

**«Методы построения калибровочных моделей
для прогнозирования свойств индивидуальных
углеводородов»**

Интеллектуальный анализ данных

Интеллектуальный анализ данных (ИАД) или Data mining:

Исследование и обнаружение вычислительной машиной (алгоритмами, средствами искусственного интеллекта) в сырых данных **скрытых знаний**, которые ранее были **не известны, нетривиальны, практически полезны, доступны для интерпретации человеком.**

Примеры применения

Казначейство США

- Финансовые преступления - поиск подозрительных транзакций

Управление по борьбе с наркотиками

- Анализ телефонных переговоров подозреваемых

HSBC

- Снижение затрат на рассылки за счет анализа клиентской базы

Финансовая группа Capital One

- Анализ клиентской базы и подбор соответствующих продуктов и идентификация мошенников

BBC

- Использование ИАД для составления сетки вещания

Boeing

- Улучшение тех. процессов

Southern California Gas Company

- Применение ИАД как средства стратегического маркетинга

Интеллектуальный анализ данных в нефтегазовой отрасли

Основное направление – анализ геологических данных.

- Классификация коллектор – не коллектор.
- Анализ изображений в литологии.

Регрессионный анализ и деревья решений

- Зависимости между уплотнением сетки скважин и характеристиками месторождения.
- Зависимости между оценками конечной (суммарной) добычи и геологическими/пластовыми характеристиками.

Нейронные сети

- Идентификация литофаций с помощью карт Кохонена.
- Прогнозирование оценки конечной (суммарной) добычи из геологических характеристик.

Визуализация и поиск трендов

Хемометрика

- Контроль качества
- Экспресс-методы анализа

Хемометрика

Хемометрика – это химическая дисциплина, применяющая математические, статистические и другие методы, для построения или отбора оптимальных методов измерения и планов эксперимента, а также для извлечения наиболее важной информации при анализе экспериментальных данных

Методы определения ОЧ

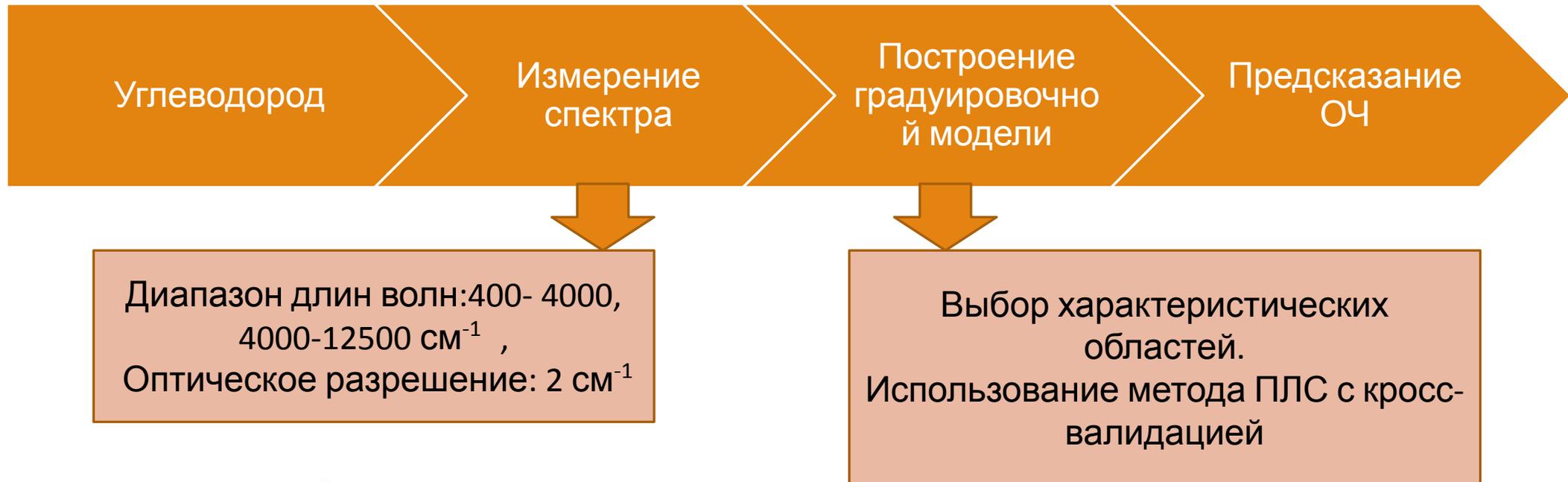
Арбитражные

- Моторные методы:
- ГОСТ Р 52946-52947-2008
- ГОСТ 8226-82
- ГОСТ 511-82
- ASTM D2699 – 2700
- EN ISO 5163 - 5164:2005

Экспресс

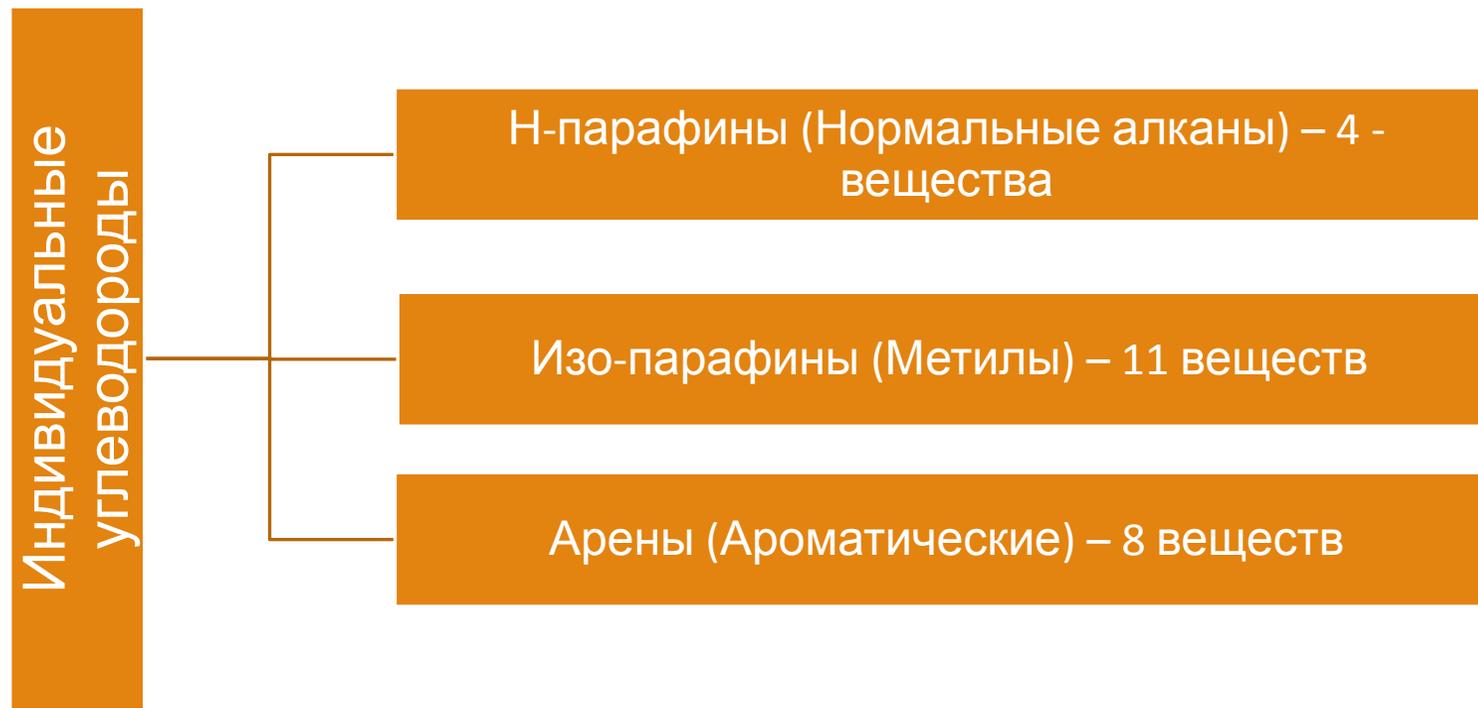
- Хроматография
- По диэлектрической проницаемости
- **ИК-Фурье-спектрометрия**
- По показателю преломления

