



QUALITY MONITORING

SELEZNEV ARTEM
DS TEAM LEADER @ SBER

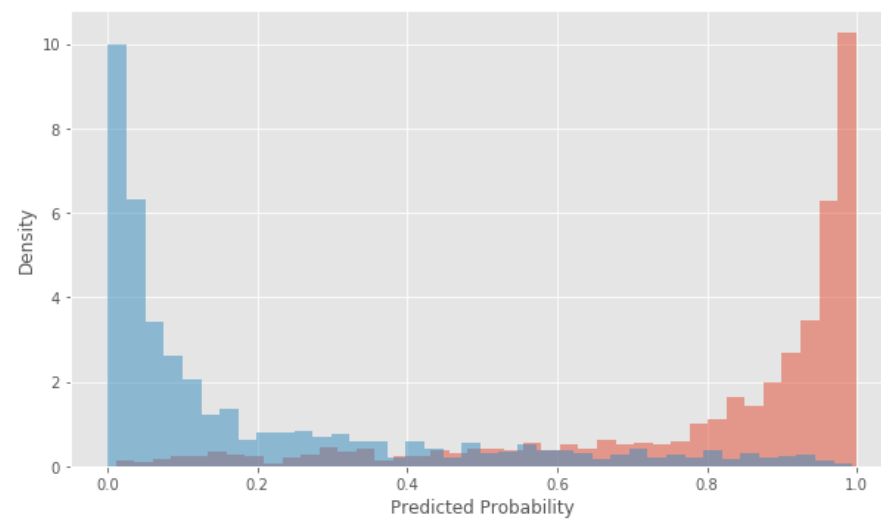
НА СЕГОДНЯ

- **Model Quality**



НА СЕГОДНЯ

- **Model Quality**
- **Основные метрики**



НА СЕГОДНЯ

- **Model Quality**
- **Основные метрики**
- **Мониторинг + Streamlit**



MODEL / DATA QUALITY



ЗАЧЕМ?

Качество	Описание
Данные	Данные полезные и качественные?
Стойкость модели	Готова модель к изменения? На сколько большим?
Предикт модели	Правильно модель делает предсказания?
Сравнение моделей	Новая модель лучше, чем старая или альтернативная? На сколько?

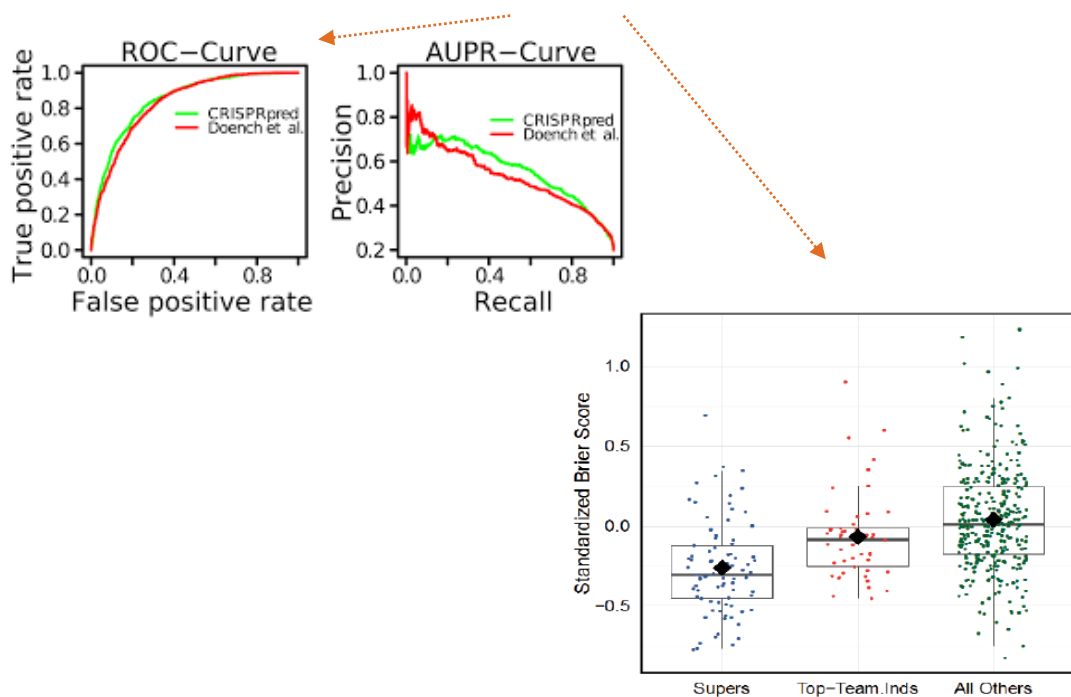
???

		True condition				
Total population		Condition positive	Condition negative	Prevalence = $\frac{\Sigma \text{Condition positive}}{\Sigma \text{Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{True positive} + \Sigma \text{True negative}}{\Sigma \text{Total population}}$	
Predicted condition	Predicted condition positive	True positive, Power	False positive, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{True positive}}{\Sigma \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{False positive}}{\Sigma \text{Predicted condition positive}}$	
	Predicted condition negative	False negative, Type II error	True negative	False omission rate (FOR) = $\frac{\Sigma \text{False negative}}{\Sigma \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{True negative}}{\Sigma \text{Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{True positive}}{\Sigma \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{False positive}}{\Sigma \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR-}}$	F ₁ score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{False negative}}{\Sigma \text{Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{True negative}}{\Sigma \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$		

