

# Отчет по лабораторной работе No1 по курсу «Искусственный интеллект»

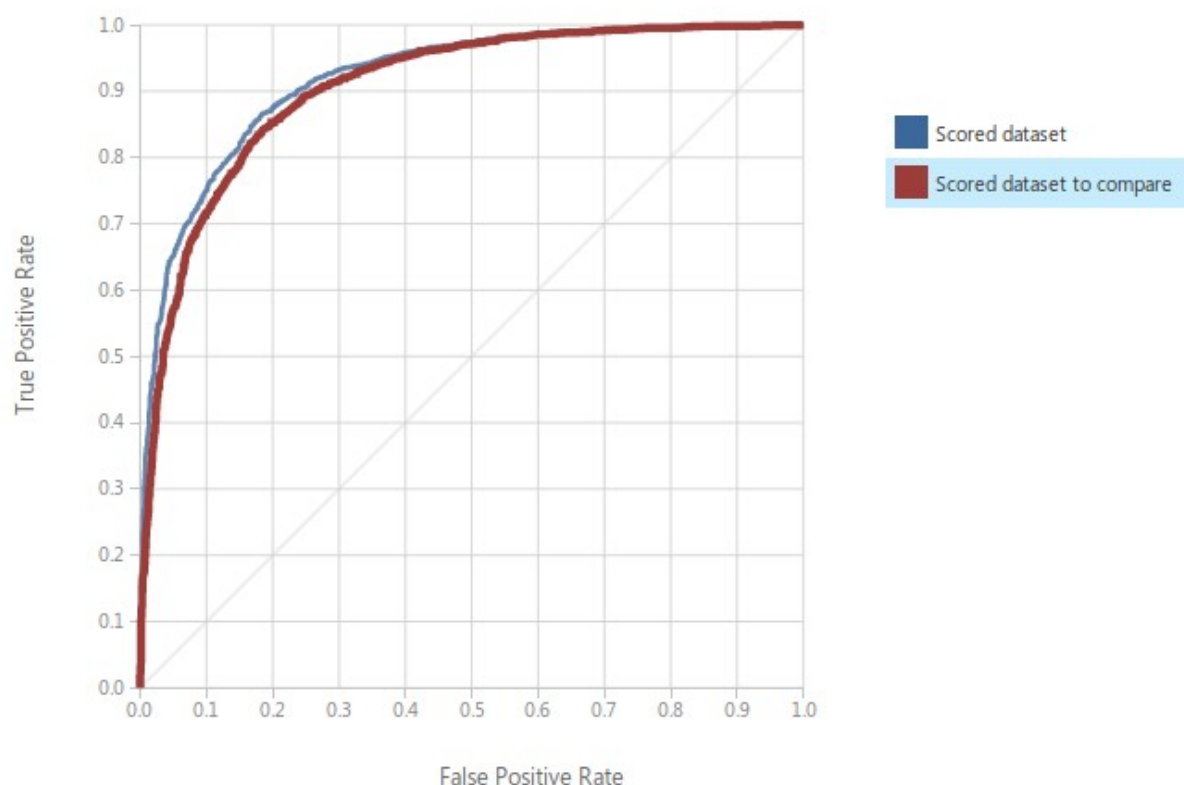
Студент группы 8О-407 Зайцев Никита, No по списку 7

- Текстовый датасет:

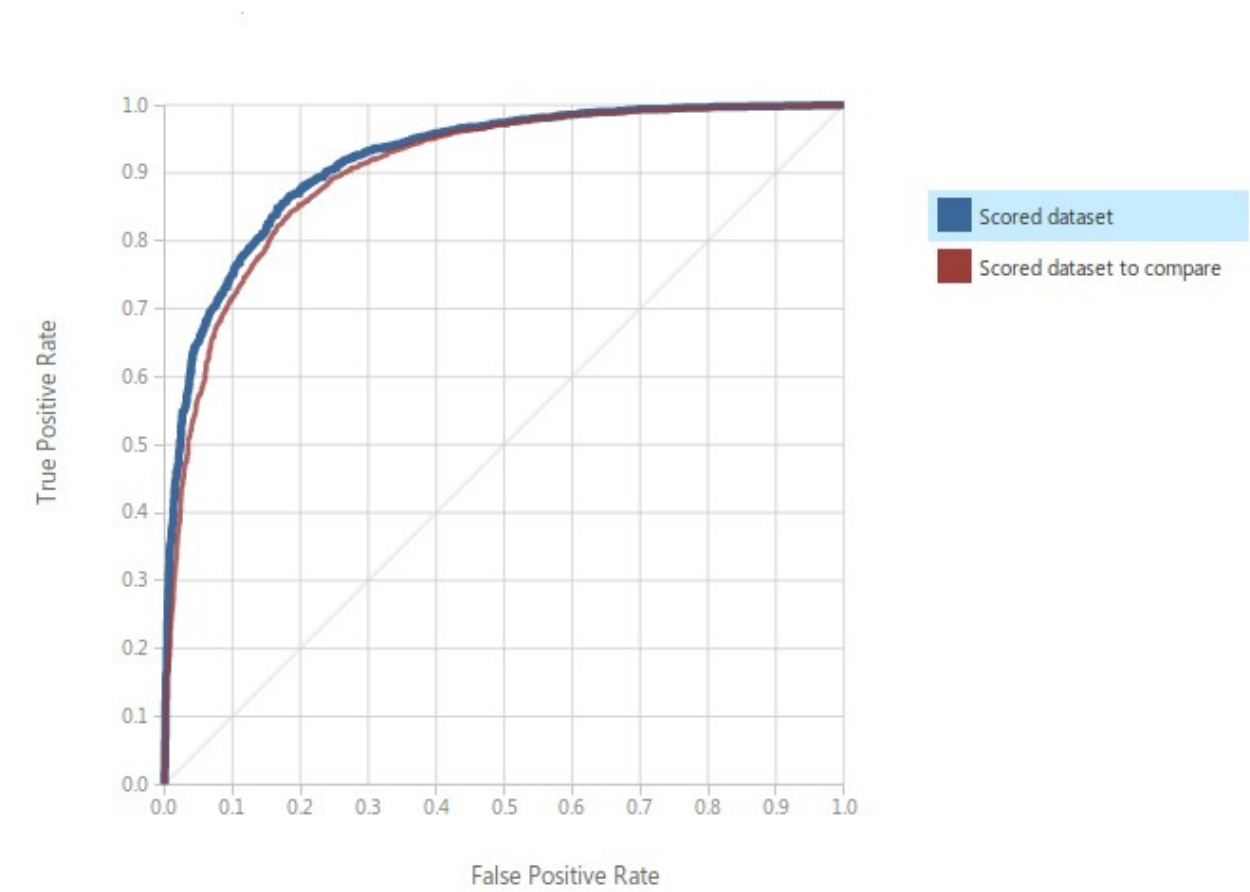
Несмотря на то, что решение выглядит достаточно глубоким, его можно значительно сократить, используя больше готовых конструкций.

Изначальные данные приводятся к удобному для применения виду, затем идёт обработка данных (убирание лишних слов, знаком препинания и т. д.), данные делятся на 2 части, строятся tf и tf-idf вектора и обучаются с помощью логистической регрессии (этот метод неплохо работает с обработкой естественного языка). Потом уже идёт оценка результатов. В целом результаты обучения получились высокими

