

# Лабораторная работа 1

Студент: Введенская П.К.

Группа: 80-304Б

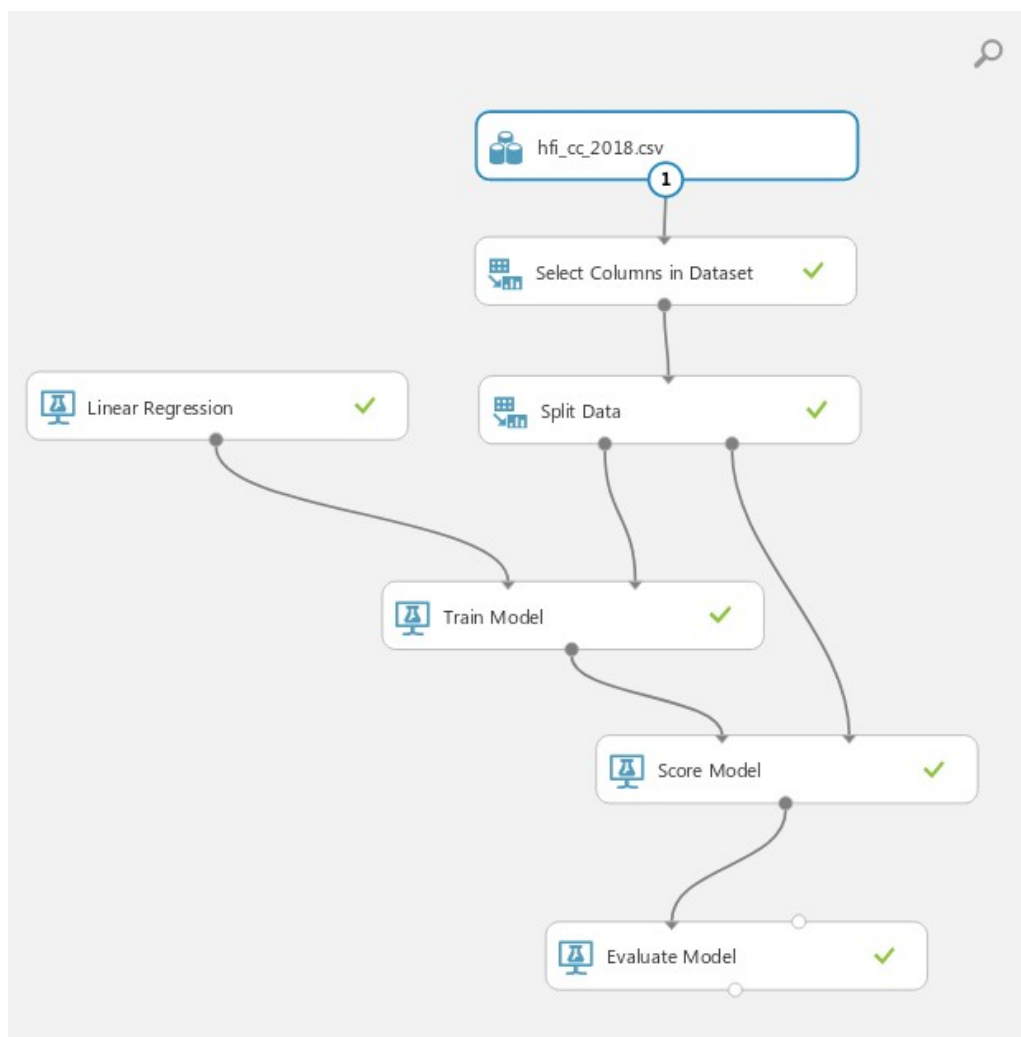
## Постановка задачи

Познакомиться с платформой Azure Machine Learning, реализовав полный цикл разработки решения задачи машинного обучения, используя три различных алгоритма, реализованные на этой платформе.



## Задание 1.

Зависимость индекса человеческой свободы от индексов свободы слова (pf\_expression\_killed, pf\_expression\_jailed, pf\_expression\_influence, pf\_expression\_control, pf\_expression\_cable, pf\_expression\_newspapers, pf\_expression\_internet)

Использован алгоритм множественной линейной регрессии. Из датасета выделено 70% для обучающей выборки.