НАЧНЕМ С ДАННЫХ

MODEL QUALITY

SCORE сравнение

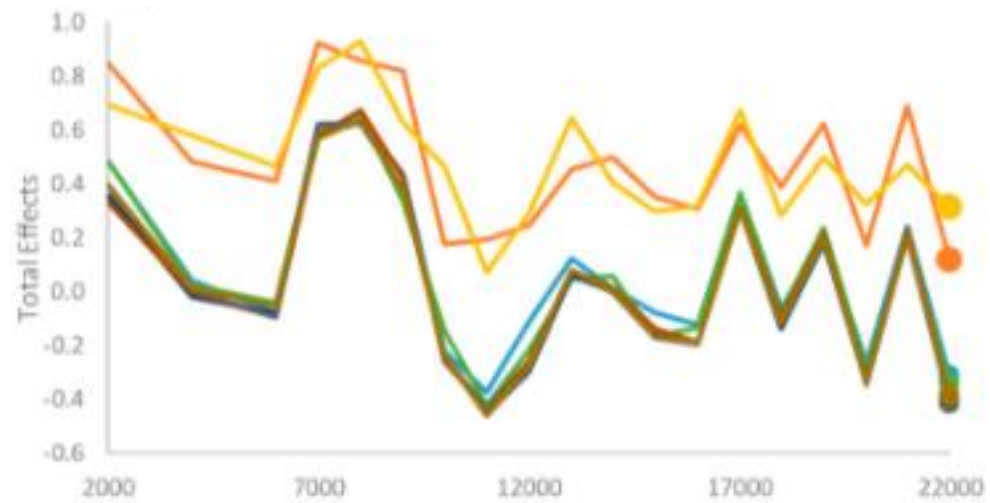


Ключевые факторы сравнение

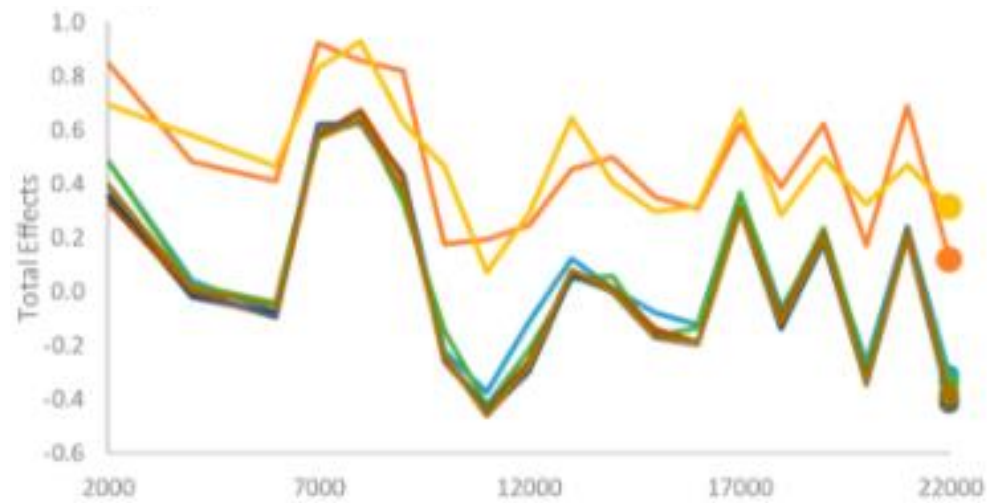
KEY Factors
var / mean

KEY Factors
avg

MQ - ЧУВСТВИТЕЛЬНОСТЬ



MQ - ЧУВСТВИТЕЛЬНОСТЬ



Ваши данные

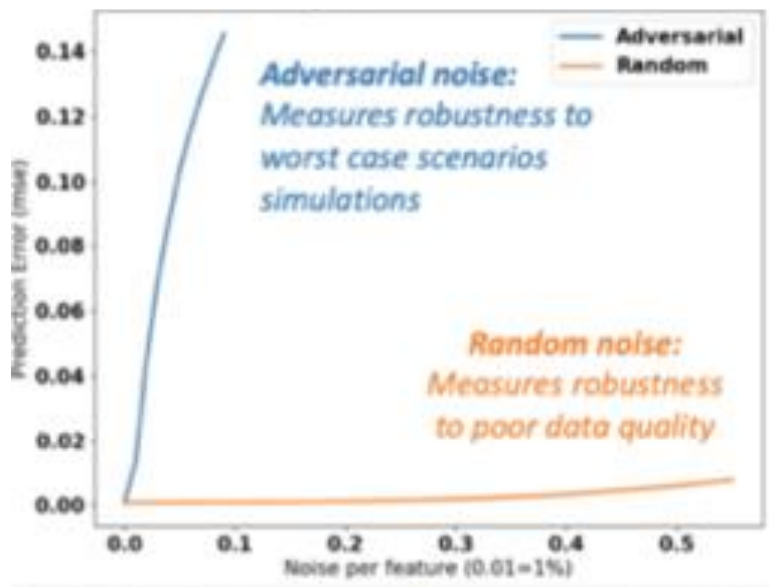
*

coef

=

‘Ваши данные

MQ - ЧУВСТВИТЕЛЬНОСТЬ



Ваши данные

*

coef

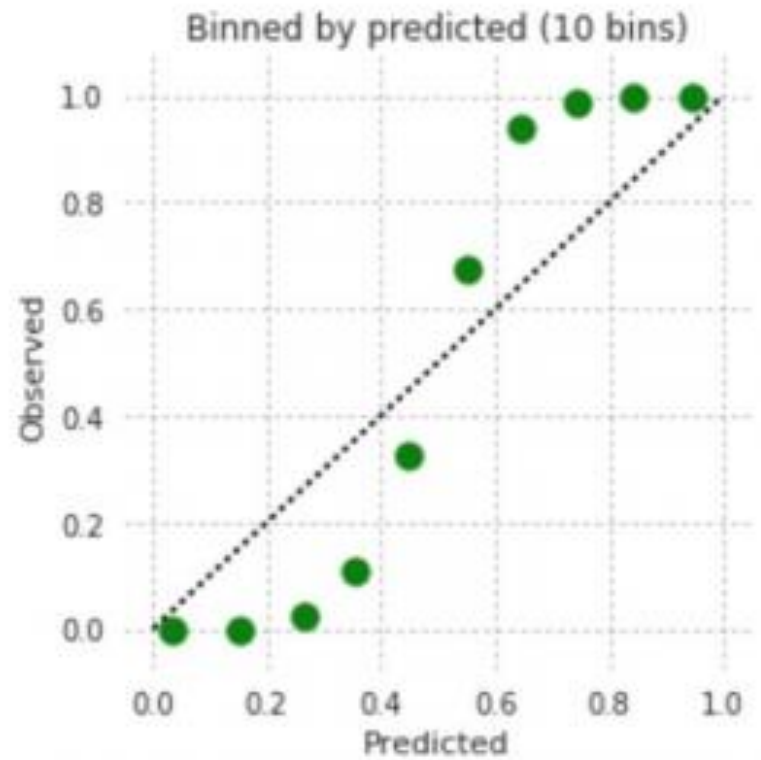
=

‘Ваши данные

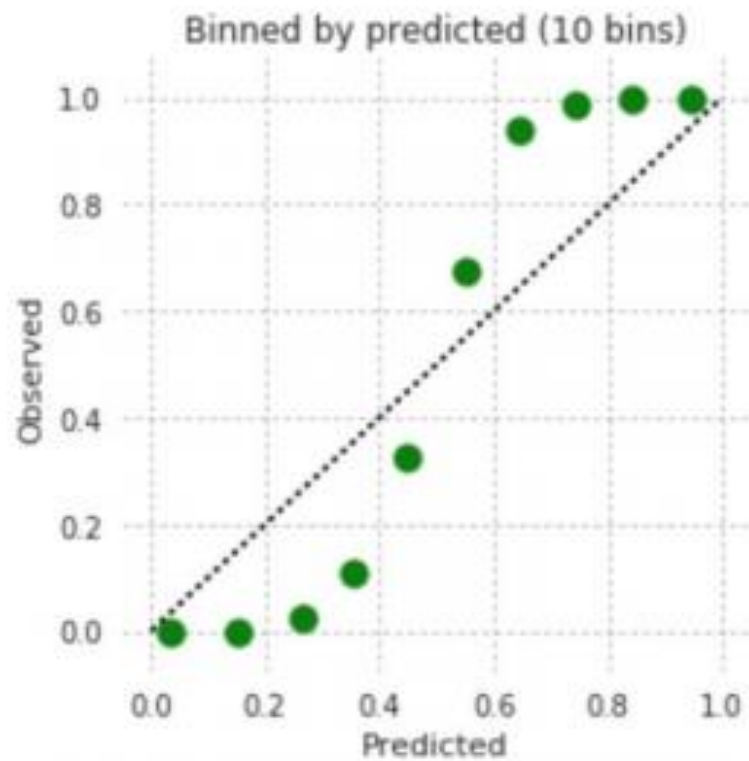
- Рандом
- Осмысленное искажение

MODEL QUALITY - ПРИМЕР

MODEL QUALITY - ДОПОЛНЕНИЕ

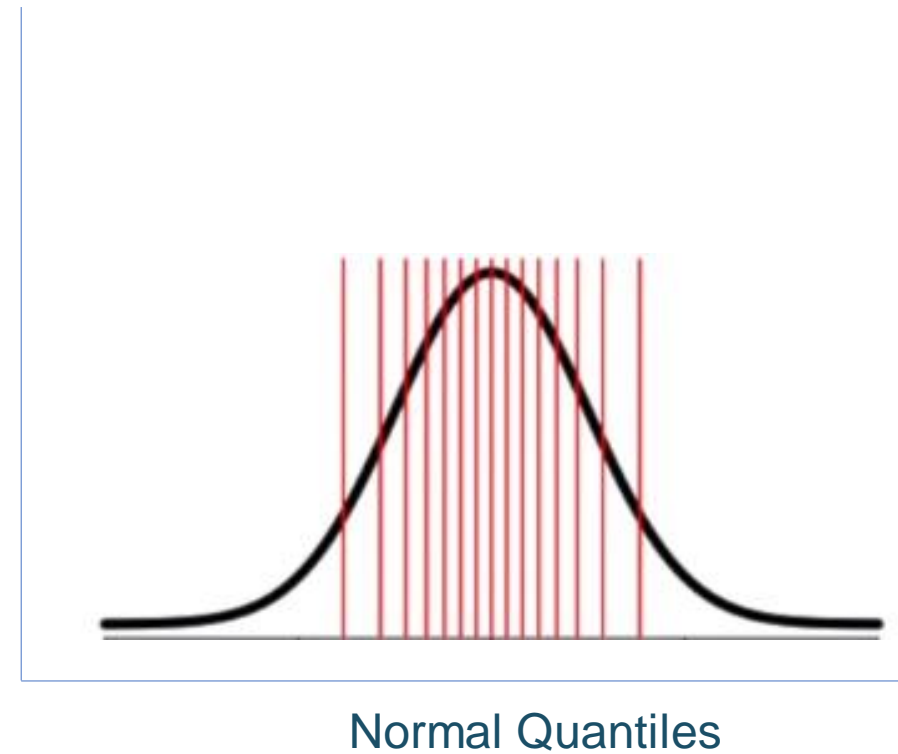
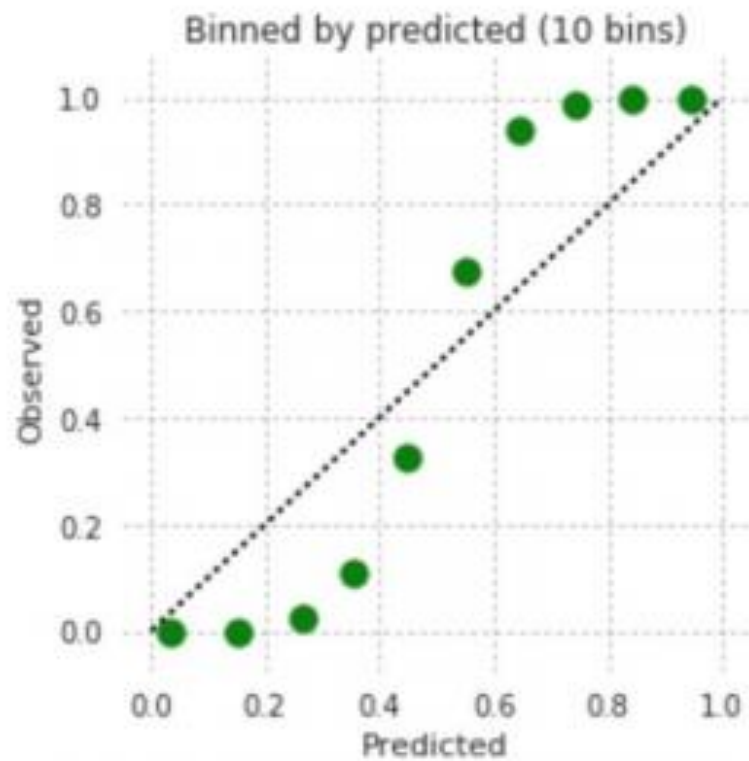


