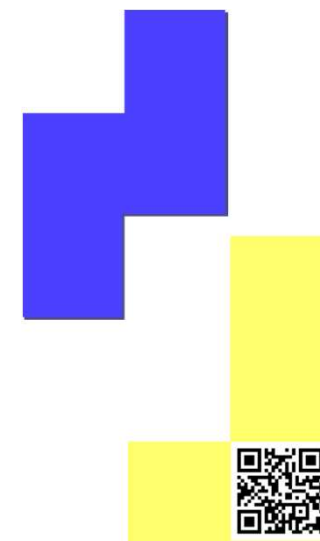




ЦЕНТР НЕПРЕРЫВНОГО ОБРАЗОВАНИЯ
ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК

Сопоставление названий товаров из ассортимента аптек



Исходные данные



ЦЕНТР НЕПРЕРЫВНОГО ОБРАЗОВАНИЯ
ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК

Out[3]:

	title	manuf	ean	title_cat	manuf_cat	is_drug
0	NaN	NaN	NaN	NaN	NaN	False
1	□б/л чай "мелисса,чабрец,фенхель"успокаив	ОО "Фитопром"	4600728005446	NaN	NaN	False
2	! Accu-chek перформа глюкометр комплект N 1	Общество с ограниченной ответственностью "Годо...	4015630063598	NaN	Roche	False
3	! Blend-a-med про эксперт паста зубная отбелив...	Общество с ограниченной ответственностью "ПУЛЬ...	4605052001591	БЛЕНД-А-МЕД ЗУБНАЯ ПАСТА ПРО ЭКСПЕРТ КРЕПКИЕ З...	Проктер энг Гэмбл	False
4	! Hartmann cosmos hyd act платырь от ожогов N 3	ФИЛИАЛ АО НПК КАТРЕН в г. ТЮМЕНИ	2000002047940	NaN	Хартманн	False
...
21376721	№9 Конфеты "ФРУКТОЗОВ" МИРАБЕЛЛА гл/ фр 185г (1...	ООО Сладкая сказка	NaN	NaN	NaN	False
21376722	№9 сбор травян иммунитет плюс 70г пак.	Центр Ароматерапии Ирис	NaN	NaN	NaN	False
21376723	№90 капсулы Витамин Д3	NaN	NaN	ВИТАМИН-Д3 КАПС №90	NaN	True
21376724	№Интер Лозартан таб. п.п.о. 100мг №90	Пранафарм	4607020333543	ЛОЗАРТАН ТАБ 100 МГ №90	Пранафарм	True
21376725	№Интер Лозартан таб. п.п.о. 100мг №90	Пранафарм ООО	4607020333543	ЛОЗАРТАН ТАБ 100 МГ №90	Пранафарм	True

Постановка задачи



ЦЕНТР НЕПРЕРЫВНОГО ОБРАЗОВАНИЯ
ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК

```
pharmacy_df[:5]
```

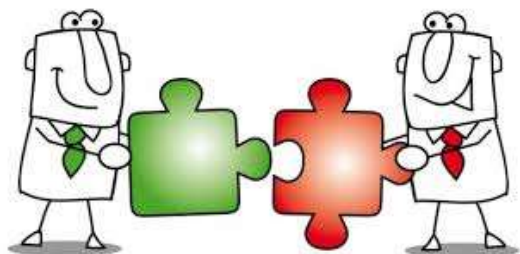
Out[17]:

title
5.нок 50мг таб.п.о x50
абакавир канон таб.п.пл.об . 300мг .60
абакавир канон . тбл п.п.о 600мг .30
абилифай 10мг .28
агалатес 0.5 мг таб. . 8

```
supplier_df[:5]
```

Out[19]:

title
5.нок тб 50мг .50
абакавир.авс табл. п.п.о . 300мг бл. п60 р.фарм
абакавир канон тбл п.п.о 600мг .30 банки
абилифай таб 10мг .28
агалатес тб 0.5мг .8



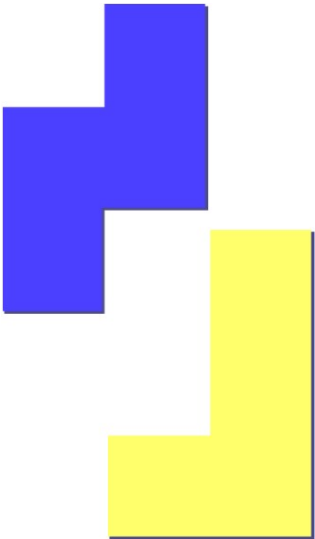
```
title_to_title_true_df[:5]
```

Out[22]:

	Ключ	Значение
0	5.нок 50мг таб.п.о x50	5.нок тб 50мг .50
1	абакавир канон таб.п.пл.об . 300мг .60	абакавир.авс табл. п.п.о . 300мг бл. п60 р.фарм
2	абакавир канон . тбл п.п.о 600мг .30	абакавир канон тбл п.п.о 600мг .30 банки
3	абилифай 10мг .28	абилифай таб 10мг .28
4	агалатес 0.5 мг таб. . 8	агалатес тб 0.5мг .8

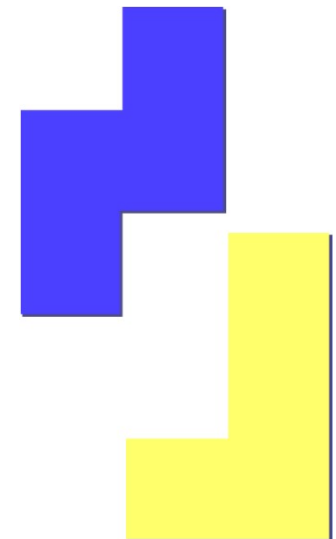
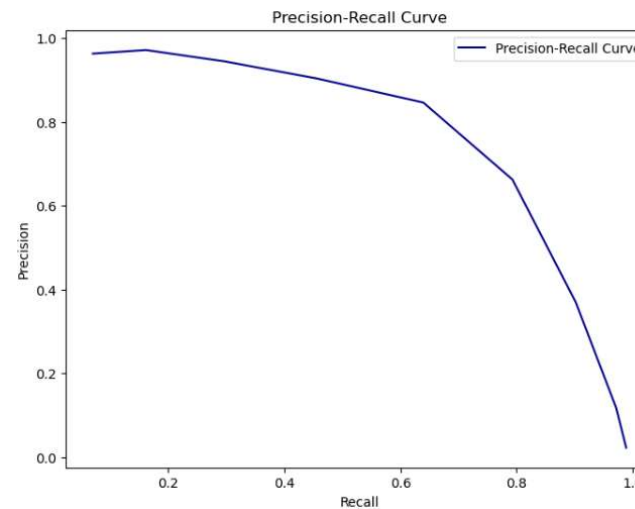
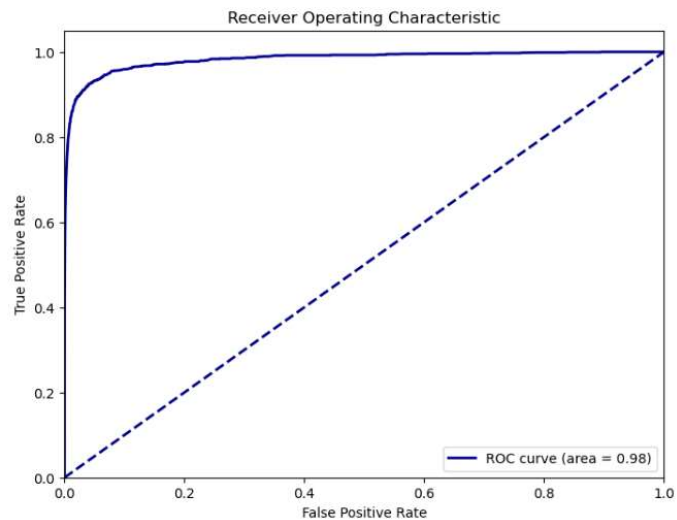
Классификация

1 500 категорий (обучающая выборка 24 000 строк)		NBC	DT	RF	SVM	KNN
		3 секунды	490 секунд	370 секунд	196 секунд	1 секунда
		207 мб	59 мб	7 112 мб	464 мб	11 мб
список аптеки	accuracy	0,87	0,96	0,96	0,95	0,93
	precision	0,81	0,95	0,95	0,94	0,90
	recall	0,87	0,96	0,96	0,95	0,93
	F1-score	0,83	0,95	0,95	0,94	0,91
список поставщика	accuracy	0,87	0,96	0,96	0,96	0,94
	precision	0,81	0,95	0,95	0,94	0,92
	recall	0,87	0,96	0,96	0,96	0,94
	F1-score	0,83	0,95	0,95	0,95	0,92
процент сопоставлений		0,90	0,95	0,95	0,94	0,92



TF-IDF и косинусное сходство

Изменение порога	0,2	0,3	0,4	0,5	0,6	0,7
precision	0.1179	0.3692	0.6628	0.8464	0.9037	0.9448
recall	0.9713	0.9020	0.7927	0.6393	0.4567	0.2967
F1-score	0.2103	0.5239	0.7219	0.7284	0.6067	0.4515



Расстояние Левенштейна

Изменение порога	0,2	0,3	0,4	0,5	0,6	0,7
precision	0.9111	0.6572	0.1943	0.0333	0.0044	0.0010
recall	0.1913	0.4013	0.6693	0.8793	0.9680	0.9967
F1-score	0.3163	0.4983	0.3012	0.0641	0.0088	0.0019