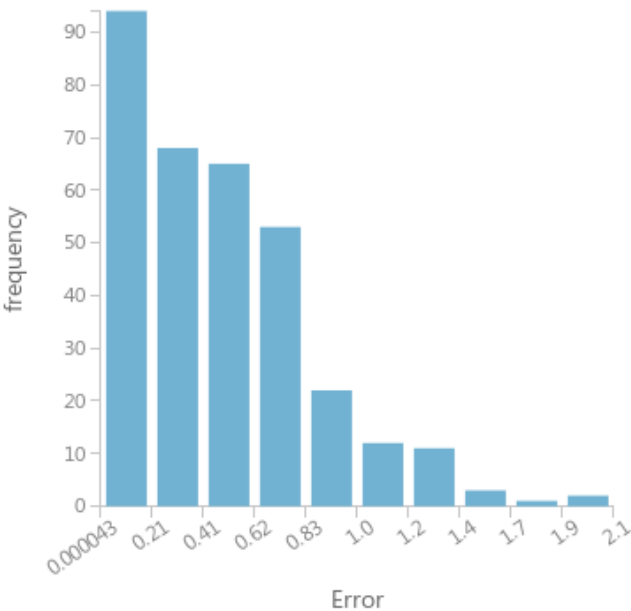


Полученные оценки (Score Model):

hf_score	Scored Labels	5.467683	
		6.324275	5.5001
		5.131931	
6.608429		7.249368	7.940259
8.247249	7.765416	5.880785	6.695708
		7.863733	7.950982
		8.87689	8.09272
8.509303	8.354282		
6.760757	6.849714	6.183304	6.36133
6.190287	6.498356	5.895044	6.220446
5.671812	6.516513	5.693494	
7.14045	7.560271	6.513792	6.467976
8.565863	8.147271	6.422788	5.939383
6.585666		...	
7.73698	7.950982		

Средняя абсолютная ошибка, средняя квадратическая ошибка, относительная абсолютная ошибка, относительная квадратическая ошибка, коэффициент корреляции (Evaluate Model):