MODEL QUALITY - ДОПОЛНЕНИЕ

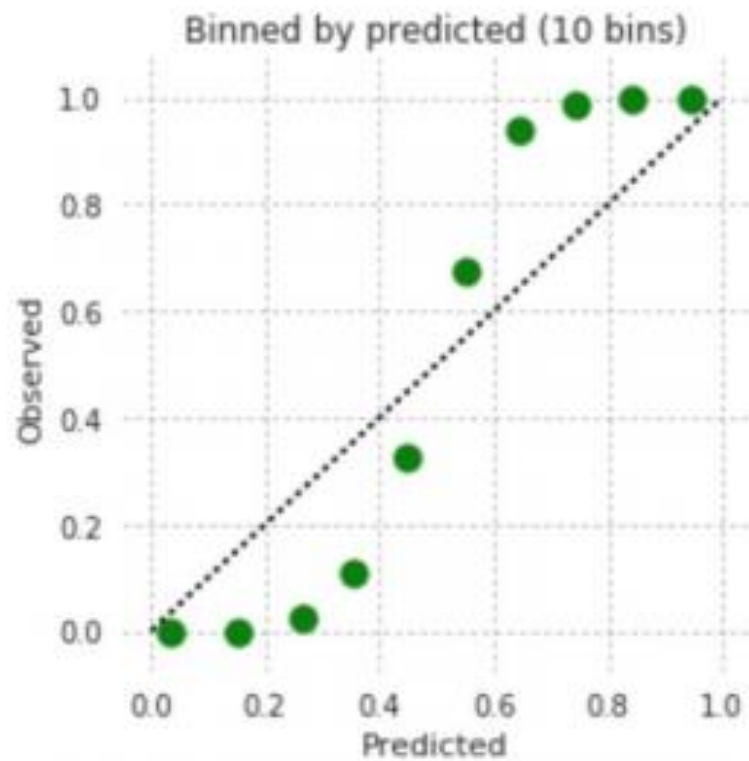


Normal Quantiles

MODEL QUALITY - ДОПОЛНЕНИЕ



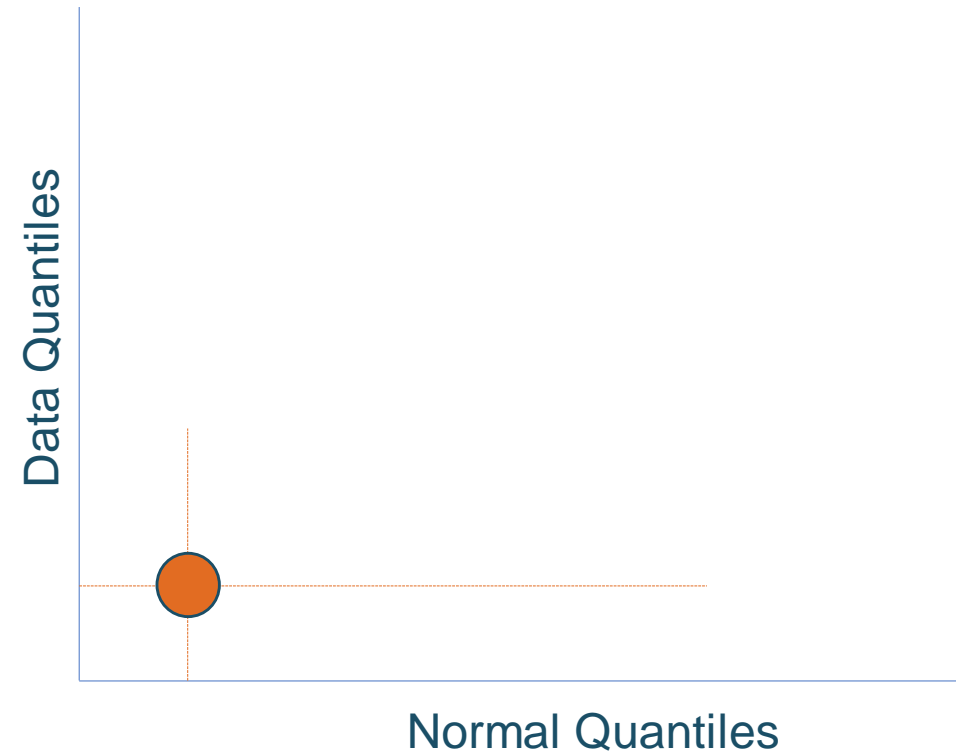
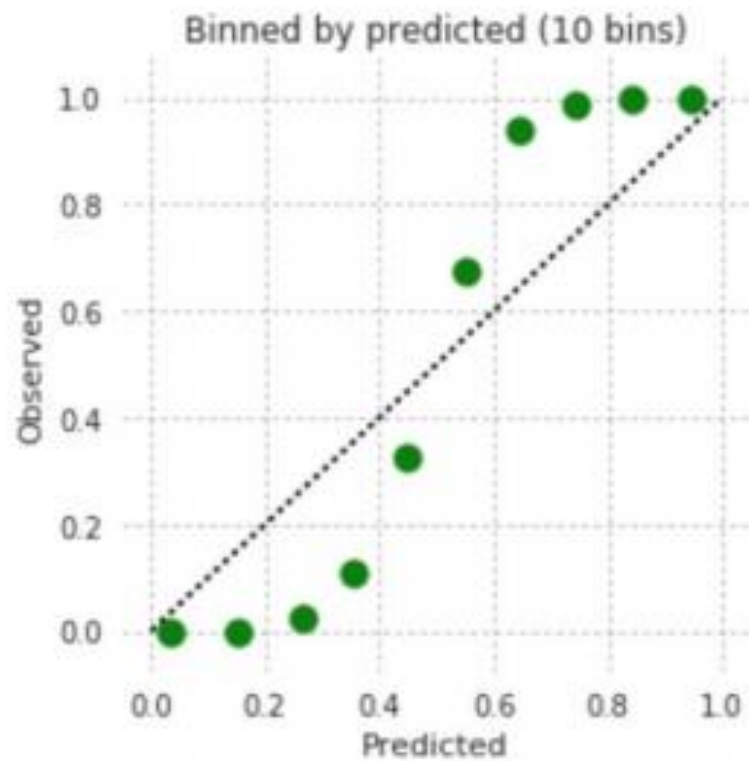
MODEL QUALITY - ДОПОЛНЕНИЕ



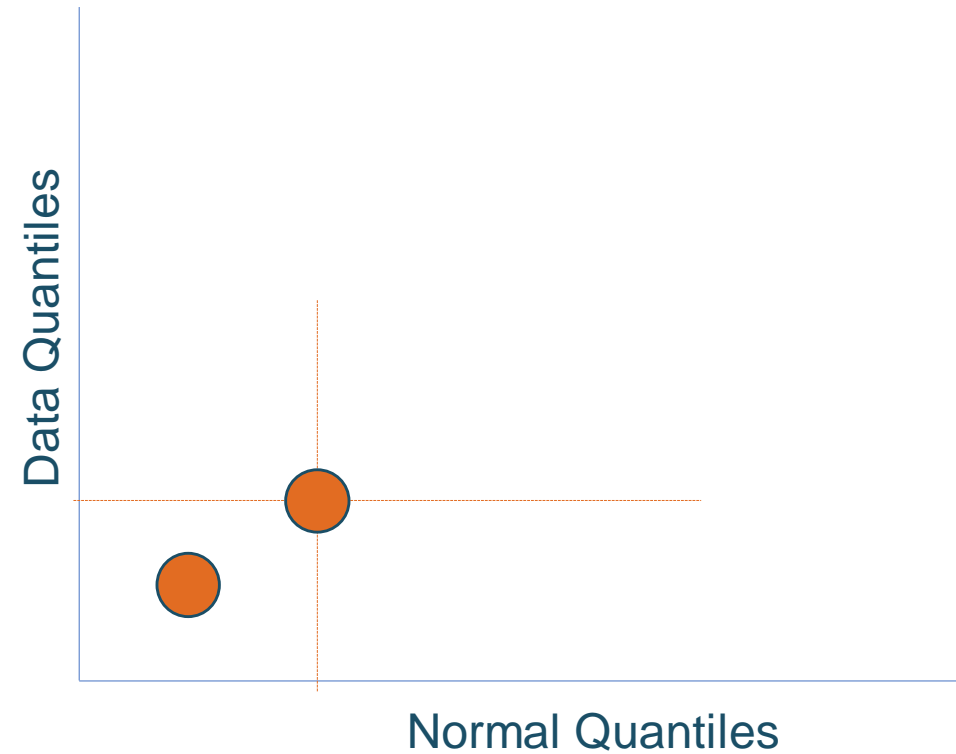
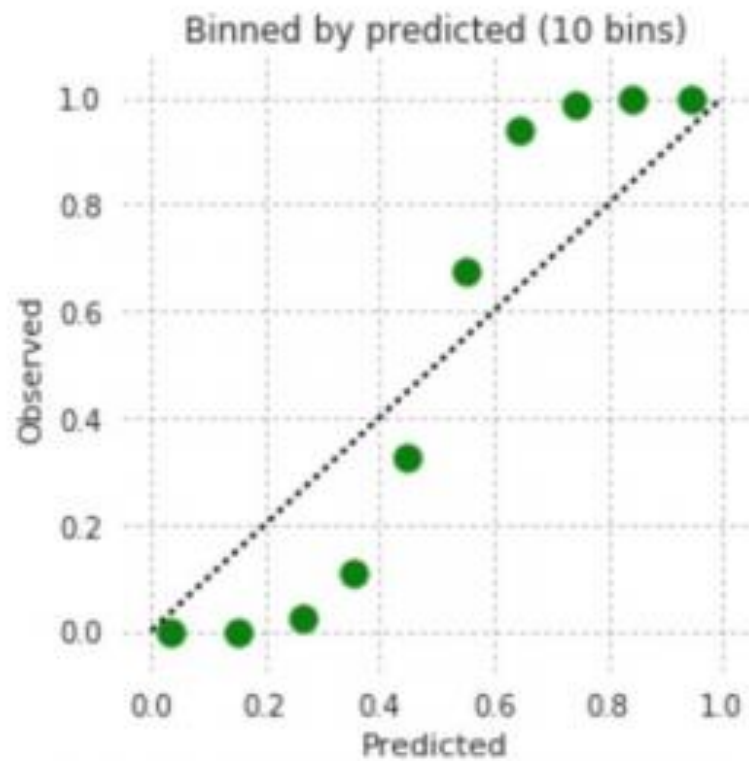
Data Quantiles

Normal Quantiles

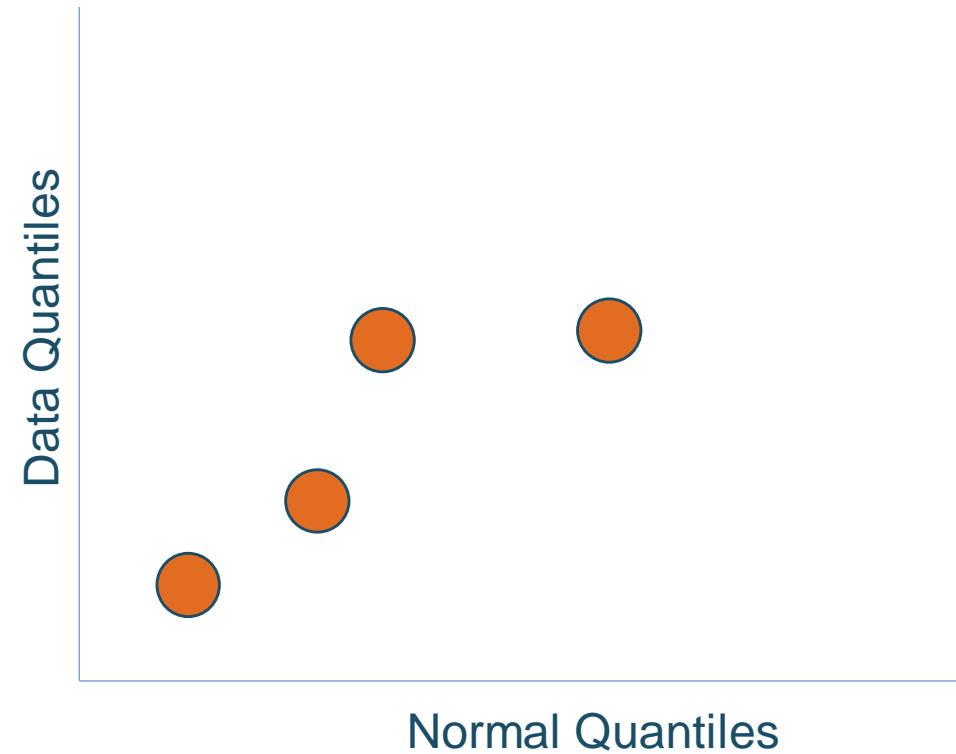
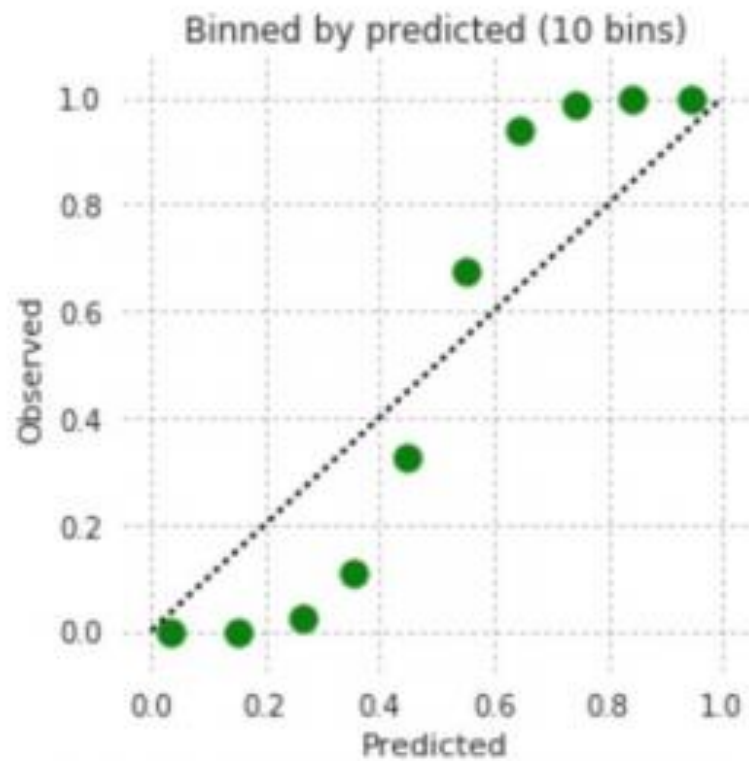
MODEL QUALITY - ДОПОЛНЕНИЕ



