Отделили 250 строк

Partition and Sample

Partition or sample mode

Head ⌄

Number of rows to select ☰

250

Properties   Project

▲ Preprocess Text

Препроцессинг

Language

| English | ⌄ |

Remove by part of speech

| False | ⌄ |

Text column to clean

Selected columns:
Column names: review

Launch column selector

☑ Remove stop words   ≡

☑ Lemmatization   ≡

☑ Detect sentences   ≡

☑ Normalize case to low...   ≡

☑ Remove numbers   ≡

☑ Remove special chara...   ≡

☑ Remove duplicate cha...   ≡

☑ Remove email address...   ≡

☑ Remove URLs   ≡

☑ Expand verb contracti...   ≡

☑ Normalize backslashe...   ≡

☑ Split tokens on special... ≡

Custom regular expression   ≡

Custom replacement string   ≡

◢ Extract N-Gram Features fro...

Text column

Selected columns:
Column names:
Preprocessed review

[Launch column selector]

Vocabulary mode ☰

| Create ▾ |

N-Grams size ☰

| 1 |

K-Skip size ☰

| 0 |

Weighting function ☰

| Binary Weight ▾ |

Minimum word length ☰

| 3 |

Maximum word length ☰

| 25 |

Minimum n-gram doc... ☰

| 5 |

Maximum n-gram doc... ☰

| 1 |

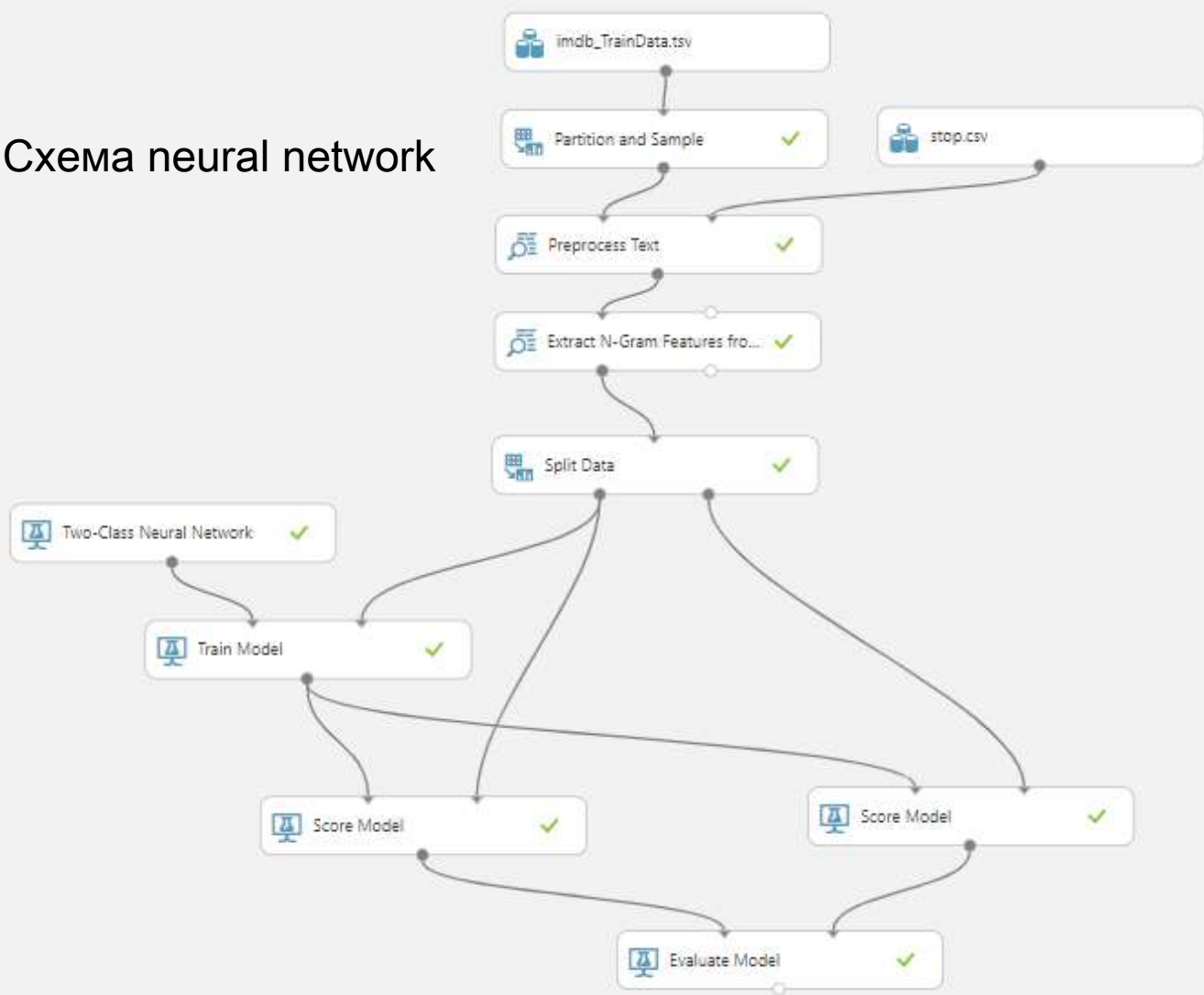☑ Detect out-of-voc... ☰

