

Инжиниринг управления данными

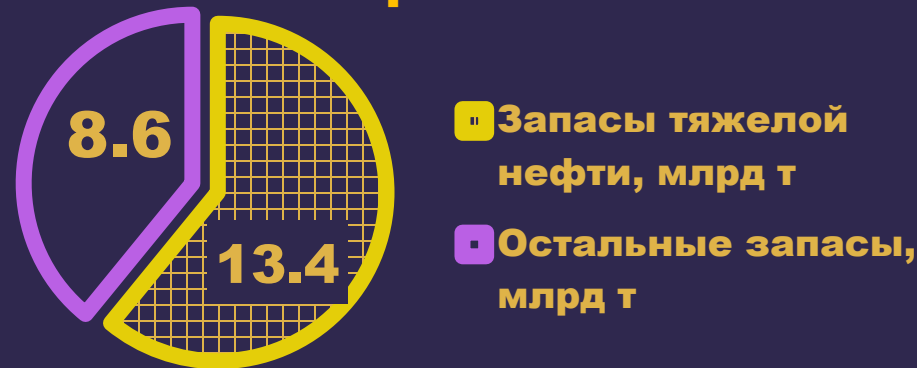
Проект ККМ ПАВ

Хасанов Артём Фаридович
Мяльдзин Никита Кириллович

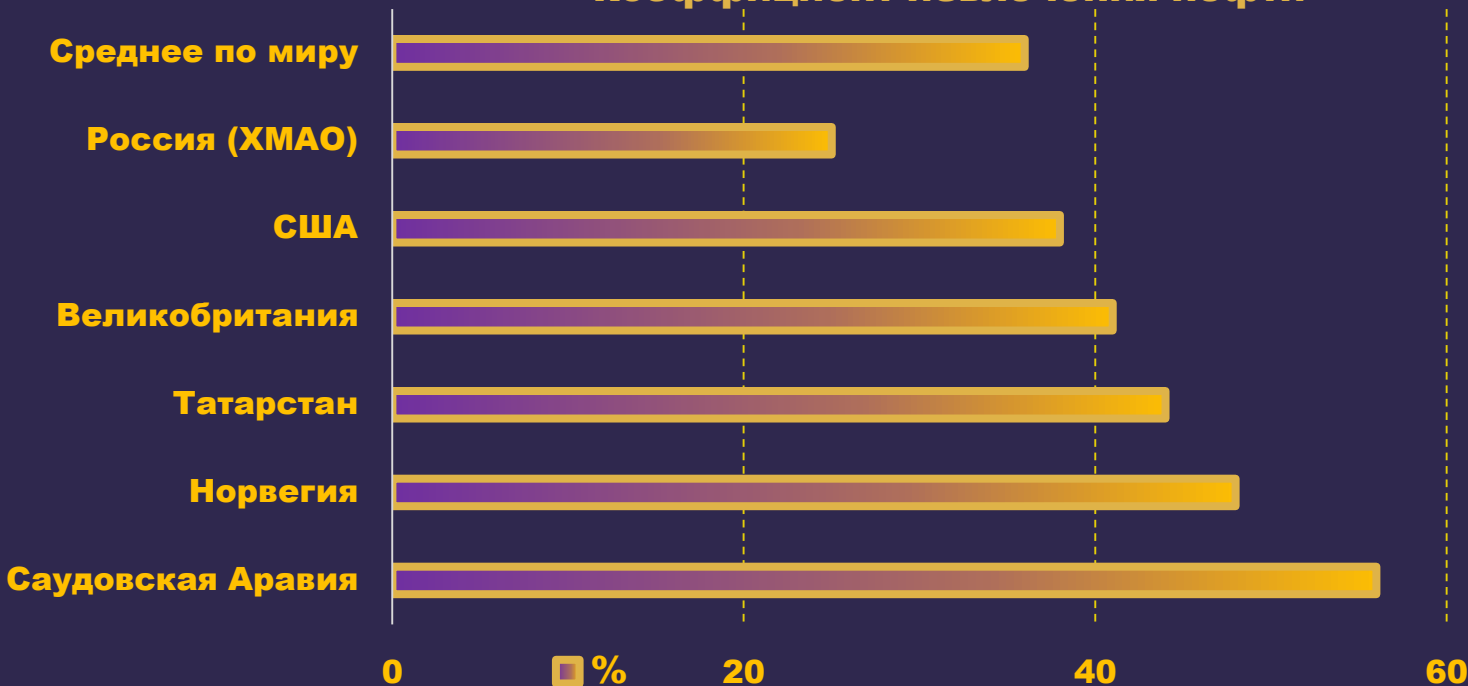


Анализ проблемы и актуальность:

Запасы нефти РФ за 2024



Коэффициент извлечения нефти



Применение ПАВ для увеличения нефтеотдачи



Изменение смачиваемости



Обработка призабойной зоны



Интенсификация добычи



Улучшение реологических свойств



Снижение трения оборудования

Суть идеи:

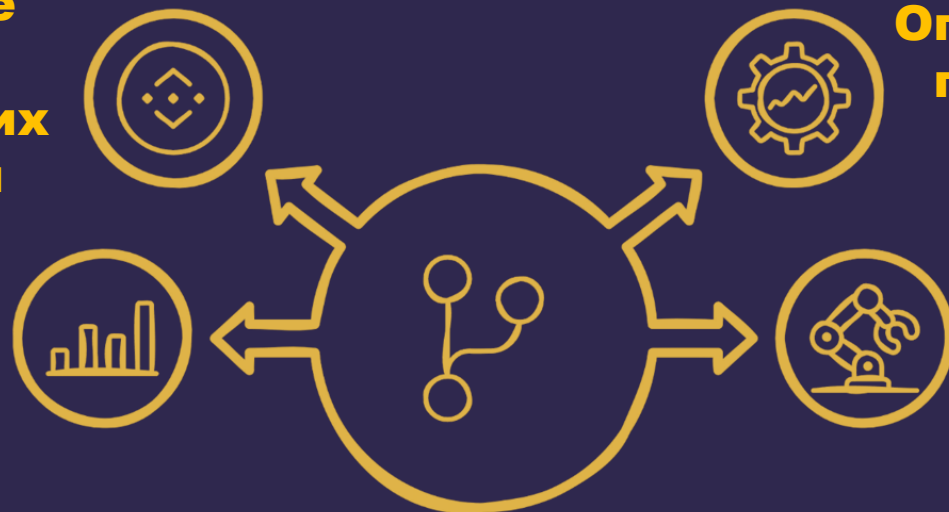
Разработка **новых поверхностно-активных веществ (ПАВ)** для увеличения нефтеотдачи на основе методов машинного обучения

Области реализации:



**Цифровые
двойники
добывающих
компаний**

**Анализ
химических
данных**



**Оптимизация
процессов
добычи**

**Автоматизация
проектных
решений**



Ріреліне проєкта:

**Сбор данных о
концентрации
мицеллообразования
молекул ПАВ**



**Синтез расчётных
дескрипторов для
молекул**



**Создание
Dashboard**



**Предобработка
собранных данных**



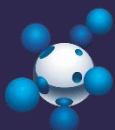
**Анализ данных и
визуализация**





Сбор базы данных молекул ПАВ

PubChem



SASOL

R⁶



Критическая
концентрация
мицеллообразования



Тип ПАВ



Температура
эксперимента



Запись молекулы
SMILES



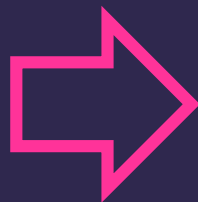
Предобработка после сбора:

Привели
концентрацию к
единой величине

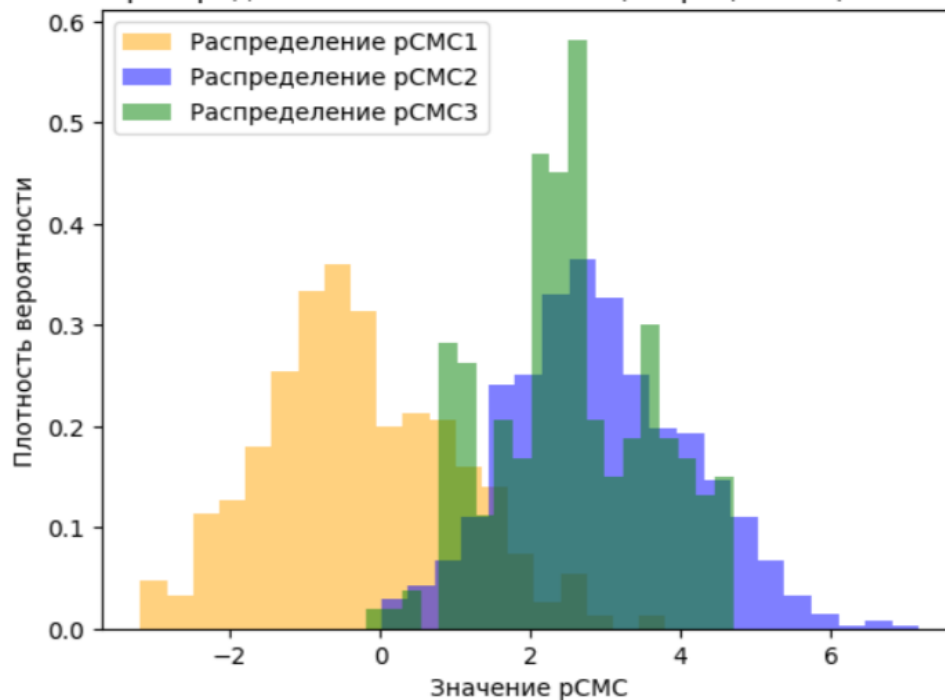
Уход от
логарифмического
распределения

Объединение данных

Единый тип данных



Сравнение распределений показателя концентраций мицеллообразования





Синтез расчётных дескрипторов для молекул



ККМ ~ 1 267 уникальных строк по ПАВ и T°



Библиотеки:



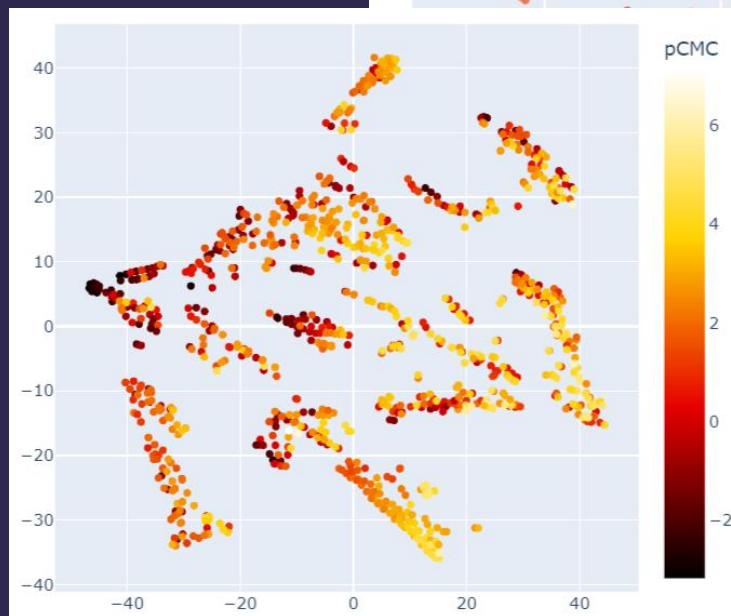
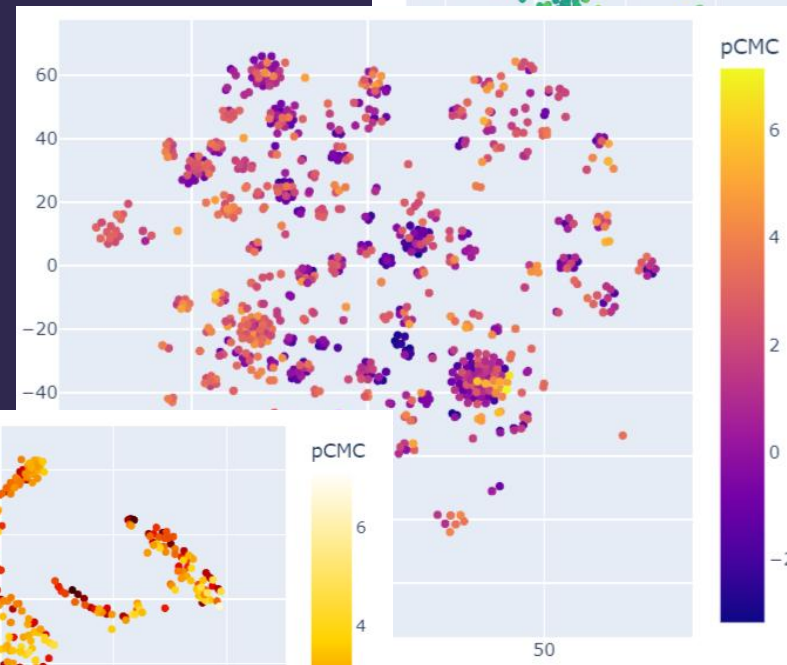
RDKit Descriptors ~ 208



