



Разработка бизнес-приложений
на платформе Lexema

ОПЫТ ПРИМЕНЕНИЯ АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РАБОТЕ РОБОТА-ВРАЧА





Проект: ООО «Лаборатория гемодиализа» – сеть из 27 гемодиализных клиник, работают на ERP-Lexema более 5 лет, врачи в системе делают назначения пациентам, фиксируют результаты анализов.



Проблема:



Недостаточная квалификация врачей на периферии в гемодиализных центрах

Неэффективные назначения врачей

Оперативно после получения актуальных анализов пациента, выдавать в системе рекомендации по назначению лечения.

Рекомендации системы должны быть оптимальными с точки зрения эффективности лечения для конкретного пациента



**ГЛАВНЫЙ КРИТЕРИЙ ЭФФЕКТИВНОСТИ ЛЕЧЕНИЯ – УЛУЧШЕНИЕ
ПОКАЗАТЕЛЕЙ АНАЛИЗОВ КРОВИ ПАЦИЕНТА**

ОБОСНОВАННОСТЬ ПРИМЕНЕНИЯ МЕТОДОВ ИИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ

ВАРИАНТ №1:

Применение линейных
алгоритмов

НЕ ПОДОШЕЛ

**ВАРИАНТ №2:**

Применение
методов ИИ

ВЫБРАН



ПРИЧИНА: большое количество факторов, от которых зависит выбор назначения и, как следствие, многовариантность назначений

1. Разметка данных
2. Сбор репрезентативных данных
3. Выравнивание выборок
4. Формирование обучающей и тестовой выборок
5. Отбор признаков и факторов, влияющих на лечение
6. Выбор алгоритмов ML
7. Выбор метрик качества
8. Проверка результатов на тестовой выборке
9. Подбор эффективной стратегии лечения
10. Встраивание алгоритмов в медицинскую ERP-систему

На сегодняшний день в диализных центрах ООО «Лаборатория гемодиализ» используется ERP-система Lexema для учета данных о пациентах на диализе



Собрана **база данных**, позволяющая оперативно извлекать данные в динамике по любому пациенту диализного центра.



Получены размеченные данные, относящие противоанемическую терапию (ААТ) и терапию для восстановления фосфорно-кальциевого обмена (ФКО) к эффективной или нет.



Особенность: избыточная терапия относилась также к неэффективной.



Была подтверждена гипотеза о согласованности мнений экспертов



Основными критериями эффективности терапий ФКО и ААТ считалось:

- 1) **достижение целевых показателей анализов крови пациентов**
- 2) **наличие устойчивой положительной динамики показателей в течении 3-х месяцев**



**для противоанемической терапии (ААТ) было получено – более 9 тысяч записей,
из них 39% эффективная терапия, 61% - неэффективная,
из неэффективной терапии: избыточная 22%, недостаточная и неэффективная – 78%**



**для противоанемической терапии (ФКО) было получено – около 9 тысяч записей,
из них 55% эффективная терапия, 45% - неэффективная,
из неэффективной терапии: избыточная 8%, недостаточная и неэффективная – 92%**

Решение проблемы отсутствия сбалансированности – сэмплирование выборки (изменение пропорции классов за счет увеличения меньшего класса, или уменьшения большего класса)



Алгоритм оверсэмплинга: размножение меньшего класса
Библиотека в Python: `imblearn.over_sampling`.



Алгоритм RUS (Random undersampling): уменьшение большего класса случайного образом
Библиотека в Python: `imblearn.over_sampling (RUS)`.



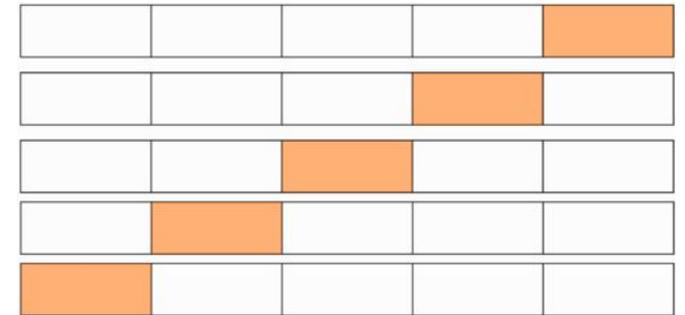
Алгоритм SMOTE: добавление в меньший класс «искусственных» похожих примеров (образцов)
Библиотека в Python: `imblearn.over_sampling (SMOTE)`.