☐ Mark begin-of-sen... ☰

☑ Normalize n-gram... ☰

Use filter-based feature sel...

| False ▾ |

Векторная модель по униграммам

Схема neural network

◢ Two-Class Neural Network

Параметры neural network

Create trainer mode

Single Parameter ⌄

Hidden layer specification

Fully-connected case ⌄

Number of hidden nodes ☰

100

Learning rate ☰

0.001

Number of learning iterati... ☰

1000

The initial learning weight... ☰

0.1

The momentum ☰

0

The type of normalizer ☰
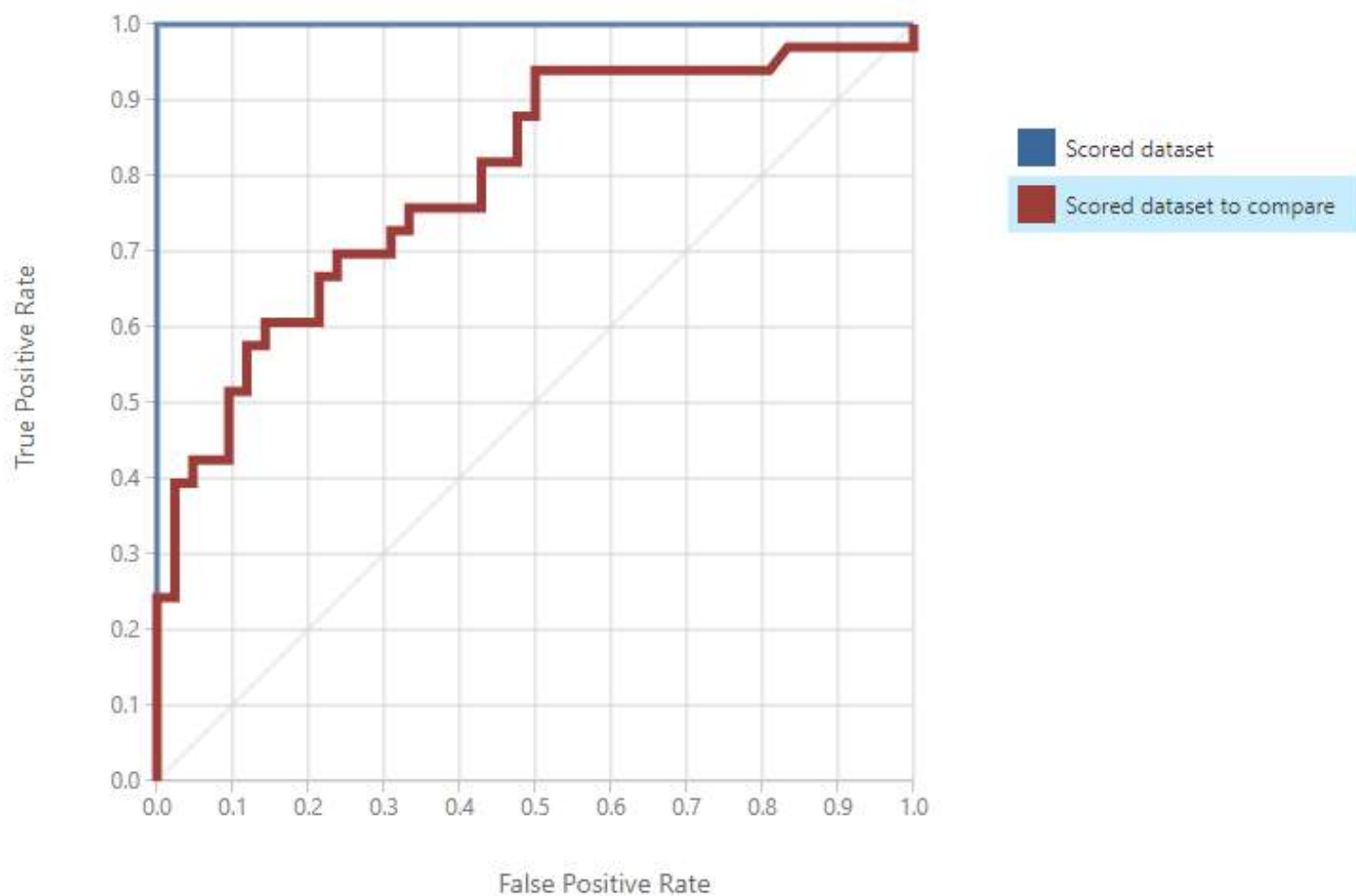
Min-Max normalizer ⌄

