











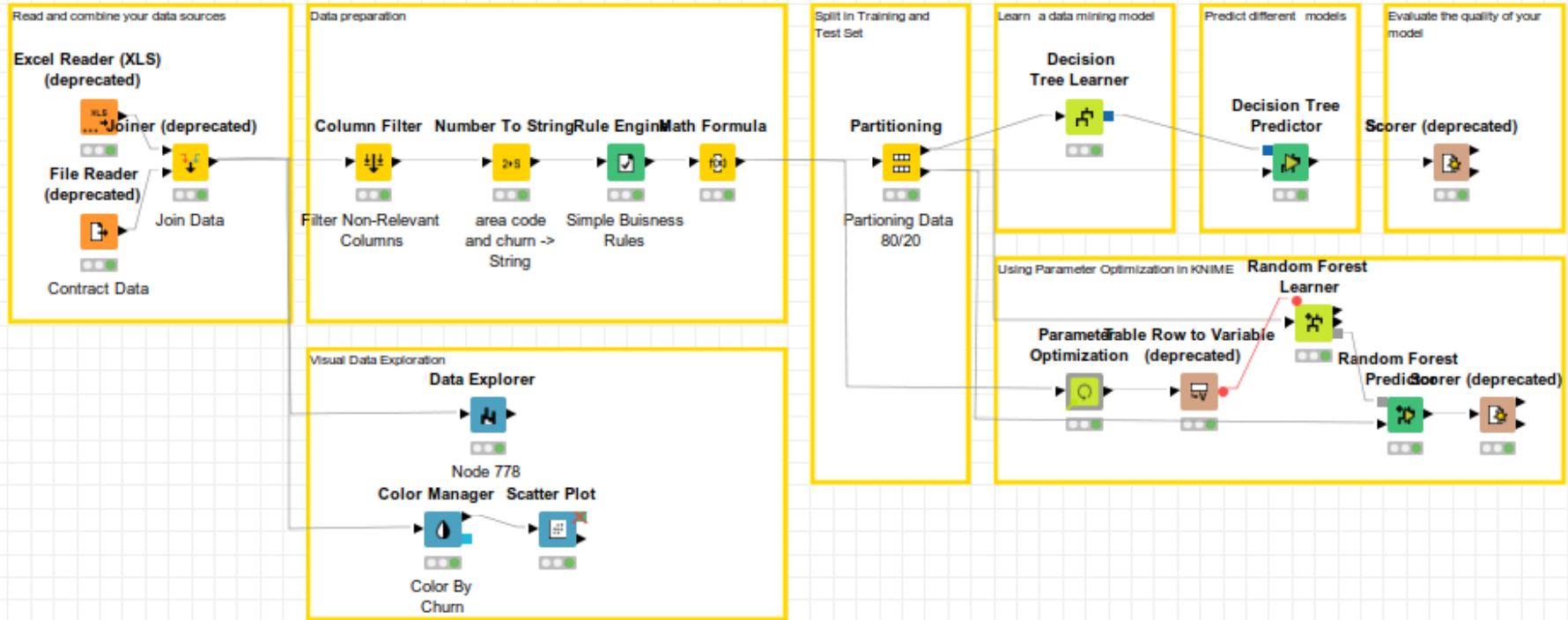
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«МИРЭА – Российский Технологический Университет»

Отчёт по практической работе №5:
«Реализация алгоритма для прогнозирования оттока
клиентов»

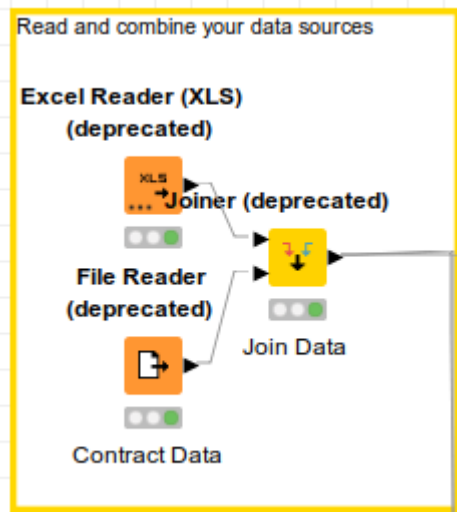
Подготовила студентка
группы ББМО-01-21
Концевая Юлия

