Sales channel classifier

April 16, 2021

Gerasimov Alexander
Data analyst – BAI Russia

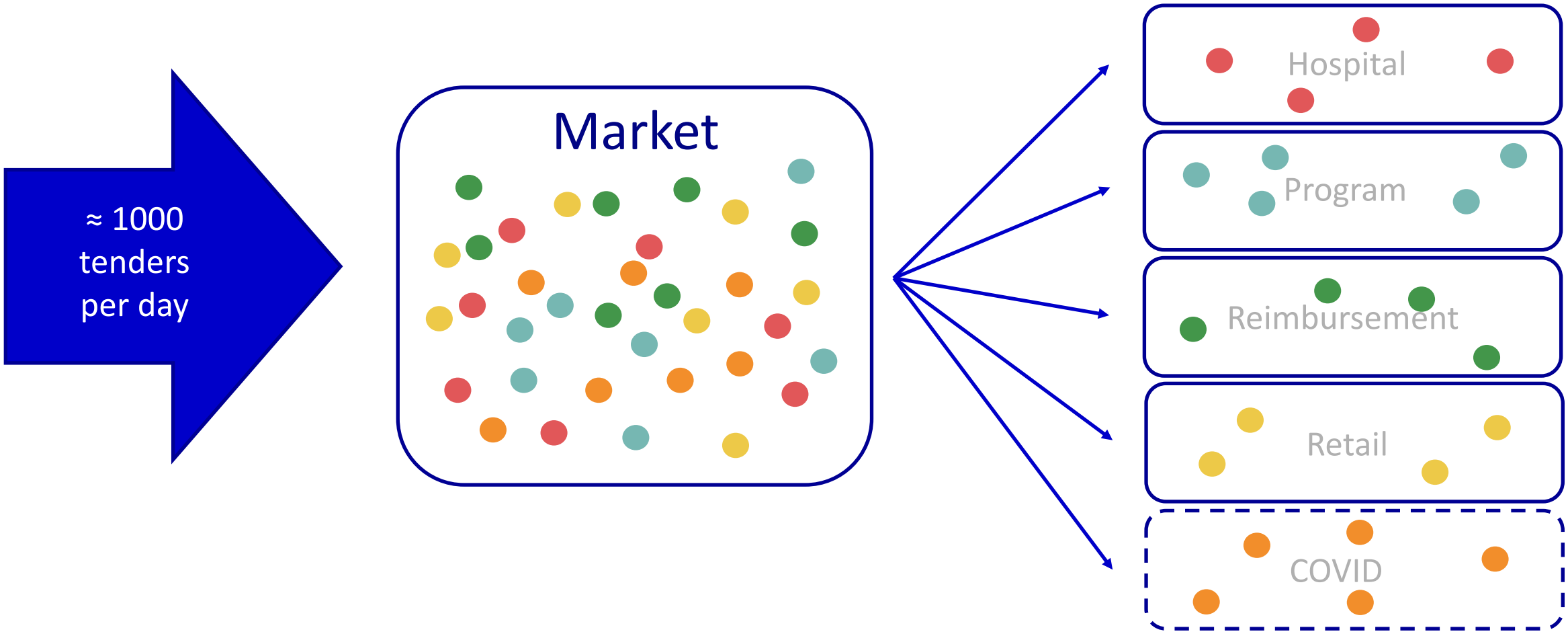
Agenda

01 Problem definition and solution approach

02 Solving the problem and how it works

03 Results and prospects of use

Background


≈ 1000
tenders
per day
Market
Hospital
Program
Reimbursement
Retail
COVID