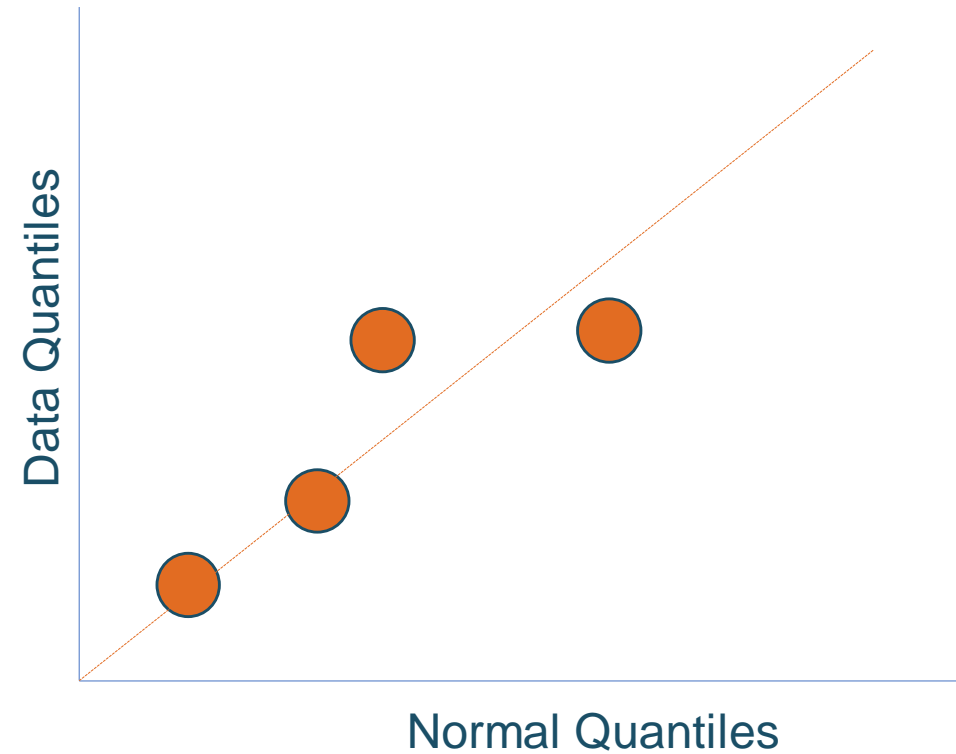
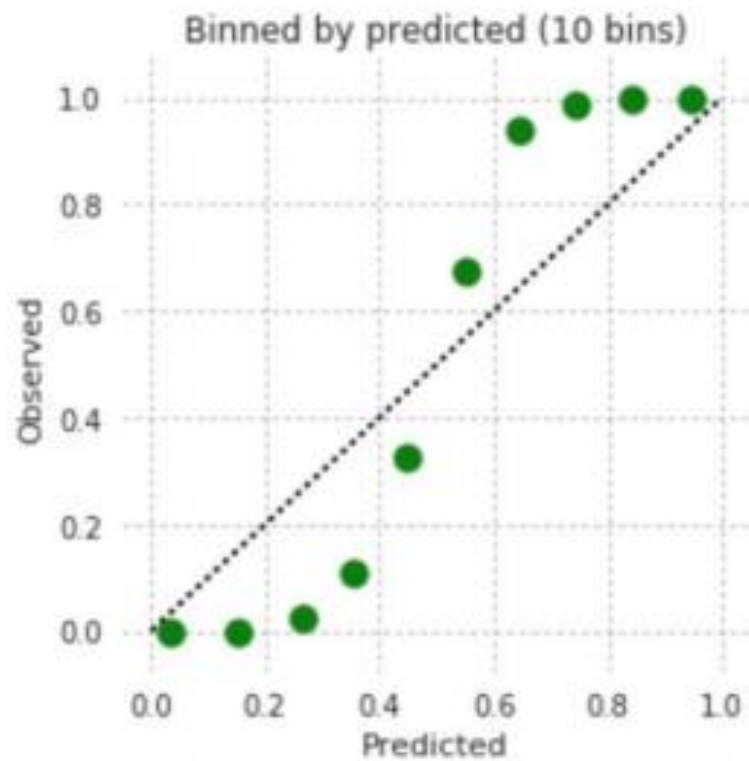
MODEL QUALITY - ДОПОЛНЕНИЕ



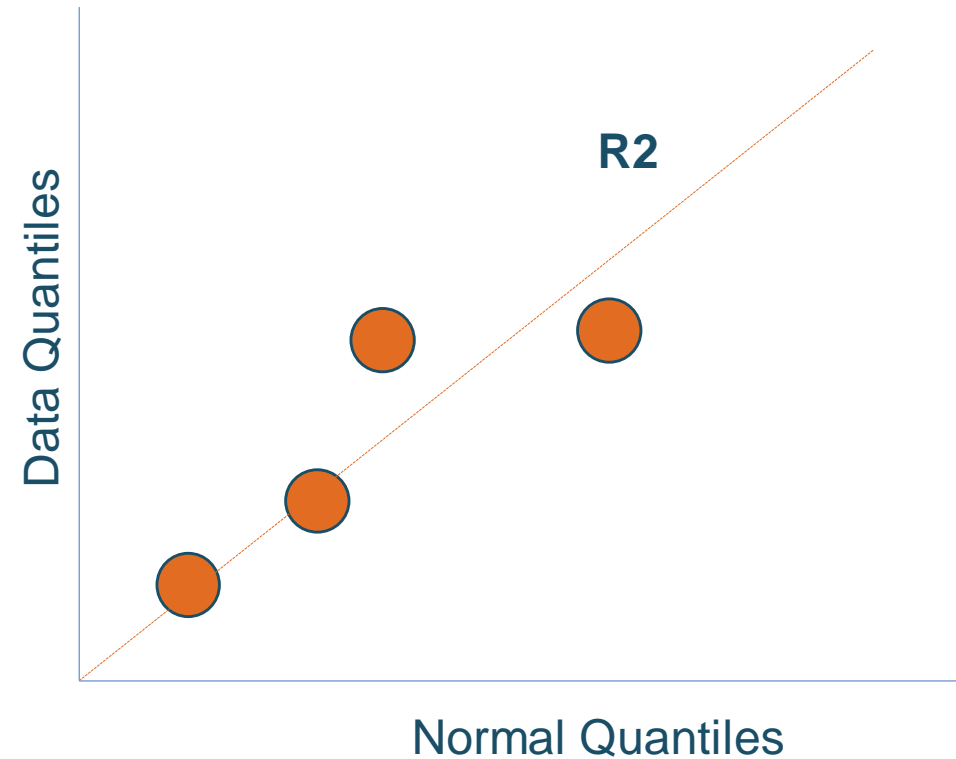
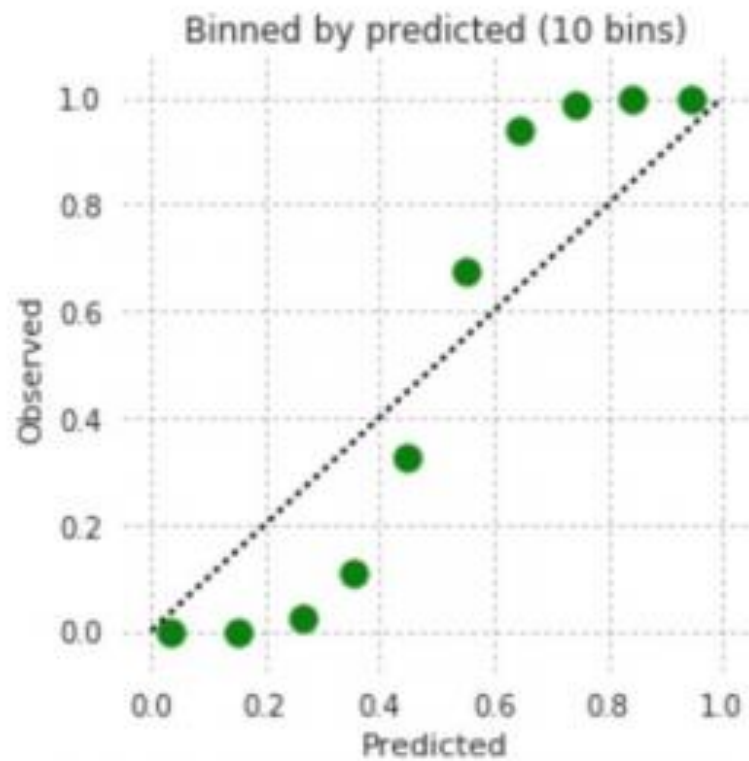
MODEL QUALITY - ДОПОЛНЕНИЕ