Mordred Descriptors ~ 855



Fingerprints ~ 2048



Визуализация пространства молекул: PCA, T-SNE

Анализ данных



Метрики молекул



Число уникальных молекул: 1111



Процент молекул с ККМ менее 0.01 моль/л: 25.18%



Распределение не похоже на нормальное



Статистически значимая корреляция (T° /pCMC)

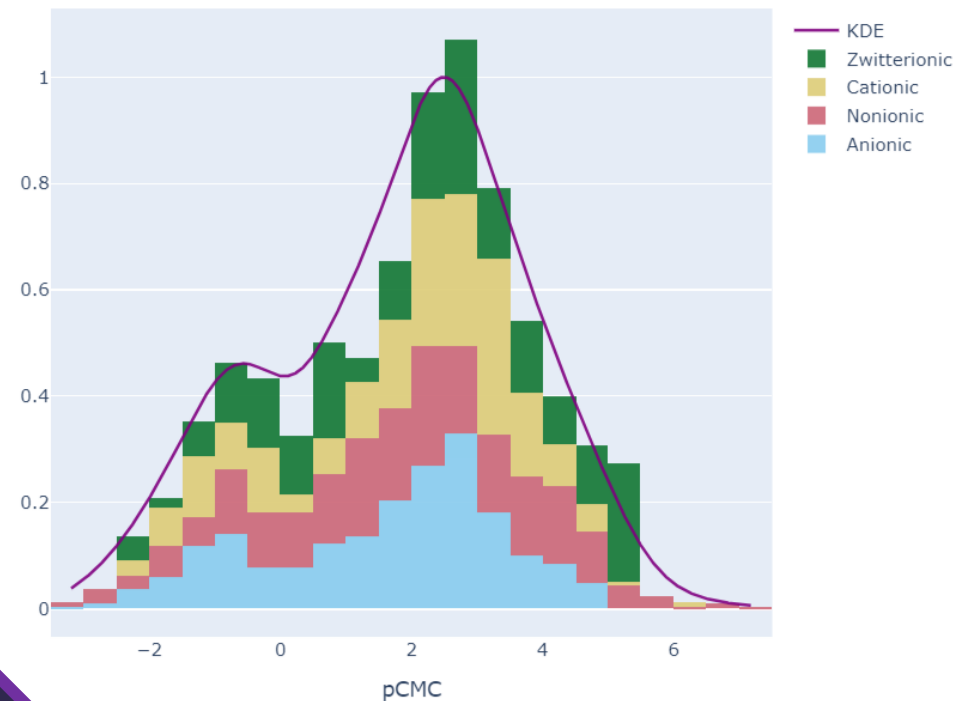


Класс ПАВ не влияет на показатель ККМ

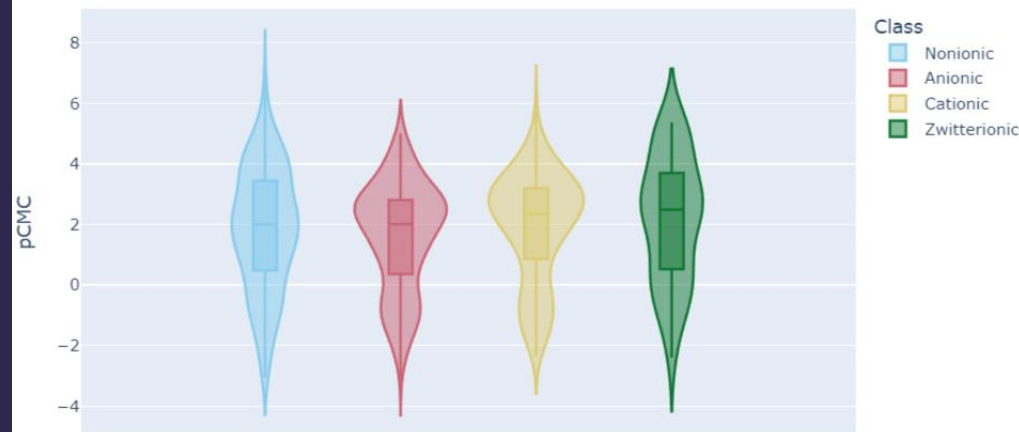


Создание Dashboard

Class distribution of pCMC with density estimation



Violin plot of classes of surfactants



Dashboard

GitHub репозиторий



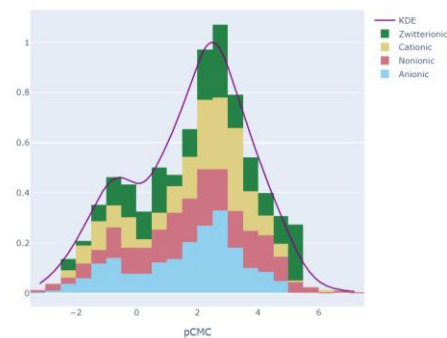