☑ Shuffle examples ☰

Random number seed ☰

1

☑ Allow unknown categ... ☰

ROC PRECISION/RECALL LIFT

Результаты neural network

				Threshold		AUC
True Positive	False Negative	Accuracy	Precision			
20	13	0.707	0.690	0.5		0.795
False Positive	True Negative	Recall	F1 Score			
9	33	0.606	0.645			

Схема SVM
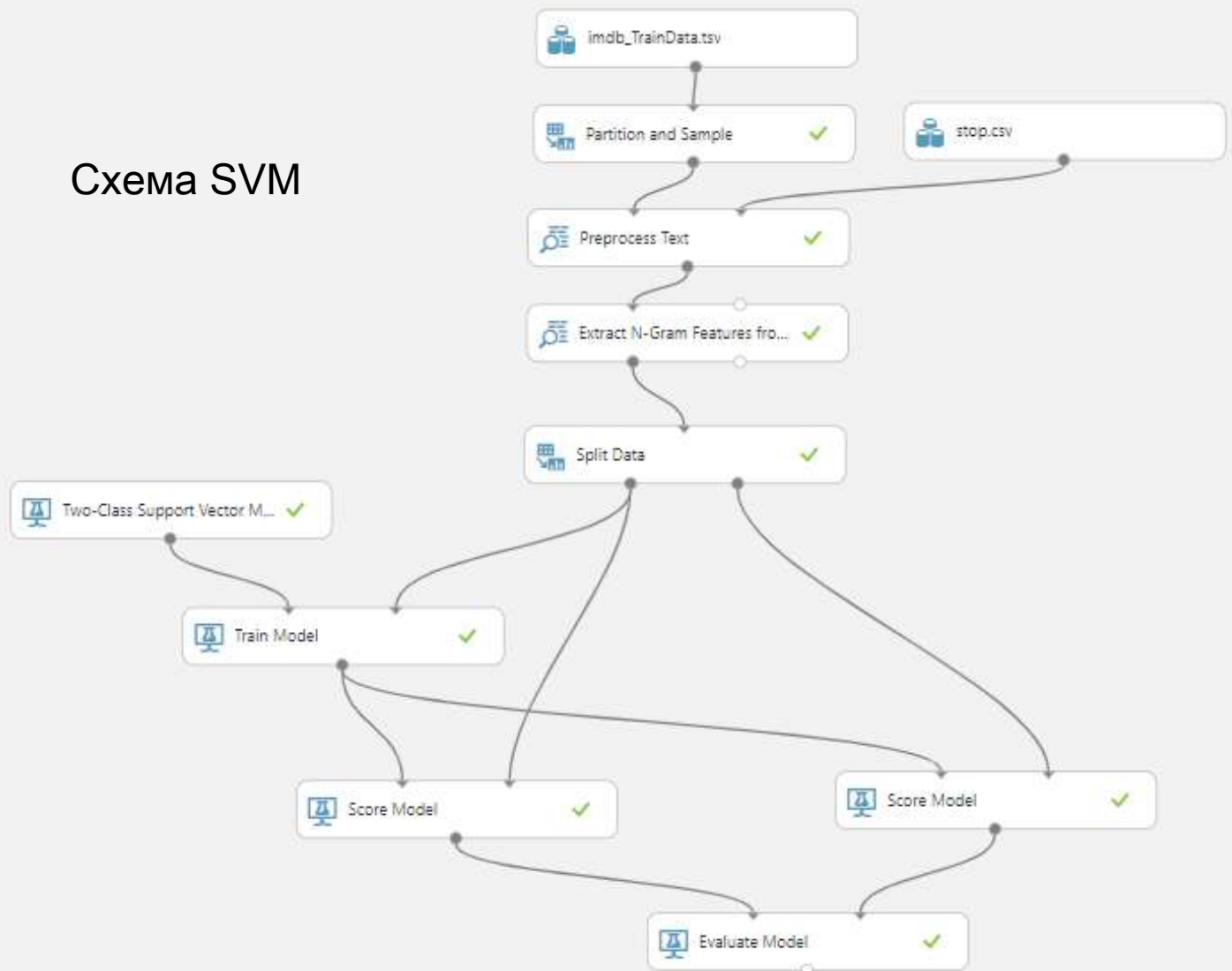
Параметры SVM

◢ Two-Class Support Vector Mac...

Create trainer mode

Parameter Range ⌄

Number of iterations ☰

☐ Use Range Builder

1, 10, 100, 1000

Lambda ☰

☐ Use Range Builder

0.00001,0.0001,0.001,0.01

☑ Normalize features ☰

☐ Project to the unit-sp... ☰

Random number seed ☰

1

☑ Allow unknown categ... ☰

START TIME   11/24/2021...

END TIME   11/24/2021...