Методика определения ОЧ



Компоненты модельных смесей бензинов

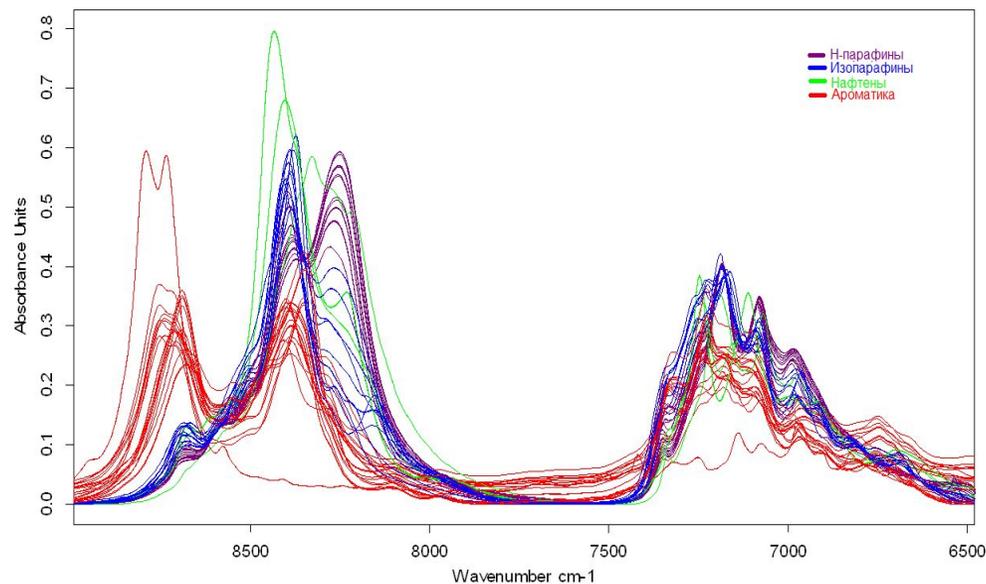
Для предсказания ОЧ использовались индивидуальные углеводороды, потенциально входящие в состав бензинов (23 соединения).