Error Histogram



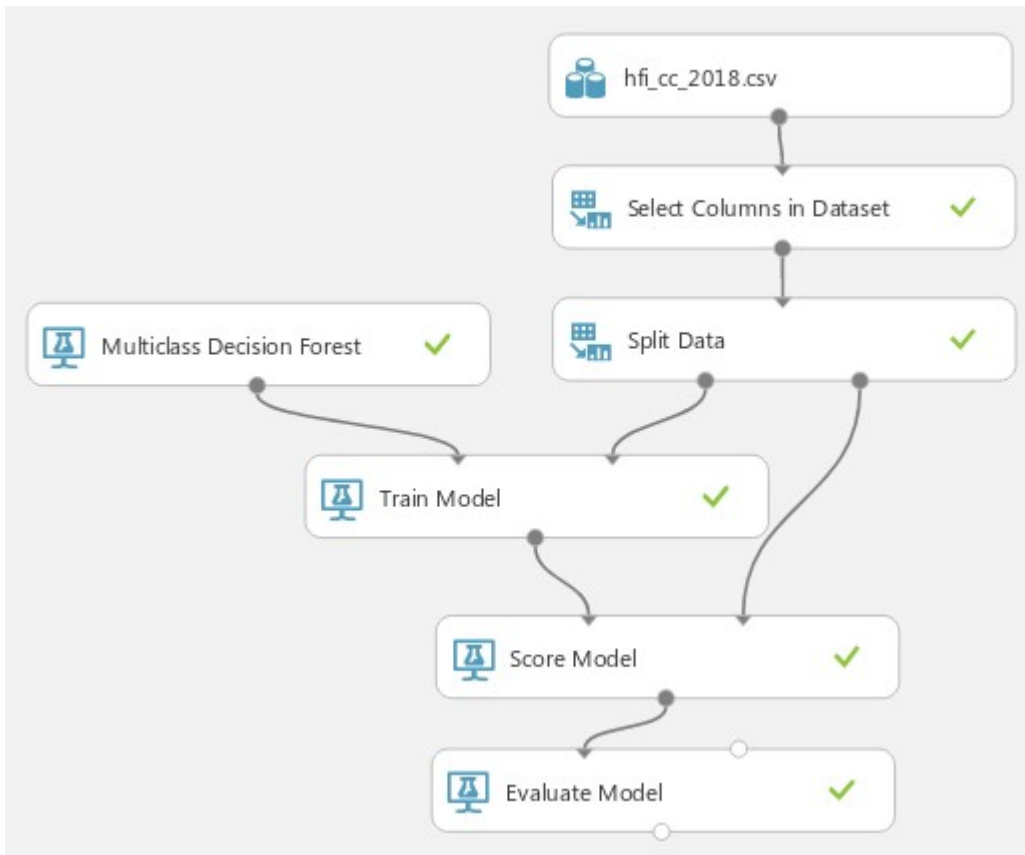
Metrics

Mean Absolute Error	0.481628
Root Mean Squared Error	0.607559
Relative Absolute Error	0.539952
Relative Squared Error	0.330961
Coefficient of Determination	0.669039

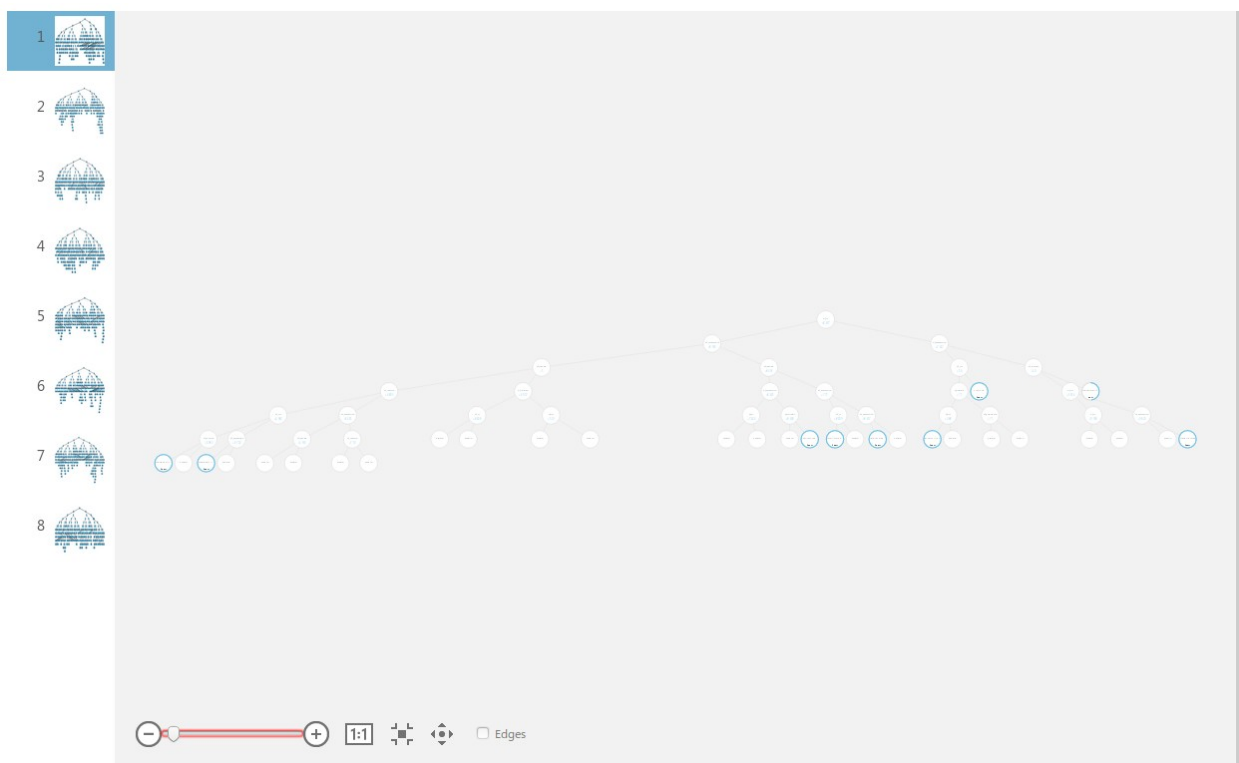
## Задание 2.

Мультиклассовая классификация региона по индексам индивидуальной свободы.

Используется лес деревьев решений. Из датасета выделено 70% для обучающей выборки.