MODEL QUALITY - ДОПОЛНЕНИЕ



MODEL QUALITY - ДОПОЛНЕНИЕ



MQ – БОЛЬШЕ ДЛЯ РЕГРЕССИИ

SUMMARY OUTPUT

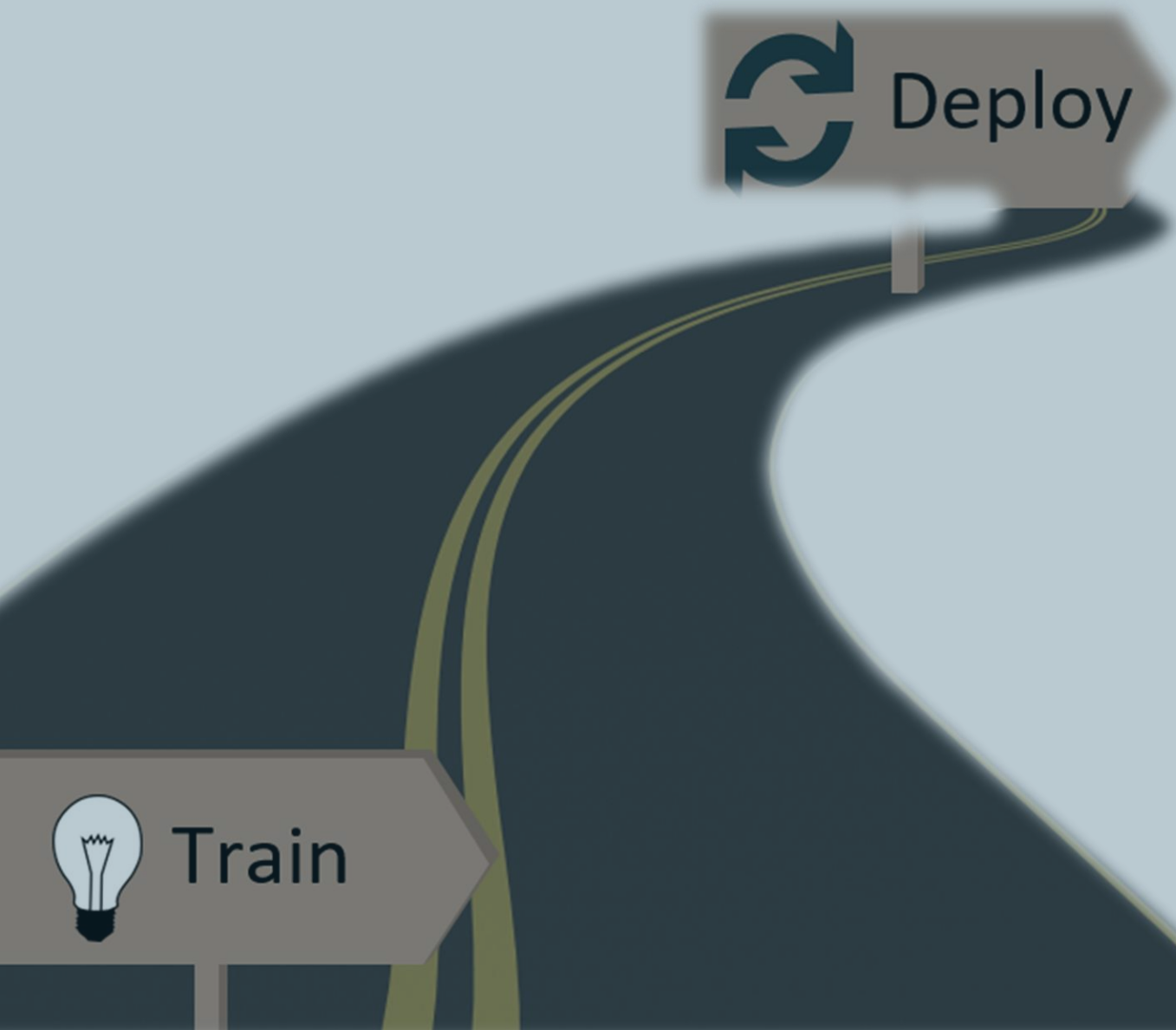
Regression Statistics	
Multiple R	0.987226765
R Square	0.974616686
Adjusted R Square	0.954310036
Standard Error	4.541359081
Observations	10

Sum of Squares Mean Sum of Squares F statistic

ANOVA

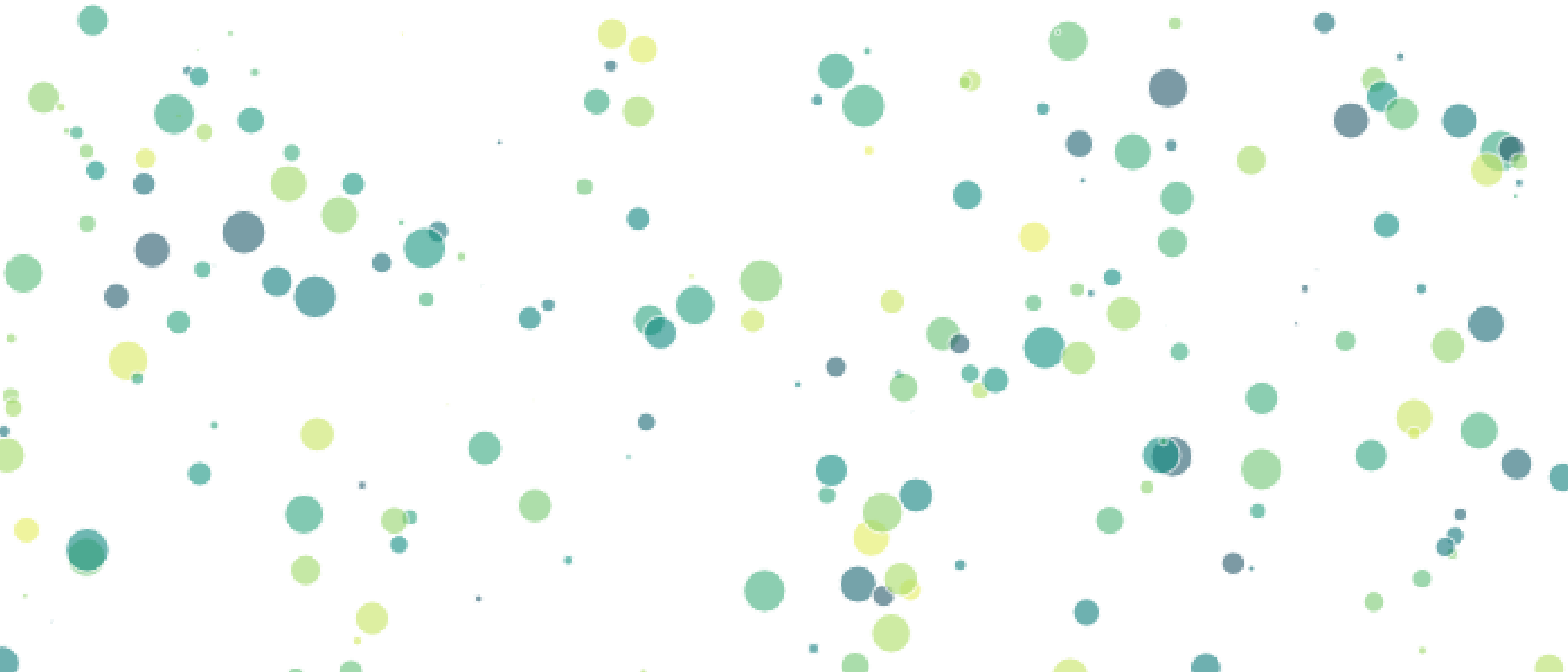
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	3959.4	989.8450721	47.99494964	0.000352771
Residual	5	103.1197115			
Total	9	4062.5			

Coefficients	Standard Error	t Stat	P-value	Interval	Lower 95%	Upper 95%
1.945880922	7.18396365	0.270864528	0.797321326	18.46696647	-16.52108554	20.41284739
8.039875895	3.545856848	2.267400022	0.072674984	9.114915204	-1.075039309	17.1547911
0.068115307	0.182217191	0.373813838	0.723869688	0.4684042	-0.400288893	0.536519507
0.816434539	0.096317968	8.476450995	0.000375456	0.24759322	0.568841319	1.06402776
8.683514224	3.505401878	2.477180799	0.056034438	9.010922394	-0.32740817	17.69443662



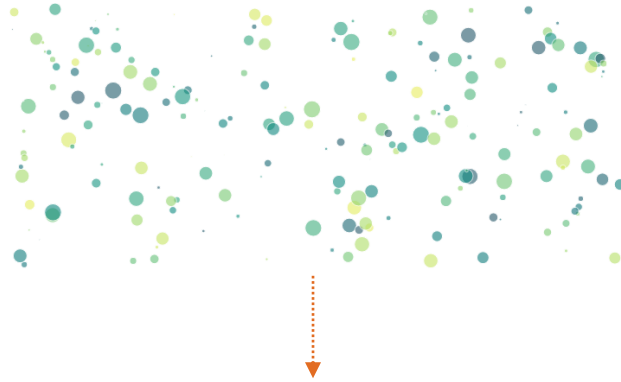
G/B DEPLOY

СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ



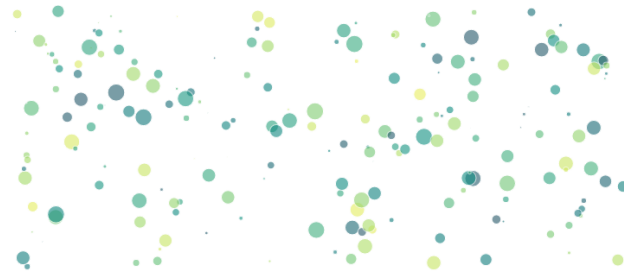
СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users



СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users

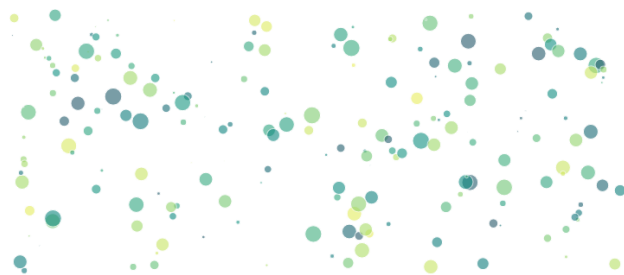


model



СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

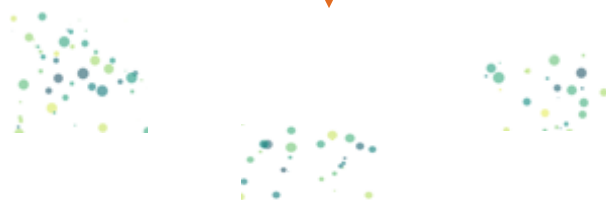
users



model

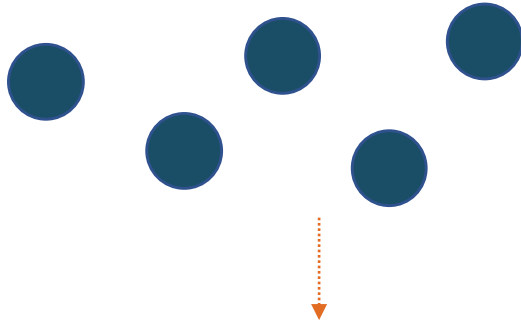


result



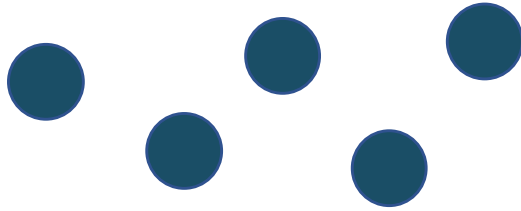
СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users