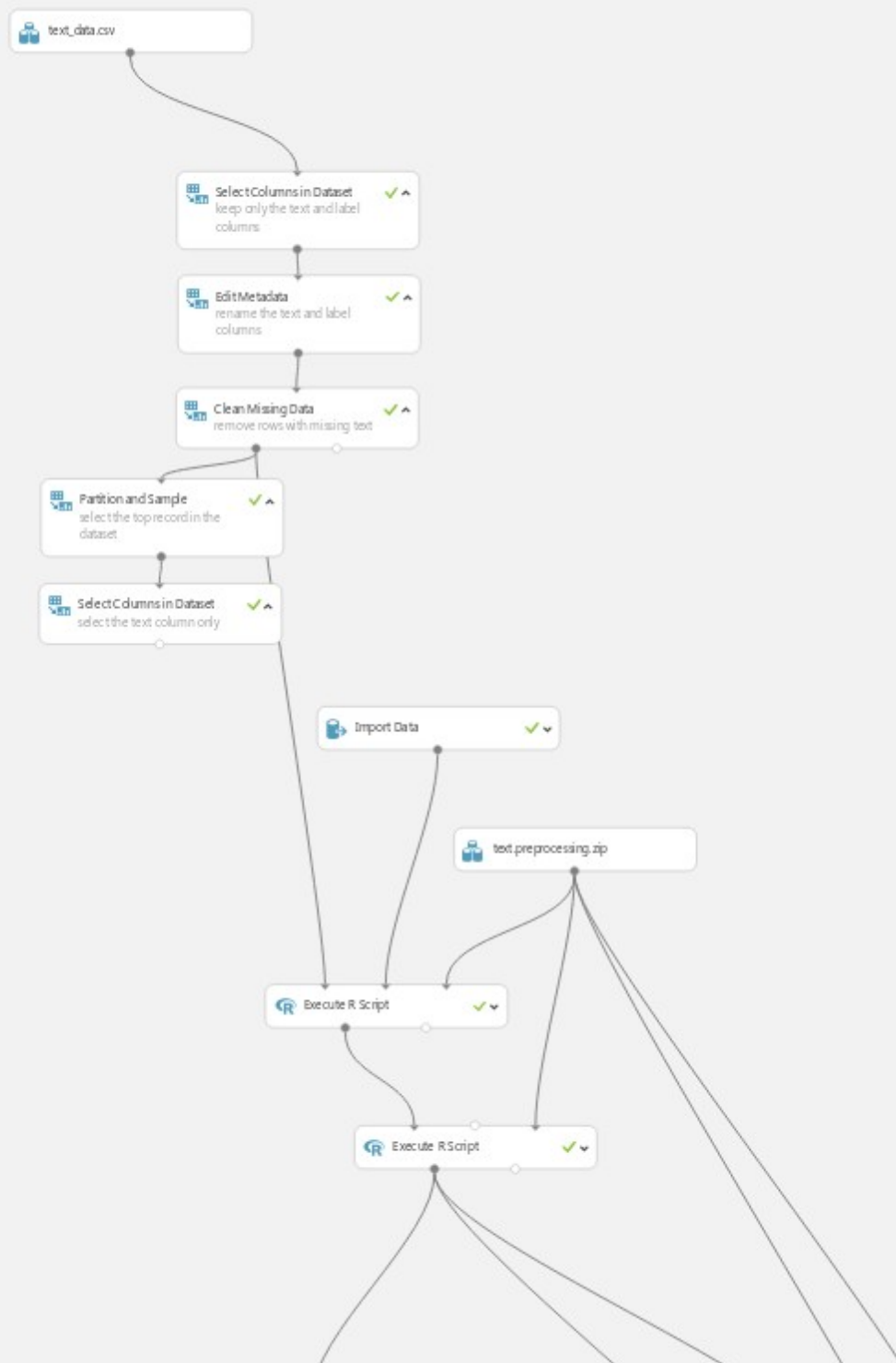
Lab1-Text > Evaluate Model > Evaluation results