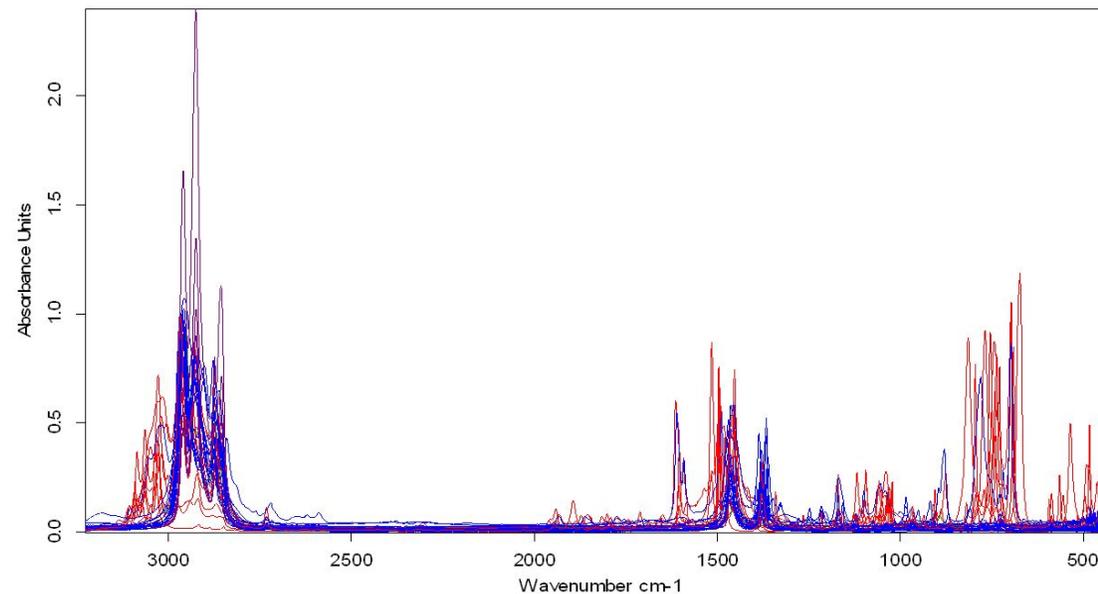
Исходные данные

	A	B	C	D	E	F	G	H	I
1	2,2-диметилбутан	91,8	-0,00657	-0,00655	-0,00654	-0,00653	-0,00649	-0,00644	-0,00637
2	2,2-диметилбутан	91,8	-0,0076	-0,00759	-0,00758	-0,00758	-0,00755	-0,0075	-0,00744
3	2,2-диметилбутан	91,8	-0,00888	-0,00887	-0,00886	-0,00886	-0,00883	-0,0088	-0,00874
4	2,3-диметилбутан	101,7	-0,00954	-0,00953	-0,00953	-0,00953	-0,00951	-0,00948	-0,00944
5	2,3-диметилбутан	101,7	-0,00964	-0,00964	-0,00964	-0,00964	-0,00963	-0,0096	-0,00955
6	2,3-диметилбутан	101,7	-0,01097	-0,01096	-0,01096	-0,01096	-0,01093	-0,0109	-0,01085
7	2,3-диметилпентан	91,1	-0,02406	-0,02404	-0,02404	-0,02403	-0,02401	-0,02397	-0,02392
8	2,3-диметилпентан	91,1	-0,02529	-0,02528	-0,02528	-0,02527	-0,02523	-0,02518	-0,02512
9	2,3-диметилпентан	91,1	-0,01874	-0,01873	-0,01873	-0,01872	-0,01869	-0,01865	-0,01859
10	2,4-диметилгексан	65,2	-0,00319	-0,00319	-0,00321	-0,00322	-0,00322	-0,00319	-0,00315
11	2,4-диметилгексан	65,2	-0,00338	-0,00338	-0,00339	-0,00341	-0,0034	-0,00338	-0,00334
12	2,4-диметилгексан	65,2	-0,00371	-0,00371	-0,00372	-0,00374	-0,00373	-0,0037	-0,00365
13	2,4-диметилпентан	83,1	-0,01353	-0,01353	-0,01354	-0,01354	-0,01354	-0,01351	-0,01347
14	2,4-диметилпентан	83,1	-0,01441	-0,01441	-0,01443	-0,01444	-0,01443	-0,0144	-0,01435
15	2,4-диметилпентан	83,1	-0,01457	-0,01457	-0,01458	-0,01458	-0,01457	-0,01454	-0,01449
16	2-метилгексан	42,4	-0,00312	-0,00312	-0,00313	-0,00313	-0,00313	-0,00312	-0,00309
17	2-метилгексан	42,4	-0,00356	-0,00356	-0,00357	-0,00358	-0,00358	-0,00356	-0,00352

Результаты спектроскопии



Спектры углеводородов в БИК-диапазоне
(NIR)(4000-12000 см⁻¹)



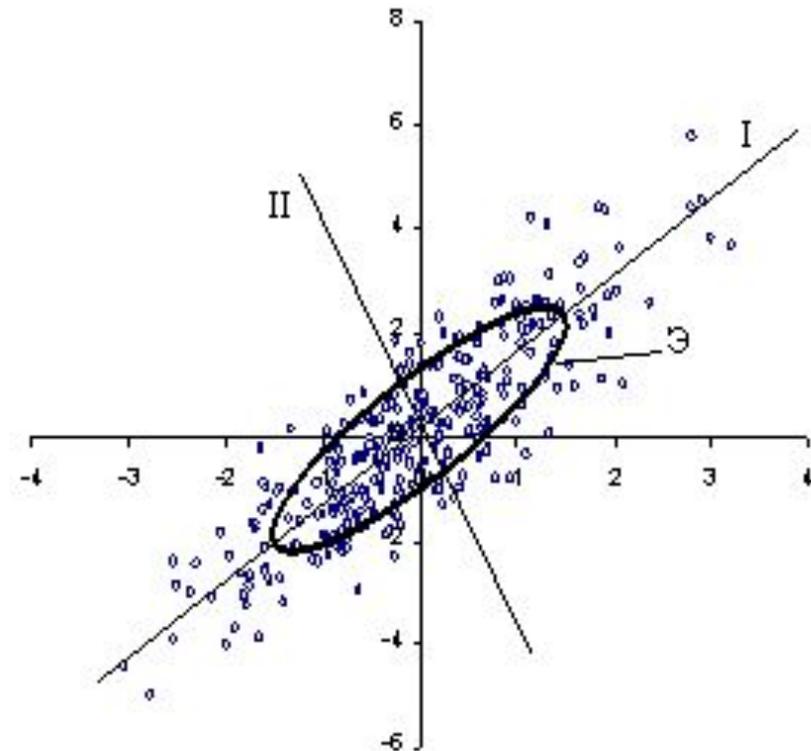
Спектры углеводородов в среднем ИК-
диапазоне(MIR) (400-4000 см⁻¹)

Подготовка данных

- Импорт данных в RapidMiner Studio
- Использование Метода главных компонент

Метод Главных Компонент (англ. Principal Components Analysis, PCA) — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации.

Comp...	Standard Dev...	Proportion of Var...	Cumulative Var...
PC 1	1.97095473	0.91499846	0.91499846
PC 2	0.57089964	0.07676913	0.99176759
PC 3	0.15742970	0.00583768	0.99760527
PC 4	0.06822321	0.00109630	0.99870158
PC 5	0.05302030	0.00066214	0.99936372
PC 6	0.04157378	0.00040710	0.99977082
PC 7	0.02322053	0.00012700	0.99989783
PC 8	0.01727698	0.00007031	0.99996813