ELAPSED TIME   0:00:02.031
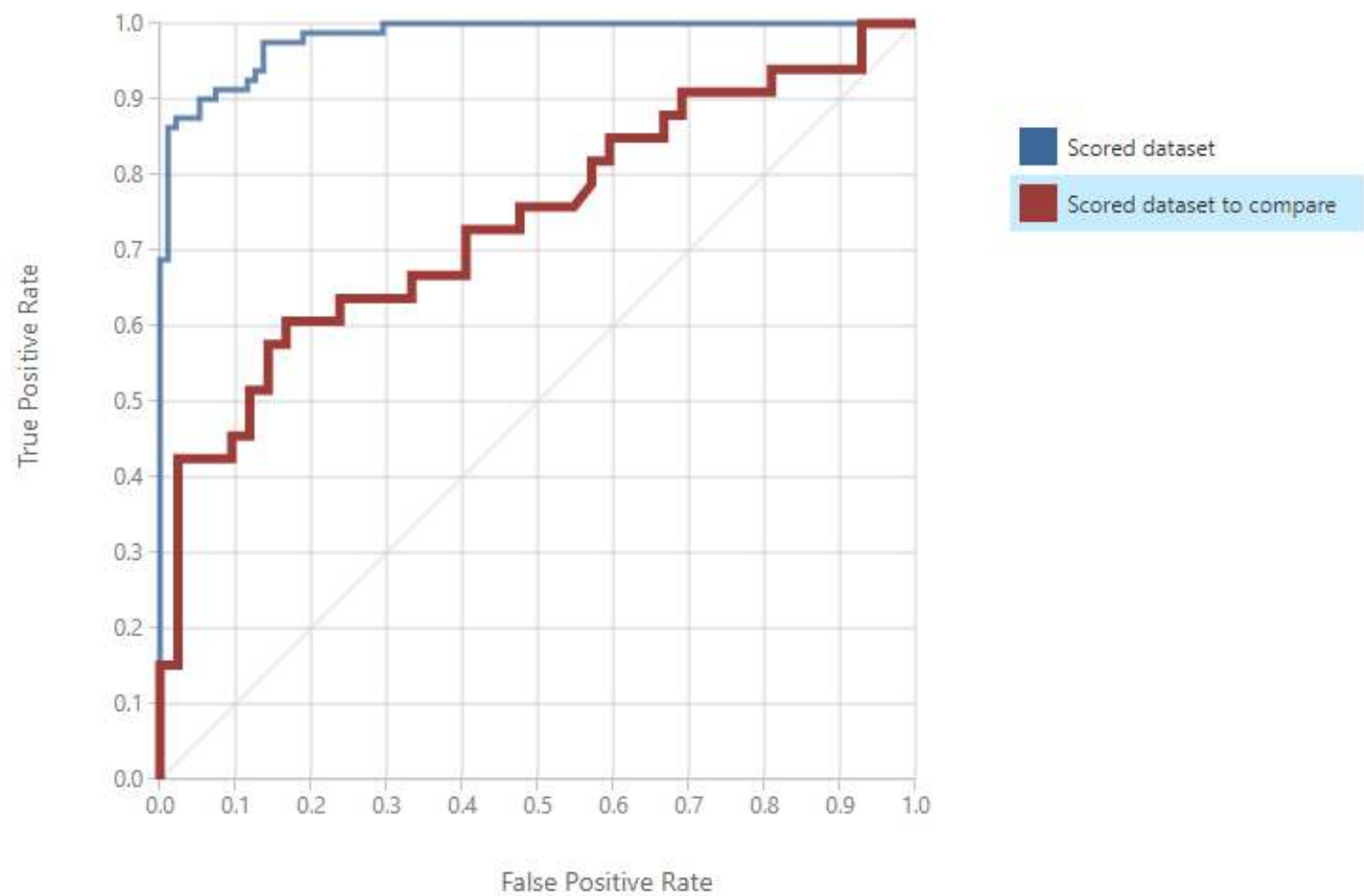