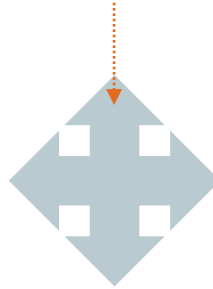


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users

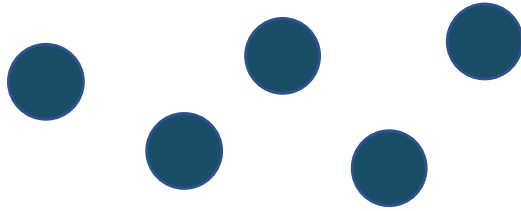


switch

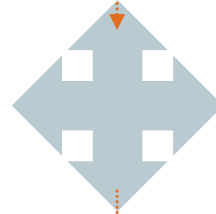


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

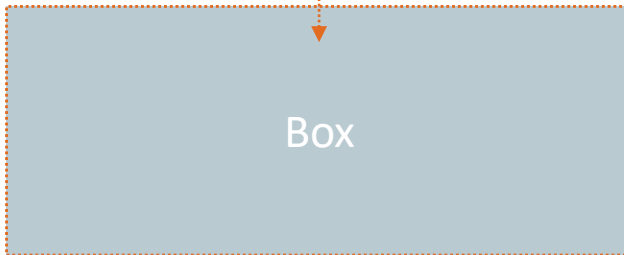
users



switch

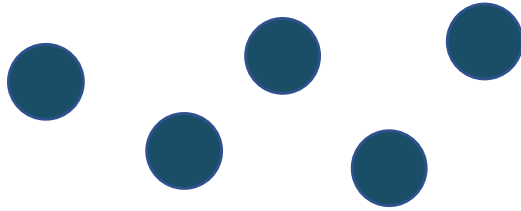


model

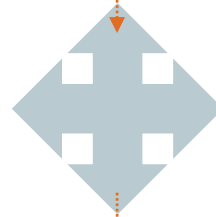


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

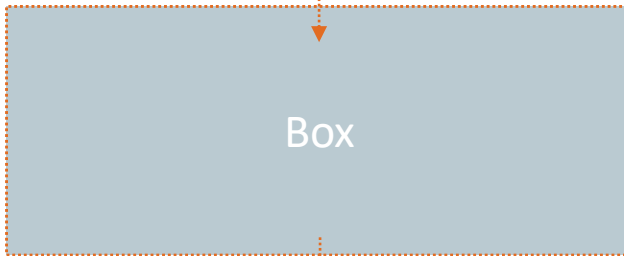
users



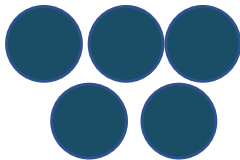
switch



model

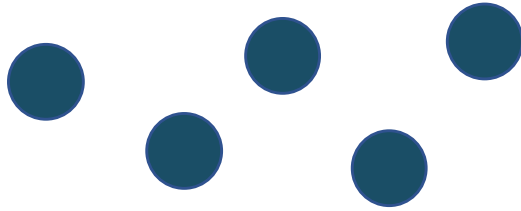


result

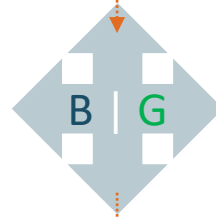


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users



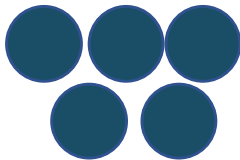
switch



model

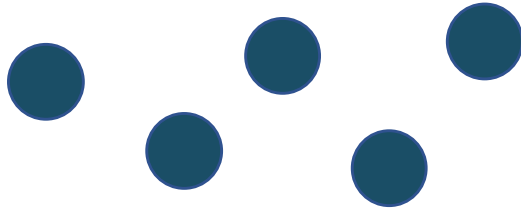


result

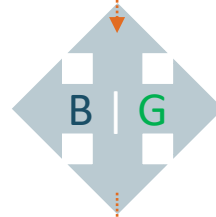


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users



switch



model

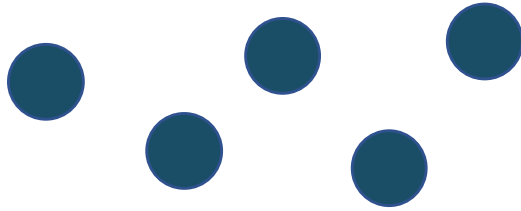


result

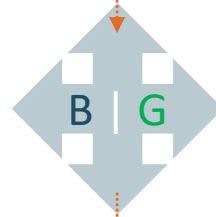


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users



switch



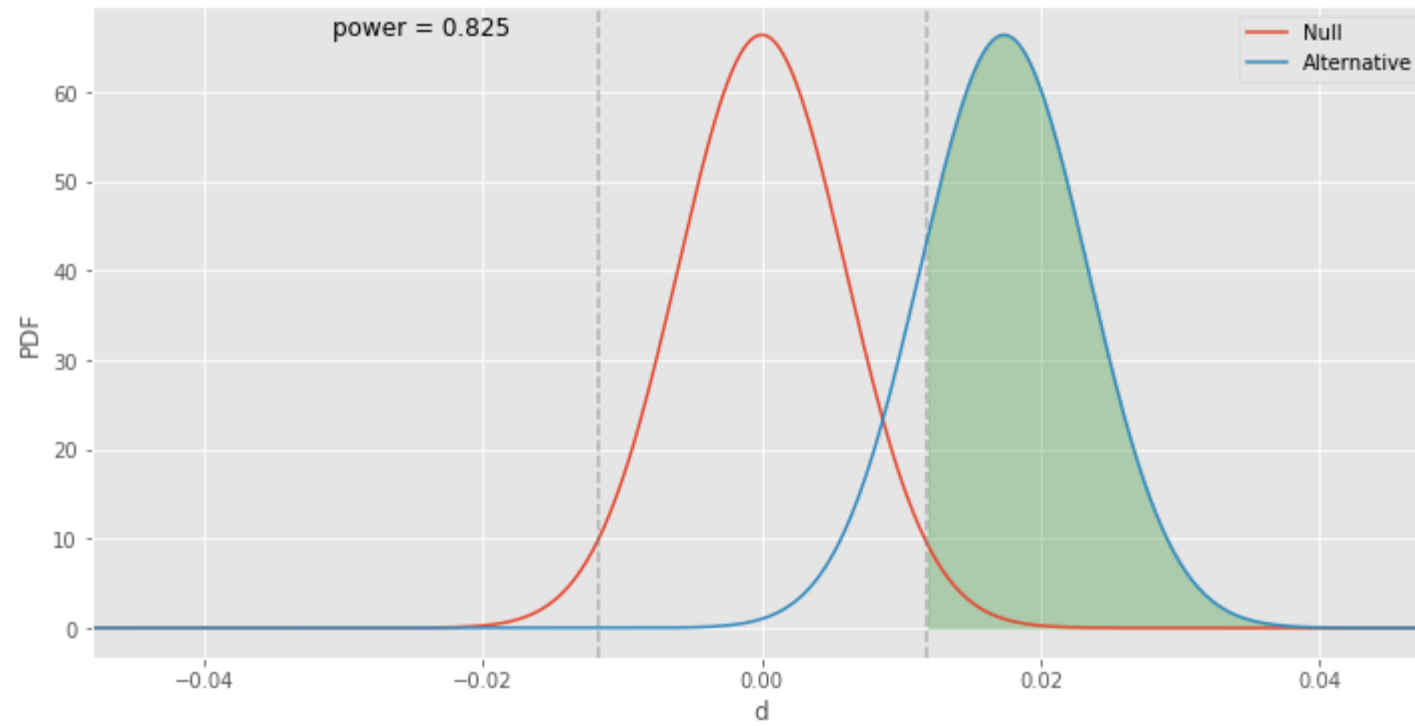
model



result + AB Test

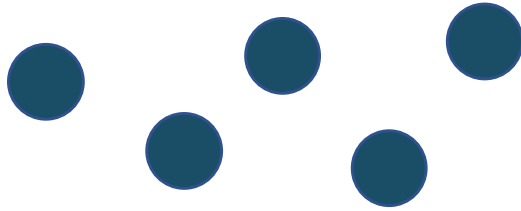


ГИПОТЕЗЫ – АВ ТЕСТЫ

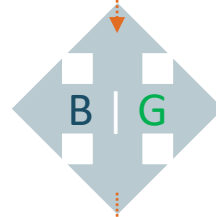


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

users



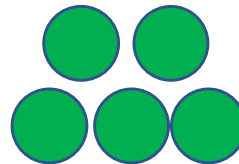
switch



model

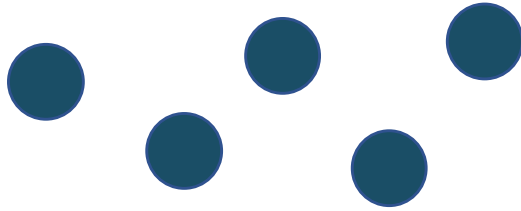


result + **True**

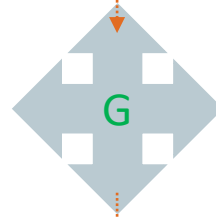


СИНЕ-ЗЕЛЕНЫЙ ДЕПЛОЙ

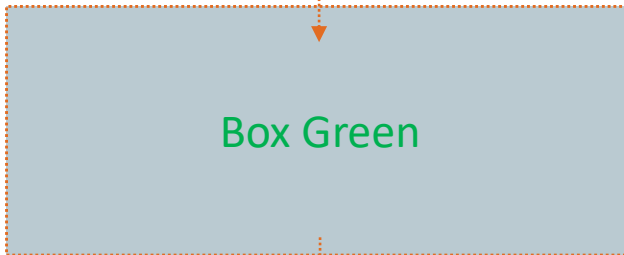
users



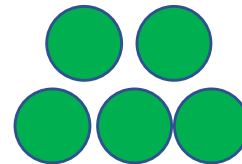
switch



model



result + **True**



МЕТРИКИ – ЭТО ПРОКСИ НА БИЗНЕС

AUC	z	d	r_{pb} $p = .50$	r_{pb} $p = .3 \& q = .7$	r_{pb} $p = .2 \& q = .8$	r_{pb} $p = .1 \& q = .9$
0.566	0.166	0.235	0.117	0.107	0.093	0.070
0.567	0.168	0.238	0.118	0.109	0.095	0.071
0.568	0.171	0.242	0.120	0.110	0.096	0.072
0.569	0.174	0.245	0.122	0.112	0.098	0.073
0.570	0.176	0.249	0.124	0.113	0.099	0.074
0.571	0.179	0.253	0.125	0.115	0.101	0.076
0.572	0.181	0.256	0.127	0.117	0.102	0.077
0.573	0.184	0.260	0.129	0.118	0.103	0.078
0.574	0.186	0.263	0.131	0.120	0.105	0.079
0.575	0.189	0.267	0.132	0.121	0.106	0.080
0.576	0.191	0.271	0.134	0.123	0.108	0.081
0.577	0.194	0.274	0.136	0.125	0.109	0.082
0.578	0.196	0.278	0.138	0.126	0.110	0.083
0.579	0.199	0.281	0.139	0.128	0.112	0.084
0.580	0.202	0.285	0.141	0.130	0.113	0.085
0.581	0.204	0.289	0.143	0.131	0.115	0.086
0.582	0.207	0.292	0.145	0.133	0.116	0.087
0.583	0.209	0.296	0.146	0.134	0.118	0.088
0.584	0.212	0.300	0.148	0.136	0.119	0.090
0.585	0.214	0.303	0.150	0.138	0.120	0.091
0.586	0.217	0.307	0.152	0.139	0.122	0.092
0.587	0.219	0.310	0.153	0.141	0.123	0.093
0.588	0.222	0.314	0.155	0.142	0.125	0.094
0.589	0.225	0.318	0.157	0.144	0.126	0.095
0.590	0.227	0.321	0.159	0.146	0.127	0.096
0.591	0.230	0.325	0.160	0.147	0.129	0.097