Деревья решений:

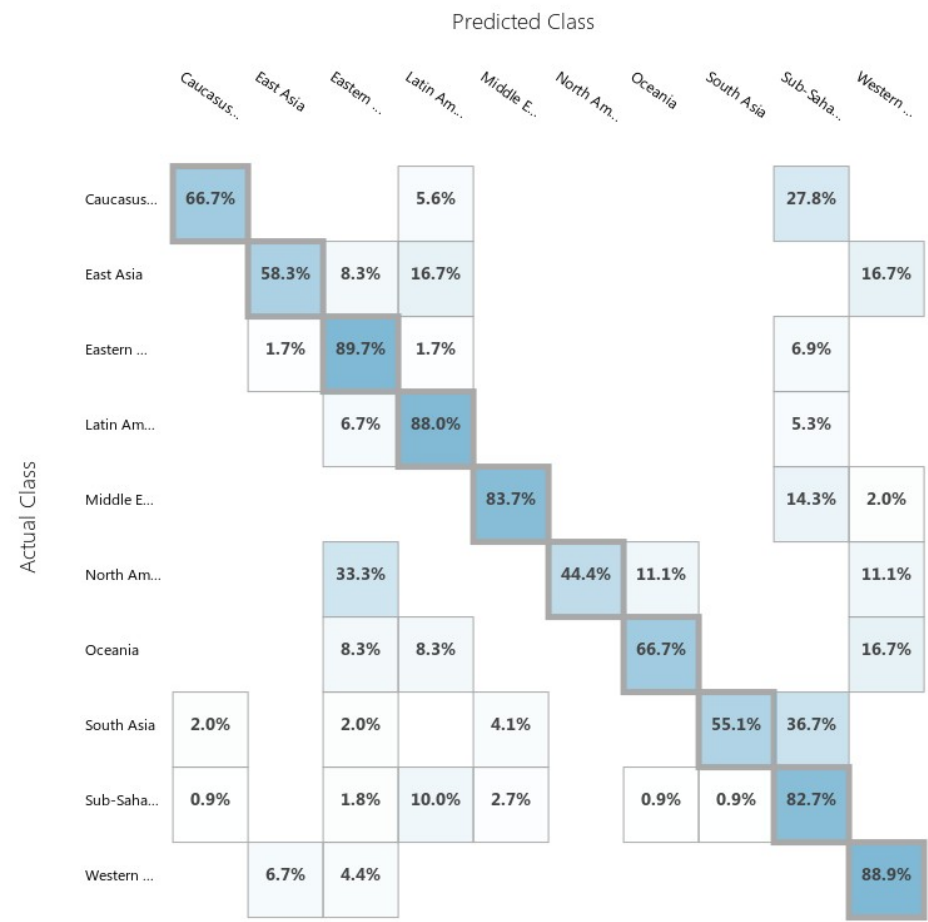


Вероятности каждого класса для каждой позиции (Score Model):

rows437columns19

Scored Probabilities for Class "Caucasus & Central Asia"	Scored Probabilities for Class "East Asia"	Scored Probabilities for Class "Eastern Europe"	Scored Probabilities for Class "Latin America & the Caribbean"	Scored Probabilities for Class "Middle East & North Africa"	Scored Probabilities for Class "North America"	Scored Probabilities for Class "Oceania"	Scored Probabilities for Class "South Asia"	Scored Probabilities for Class "Sub-Saharan Africa"	Scored Probabilities for Class "Western Europe"	Scored Labels
0	0	0.375	0	0	0	0	0	0.625	0	Sub-Saharan Africa
0	0.375	0.125	0	0	0.25	0	0	0	0.25	East Asia
0	0	0.09049	0.049781	0.274304	0	0	0.175377	0.410048	0	Sub-Saharan Africa
0	0	0	0	0	0	0	0	0	1	Western Europe
0	0	0	0.625	0	0	0	0.125	0.25	0	Latin America & the Caribbean
0	0	0	0	1	0	0	0	0	0	Middle East & North Africa
0	0	0	0	0	0	0	0	1	0	Sub-Saharan Africa
0	0	0	0.25	0	0	0	0.125	0.625	0	Sub-Saharan Africa

Матрица ошибок (Evaluate Model):



Метрики (общая точность, средняя точность, микро-усредненная точность, макро-усредненная точность, микро-усредненная полнота, макро-усредненная полнота):