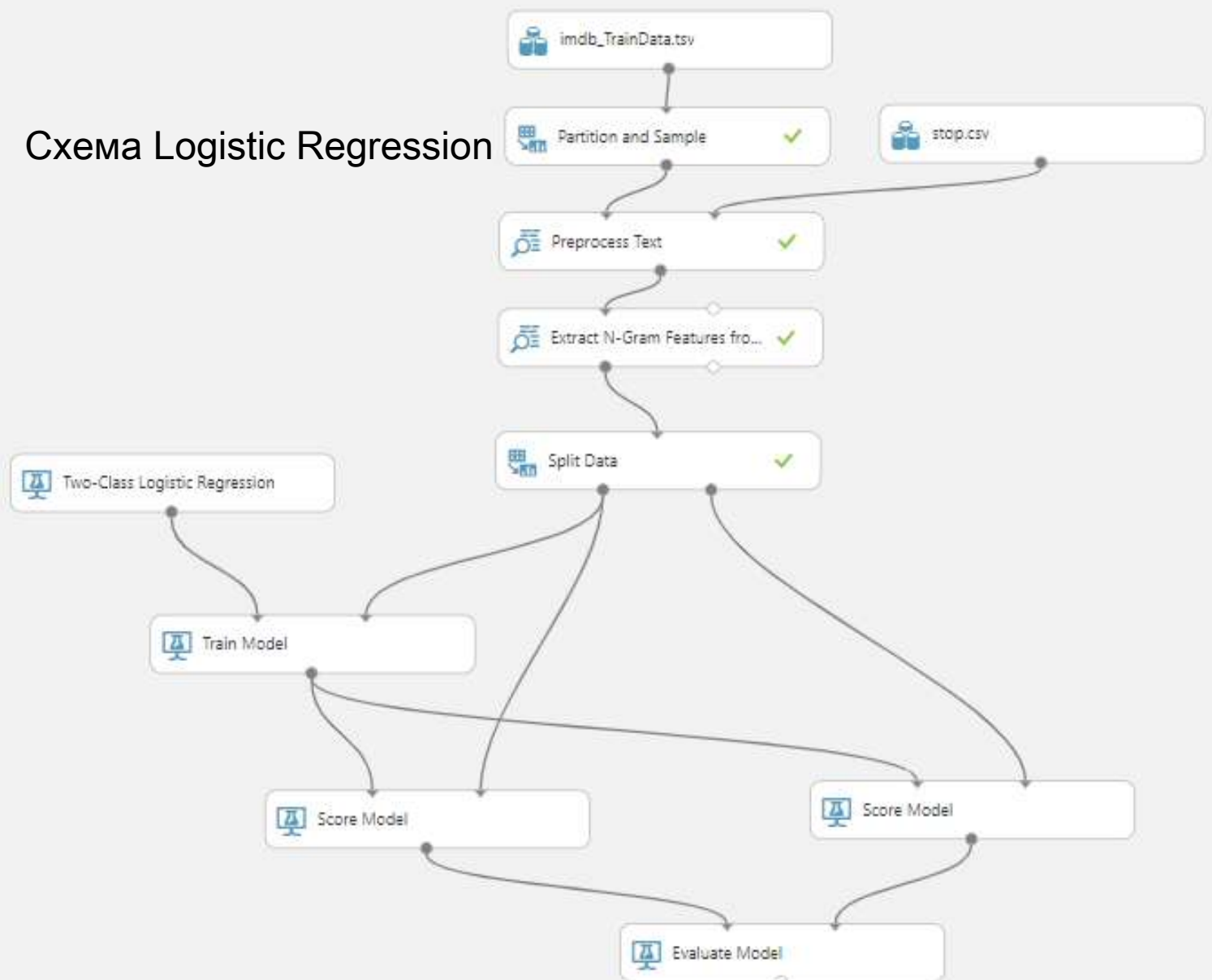STATUS CODE   Finished