Problem definition and solution approach

BUSINESS REQUEST: split 12,500 tenders by sales channel

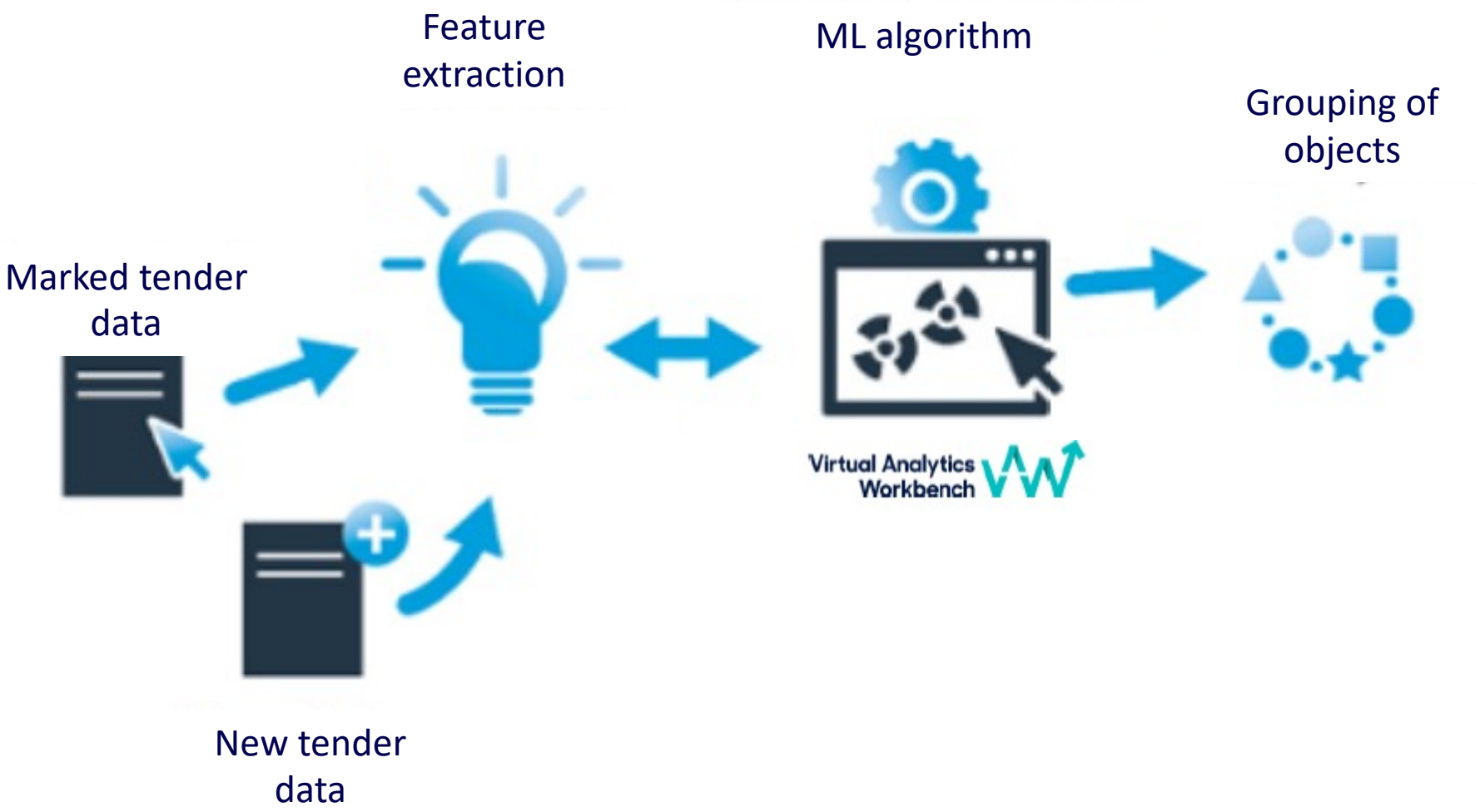
HISTORICAL APROACH

- 18 regional managers
- One-off cost: 2-3 weeks of work, plus consolidation and verification

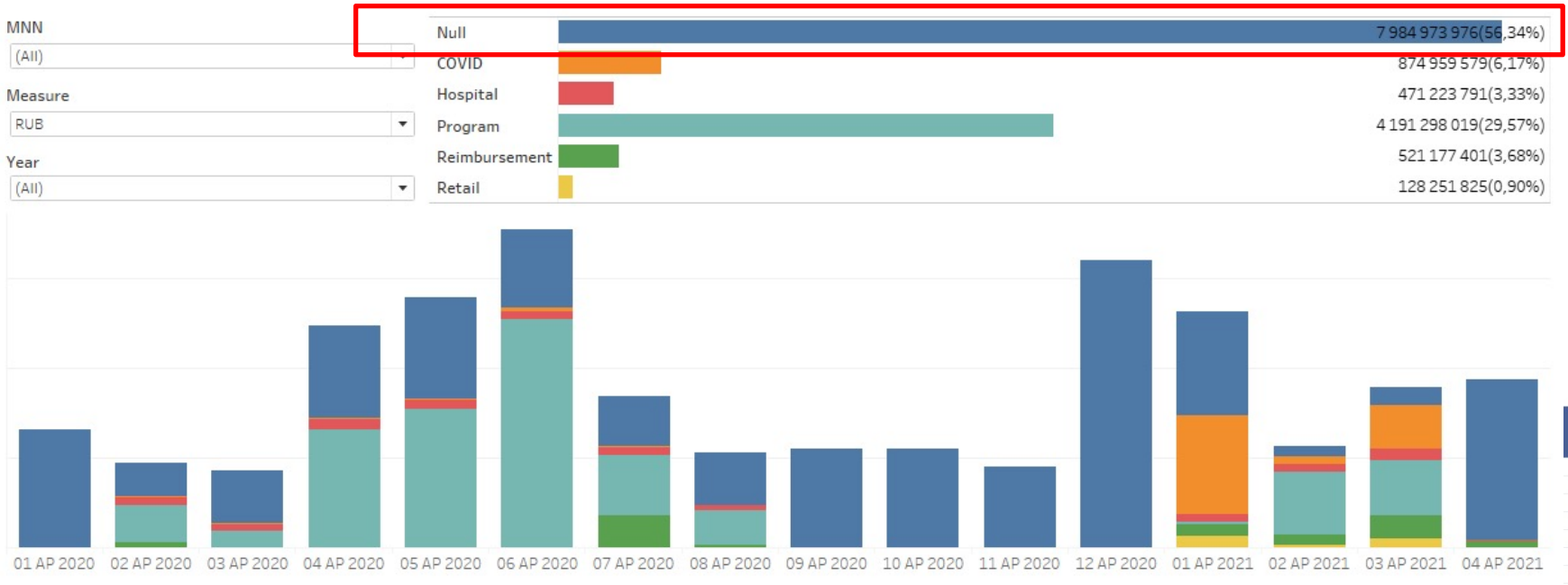
PROPOUSED APROACH
(with artificial intelligence)