Регрессия на главные компоненты

I

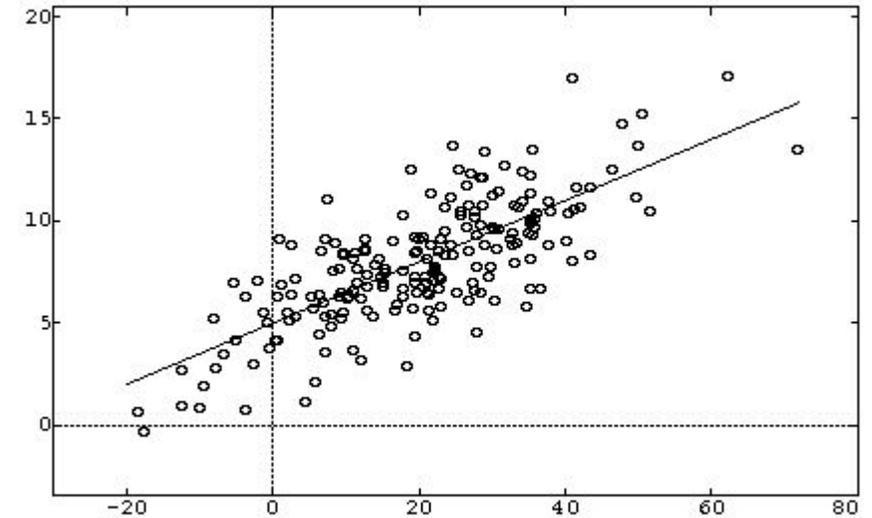
МГК

II

Многомерная
Линейная
Регрессия



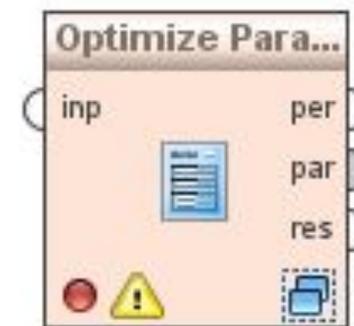
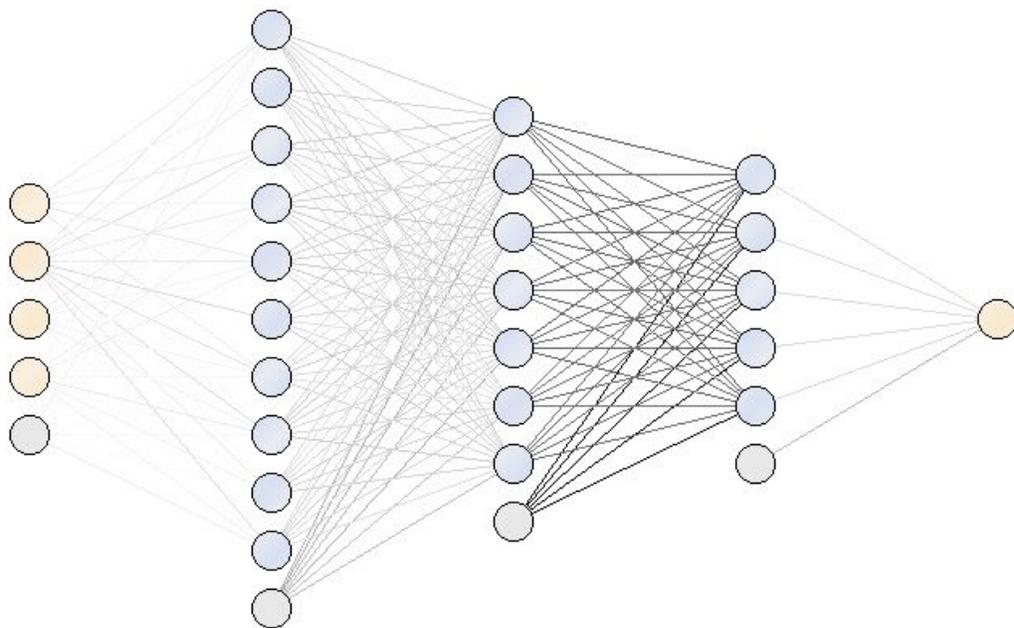
Многомерная линейная регрессия — это линейная регрессия в n -мерном пространстве. Задачей множественной линейной регрессии является построение линейной модели связи между набором непрерывных предикторов и непрерывной зависимой переменной.



Регрессионный анализ используется для прогноза, анализа временных рядов, тестирования гипотез и выявления скрытых взаимосвязей в данных.

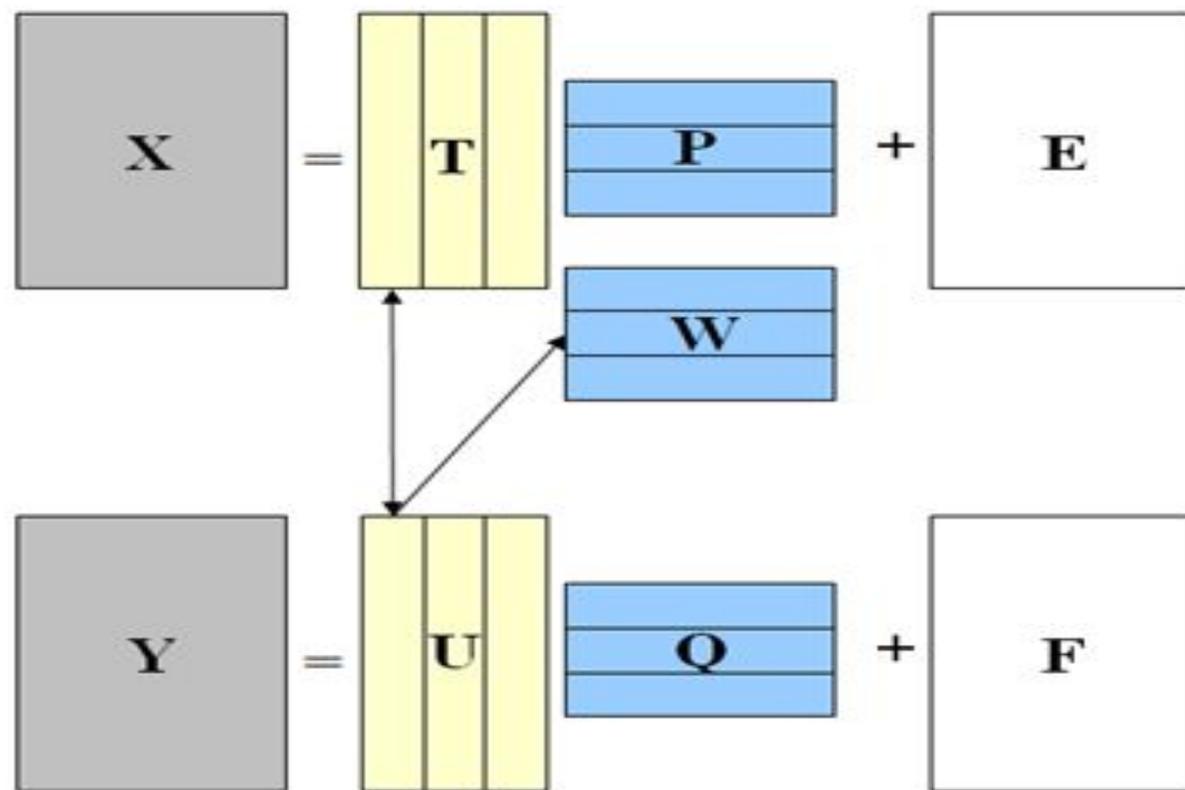
Нейронные сети

Искусственная нейронная сеть (ИНС) — математическая модель, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.



Оператор перебора параметров
Optimize parameters (Grid)