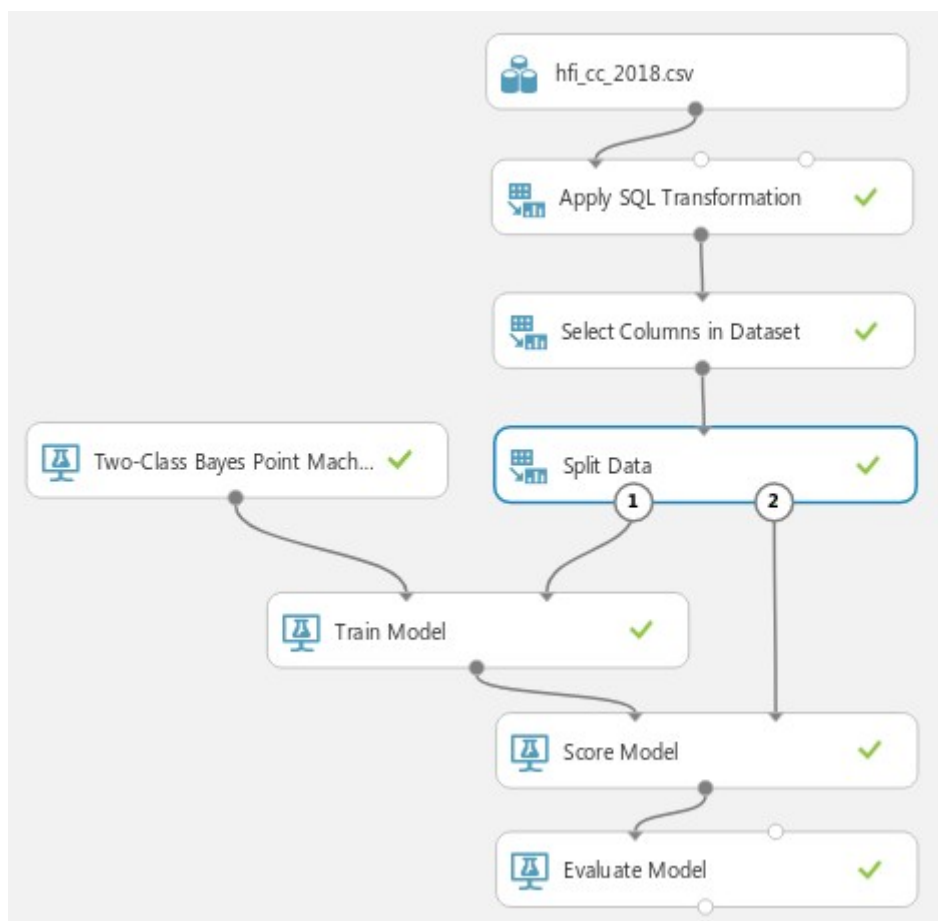
#### Metrics

Overall accuracy	0.796339
Average accuracy	0.959268
Micro-averaged precision	0.796339
Macro-averaged precision	0.830509
Micro-averaged recall	0.796339
Macro-averaged recall	0.724158

### Задание 3.

Классификация 2008 и 2016 года по индексам индивидуальной и экономической свободы.

Используется двухклассовая точечная машина Байеса. Из датасета выделено 80% для обучающей выборки. Обучение проходит 50 итераций.



Результат обучения:











Feature Weights

Feature	Weight Mean	Weight Standard Deviation
Bias	0.0878505	0.0926736
ef_money	0.078264	0.0113902
pf_rol	-0.0768042	0.0163224
pf_religion	-0.0680441	0.0120068
ef_regulation	0.0571669	0.0132482
ef_legal	-0.0239291	0.0163258
ef_government	-0.0148768	0.0141759
pf_identity	-0.0126743	0.0116609
pf_association	0.012066	0.0120561
ef_trade	0.00957802	0.0128353
pf_movement	0.00441328	0.0110458
pf_ss	0.00357042	0.010887
pf_expression	0.00268715	0.0114739