Score calibration #7

Open

NameArtem opened this issue now · 0 comments



NameArtem commented now

Калибровка скоринга

- Трансформация ROC_AUC
- Калибровка – общая информация
- Калибровка классификатора в вероятность
- Калибровка кредитного скоринга
- Калибровка для рекомендательных систем

MONITORING




МОНИТОРИНГ

- Данные в «песочнице» - не данные на проде


МОНИТОРИНГ

- Данные в «песочнице» - не данные на проде
- Ваша модель работает в «контексте», вы его не учитываете в разработке




1 из 6 фото

Стиль: классический, повседневный, деловой, спортивный, морской



1 из 6 фото

Стиль: классический, повседневный, деловой, спортивный, морской



1 из 1 фото

Стиль: классический, повседневный, деловой, спортивный, морской

☒ В продаже

Сезон

☐ весна
☐ зима
☐ лето
☐ осень

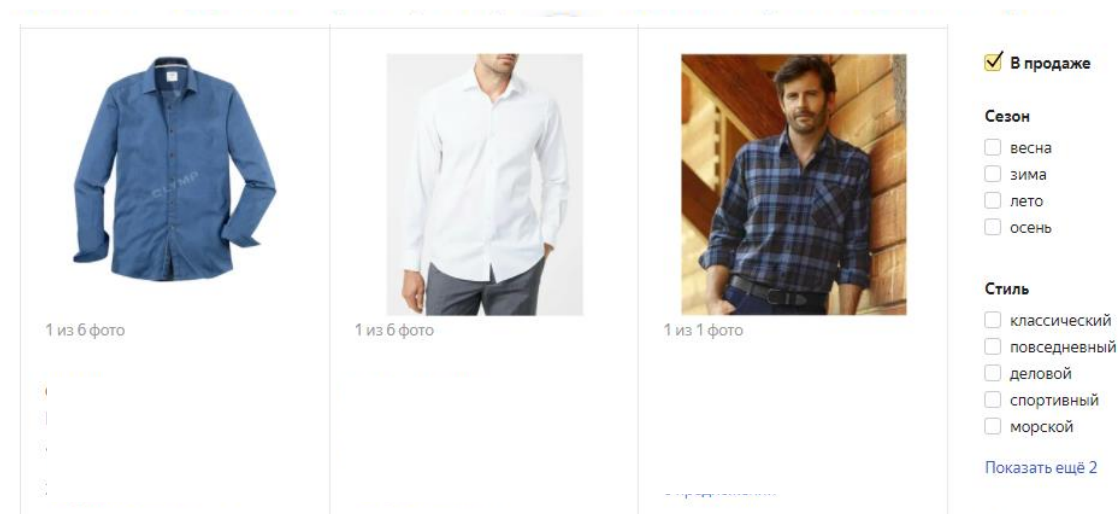
Стиль

☐ классический
☐ повседневный
☐ деловой
☐ спортивный
☐ морской

[Показать ещё 2](#)

МОНИТОРИНГ

- Данные в «песочнице» - не данные на проде
- Ваша модель работает в «контексте», вы его не учитываете в разработке

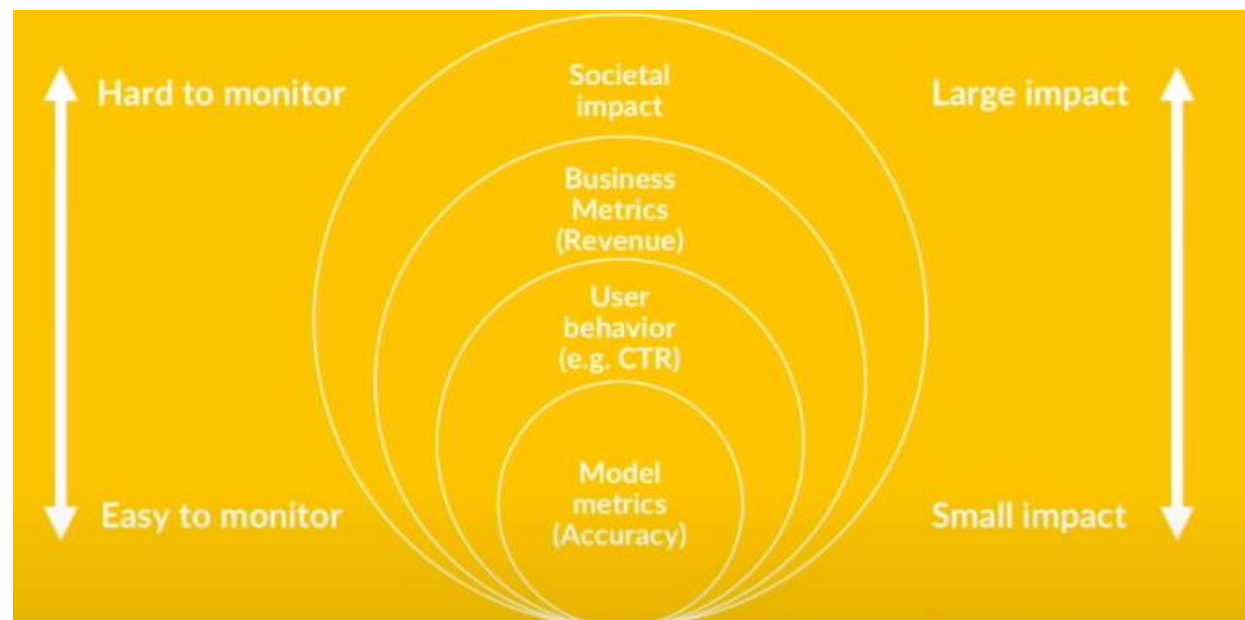


МОНИТОРИНГ

- Данные в «песочнице» - не данные на проде
- Ваша модель работает в «контексте», вы его не учитываете в разработке
- Изменение данных – это не мониторинг (train/live data = 1.56)

МОНИТОРИНГ

- Данные в «песочнице» - не данные на проде
- Ваша модель работает в «контексте», вы его не учитываете в разработке
- Изменение данных – это не мониторинг (train/live data = 1.56)
- Учитывать метрики от модели до пользователя



МОНИТОРИНГ - №1

Обучили:
RMSE – 1.25



МОНИТОРИНГ - №1

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 10.25

МОНИТОРИНГ - №1

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 10.25

- Ошибка в данных – PCI факта больше
- Неправильно сделан ETL для прода (если переписывали код)

МОНИТОРИНГ - №1

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 10.25

- Ошибка в данных – PCI факта больше
- Неправильно сделан ETL для прода (если переписывали код)
- Библиотеки!

МОНИТОРИНГ - №1

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 10.25

```
X_train['Hour_cosine'].std()
```

```
0.7206233339256797
```

```
np.std(X_train['Hour_cosine'].values)
```

```
0.7206224945074595
```

МОНИТОРИНГ - №2

Обучили:
RMSE – 1.25



МОНИТОРИНГ - №2

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 5.25

МОНИТОРИНГ - №2

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 5.25

- История данных на выборке – не вся история
- Агрегация пользователя в виде группы - GroupKFold

МОНИТОРИНГ - №2

Обучили:
RMSE – 1.25

АБ Тест:
RMSE – 5.25

- История данных на выборке – не вся история
- Агрегация пользователя в виде группы – GroupKFold
- Работа - не Kaggle!

МОНИТОРИНГ - МЕТРИКИ

		True condition			
Total population		Condition positive	Condition negative	$Prevalence = \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	$Accuracy (ACC) = \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$
Predicted condition	Predicted condition positive	True positive , Power	False positive , Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$
	Predicted condition negative	False negative , Type II error	True negative	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{TPR}{FPR}$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{FNR}{TNR}$	
				$F_1 \text{ score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$	

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted	0	1
Actual		
0	5	2
1	0	3

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted	0	1
Actual		
0	5	2
1	0	3

MCC - 0.654653

F1 - 0.7499

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted	1
Actual	
0	2
1	3

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted 1

Actual

0 2

1 3

MCC - 0.0

F1 - 0.7499

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted 0 1

Actual

0	6	2
1	1	1

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted 0 1

Actual

0	6	2
1	1	1

MCC - 0.218217

F1 - 0.4

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted	0	1
Actual		
0	1	2
1	1	6

МОНИТОРИНГ - МЕТРИКИ

Matthews correlation coefficient

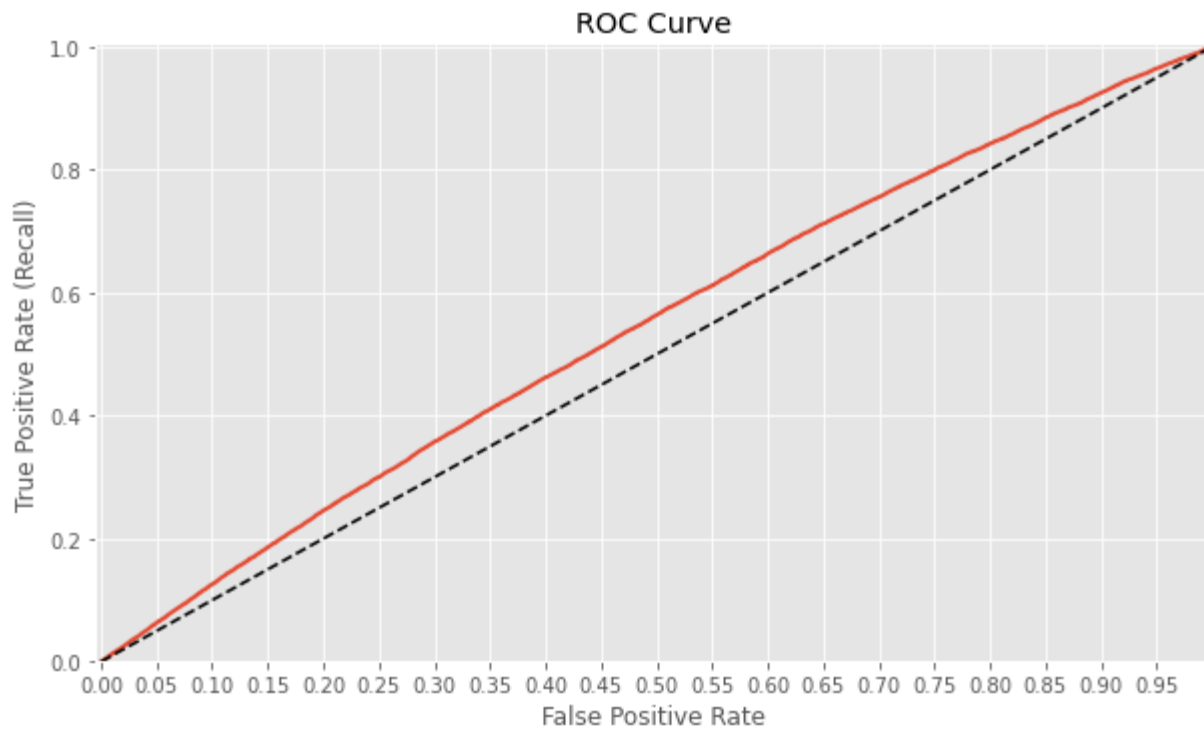
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Predicted	0	1
Actual		
0	0	1
1	1	6