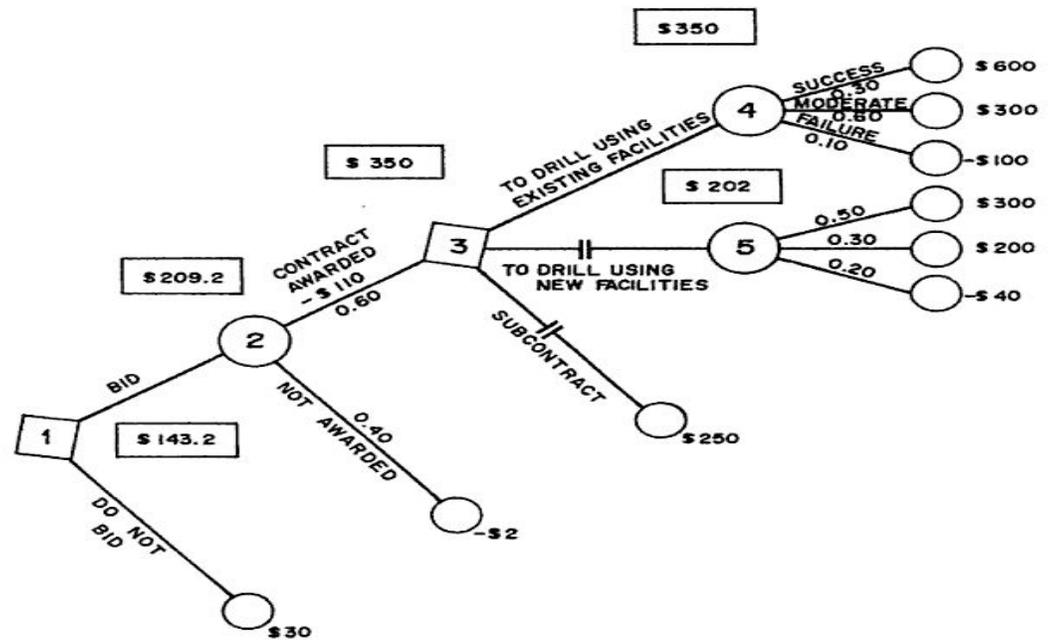
PLS



Графическое представление PLS-регрессии

Деревья решений (регрессии)

Дерево решений – бинарное дерево, где в каждом узле происходит проверка условия. Каждая ветвь представляет собой результат такой проверки. В каждом листе получается предсказанное значение.



Сравнение результатов

Октановое число для 3-х классов веществ и всех веществ, в среднем ИК диапазоне

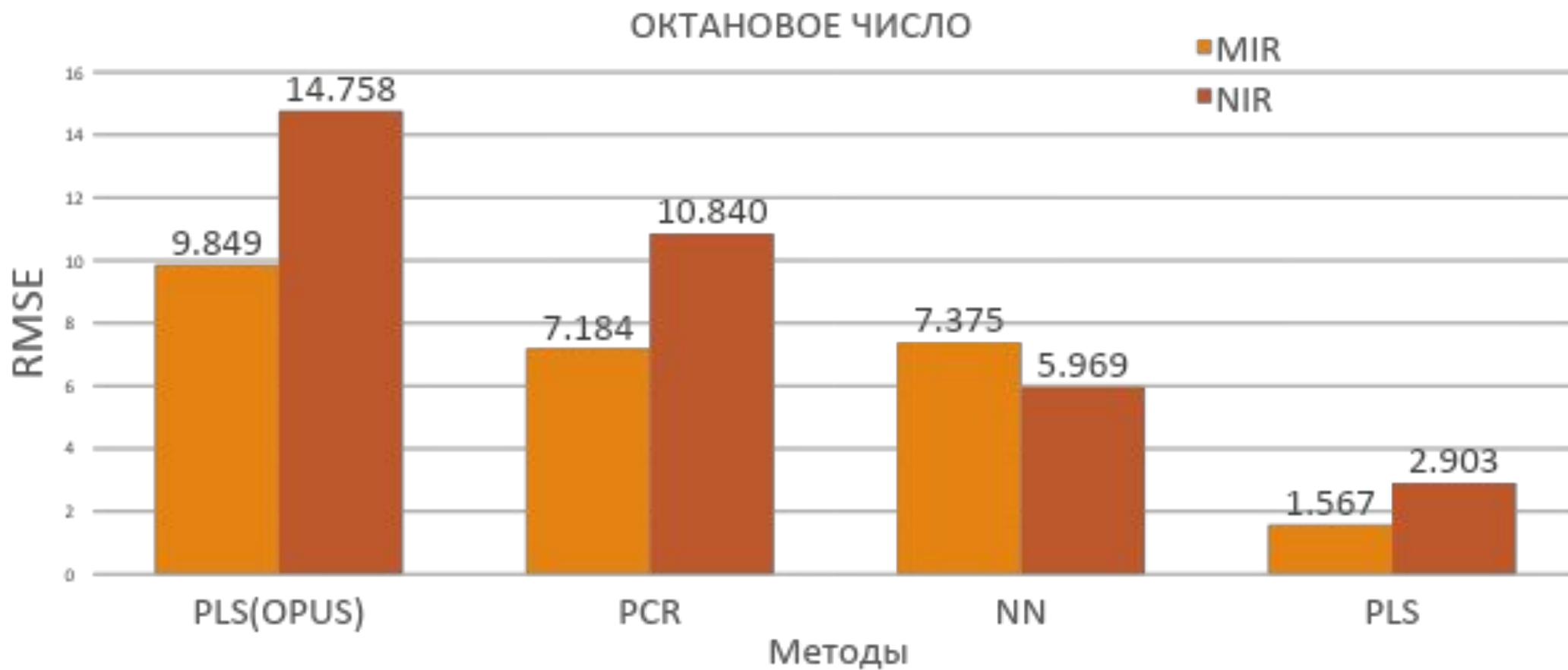
Группы веществ	PLS	Деревья решений
изо-алканы	2,47	31,13
Ароматические углеводороды	0,02	2,97
алканы	0,48	177,20
все вещества	1,57	72,64