- 2-3 hours of analytical FTE
- Can be used for new data

General idea



Input data



More then 50% is not classified

Each tender has about 200 numerical and text characteristics

Канал финансирования	Регион размещения заказа	Заказчик
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Госпитальные продажи	Москва	ПАО Газпром
Региональная льгота	Вологодская область	Департамент здравоохранения Волого...
Региональная льгота	Вологодская область	Департамент здравоохранения Волого...
Региональная льгота	Вологодская область	Департамент здравоохранения Волого...
Госпитальные продажи	Астраханская область	ЧУЗ МЕДИКО-САНИТАРНАЯ ЧАСТЬ
Региональная льгота	Республика Башкортостан	УФСБ РФ по Республике Башкортостан
Госпитальные продажи	Республика Башкортостан	ГБУЗ Республиканский кардиологичес...
Региональная льгота	Республика Башкортостан	ГКУ УЧРЕЖДЕНИЕ МАТЕРИАЛЬНО-ТЕХ...
Региональная льгота	Республика Башкортостан	ГКУ УЧРЕЖДЕНИЕ МАТЕРИАЛЬНО-ТЕХ...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...
Региональная льгота	Республика Бурятия	Министерство здравоохранения Респу...

Data preparation

Наименование товаров (контракт)	Характеристика товара (контракт)	EphMRA (контракт)	ATC (контракт)	New Form Classification (контракт)	ЖНВЛП (контракт)	ОНЛС (контракт)	Rx/OTC (контракт)	Оригинальный препарат/Дженерик (контракт)	Лекарственная форма (контракт)
апиксобан	таб 5 мг №60	Direct facto...	B01AF02 A...	AAA Oral Solid...	ДА	ДА	PRESC	NON GENERIC PRO...	таблетки

таб 5 мг № 60 многопозиционный смешать поставка медикамент тюменский область декабрь поставка товар осуществляться течение 1 календарный деньменее 12 месяц момент поставка товар заказчик закупка единственный поставщик госпитальный закупка госпитальный продажа уральский фо тюменский область иной средство территориальный фонд обязательный медицинский страхование 2020г средство предпринимательский деятельность 2020г поставка медикамент 223 фз гауз лечебный реабилитационный центр градостроитель канал определять apixaban...

[0.0720720720720207, 0.06756756756756757, 0.036036036036036036, ..., 0.00450450450450450 45]

Numerical value of one word

Data preparation

Put array into the model

array([[0.56891668, 0.28837225, …, 0.41496209],

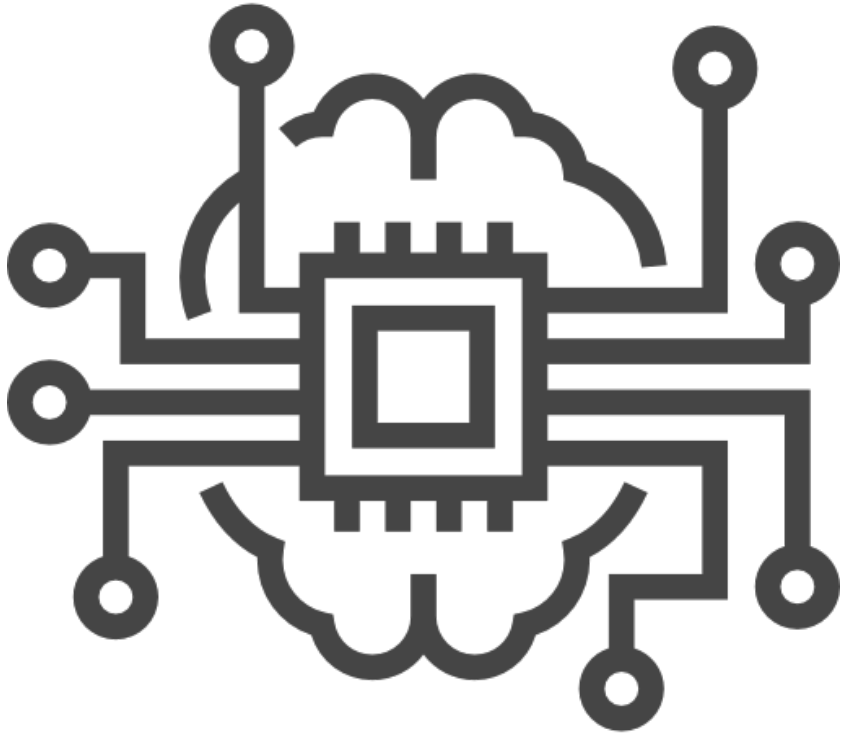
[0.50590898, 0.51392183, …, 0.60730685],

