

Искусственный интеллект в химии

Модель прогнозирования ZOI

DataCon, 2023

Работа команды №9
Гавенчук Антон
Головкин Иван
Смирнова Алина
Чистяков Денис

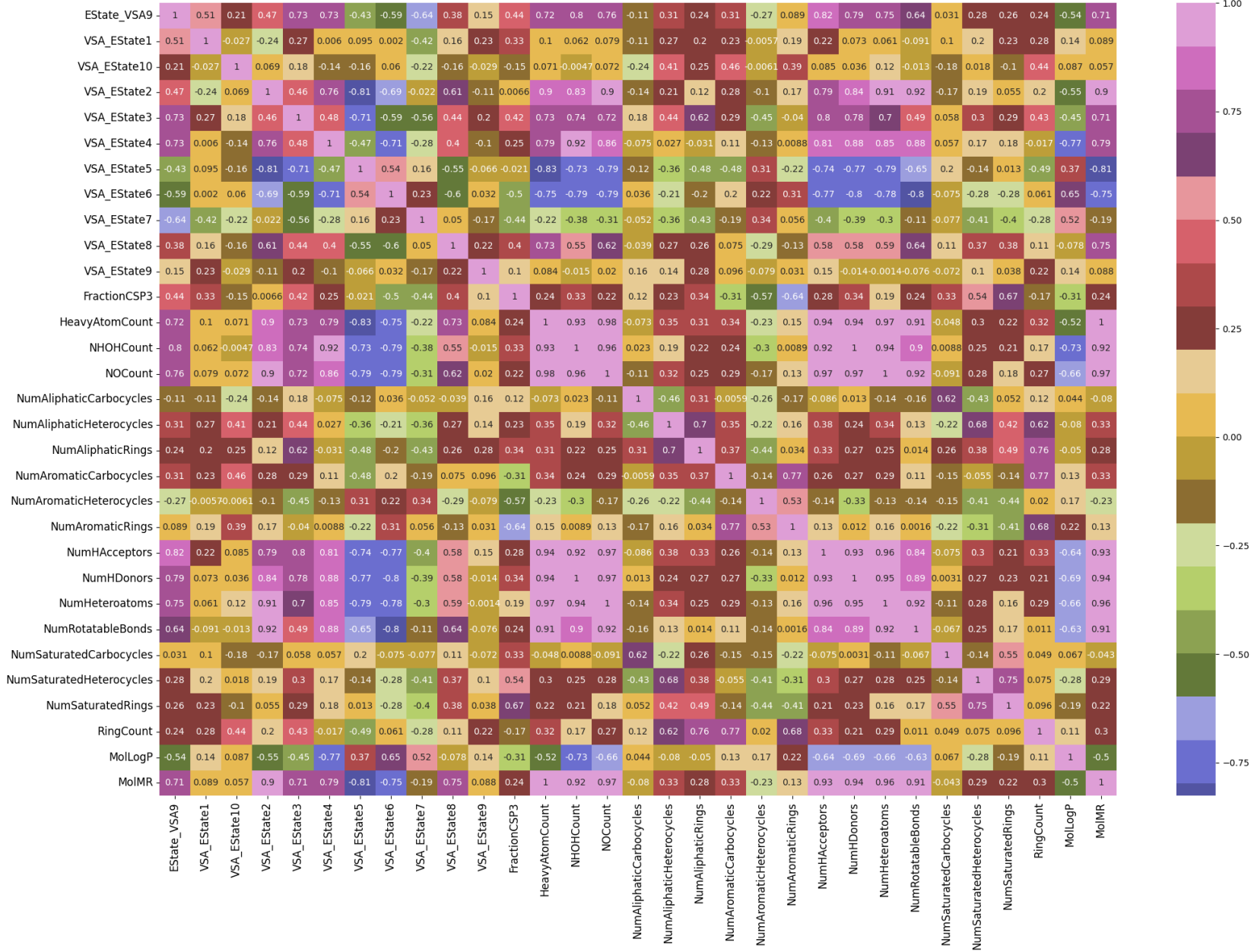
Executive summary

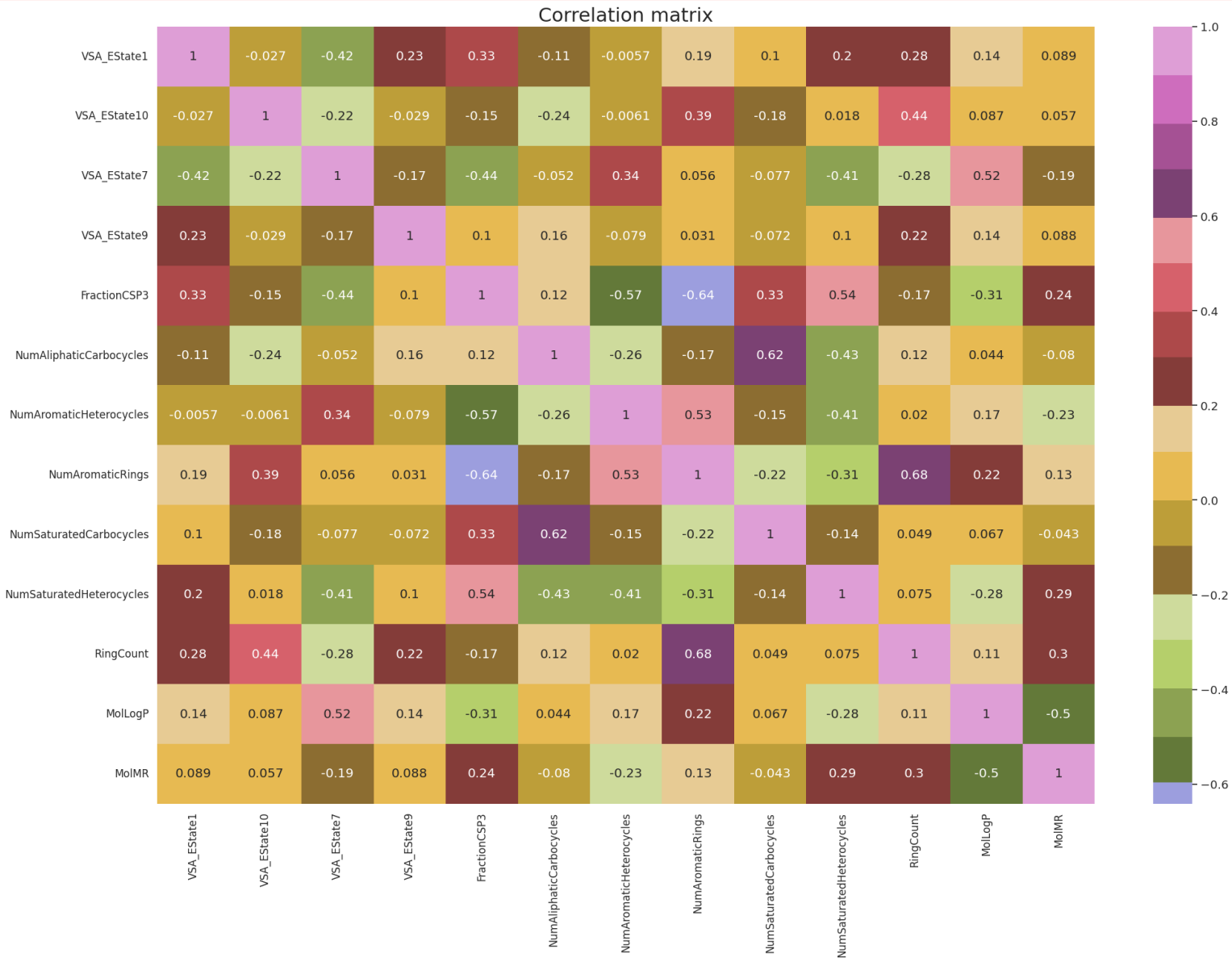
Какая задача перед нами стояла?

