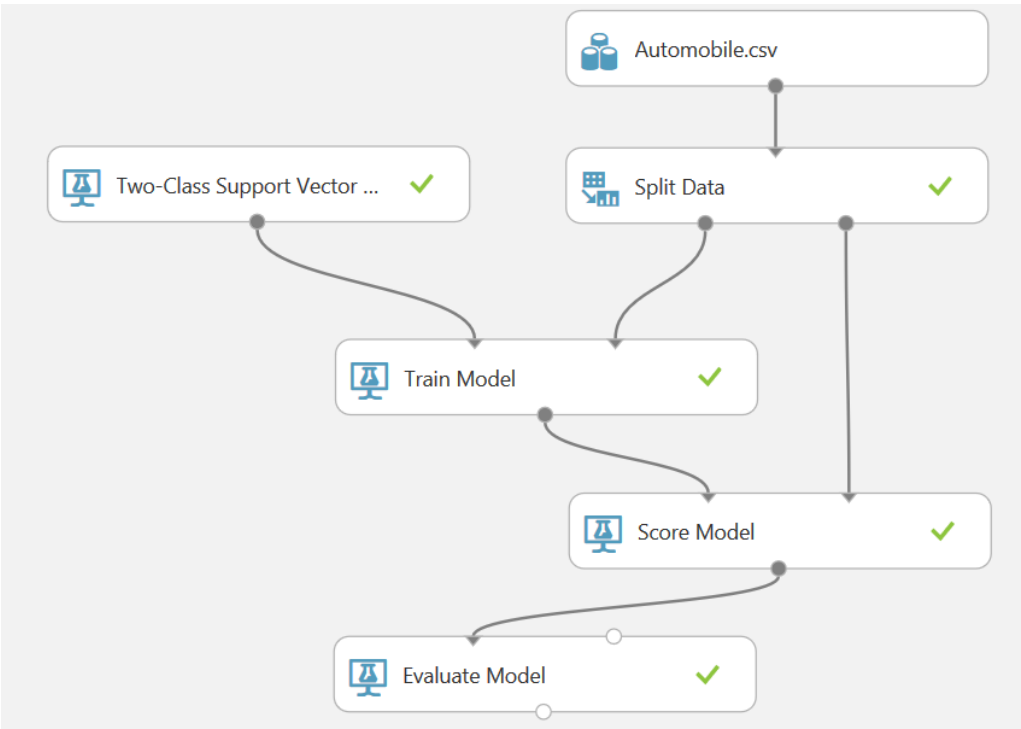


Рис 10. Модель эксперимента










rows 60	columns 28								
stroke	compression_ratio	horsepower	peak_rpm	city_mpg	highway_mpg	price	Scored Labels	Scored Probabilities	
									
3.29	9.4	69	5200	31	37	7499	four	0.11311	
3.23	9.4	68	5500	31	38	7609	four	0.348712	
3.54	8.7	92	4200	29	34	8948	four	0.017442	
3.47	21.9	55	4800	45	50	7099	four	0.013165	
3.1	8.3	155	4750	16	18	34184	four	0.142905	
3.39	7.6	102	5500	24	30	7689	two	0.925762	
3.23	9.4	68	5500	31	38	6229	two	0.936669	
3.39	8.6	84	4800	26	32	10245	four	0.066718	
3.15	9	68	5000	31	38	6095	two	0.874707	