STATUS DETAILS   None

View output log

Результаты SVM

ROC PRECISION/RECALL LIFT

					AUC
True Positive	False Negative	Accuracy	Precision	Threshold	
21	12	0.653	0.600	0.5	0.739
False Positive	True Negative	Recall	F1 Score		
14	28	0.636	0.618		

Схема Logistic Regression

◢ **Two-Class Logistic Regression**

Параметры
Logistic
Regression

Create trainer mode

Parameter Range ⌄

Optimization tolerance ☰

☐ Use Range Builder

0.0001, 0.0000001

L1 regularization weight ☰

☐ Use Range Builder

0.0, 0.01, 0.1, 1.0

L2 regularization weight ☰

☐ Use Range Builder

0.01, 0.1, 1.0

Memory size for L-BFGS ☰

☐ Use Range Builder
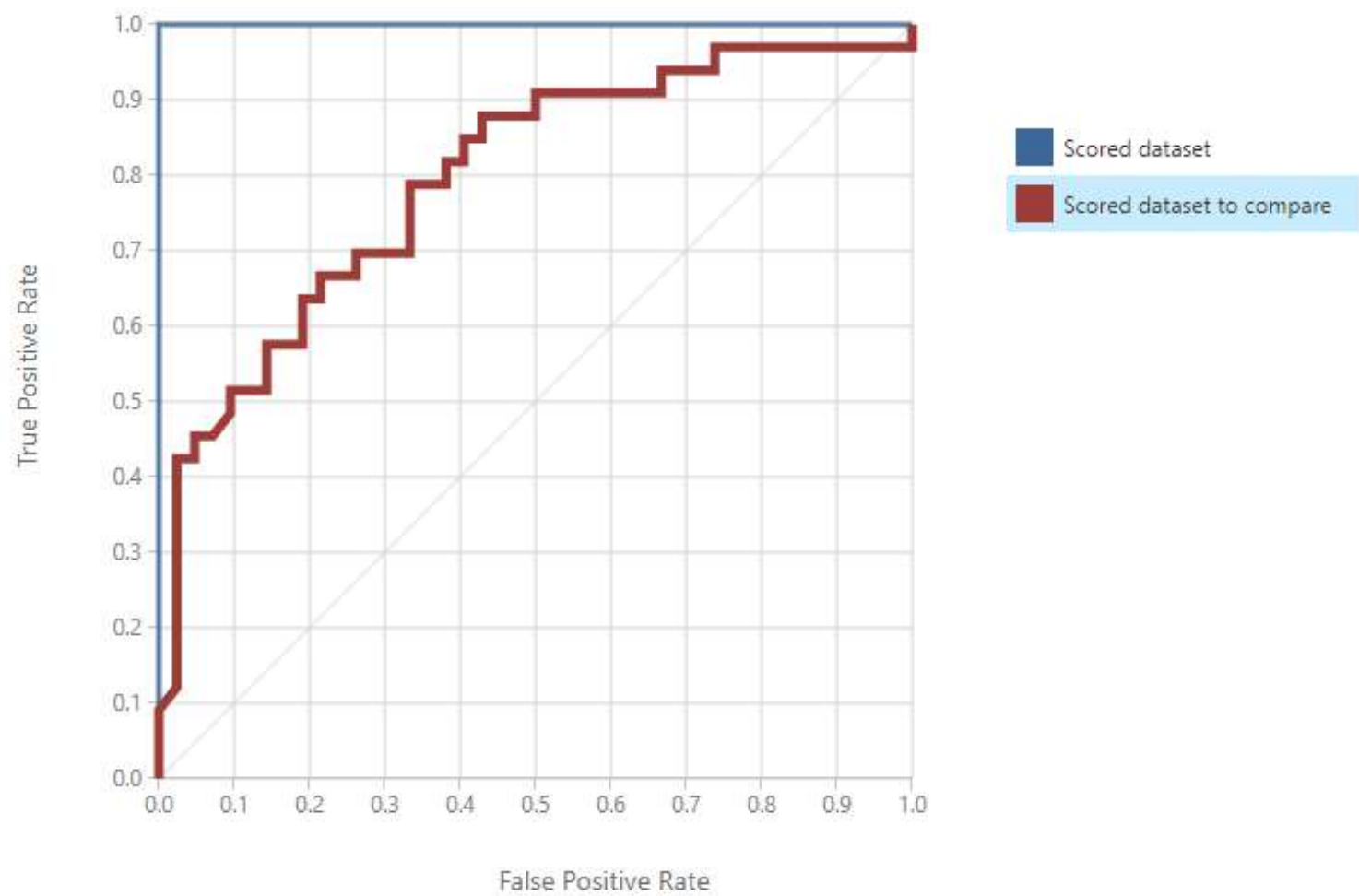
5, 20, 50

Random number seed ☰

☑ Allow unknown categ... ☰

Результаты Logistic Regression

ROC   PRECISION/RECALL   LIFT

	Scored dataset
	Scored dataset to compare

True Positive	False Negative	Accuracy	Precision
22	11	0.733	0.710
False Positive	True Negative	Recall	F1 Score
9	33	0.667	0.688

Threshold		AUC
0.5		0.797