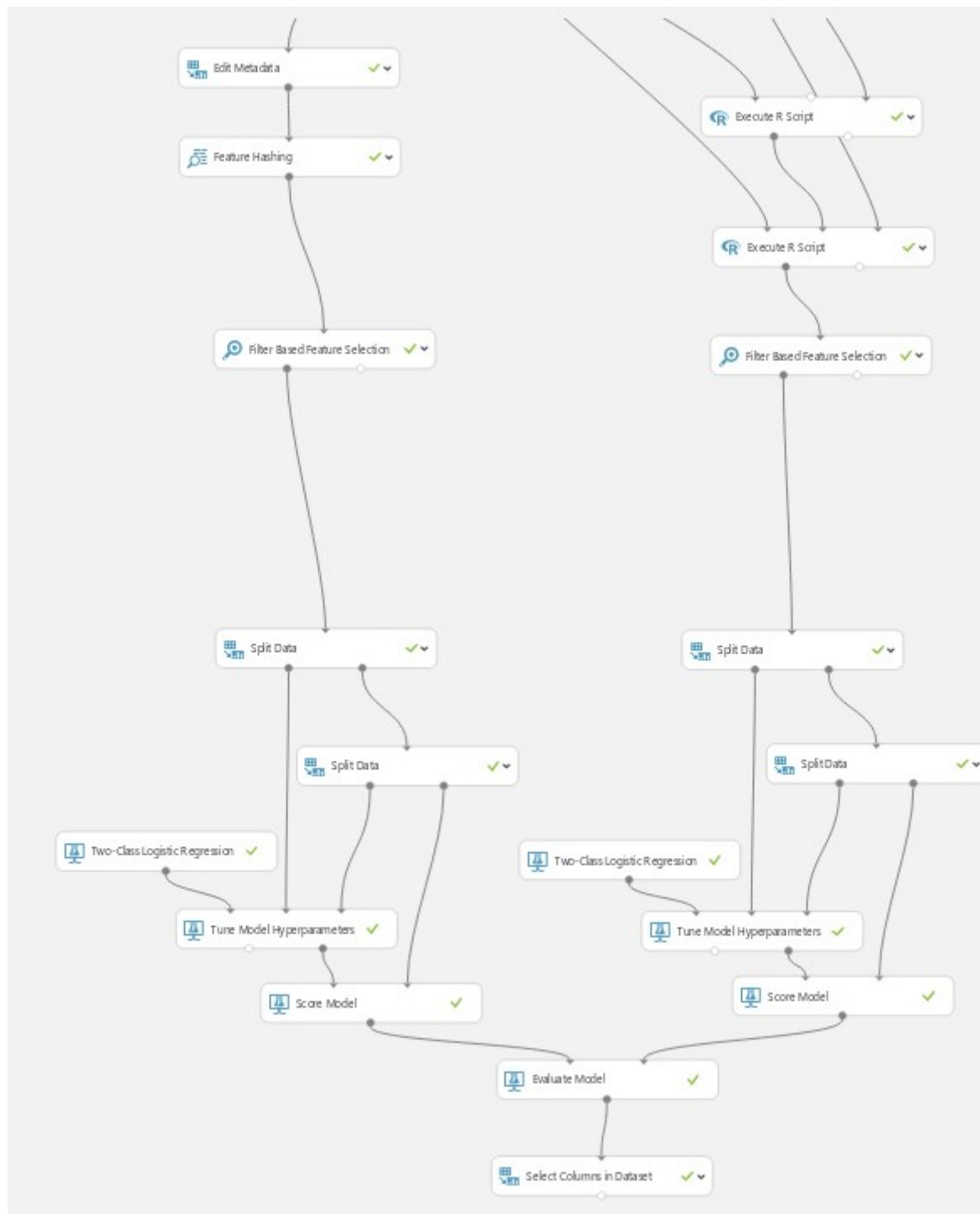


True Positive	False Negative	Accuracy	Precision	Threshold	AUC
2938	358	0.833	0.828	0.5	
False Positive	True Negative	Recall	F1 Score		
610	1876	0.891	0.859		
Positive Label	Negative Label				
1	0				