[0.95103629, 0.08641472, …, 0.81941185],

[0.24976724, 0.9070189 , …, 0.65477065],

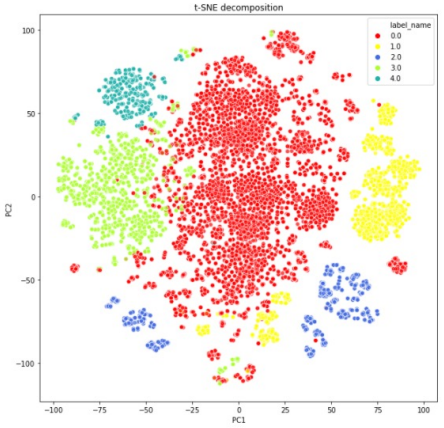
[0.90209274, 0.81263531, …, 0.96802075]])

One tender

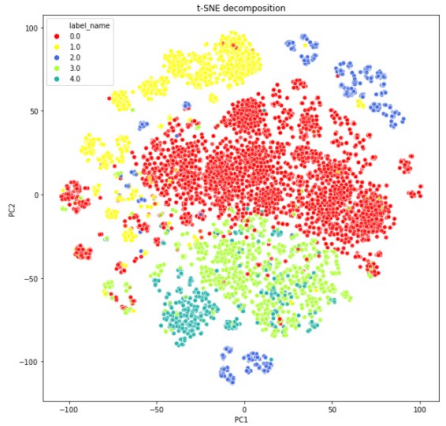


Model development

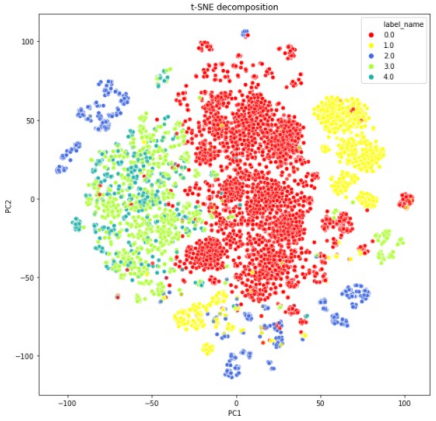
SVC



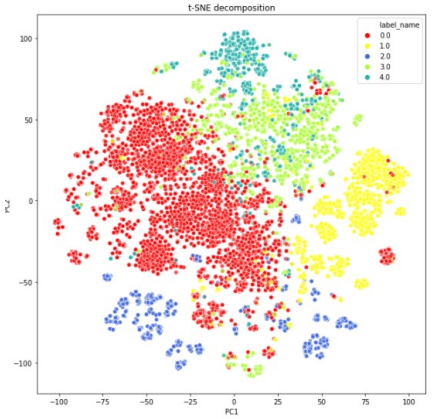
GradientBoostingClassifier



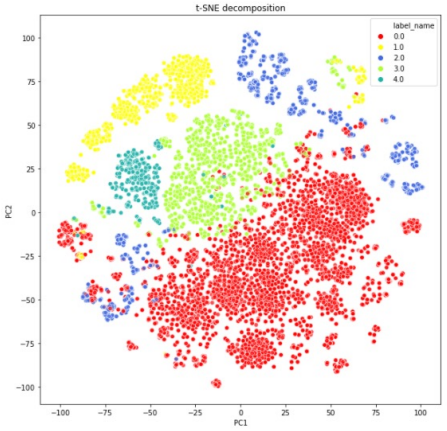
RandomForestClassifier



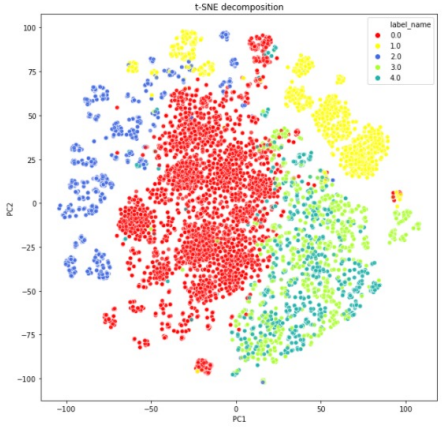
KNN



MultinomialNB

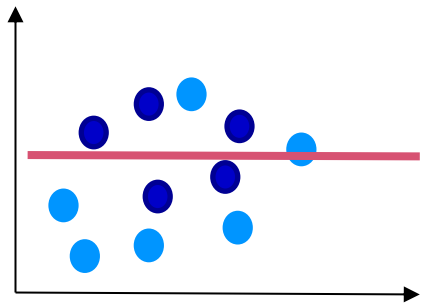


LogisticRegression

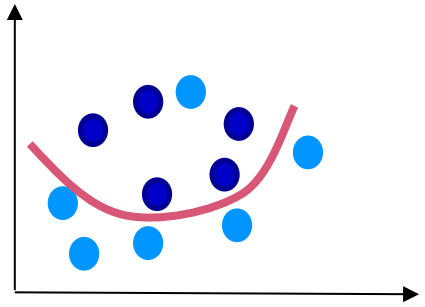


Model	Training Set Accuracy	Test Set Accuracy
Random Forest	1.0	0.97289
Gradient Boosting	1.0	0.97285
KNN	1.0	0.96380
Logistic Regression	0.94456	0.94578
SVM	0.90161	0.89140
Multinomial Naïve Bayes	0.88608	0.88084