- ▼  EXAMPLES (knime@api.hub.knime.com)
 - ▶  00_Components
 - ▶  01_Data_Access
 - ▶  02_ETL_Data_Manipulation
 - ▶  03_Visualization
 - ▼  04_Analytics
 - ▶  01_Preprocessing
 - ▶  02_Statistics
 - ▶  03_Clustering
 - ▼  04_Classification_and_Predictive_Modelling
 - ▲ 01_Example_for_Learning_a_Decision_Tree
 - ▲ 02_Example_for_Learning_a_Neural_Network
 - ▲ 03_Example_for_Learning_a_Naive_Bayes_Model
 - ▲ 04_Exporting_a_Decision_Tree_as_Image
 - ▲ 05_Gradient_Boosted_Trees
 - ▲ 06_Logistic_Regression
 - ▲ 07_Decision_Tree
 - ▲ 08_Regularized_Logistic_Regression
 - ▲ 09_Random_Forest
 - ▲ 10_Analyzing_Churn_Models_with_the_Binary_Classification_Inspe
 - ▲ 11_DataScience_with_KNIME_an_Introduction

DataScience_with_KNIME_an_Introduction



Первый набор данных



Preview

File Content

Preview with current settings: CallsData.xls [default]

refresh

Row ID	I VMail ...	D Day Mi...	D Eve Mi...	D Night ...	D Intl Mins	I CustS...	I Day C...	D Day C...	I Eve C...	D Eve C...	I Night ...	D Night ...	I Intl Calls	D Intl Ch...	I Area C...	S Phone
Row0	25	265.1	197.4	244.7	10	1	110	45.07	99	16.78	91	11.01	3	2.7	415	382-4657
Row1	26	161.6	195.5	254.4	13.7	1	123	27.47	103	16.62	103	11.45	3	3.7	415	371-7191
Row2	0	243.4	121.2	162.6	12.2	0	114	41.38	110	10.3	104	7.32	5	3.29	415	358-1921
Row3	0	299.4	61.9	196.9	6.6	2	71	50.9	88	5.26	89	8.86	7	1.78	408	375-9999
Row4	0	166.7	148.3	186.9	10.1	3	113	28.34	122	12.61	121	8.41	3	2.73	415	330-6626
Row5	0	223.4	220.6	203.9	6.3	0	98	37.98	101	18.75	118	9.18	6	1.7	510	391-8027
Row6	24	218.2	348.5	212.6	7.5	3	88	37.09	108	29.62	118	9.57	7	2.03	510	355-9993
Row7	0	157	103.1	211.8	7.1	0	79	26.69	94	8.76	96	9.53	6	1.92	415	329-9001
Row8	0	184.5	351.6	215.8	8.7	1	97	31.37	80	29.89	90	9.71	4	2.35	408	335-4719
Row9	37	258.6	222	326.4	11.2	0	84	43.96	111	18.87	97	14.69	5	3.02	415	330-8173
Row10	0	129.1	228.5	208.8	12.7	4	137	21.95	83	19.42	111	9.4	6	3.43	415	329-6603
Row11	0	187.7	163.4	196	9.1	0	127	31.91	148	13.89	94	8.82	5	2.46	415	344-9403
Row12	0	128.8	104.9	141.1	11.2	1	96	21.9	71	8.92	128	6.35	2	3.02	408	363-1107
Row13	0	156.6	247.6	192.3	12.3	3	88	26.62	75	21.05	115	8.65	5	3.32	510	394-8006
Row14	0	120.7	307.2	203	13.1	4	70	20.52	76	26.11	99	9.14	6	3.54	415	366-9238
Row15	0	332.9	317.8	160.6	5.4	4	67	56.59	97	27.01	128	7.23	9	1.46	415	351-7269
Row16	27	196.4	280.9	89.3	13.8	1	139	33.39	90	23.88	75	4.02	4	3.73	408	350-8884
Row17	0	190.7	218.2	129.6	8.1	3	114	32.42	111	18.55	121	5.83	3	2.19	510	386-2923
Row18	33	189.7	212.8	165.7	10	1	66	32.25	65	18.09	108	7.46	5	2.7	510	356-2992
Row19	0	224.4	159.5	192.8	13	1	90	38.15	88	13.56	74	8.68	2	3.51	415	373-2782
Row20	0	155.1	239.7	208.8	10.6	0	117	26.37	93	20.37	133	9.4	4	2.86	415	396-5800
Row21	0	62.4	169.9	209.6	5.7	5	89	10.61	121	14.44	64	9.43	6	1.54	408	393-7984
Row22	0	183	72.9	181.8	9.5	0	112	31.11	99	6.2	78	8.18	19	2.57	415	358-1958
Row23	0	110.4	137.3	189.6	7.7	2	103	18.77	102	11.67	105	8.53	6	2.08	415	350-2565
Row24	0	181.1	348.8	212.6	7.8	0	86	33.36	112	28.84	115	13.67	5	3.36	415	345-1666