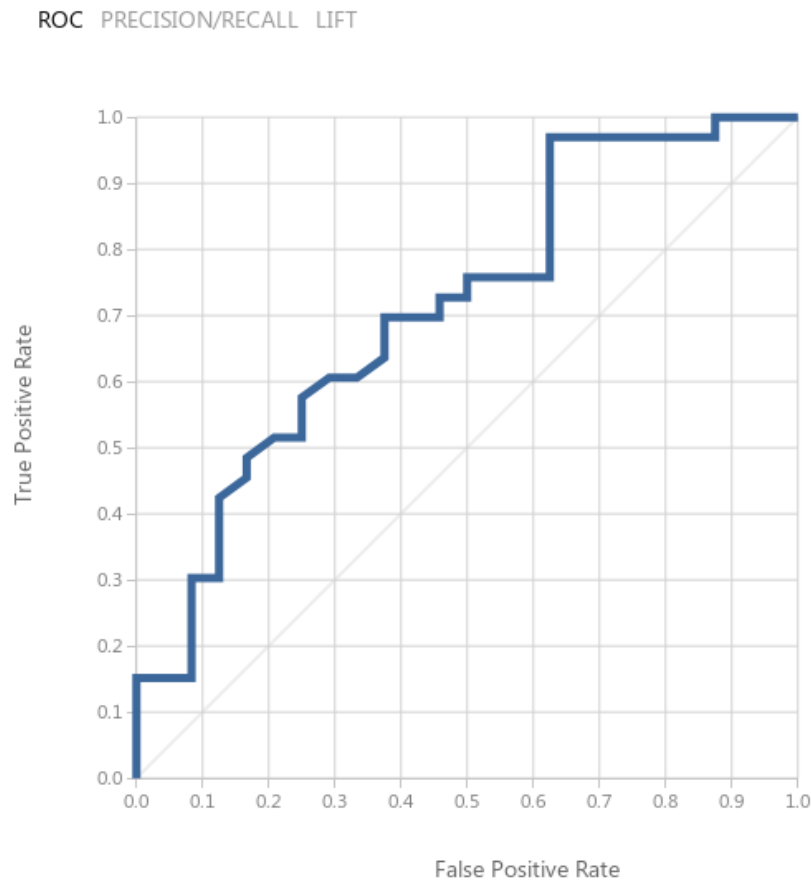
Score Model ():

rows  
65

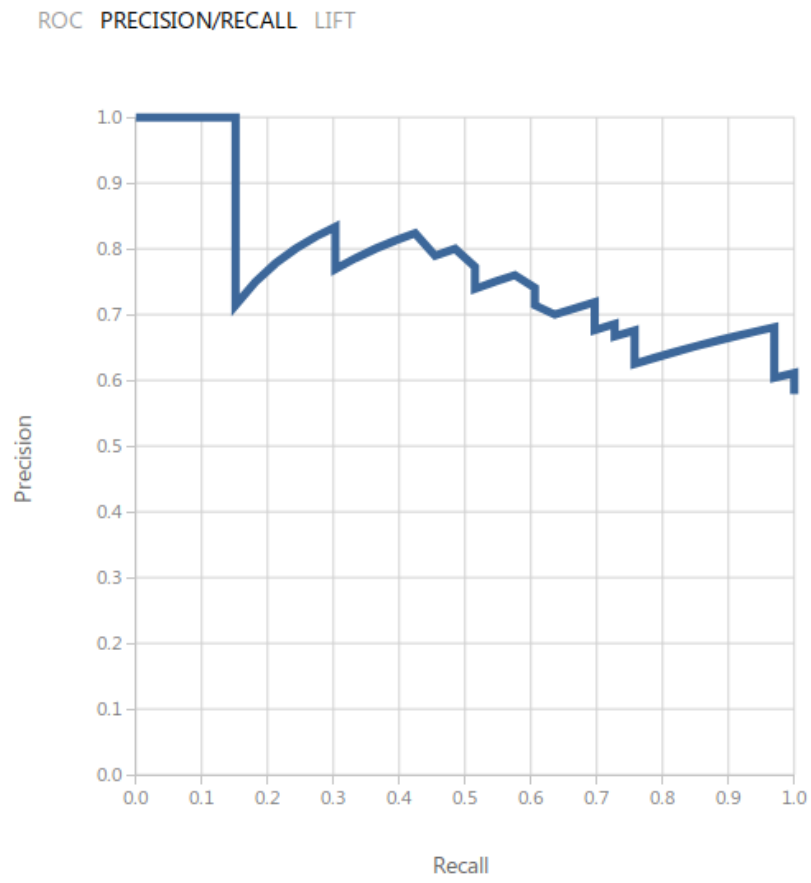
columns  
15

pf_association	pf_expression	pf_identity	ef_government	ef_legal	ef_money	ef_trade	ef_regulation	Scored Labels	Scored Probabilities
									