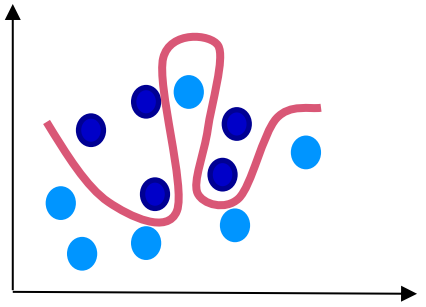
Model development



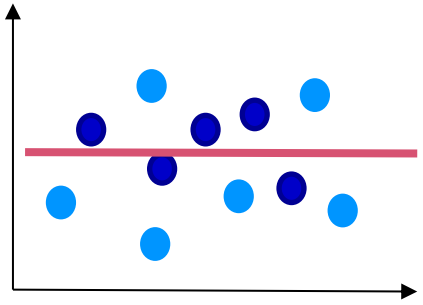
Underfitting

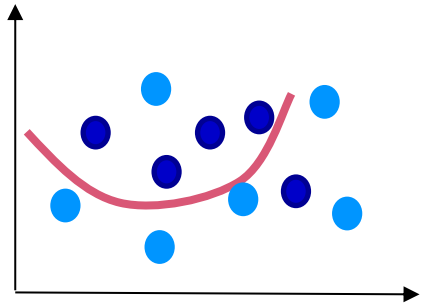


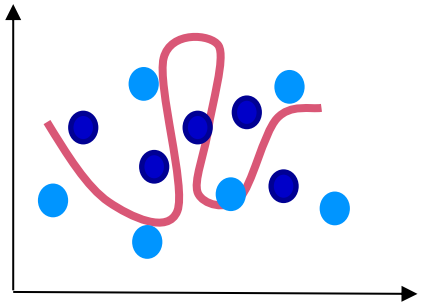