- Чистка датасетов
- Добавление дескрипторов
- Поиск наилучшей модели ML
- Интерпретация результатов



Correlation matrix





● Добавлено 13 фичей из RDKit

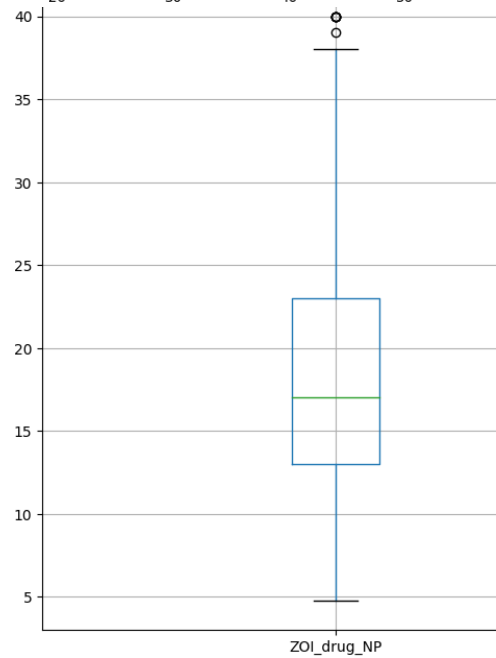
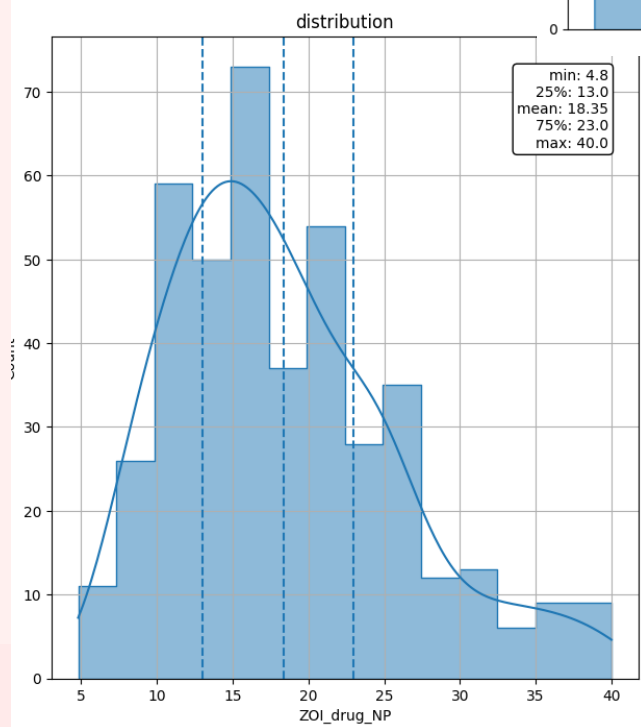
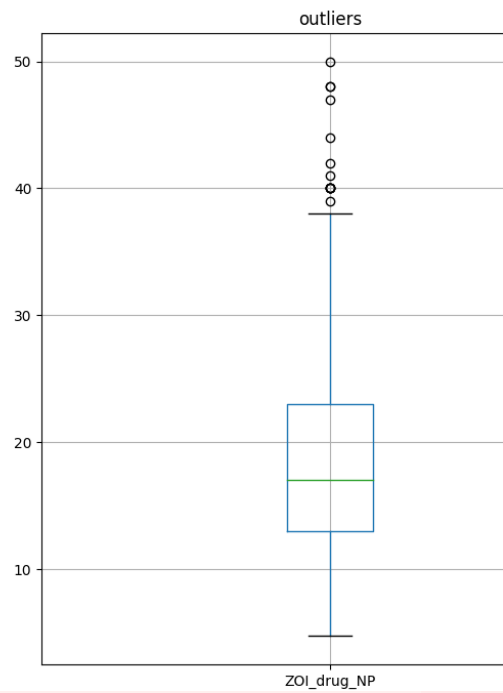
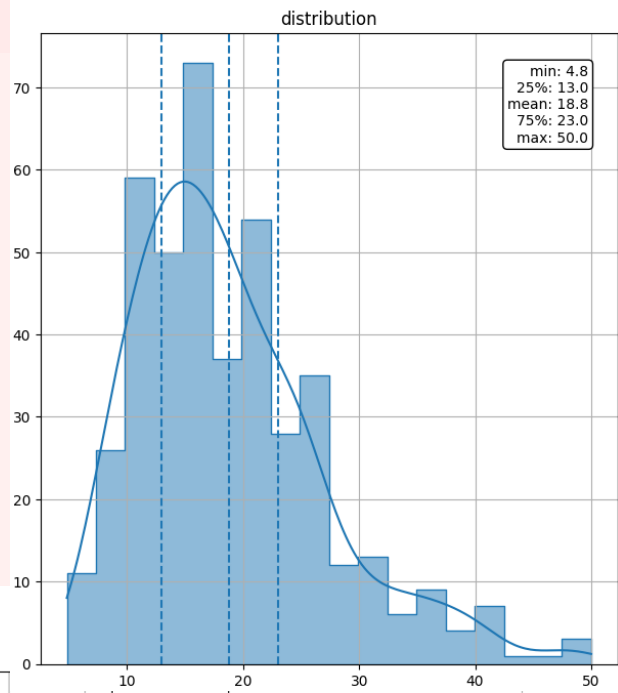
● 44 параметра