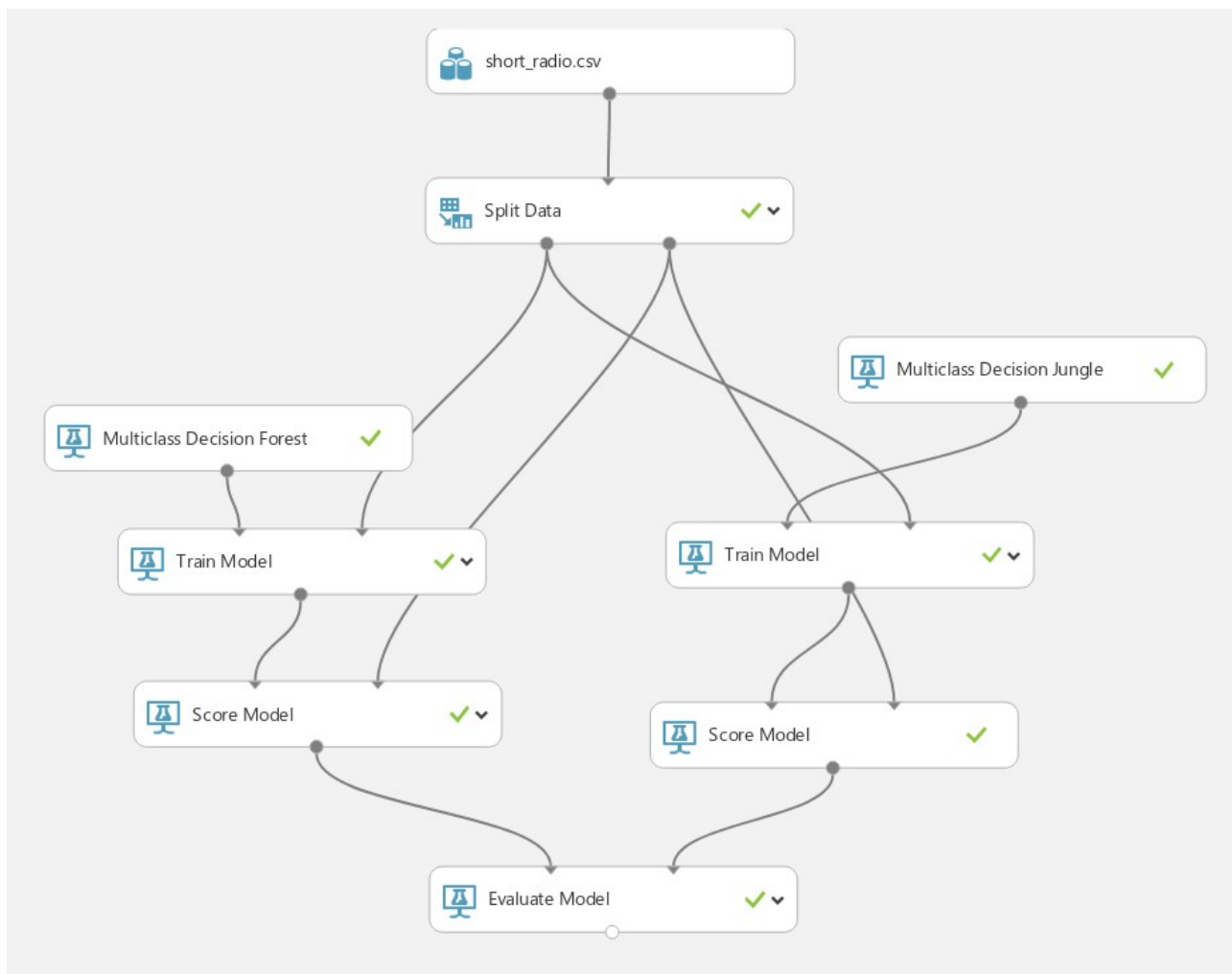
True Positive	False Negative	Accuracy	Precision	Threshold	<div></div>	AUC
2912	384	0.842	0.847	0.5		0.919
False Positive	True Negative	Recall	F1 Score			
527	1959	0.883	0.865			
Positive Label	Negative Label					
1	0					





- Численный датасет:

Данный вариант достаточно просто устроен — исходные данные разбиваются на две части, после чего происходит их тренировка и оценка на другой части



Для тренировки использованы Multiclass Decision Forest и Multiclass Decision Jungle так как они успешно справляются со своей задачей, как и многие другие методы, но не далеко не все.

Результаты оценки почти достигают 100%, это связано с тем, что в базе данных

Lab1\_dg > Evaluate Model > Evaluation results

#### Metrics

Overall accuracy	0.99592
Average accuracy	0.99728
Micro-averaged precision	0.99592
Macro-averaged precision	0.99607
Micro-averaged recall	0.99592
Macro-averaged recall	0.995773

#### Confusion Matrix

		Predicted Class		
		Corridor...	Lab139_7.1	Sport_Ha...
Actual Class	Corridor...	100.0%	0.0%	0.0%
	Lab139_7.1	0.0%	99.6%	0.3%
	Sport_Ha...	0.0%	0.9%	99.1%

#### Metrics

Overall accuracy	0.96698
Average accuracy	0.977986
Micro-averaged precision	0.96698
Macro-averaged precision	0.967727
Micro-averaged recall	0.96698
Macro-averaged recall	0.965836

#### Confusion Matrix

		Predicted Class		
		Corridor...	Lab139_7.1	Sport_Ha...
Actual Class	Corridor...	98.9%	0.9%	0.2%
	Lab139_7.1	0.3%	97.1%	2.6%
	Sport_Ha...	0.0%	6.1%	93.8%

содержится избыточное число данных покрывающих все возможные случаи, хотя для этой лабораторной работы было взято небольшое подмножество исходных данных.