

TEAM 7

X.X. и в продакшн © HTP

PRESENTED BY:

Наталья Гурьева Вероника Карпушенкова

ПРЕДОБРАБОТКА ДАННЫХ

- 1. убрали ненужные колонки
- 2. восполнили некоторые значения
- 3. удалили NαN там, где не смогли восполнить
- 4. замерджили датасеты
- 5. создали сабсет с числовым типом данных
- 6. пронормализовали
- 7. добавили дескрипторы
- 8. удалили сильно коррелирующие столбцы

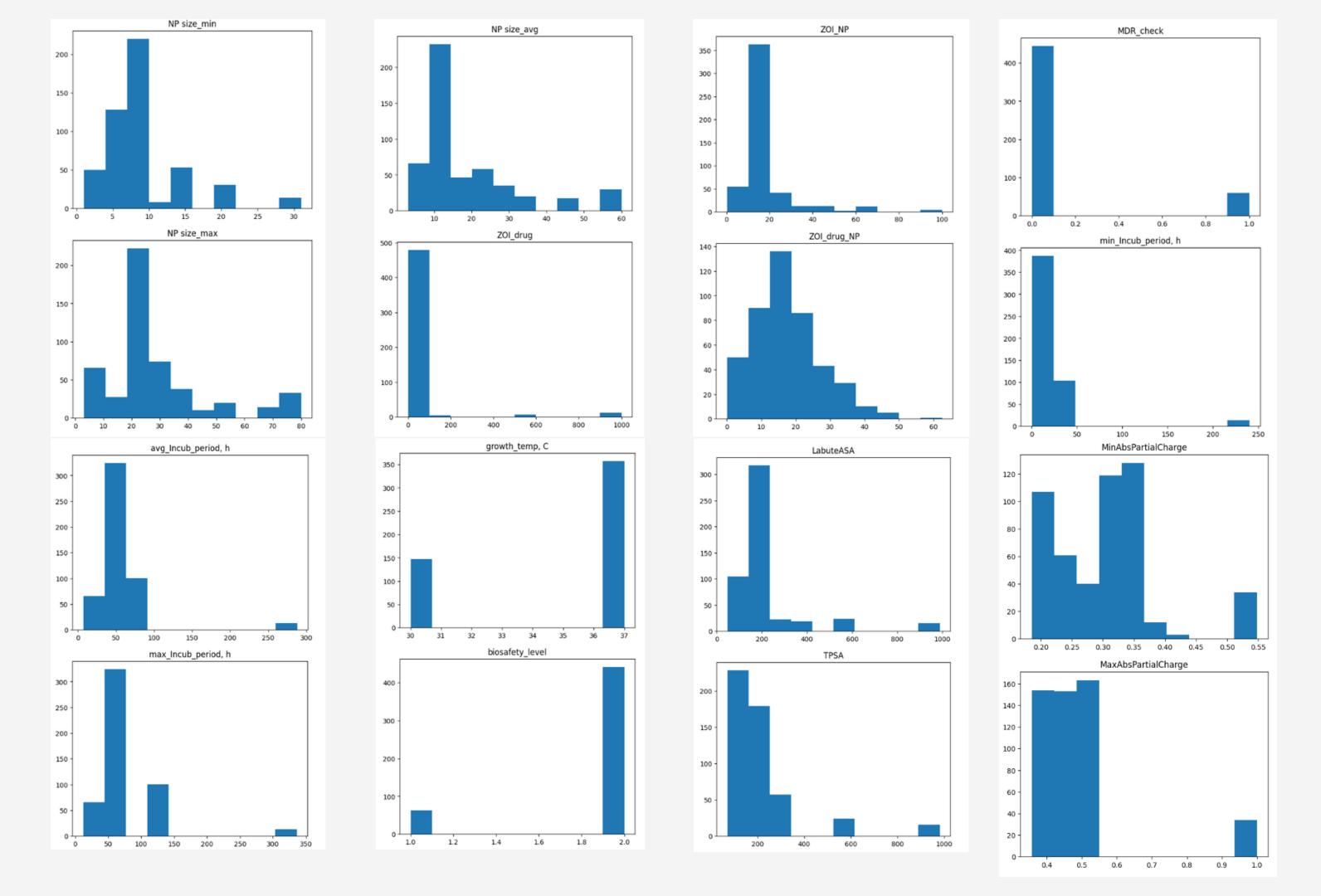
Трудом и Потом

SlovoDna

(c.) как я добиваюсь своих целей

SlovoDna*

SlovoDna.ru



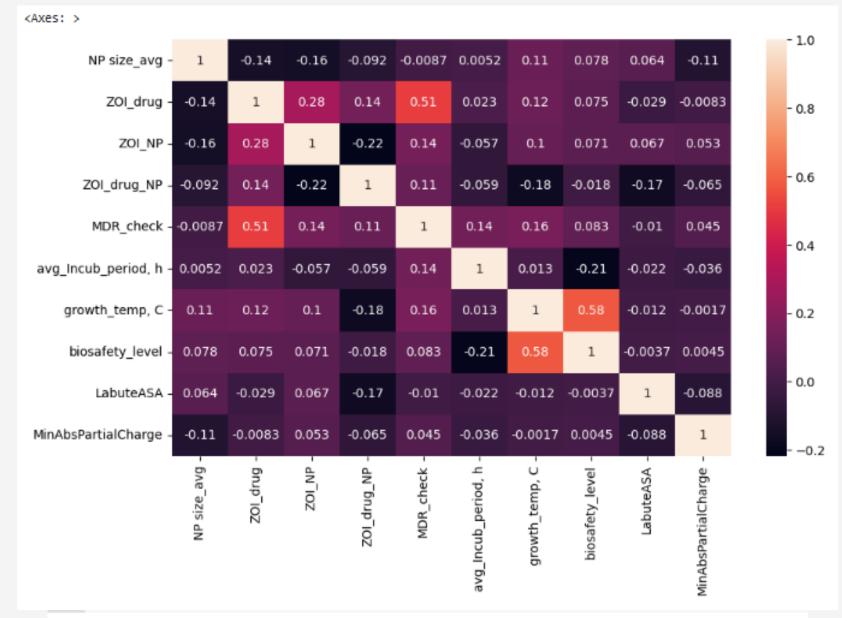
0

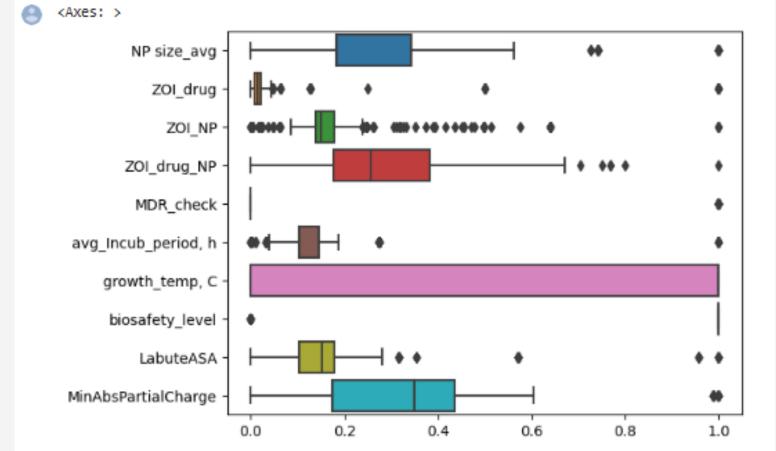
num_subset.info()

```
C < class 'pandas.core.frame.DataFrame'>
  RangeIndex: 504 entries, 0 to 503
  Data columns (total 16 columns):
    # Column Non-Null Count Dtype
```

```
NP size_min
                                     float64
                     504 non-null
                                     float64
NP size max
                     504 non-null
                                     float64
NP size_avg
                     504 non-null
ZOI_drug
                                     float64
                     504 non-null
                                     float64
ZOI NP
                     504 non-null
                                     float64
ZOI_drug_NP
                     450 non-null
MDR_check
                                     float64
                     504 non-null
                                     float64
min_Incub_period, h 504 non-null
avg_Incub_period, h 504 non-null
                                     float64
max_Incub_period, h 504 non-null
                                     float64
growth_temp, C
                                     float64
                     504 non-null
                                     float64
biosafety level
                     504 non-null
LabuteASA
                     504 non-null
                                     float64
                                     float64
TPSA
                     504 non-null
MinAbsPartialCharge 504 non-null
                                     float64
MaxAbsPartialCharge 504 non-null
                                     float64
```