Второй набор данных

Click colu							
Row ID	<input type="checkbox"/> Accou...	<input type="checkbox"/> Churn	<input type="checkbox"/> Int'l Plan	<input type="checkbox"/> VMail ...	<input checked="" type="checkbox"/> State	<input type="checkbox"/> Area C...	<input checked="" type="checkbox"/> Phone
Row0	128	0	0	1	KS	415	382-4657
Row1	107	0	0	1	OH	415	371-7191
Row2	137	0	0	0	NJ	415	358-1921
Row3	84	0	1	0	OH	408	375-9999
Row4	75	0	1	0	OK	415	330-6626
Row5	118	0	1	0	AL	510	391-8027
Row6	121	0	0	1	MA	510	355-9993
Row7	147	0	1	0	MO	415	329-9001
Row8	117	0	0	0	LA	408	335-4719
Row9	141	0	1	1	WV	415	330-8173
Row10	65	1	0	0	IN	415	329-6603
Row11	74	0	0	0	RI	415	344-9403
Row12	168	0	0	0	IA	408	363-1107
Row13	95	0	0	0	MT	510	394-8006
Row14	62	0	0	0	IA	415	366-9238
Row15	161	1	0	0	NY	415	351-7269
Row16	85	0	0	1	ID	408	350-8884
Row17	93	0	0	0	VT	510	386-2923
Row18	76	0	0	1	VA	510	356-2992
Row19	73	0	0	0	TX	415	373-2782
Row20	147	0	0	0	FL	415	396-5800
Row21	77	1	0	0	CO	408	393-7984
Row22	130	0	0	0	AZ	415	358-1958
Row23	111	0	0	0	SC	415	350-2565
Row24	122	0	0	0	VA	510	342-4606