Сравнение результатов

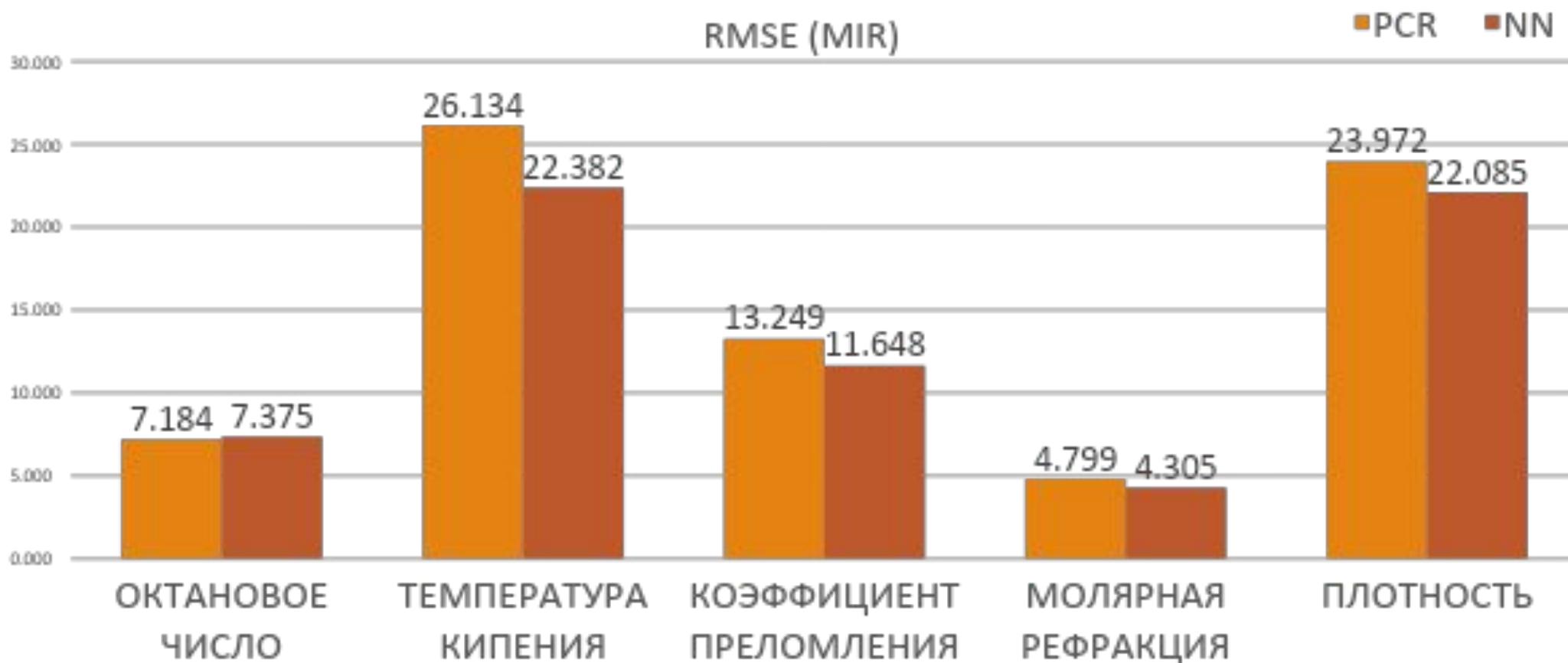
Октановое число для 3-х классов веществ и всех веществ, в ближнем ИК диапазоне

Группы веществ	PLS	Деревья решений
изо-алканы	3,17	64,71
Ароматические углеводороды	0,31	7,76
алканы	3,94	194,91
все вещества	2,90	62,25