dtypes: float64(16) memory usage: 63.1 KB





1.0		•	es esses	e) (0.000000		· · ·				• • ••
0.8		8.								
NP size_avg		.		± 1 ·	:]: :].	*** *	:
0.2		\$ \$	• (ess) • (ess) • (ess)	elificatification of grants or o					***** ********************************	1
0.0	بالمالمالية	***	(S C(E) 0 0						\$, \$7	
0.8					•	1	•	•	•	• •
6nup 10Z						1				
R 0.4						[
0.2			addition family	mot, analiseelkorougestinin en		20 20 2].].	. 44B st	8 8 49 W 2 BBB 8 8 B
		•		60	•		1		· · · · · · · · · · · · · · · · · · ·	
0.8						1				
dN 0.6		1		-,		• • • •].			
0.2				de santa de la compansa de la compan		00000000000000000000000000000000000000			**************************************	
0.0	1	**	ļ.,	8					19	
0.8		,			•]		٠
dN 0.6			********					i	***	
0.4 0Z	Sa a 40	400 400 100 100							0-0-1-1 	410 - 410 -
0.2			all'a Basson			8 476]		200 000 000 000 000 000 000 000 000 000
		••••		• @@## • 0					. 01033	• • • • • • •
0.8						†		-		
0.6 0.4 0.4]]]]		
0.2		-				-	-			
0.0				(808-(803818888(81))))))))))))))))))))))))))))))			·	1.	010000 000 0 00	
£ 0.8]]			
o.e						-				
o.4 Jucap		•>	(D 00(KBDD) (D 0	-		1			• 0000000 • • • •	• • • • • • • • • • • • • • • • • • •
0.2		· · · ·	83 (201108) 0 0 0 0 (201108) 0 0 0 (201108) 0 0 0 0 0 (201108) 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	000 0 0 000 0000 000 00 000 00000 00 00 000 00000 00 00 000 00000 00		1,11]:	200 00 00 00 00 00 00 00 00 00 00 00 00	::22 222 2 2
1.0			((CEXEXE O O O O O O O O O O O O O O O O O	(8)41-(1838)8890(898)110-00110)	•	- 000	1	•	• =====================================	
0.8						†				
o.6 0.4]				
B. 0.2						-	-			
0.0				**************************************			-			
0.8				***************************************	•				• • • • • • • • • • • • • • • • • • • •	
0.6						-				
9 0.4 9 0.4		-				1	-	-		
0.2			0:000	(EXDIBITION 00 00]	1]			
1.0		:		60 6100000	:		•	•		:
0.8		-				1	-			
LabuteASA 90 90		•	00 (00)	e emmores er].]		•
0.2	·			estamo e es estamo es es estamo es estamo es estamo es estamo es estamo estamo estamo es estamo esta		00 00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	11 1		.1	
0.0	ļ. • • · · · · · · · · · · · ·	-	###### \$ ·	o describe a sada	· · · · · · · · · · · · · · · · · · ·		<u>.</u>	-		
0.8			• (435)• • • •	en e entition e este	•	00 000 0		1	•	
tialCharge			*Gino •	* * ***					••	
inAbsPart		(10) • (1	Steam o	6 0 (5) (1(6)00 (3) (3) (3) (3) (4) (4) (4) (4) (4) (4) (4) (4) (4) (4		5			**************************************	
¥ 0.2			60 0000 0 0 0000 0 0 0000 0 0 0000 0 0 0000			00 00 00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			· ·	
		0.0 0.2 0.4 0.6 0.8 1.0 ZOI_drug	0.0 0.2 0.4 0.6 0.8 1.0 ZOI_NP	0.0 0.2 0.4 0.6 0.8 1.0 ZOI_drug_NP		0.0 0.2 0.4 0.6 0.8 1.0 avg_Incub_period, h	0.0 0.2 0.4 0.6 0.8 1.0 growth_temp, C	0.0 0.2 0.4 0.6 0.8 1.0 biosafety_level	0.0 0.2 0.4 0.6 0.8 1.0 LabuteASA	0.0 0.2 0.4 0.6 0.8 1.0 MinAbsPartialCharge