Объединение данных

ile

Joiner Settings | Column Selection | Flow Variables | Memory Policy

Join Mode

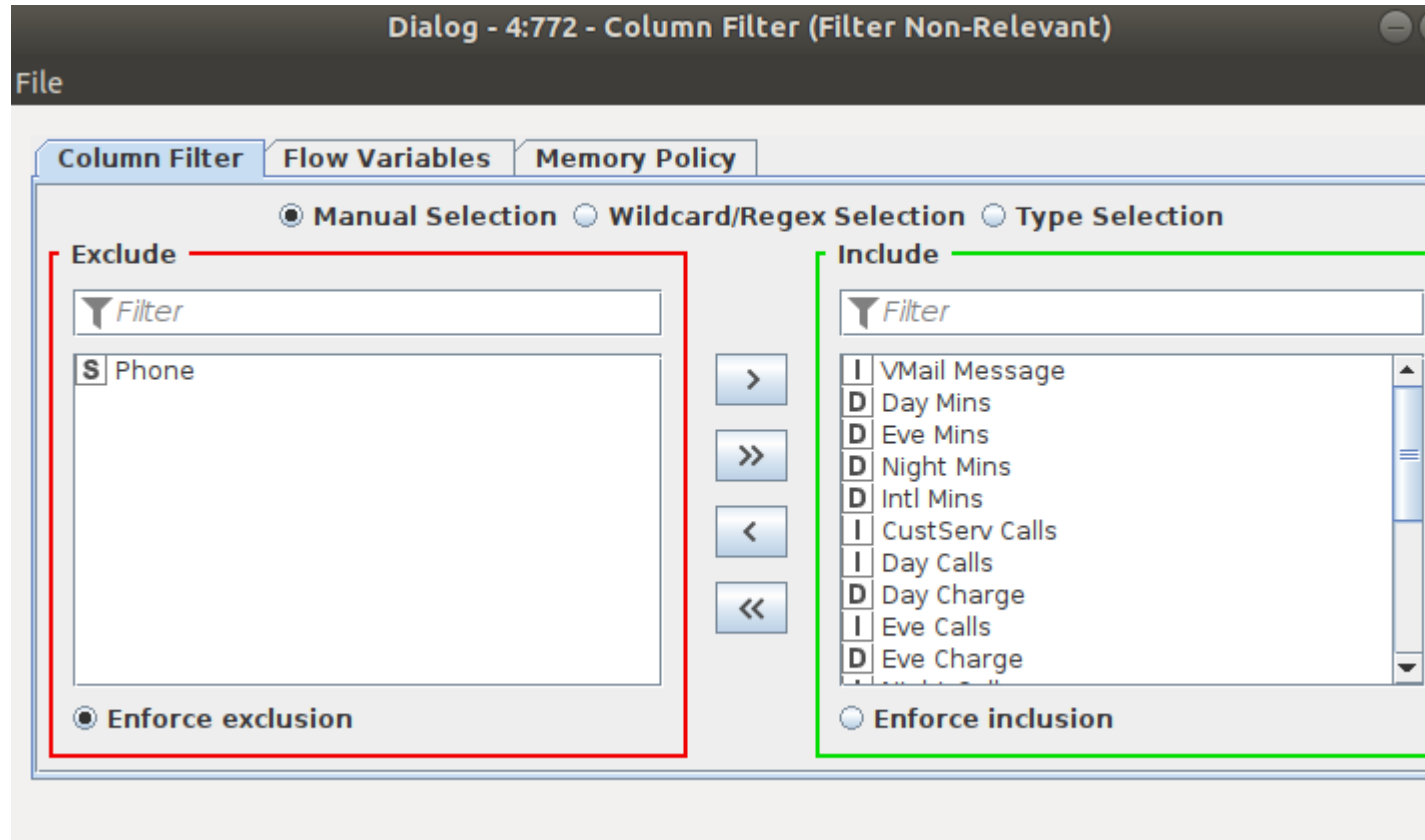
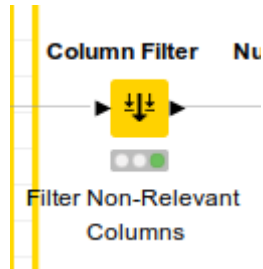
Join mode: Inner Join ▼