Balanced



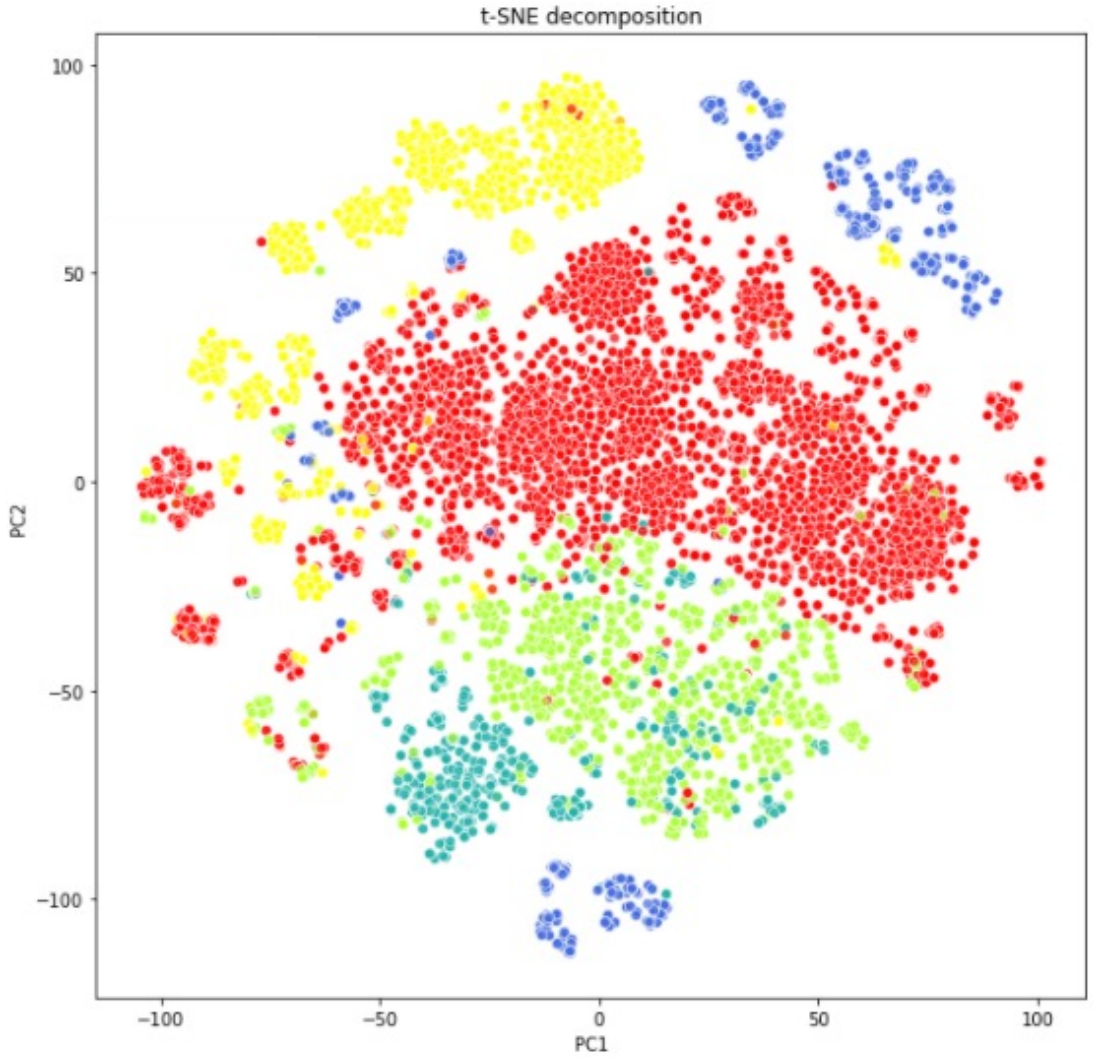
Overfitting







Model development



GradientBoostingClassifier

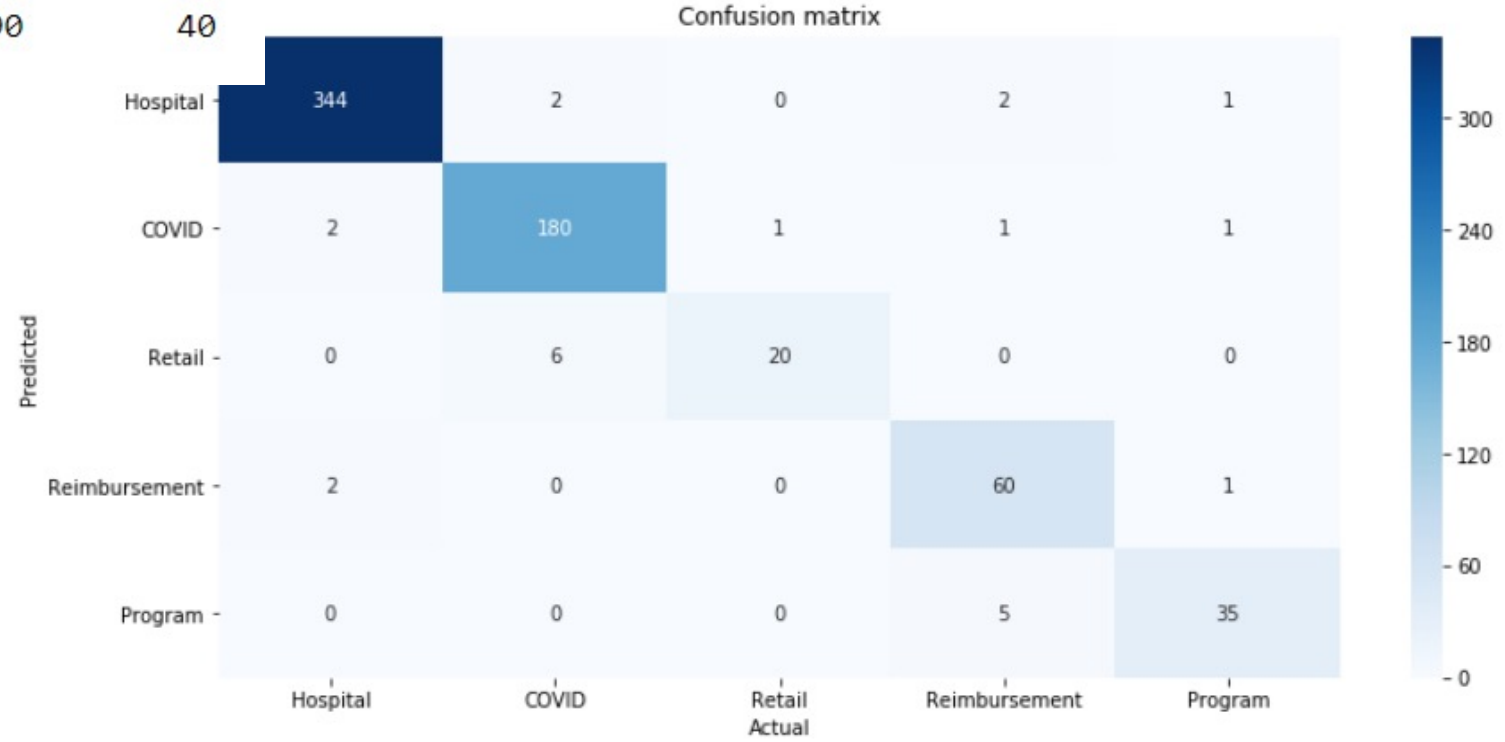
Precision:

Hospital	0.99
COVID	0.96
Retail	0.95
Reimbursement	0.88
Program	0.92

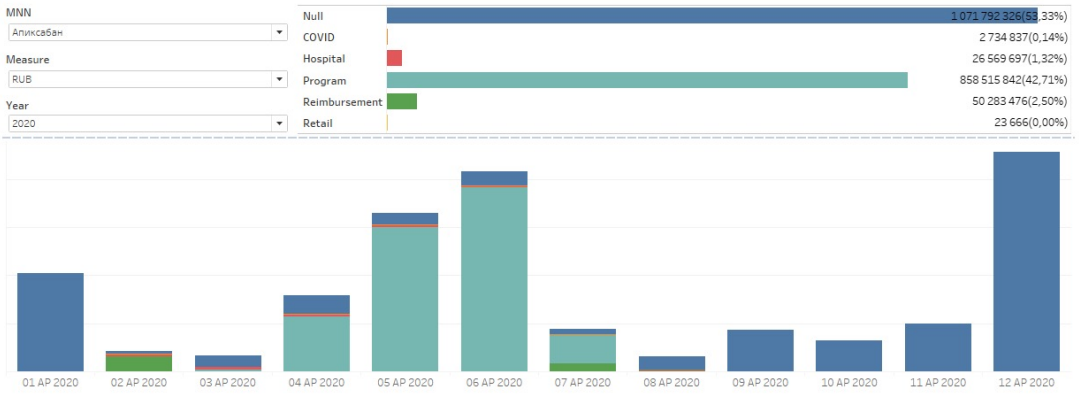
Results

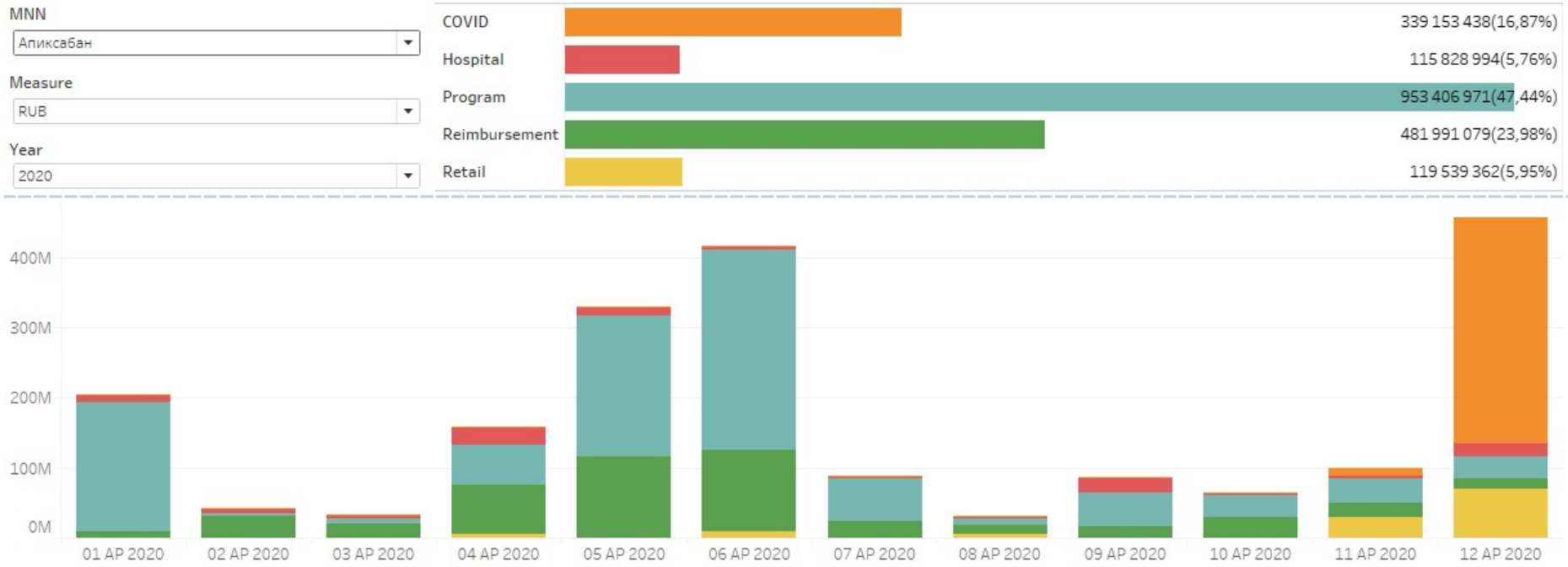
Classification report

	precision	recall	f1-score	support
Hospital	0.99	0.99	0.99	349
COVID	0.96	0.97	0.97	185
Retail	0.95	0.77	0.85	26
Reimbursement	0.88	0.95	0.92	63
Program	0.92	0.88	0.90	40



Results


MNN
Апиксабан
Measure
RUB
Year
2020
Null 1 071 792 326(53,33%)
COVID 2 734 837(0,14%)
Hospital 26 569 697(1,32%)
Program 858 515 842(42,71%)
Reimbursement 50 283 476(2,50%)
Retail 23 666(0,00%)
01 AP 2020
02 AP 2020
03 AP 2020
04 AP 2020
05 AP 2020
06 AP 2020
07 AP 2020
08 AP 2020
09 AP 2020
10 AP 2020
11 AP 2020
12 AP 2020