Рис 11. Результат обучения

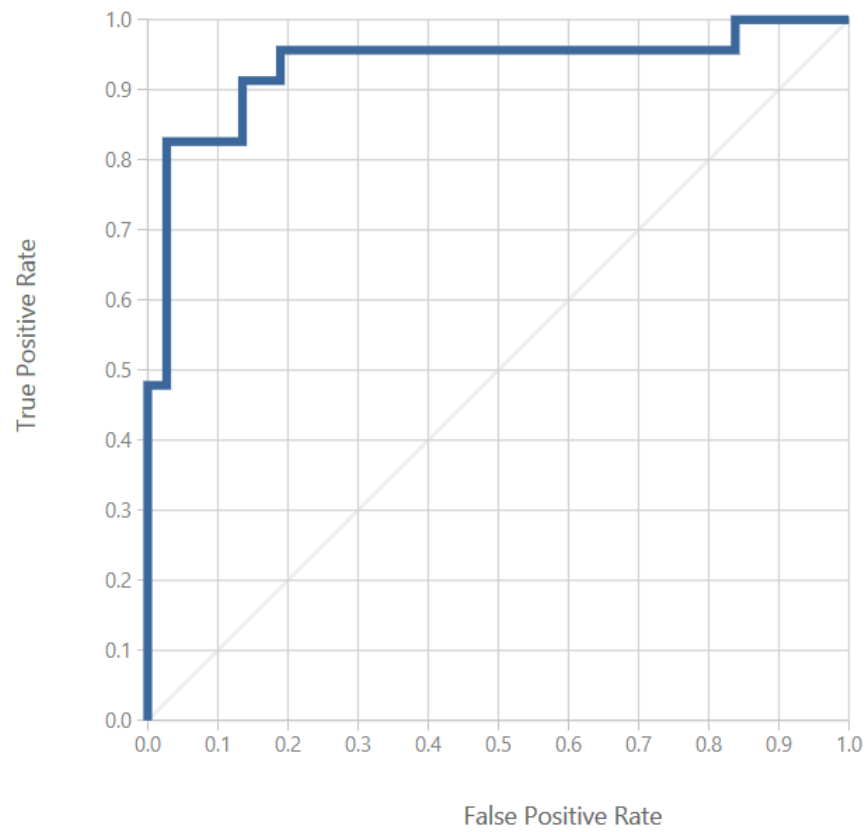


Рис 12. ROC-кривая

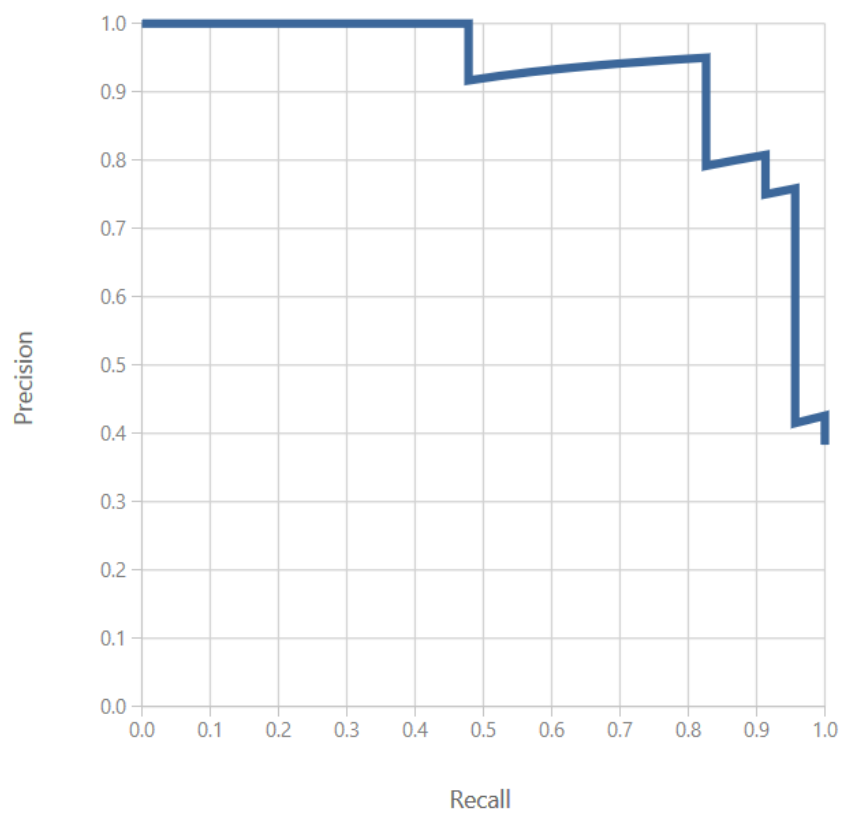


Рис 13. Кривая Precision/Recall

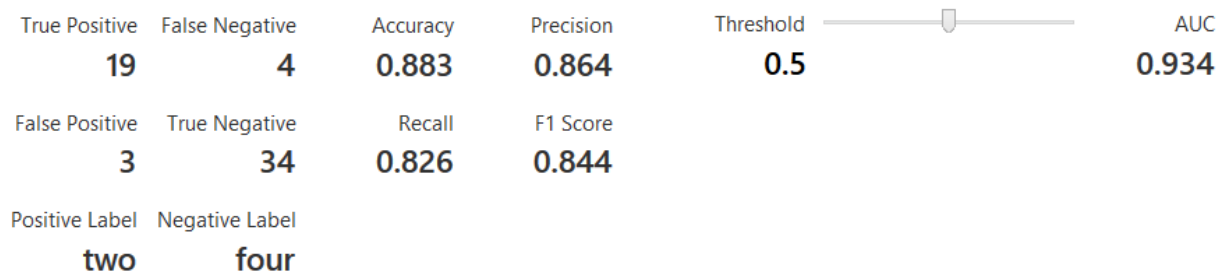


Рис 14. Метрики

Матрица ошибок, доля правильных ответов, точность, полнота, F-мера.

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	14	1	0.250	0.833	0.737	0.933	0.609	0.800	0.973	0.013
(0.800,0.900]	5	0	0.333	0.917	0.884	0.950	0.826	0.900	0.973	0.013
(0.700,0.800]	0	1	0.350	0.900	0.864	0.905	0.826	0.897	0.946	0.035
(0.600,0.700]	0	1	0.367	0.883	0.844	0.864	0.826	0.895	0.919	0.058
(0.500,0.600]	0	0	0.367	0.883	0.844	0.864	0.826	0.895	0.919	0.058
(0.400,0.500]	0	0	0.367	0.883	0.844	0.864	0.826	0.895	0.919	0.058
(0.300,0.400]	2	2	0.433	0.883	0.857	0.808	0.913	0.941	0.865	0.102
(0.200,0.300]	1	3	0.500	0.850	0.830	0.733	0.957	0.967	0.784	0.177
(0.100,0.200]	0	9	0.650	0.700	0.710	0.564	0.957	0.952	0.541	0.410
(0.000,0.100]	1	20	1.000	0.383	0.554	0.383	1.000	1.000	0.000	0.934

Рис 15. Статистика для интервалов вероятности

Выводы

Прежде чем самим писать алгоритмы машинного обучения, важно посмотреть, как они работают на готовых реализациях и что необходимо вывести для анализа точности после обработки и обучения. В Azure ML много различных алгоритмов и методов обучения, с которыми удобно и просто работать.