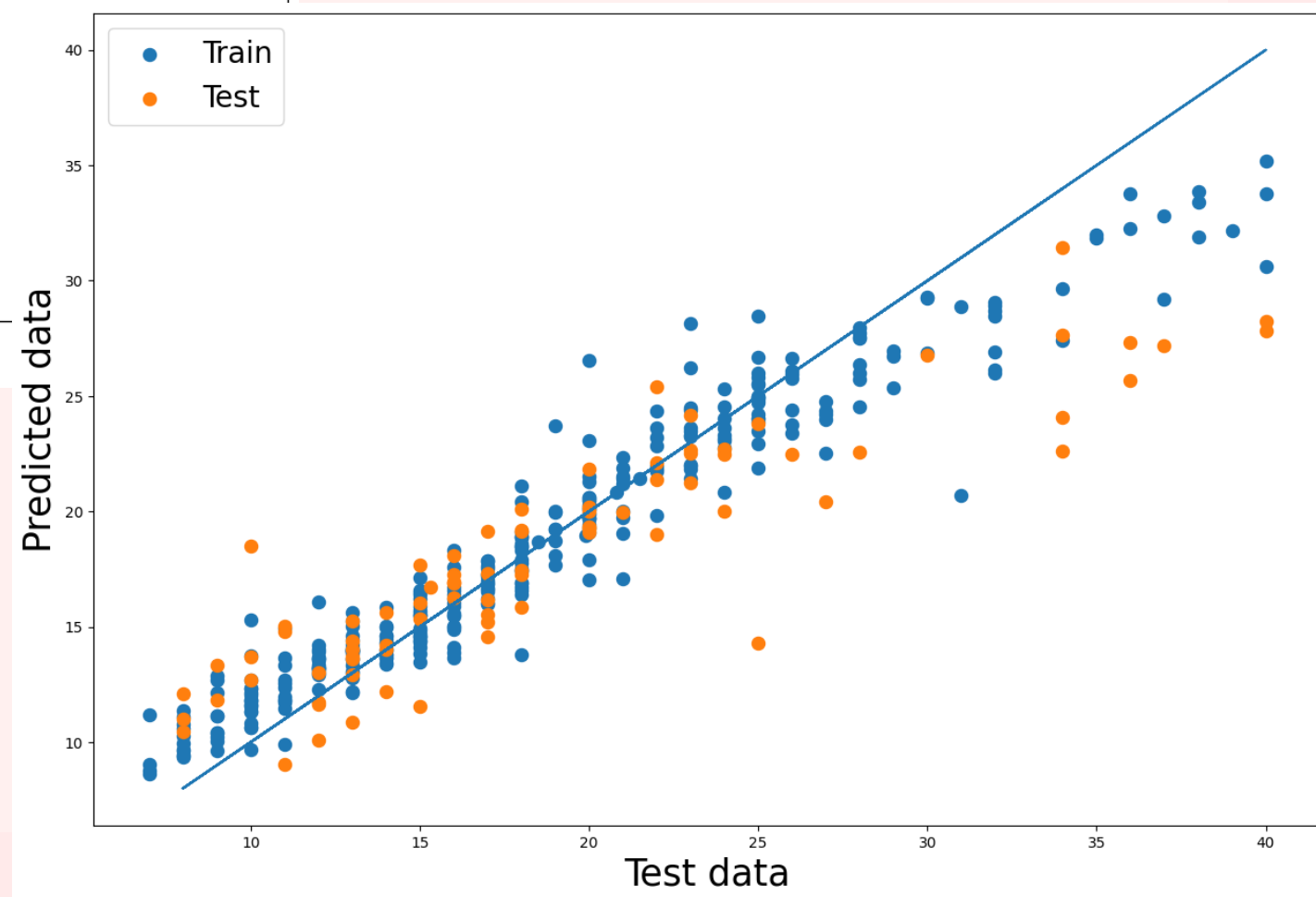