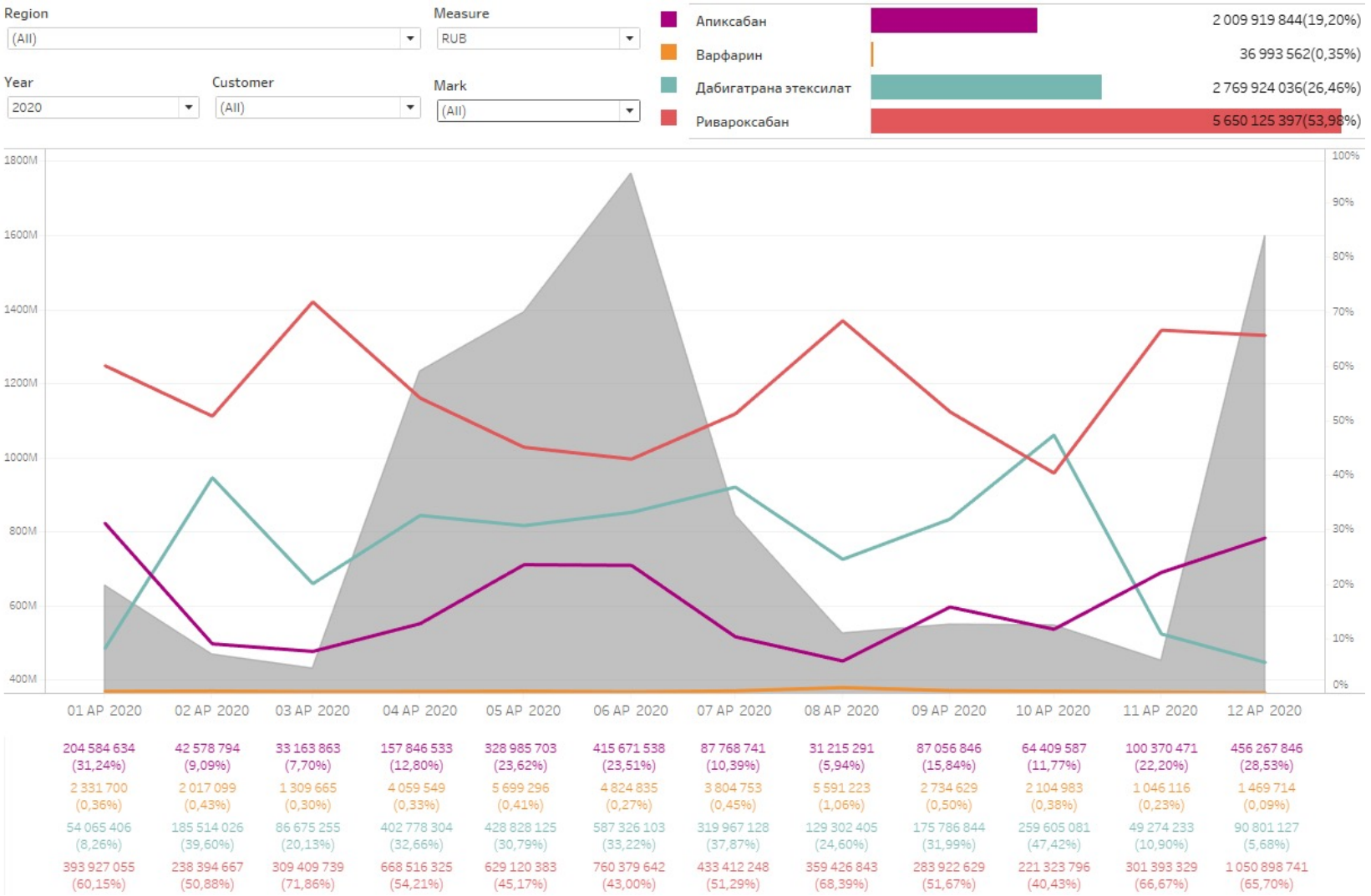

MNN
Апиксабан
Measure
RUB
Year
2020
COVID 339 153 438(16,87%)
Hospital 115 828 994(5,76%)
Program 953 406 971(47,44%)
Reimbursement 481 991 079(23,98%)
Retail 119 539 362(5,95%)
400M
300M
200M
100M
0M
01 AP 2020
02 AP 2020
03 AP 2020
04 AP 2020
05 AP 2020
06 AP 2020
07 AP 2020
08 AP 2020
09 AP 2020
10 AP 2020
11 AP 2020
12 AP 2020

Prospects of use

In general it helps to track brand strategy fulfillment:

- Set plans for FF (Field Force) by channel
- Monitor market performance by channel and areas in Tableau dashboards
- Evaluate impact of Covid on other sales both for company products & competitors

Prospects of use



Prospects of use