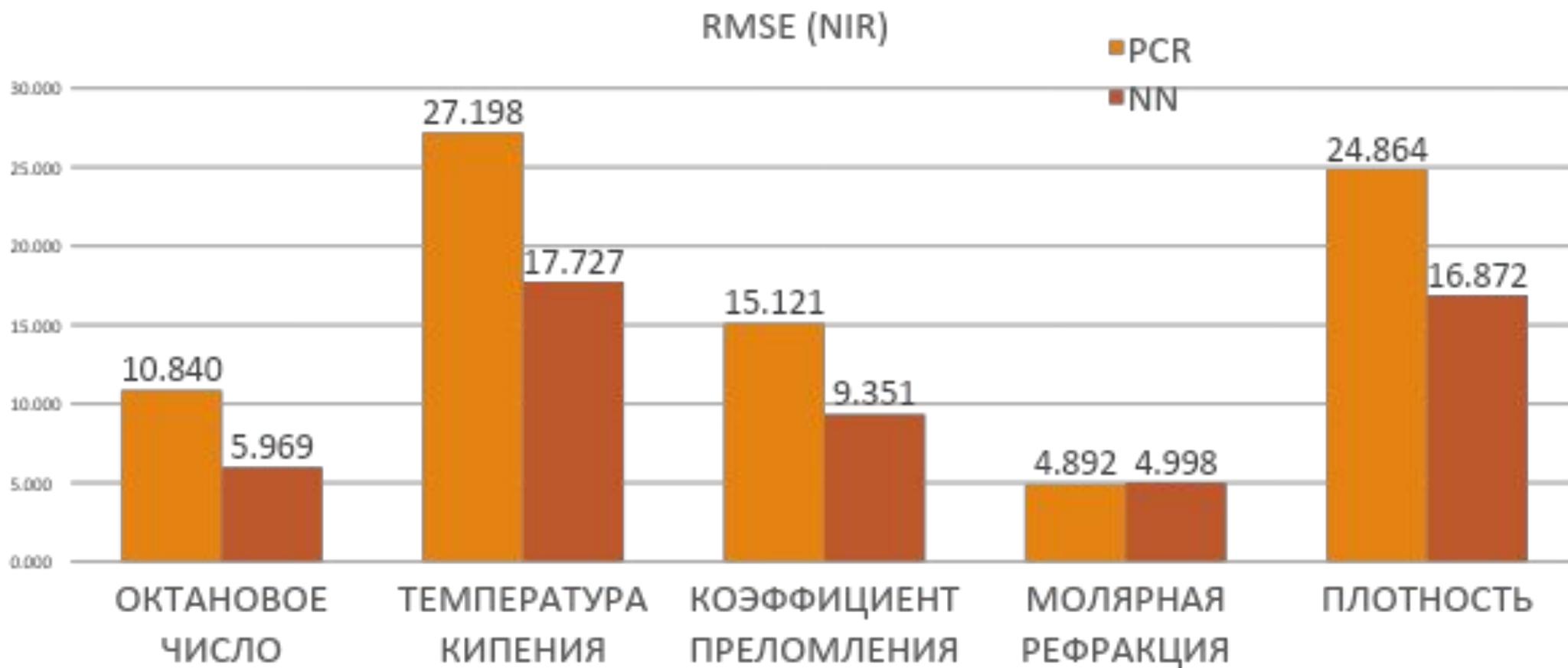
Сравнение моделей



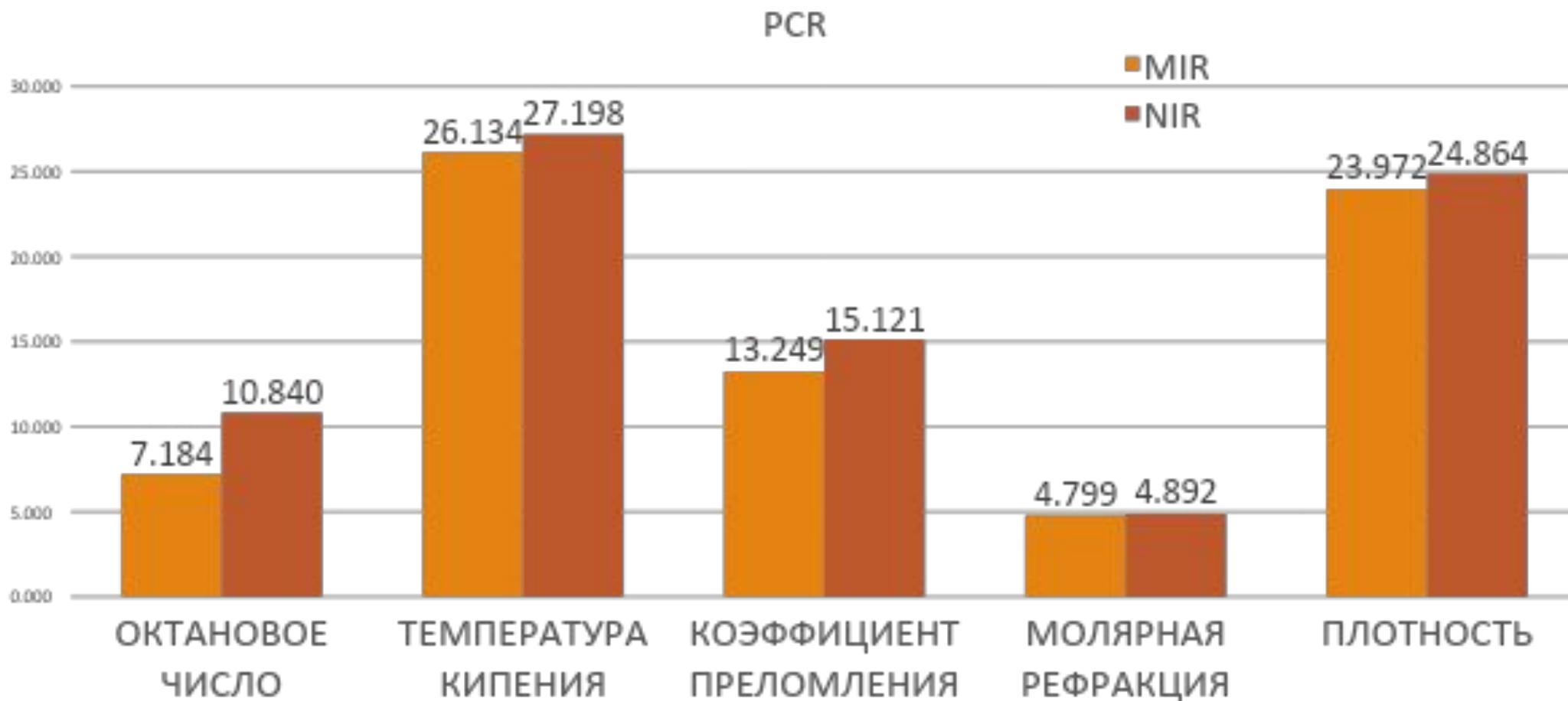
Сравнение моделей



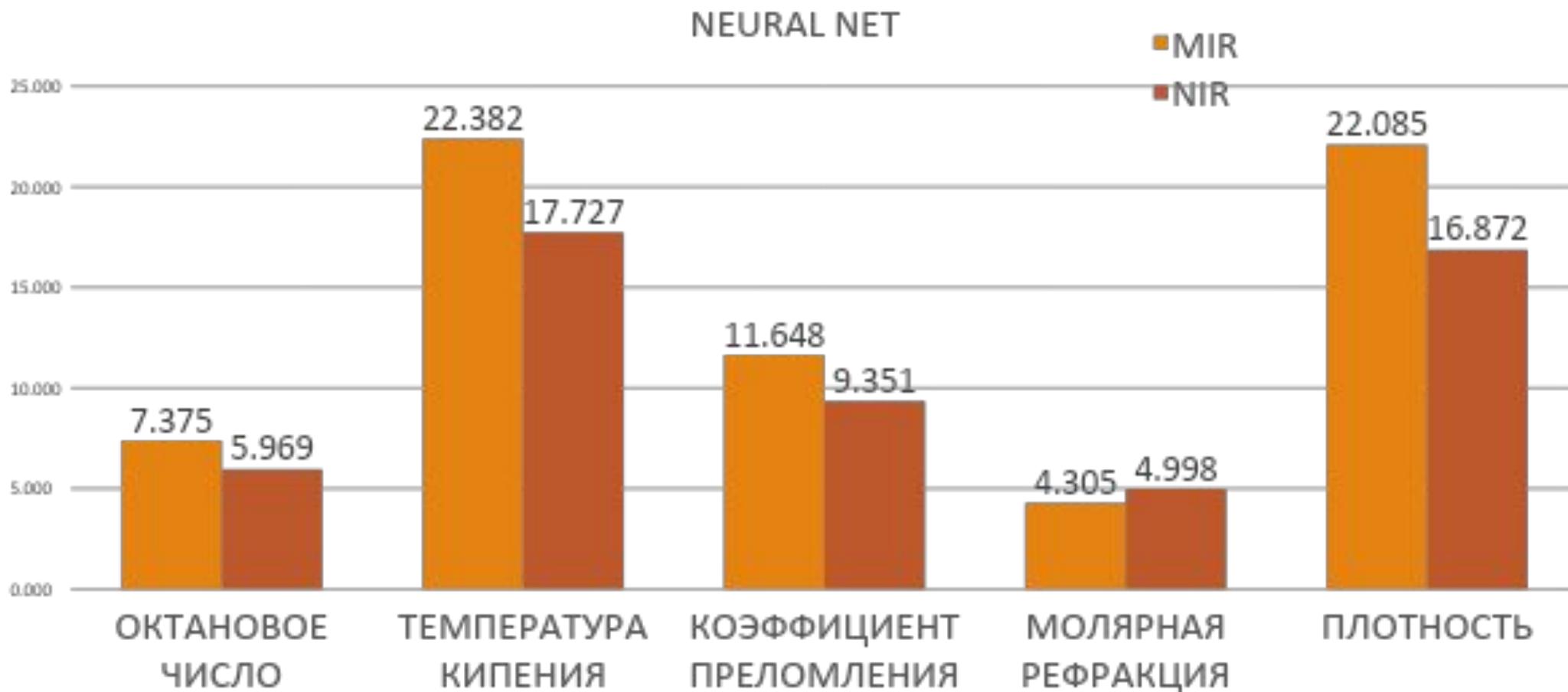
Сравнение моделей



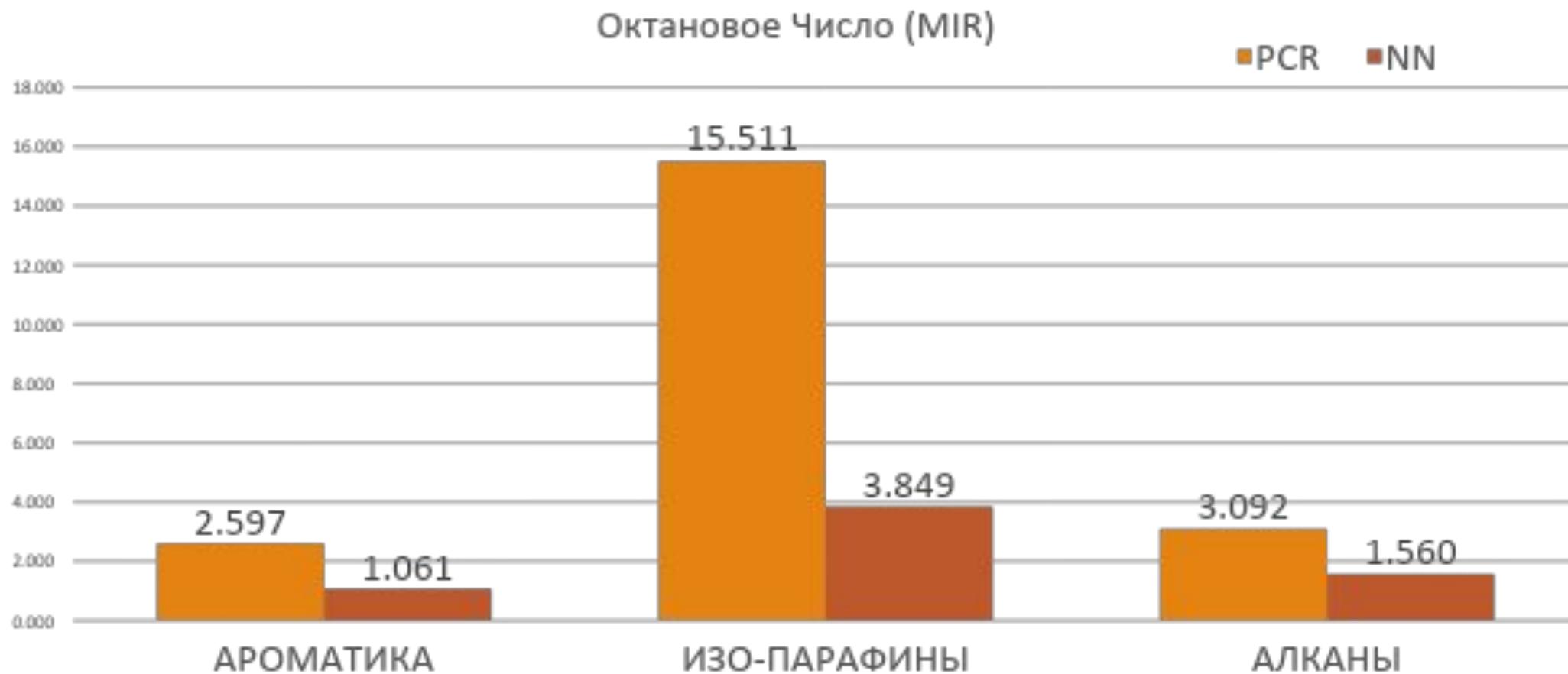
Сравнение моделей



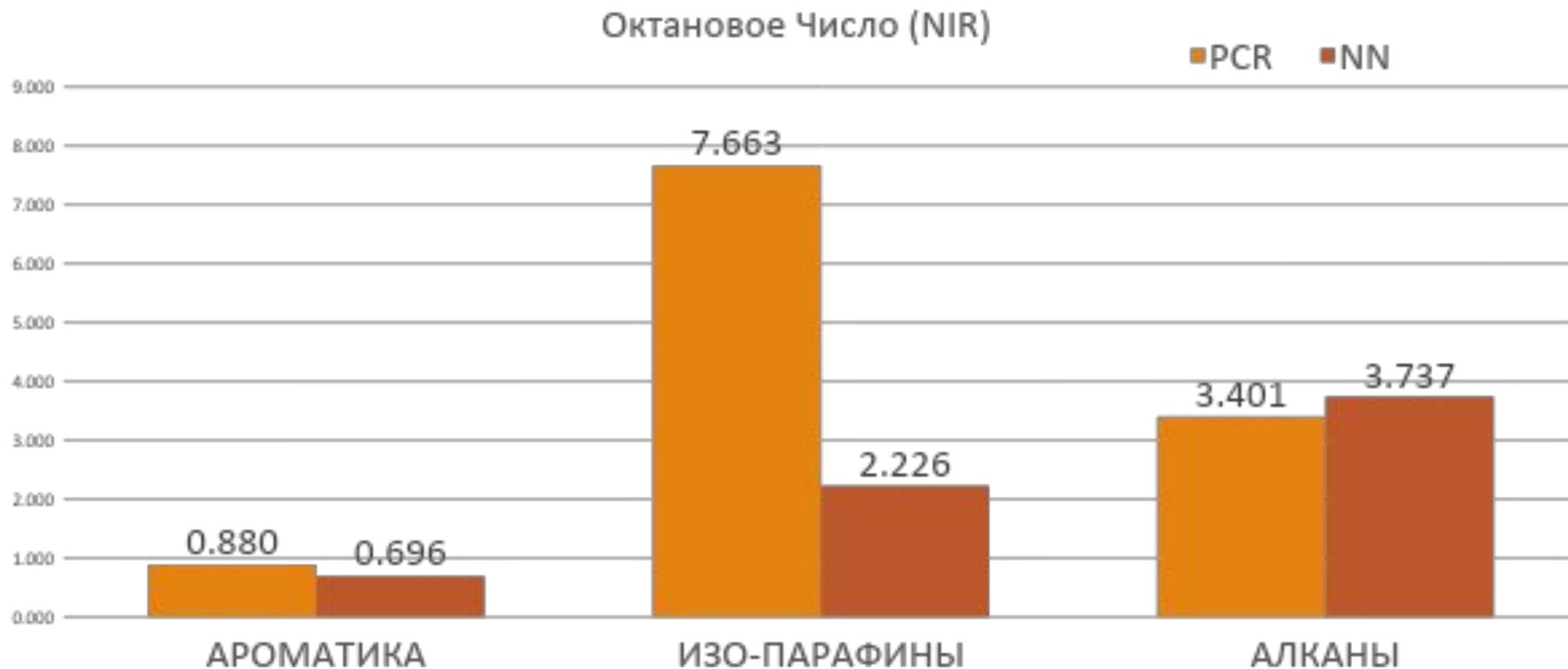
Сравнение моделей



Сравнение моделей



Сравнение моделей



Выводы

Результатом реализации рассмотренных методик стали модели, способные спрогнозировать химические и физические свойства индивидуальных углеводородов по их среднему и ближнему инфракрасному спектру. Полученные модели были исследованы и оценены, проведено детально сравнение этих моделей между собой. Данные модели могут быть применены на производстве углеводородов в качестве методов экспресс-анализа веществ и для контроля в реальном времени.