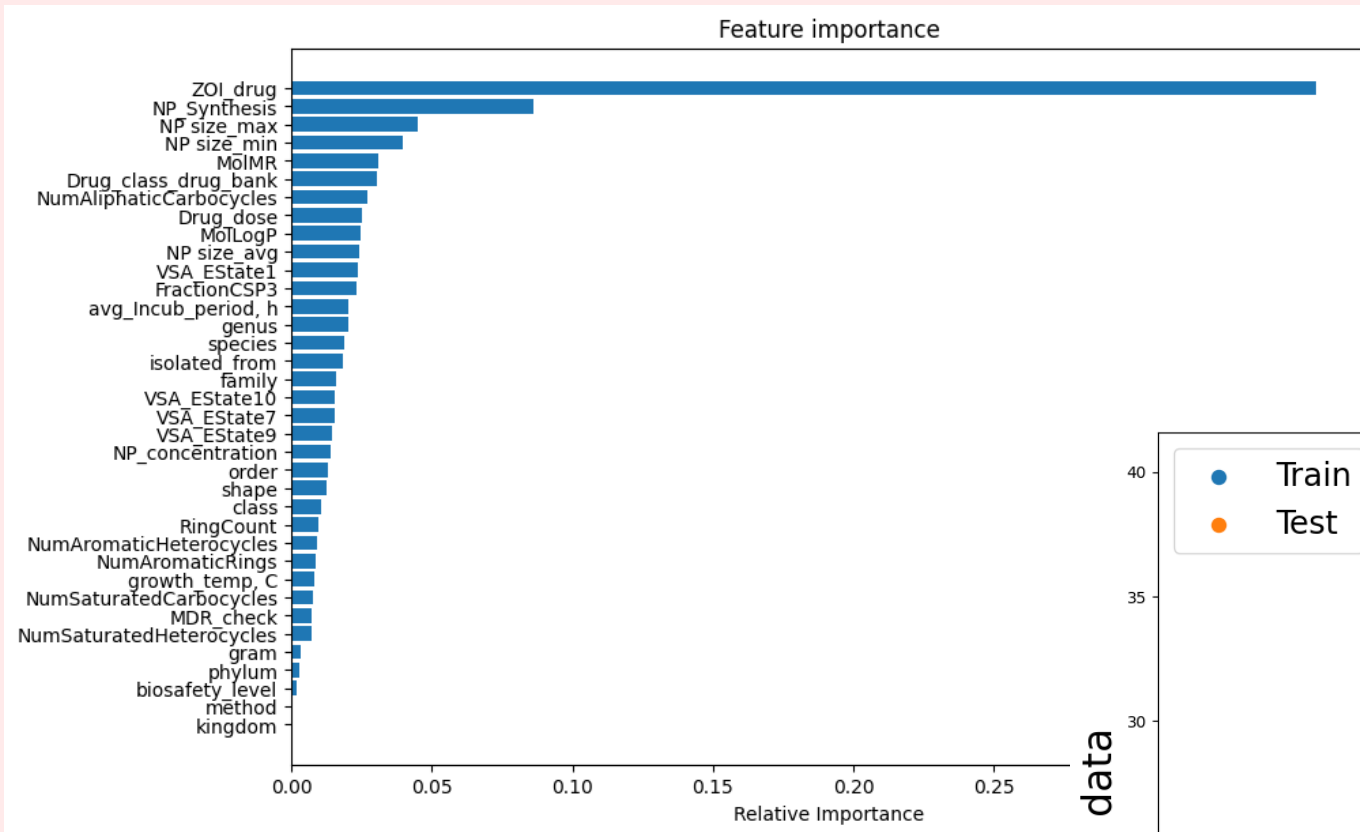
● 429 строк