ВЫБОР ДЕСКРИПТОРОВ

Доказано, что антибактериальная активность зависит от размера, площади и заряда поверхности частицы - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC826 8496/

Добавили дескрипторы этих свойств: LabuteASA, TPSA, MinAbsPartialCharge, MaxAbsPartialCharge

ОБУЧЕНИЕ МОДЕЛЕЙ

Взяли три модели:

- 1) линейная регрессия
- +: простота, интерпретируемость
- 2) градиентный бустинг
- +: высокая точность, умение работать с различными типами признаков, адаптивность
- 3) случайные леса
- +: устойчивость к выбросам и шумам

Ща всё будет

(ф.) "будет, но не ща..."

SlovoDna* | SlovoDna.ru

ЛИНЕЙНАЯ РЕГРЕССИЯ

Cross-validated scores:

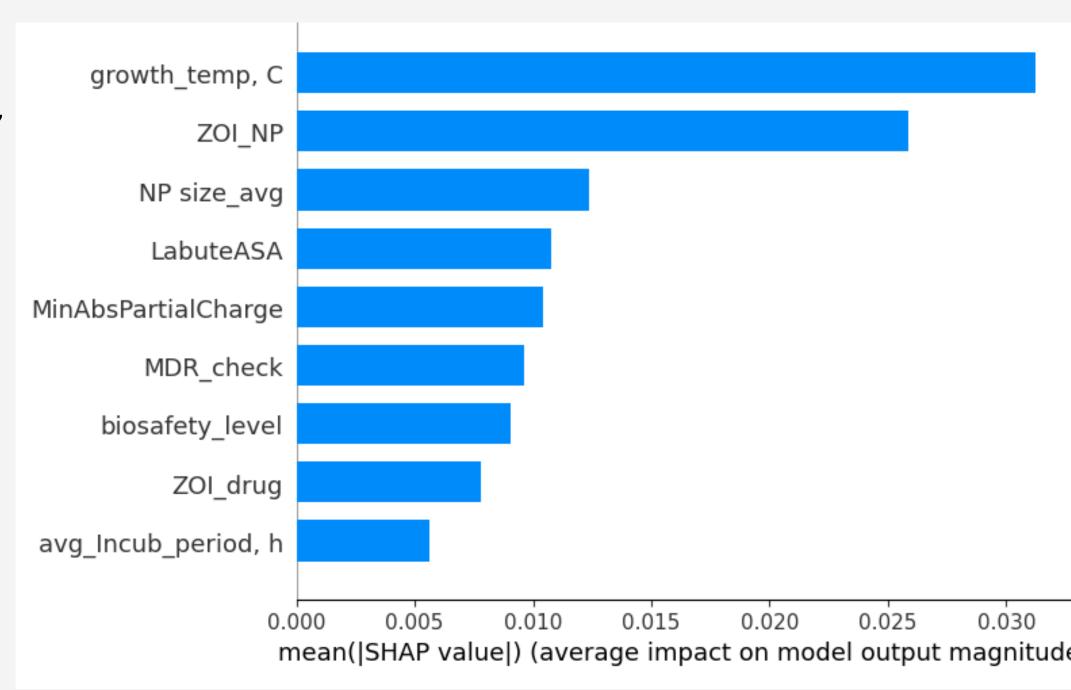
[0.17258129 0.15172355 0.02948637 0.21808306 0.02695485 0.01154977]