Анализ качества данных для предсказания концентрации мицеллообразования

Метрики молекул

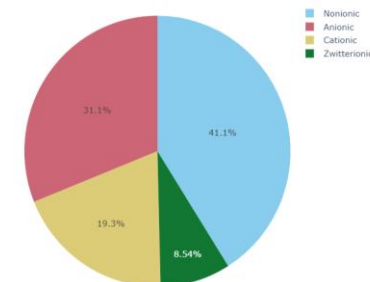
Число уникальных молекул: 1111

Процент молекул с ККМ менее 0.01 моль/л: 25.18%

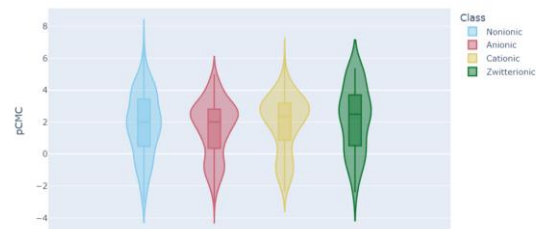
Class distribution of pCMC with density estimation



Class distribution of surfactants

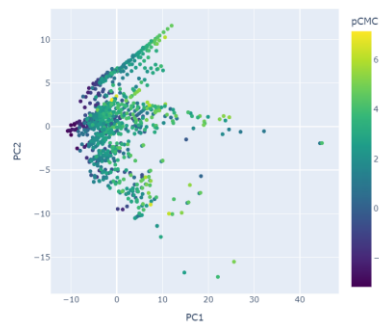


pCMC of classes of surfactants

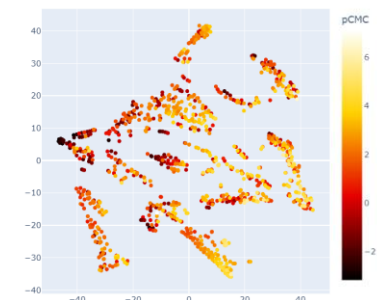


Визуализация пространства молекул в датасете

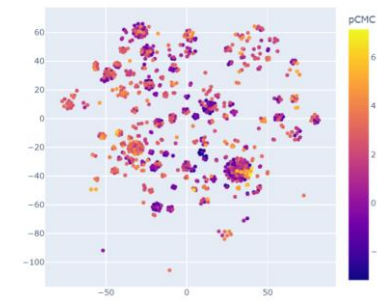
PCA for RDKit descriptors



t-SNE for RDKit descriptors



t-SNE for fingerprints



Спасибо!

Артём Хасанов
Никита Мяльдзин