0	index	429	non-null	int64
1	Bacteria	429	non-null	object
2	NP_Synthesis	429	non-null	object
3	Drug	429	non-null	object
4	Drug_class_drug_bank	429	non-null	object
5	Drug_dose	429	non-null	float64
6	NP_concentration	429	non-null	object
7	NP_size_min	429	non-null	float64
8	NP_size_max	429	non-null	float64
9	NP_size_avg	429	non-null	float64
10	shape	429	non-null	object
11	method	429	non-null	object
12	ZOI_drug	429	non-null	object
13	ZOI_drug_NP	429	non-null	float64
14	MDR_check	429	non-null	int64
15	Tax_id	429	non-null	object
16	kingdom	429	non-null	object
17	phylum	429	non-null	object
18	class	429	non-null	object
19	order	429	non-null	object
20	family	429	non-null	object
21	genus	429	non-null	object
22	species	429	non-null	object
23	gram	429	non-null	object
24	avg_Incub_period, h	429	non-null	float64
25	growth_temp, C	429	non-null	float64
26	biosafety_level	429	non-null	float64
27	isolated_from	429	non-null	object
28	chemID	429	non-null	object
29	prefered_name	429	non-null	object
30	smiles	429	non-null	object
31	VSA_EState1	429	non-null	float64
32	VSA_EState10	429	non-null	float64
33	VSA_EState7	429	non-null	float64
34	VSA_EState9	429	non-null	float64
35	FractionCSP3	429	non-null	float64
36	NumAliphaticCarbocycles	429	non-null	float64
37	NumAromaticHeterocycles	429	non-null	float64
38	NumAromaticRings	429	non-null	float64
39	NumSaturatedCarbocycles	429	non-null	float64
40	NumSaturatedHeterocycles	429	non-null	float64
41	RingCount	429	non-null	float64
42	MolLogP	429	non-null	float64
43	MolMR	429	non-null	float64

ZOI_drug_NP





В результате мы...



Random Forrest Regressor

```
'bootstrap': False  
'max_depth': None  
'max_features': 'sqrt'  
'min_samples_leaf': 3  
'min_samples_split': 9  
'n_estimators': 100
```

MSE: 15.771329712371212

R-squared: 0.7365871778963208



Спасибо за внимание!