Mean absolute error: 0.12

Mean squared error: 0.02

Root mean squared error: 0.15

r2_score: **0.25**



ГРАДИЕНТНЫЙ БУСТИНГ

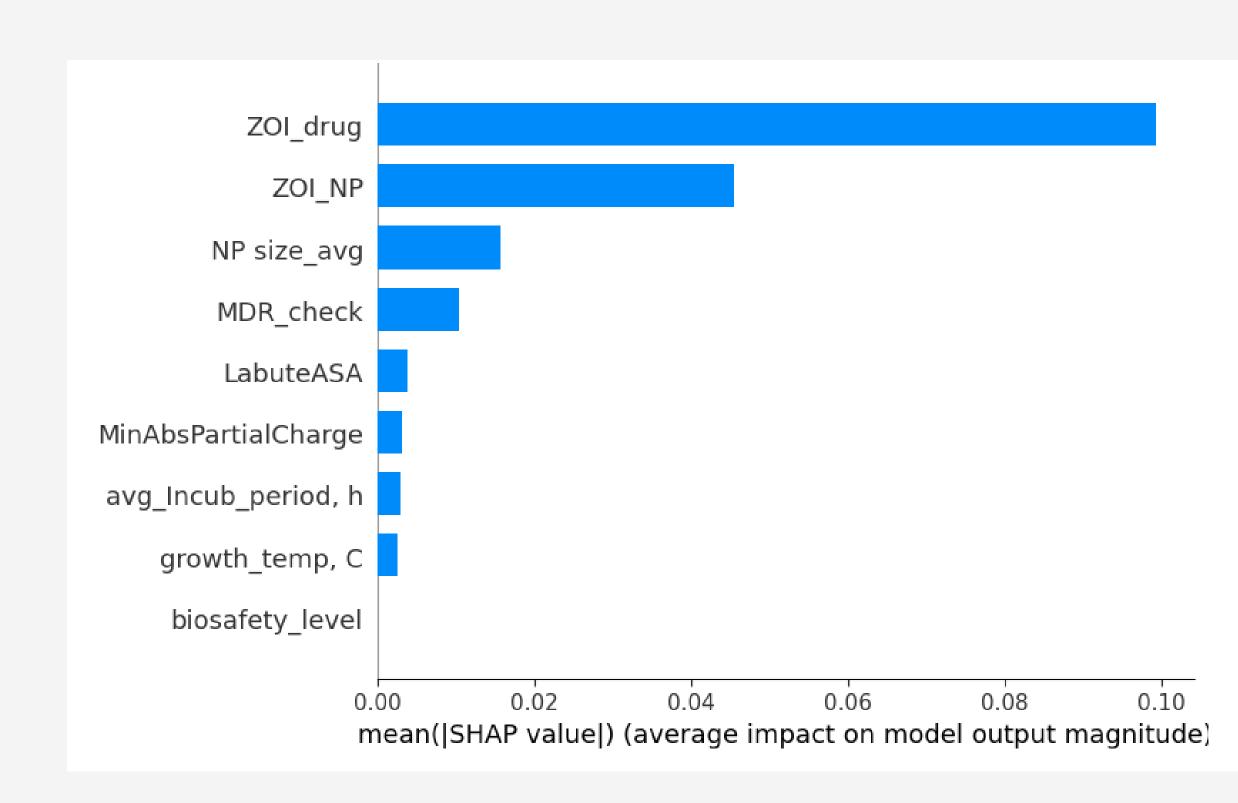
Cross-validated scores:

[0.8005852 0.81062148 0.78668551 0.73552291 0.8993295 0.68110942]

R-squared: **0.787**

MAE: **0.048**

MSE: **0.00597**

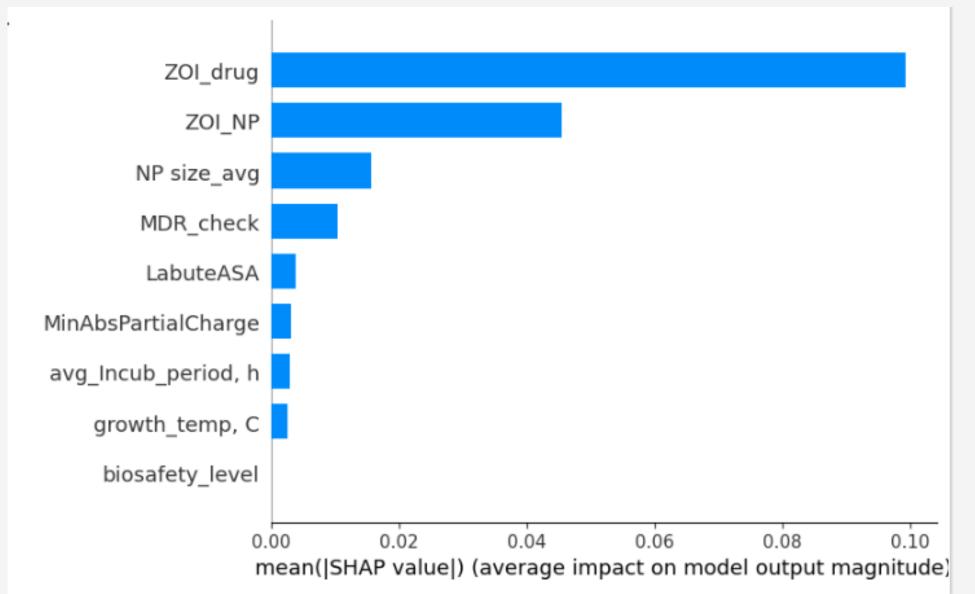


СЛУЧАЙНЫЕ ЛЕСА

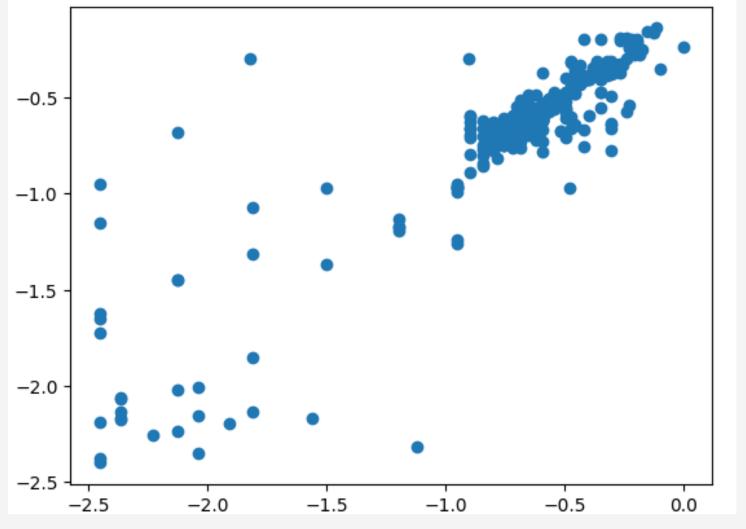
R-squared: **0.794**

MAE: **0.044**

MSE: **0.00580**







TEAM 7

PRESENTED BY:

THE PARTIE ЗАПАТЬ BOIPOC

Наталья Гурьева Вероника Карпушенкова