8.5	6.952381	9.166667	5.833377	3.244861	5.603964	4.883225	5.551182	2008	0.454523
8	7.452381	7.5	8.429052	3.576272	9.808379	7.678363	6.243651	2016	0.579195
9.416667	9.035714	10	9.410247	8.275295	9.536101	9.56153	9.175875	2008	0.488334
10	9.654762	10	7.592385	8.504601	9.399279	7.32712	8.364623	2008	0.475386
7.833333	8.392857	7.5	7.449106	3.942016	7.548048	6.899523	6.136228	2016	0.594035
9.5	9.595238	6.875	7.609155	5.891155	9.779738	8.003562	6.449131	2016	0.542173
10	9.297619	10	7.31565	5.62389	8.447859	8.05502	6.16214	2008	0.455977
7.666667	8.845238	7.5	4.97451	4.614822	6.638917	5.948205	7.59019	2016	0.517682
9.666667	9.714286	10	4.197427	7.141006	9.479876	8.162656	7.512504	2016	0.522195
	6.708333	10	7.374523	6.144341	7.221813	7.094155	6.70426		
7.5	9.190476	9.25	4.535268	7.586987	9.524659	7.495839	8.197	2016	0.523747
4	7.464286	1.666667	6.237318	4.922445	8.342266	6.84041	7.430758	2016	0.650422
5.5	6.998576	5	7.111628	3.199561	7.222024	5.054771	6.530047	2016	0.570394
5.333333	7.962447	5	6.291508	2.502268	6.397586	6.305856	6.020162	2016	0.530679
10	9.666667	9.25	4.15108	8.798181	9.360551	8.232927	7.717924	2008	0.49095
10	9.071429	10	5.886963	6.031965	9.598398	7.624509	6.365504	2016	0.51641
	7.666667	10	6.326847	5.670494	9.001838	7.7597	7.389053		
8.333333	7.885088	5	7.768612	4.156155	7.682917	6.778815	7.451936	2016	0.566945
10	8.904762	9.25	6.043834	5.185801	8.081938	6.777461	7.168277	2008	0.49886

ROC:



Precision / Recall:



Метрики (матрица ошибок, доля правильных ответов, точность, полнота, F-мера) и статистика интервалов вероятности:

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
26	7	0.614	0.634	0.5	0.710
False Positive	True Negative	Recall	F1 Score		
15	9	0.788	0.703		
Positive Label	Negative Label				
2016	2008				

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	0.421	0.000	1.000	0.000	0.421	1.000	0.000
(0.800,0.900]	0	0	0.000	0.421	0.000	1.000	0.000	0.421	1.000	0.000
(0.700,0.800]	1	0	0.018	0.439	0.059	1.000	0.030	0.429	1.000	0.000
(0.600,0.700]	8	2	0.193	0.544	0.409	0.818	0.273	0.478	0.917	0.013
(0.500,0.600]	17	13	0.719	0.614	0.703	0.634	0.788	0.563	0.375	0.342
(0.400,0.500]	7	9	1.000	0.579	0.733	0.579	1.000	1.000	0.000	0.710
(0.300,0.400]	0	0	1.000	0.579	0.733	0.579	1.000	1.000	0.000	0.710
(0.200,0.300]	0	0	1.000	0.579	0.733	0.579	1.000	1.000	0.000	0.710
(0.100,0.200]	0	0	1.000	0.579	0.733	0.579	1.000	1.000	0.000	0.710
(0.000,0.100]	0	0	1.000	0.579	0.733	0.579	1.000	1.000	0.000	0.710

## Вывод

Даже пользователь, имеющий очень общее представление об методах машинного обучения, может наглядно реализовать один из них в Azure ML. Но достигнуть хотя бы приемлемой точности гораздо сложнее. Качественная реализация требует понимания особенностей каждого алгоритма и тщательной настройки.