

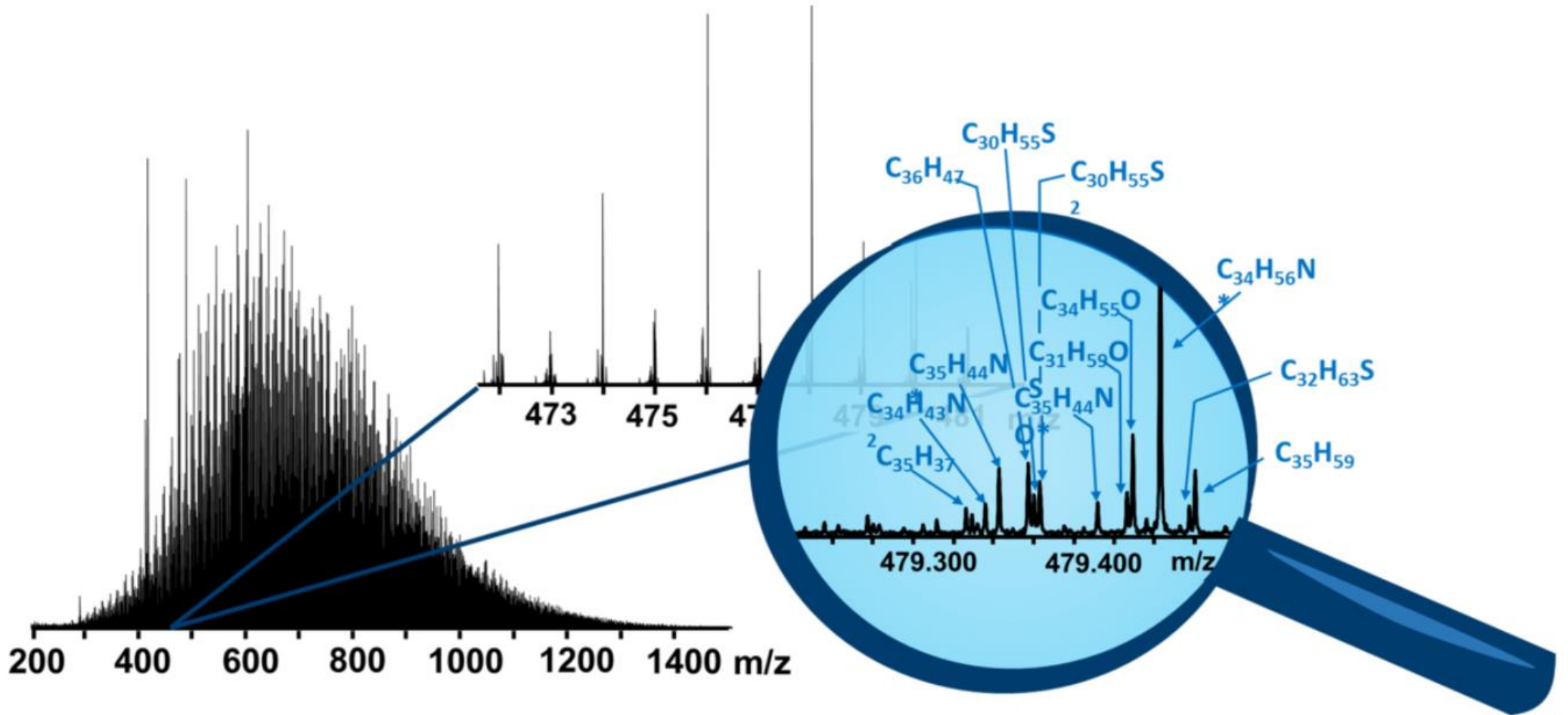
*Распознавание металлоорганических  
ионов в масс-спектрах высокого  
разрешения методами классического  
машинного обучения*

---

Выполнил: Нерсесян Лев Эрикович, ученик 11 «Х» СУНЦ МГУ

Научный руководитель: Бойко Даниил Александрович,  
аспирант ИОХ РАН им. Зелинского

# Проблема



<https://www.zimmermann.chemie.uni-rostock.de/en/research/advanced-mass-spectrometry/high-resolution-mass-spectrometry/>

# Цели и задачи

## *Цели:*

1. Создание модели классического машинного обучения, позволяющей с высокой точностью определить присутствие заданных элементов в масс-спектре вещества по характерным пиками изотопных распределениях ионов в масс-спектрах

## *Задачи:*

1. Генерация набора данных
2. Выбор моделей машинного обучения
3. Обучение моделей на сгенерированном датасете
4. Оценка работы полученных моделей

# Генерация набора данных



# Выбор и обучение моделей



	Hyperparametrs	Best Score, AUC ROC
XGBoost	<i>n_estimators</i> : <b>300</b> , 500, 1000, 1500 <i>max_depth</i> : 6, 8, 10, <b>12</b>	<b>0.99741</b>
SVC	<i>C</i> : $10^{-4}$ , $10^{-3}$ , $10^{-2}$ , 0.1, <b>1</b> <i>kernel</i> : linear, poly, <b>rbf</b>	<b>0.97175</b>
Random Forest	<i>n_estimators</i> : 50, 100, 300, 700, <b>900</b> <i>max_depth</i> : 6, 8, 10, <b>12</b>	<b>0.99608</b>
CatBoost	<i>n_estimators</i> : 300, 500, 1000, <b>1500</b> <i>max_depth</i> : 6, 8, 10, <b>12</b>	<b>0.99755</b>
Multi-layer Perceptron	<i>solver</i> : sgd, <b>adam</b> <i>lear. rate</i> : constant, <b>adaptive</b> <i>activation</i> : logistic, tanh, <b>relu</b> <i>learning rate init</i> : $10^{-5}$ , $10^{-4}$ , $10^{-3}$ , <b><math>10^{-2}</math></b> , $10^{-1}$	<b>0.99253</b>

Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Test				
	Test			
		Test		
			Test	
				Test

## Кросс-валидация



# Итоги

	AUC ROC					
	Pd	Ni	Cl	Br	Cu	Ag
XGBoost	0.785	0.976	0.719	0.757	0.719	<b>0.916</b>
SVC	0.749	0.978	0.837	0.791	0.493	0.5
Random Forest	0.677	<b>0.978</b>	0.716	0.687	<b>0.987</b>	0.499
CatBoost	<b>0.786</b>	0.978	0.762	0.715	0.983	0.846
MLP	0.714	0.965	<b>0.867</b>	<b>0.847</b>	0.697	0.854

# Ссылки

1. Boiko D.A., Kozlov K.S., Burykina J.V., Ilyushenkova V.V., Ananikov V.P., "Fully Automated Unconstrained Analysis of High-Resolution Mass Spectrometry Data with Machine Learning", J. Am. Chem. Soc. 2022, 144, 32, 14590–14606
2. Масс-спектрометрия в органической химии / А. Т. Лебедев. - М.: БИНОМ. Лаборатория знания, 2003.