Joining Columns

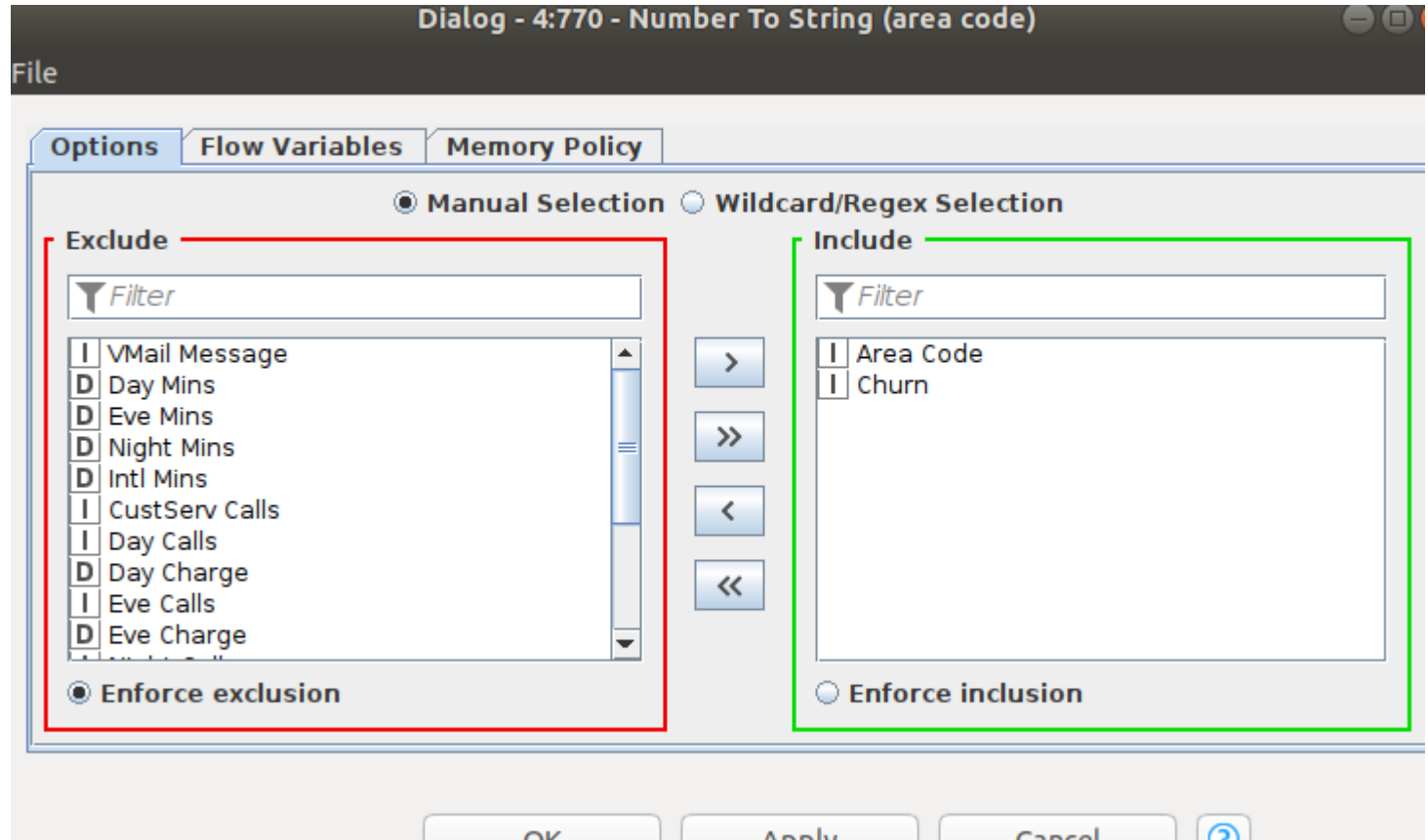
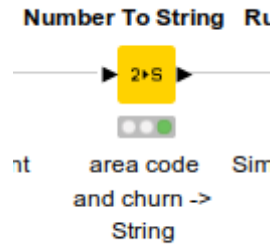
☒ Match all of the following ☐ Match any of the following

Top Input ('left' table)	Bottom Input ('right' table)		
I Area Code ▼	I Area Code ▼	+	-
S Phone ▼	S Phone ▼	+	-
		+	

Фильтрация ненужных значений

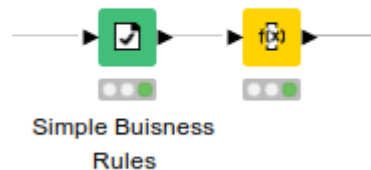


Преобразование типов данных



Предварительная обработка данных

g Rule EngineMath Formula



Expression

```

1 // enter ordered set of rules, e.g.:
2 // $double column name$ > 5.0 => "large"
3 // $string column name$ LIKE "*blue*" => "small and blue"
4 // TRUE => "default outcome"
5 $Eve Calls$ < 0.5 => "early bird"
6 TRUE => "night owl"
    
```

Append Column: TimeClassColumn

Expression

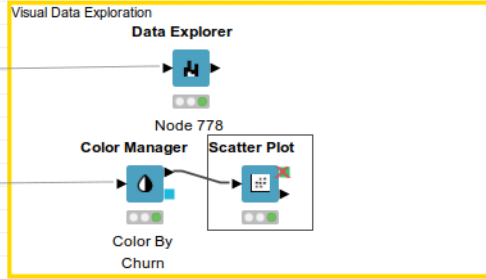
```

1 $Day Mins$+$Eve Mins$+$Night Mins$
    
```