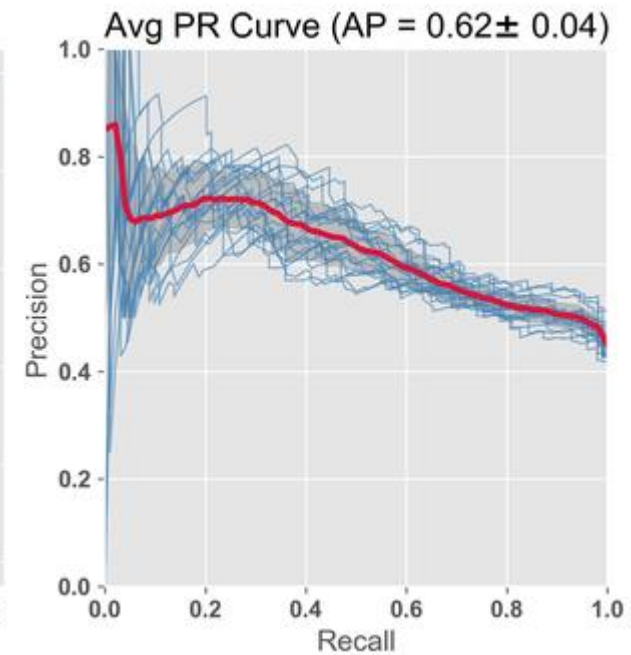
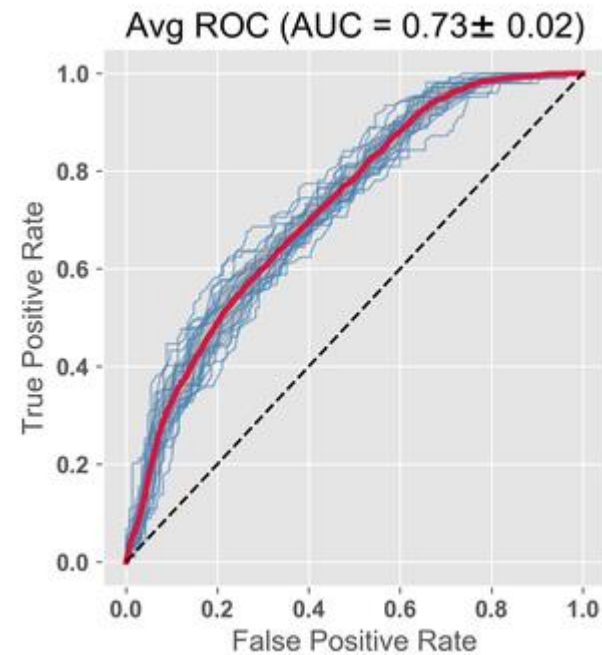
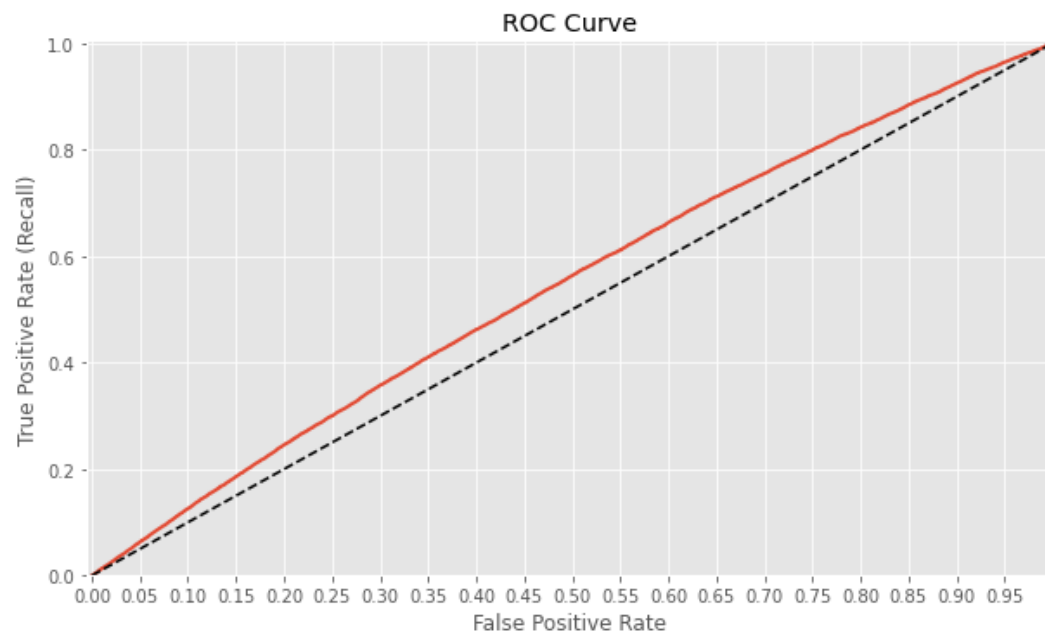
MCC - 0.218217

F1 - 0.7499

МОНИТОРИНГ - AUC



МОНИТОРИНГ - AUC



МОНИТОРИНГ - MMWA

Predicted	0	1	
Actual	0	1	2
1	1	6	

МОНИТОРИНГ - MMWA

Predicted	0	1
Actual		
0	1	2
1	1	6

```
f1_score(df['y_Actual'], df['y_Predicted'])
```

```
0.7999999999999999
```

```
f1_score(df['y_Actual'], df['y_Predicted'], average='macro')
```

```
0.6
```

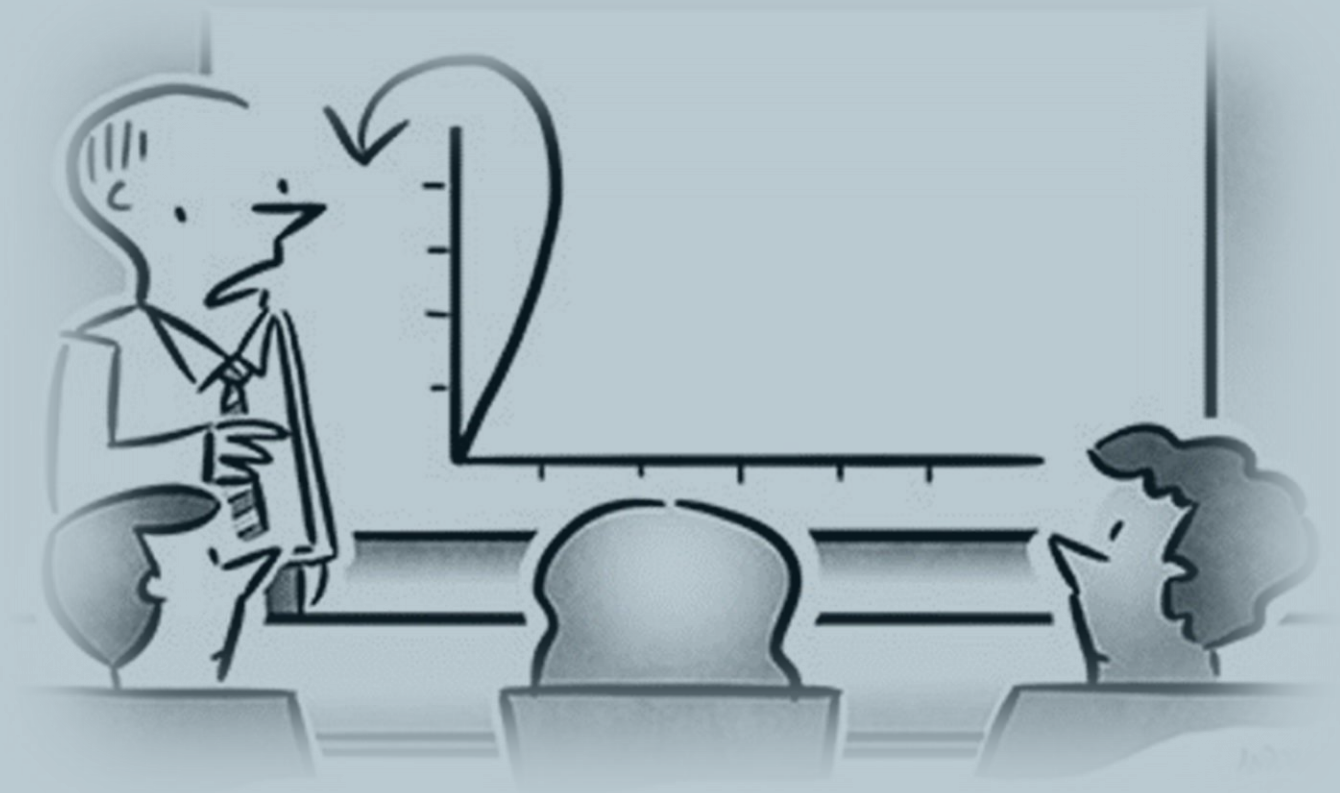
```
f1_score(df['y_Actual'], df['y_Predicted'], average='micro')
```

```
0.7
```

```
f1_score(df['y_Actual'], df['y_Predicted'], average='weighted')
```

```
0.6799999999999999
```

STREAMLIT



FINAL

ЭТО БЫЛО ОТЛИЧНОЕ ПУТЕШЕСТВИЕ!



1.0.0-beta1
Major Minor Patch Label



ЗАДАНИЕ



ОСТАВЬТЕ ОБРАТНУЮ СВЯЗЬ

- Это был первый запуск курса, помогите сделать лучше и интересней.
Заполните форму обратной связи:

<https://forms.gle/tjVHYVzB258bAj9n6>