Append Column: TotalNrOfMinutes

J...	S	Churn	I	Int'l Plan	I	S	State	D	TotalNrOfMinutes	S	TimeClassColumn
0	0	0	1	KS	707.2				707.2		night owl
0	0	0	1	OH	611.5				611.5		night owl
0	0	0	0	NJ	527.2				527.2		night owl
0	1	0	0	OH	558.2				558.2		night owl
0	1	0	0	OK	501.9				501.9		night owl
0	1	0	0	AL	647.9				647.9		night owl
0	0	0	1	MA	779.3				779.3		night owl
0	1	0	0	MO	471.9				471.9		night owl
0	0	0	0	LA	751.9				751.9		night owl
0	1	1	1	WV	807				807		night owl
1	0	0	0	IN	566.4				566.4		night owl
0	0	0	0	RI	547.1				547.1		night owl
0	0	0	0	IA	374.8				374.8		night owl
0	0	0	0	MT	596.5				596.5		night owl
0	0	0	0	IA	630.9				630.9		night owl
1	0	0	0	NY	811.3				811.3		night owl
0	0	0	1	ID	566.6				566.6		night owl
0	0	0	0	VT	538.5				538.5		night owl
0	0	0	1	VA	568.2				568.2		night owl
0	0	0	0	TX	576.7				576.7		night owl
0	0	0	0	FL	603.6				603.6		night owl
1	0	0	0	CO	441.9				441.9		night owl
0	0	0	0	AZ	437.7				437.7		night owl
0	0	0	0	SC	437.3				437.3		night owl
0	0	0	0	VA	563.3				563.3		night owl
0	0	0	0	NE	652.1				652.1		night owl
0	0	0	1	WY	586.8				586.8		night owl
0	0	0	0	MT	391.9				391.9		night owl
0	0	0	0	MO	629.7				629.7		night owl
0	0	0	0	HI	513.1				513.1		night owl
0	0	0	0	IL	472				472		night owl
0	0	0	0	NH	673.8				673.8		night owl
0	0	0	0	LA	536.5				536.5		night owl
1	0	0	0	AZ	782.2				782.2		night owl

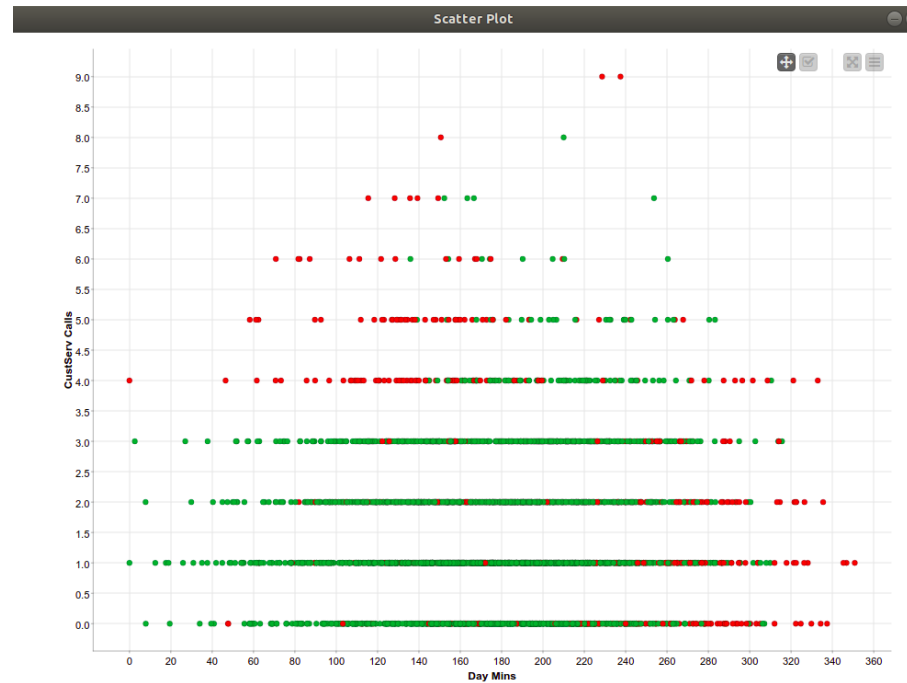
Предварительная оценка данных



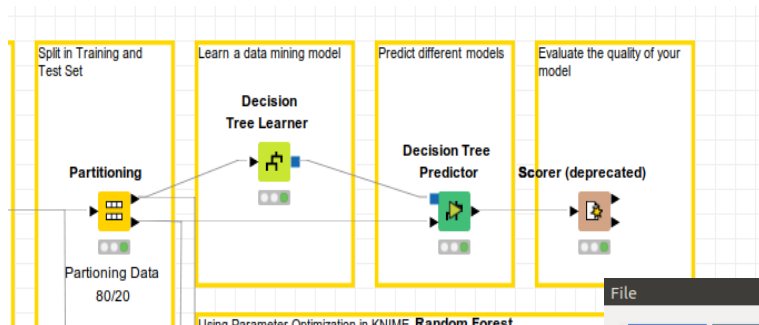
Numeric Nominal Data Preview

Search:

Column	Exclude Column	Minimum	Maximum	Mean	Standard Deviation	Variance	Skewness	Kurtosis
VMail Message	<input type="checkbox"/>	0	51	8.099	13.688	187.371	1.265	-0.051
Day Mins	<input type="checkbox"/>	0	350.800	179.775	54.467	2966.696	-0.029	-0.020
Eve Mins	<input type="checkbox"/>	0	363.700	200.980	50.714	2571.894	-0.024	0.026
Night Mins	<input type="checkbox"/>	23.200	395	200.872	50.574	2557.714	0.009	0.086
Intl Mins	<input type="checkbox"/>	0	20	10.237	2.792	7.794	-0.245	0.609
CustServ Calls	<input type="checkbox"/>	0	9	1.563	1.315	1.731	1.091	1.731
Day Calls	<input type="checkbox"/>	0	165	100.436	20.069	402.768	-0.112	0.243
Day Charge	<input type="checkbox"/>	0	59.640	30.562	9.259	85.737	-0.029	-0.020
Eve Calls	<input type="checkbox"/>	0	170	100.114	19.923	396.911	-0.056	0.206
Eve Charge	<input type="checkbox"/>	0	30.910	17.084	4.311	18.582	-0.024	0.025



Построение модели и оценка точности



File

Options PMMLSettings Flow Variables

General

Class column

Quality measure

Pruning method

☒ Reduced Error Pruning

Min number records per node

Number records to store for view

☒ Average split point

Number threads

☒ Skip nominal columns without domain information

Root split

☐ Force root split column

Root split column

Binary nominal splits

☐ Binary nominal splits

Max #nominal

☐ Filter invalid attribute values in child nodes

Confusion Matrix - 4:773 - Scorer (deprecated)

File Hilite

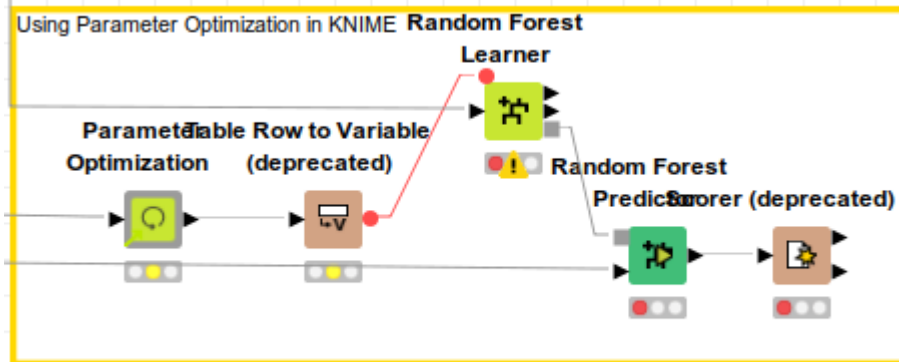
Churn \ P...	0	1
0	558	12
1	30	67

Correct classified: 625 Wrong classified: 42

Accuracy: 93,703 % Error: 6,297 %

Cohen's kappa (κ) 0,726

Оптимизация гиперпараметров



Confusion Matrix - 4:777 - Scorer (deprecated)

Churn \ P...	0	1
0	567	3
1	34	63

Correct classified: 630 Wrong classified: 37
 Accuracy: 94,453 % Error: 5,547 %
 Cohen's kappa (κ) 0,743

File Edit Hilite Navigation View			
Table "default" - Rows: 1		Spec - Columns: 3	Properties Flow Variables
Row ID	num_of_models	max_tree_depth	Accuracy
Best para...	11	50	0.939

Спасибо за внимание!