Слепая валидация:

70% данных — обучающая выборка, 30% данных — тестовая выборка

Проверка работы методов ИИ на тестовых данных и сравнение результатов работы методов с данными экспертов



Кросс валидация:

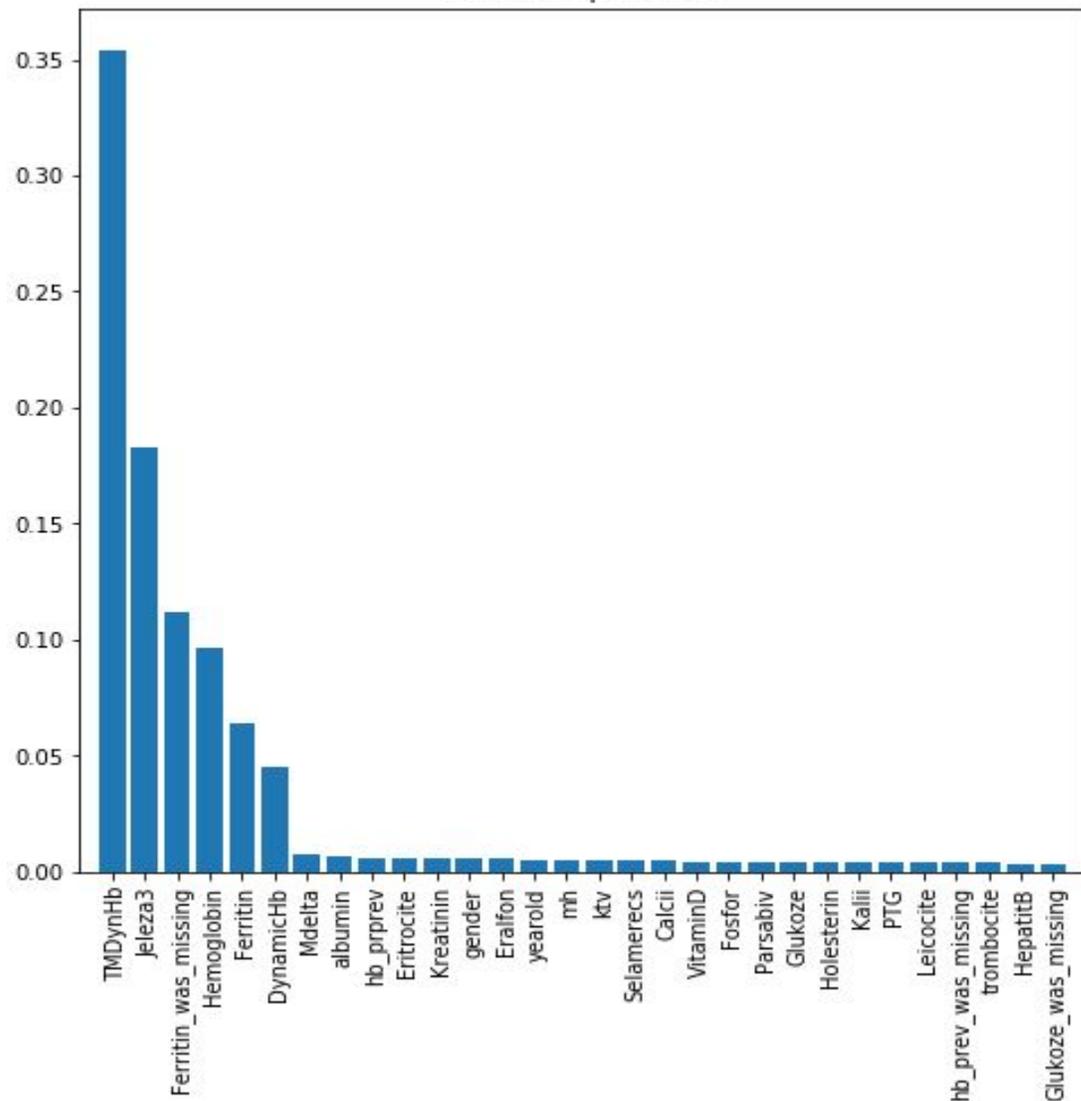
Кросс-валидация для всех 4 моделей проводилась на 5 фолдах.

Принцип построения моделей классификации — воронка

I Этап – прогнозирование эффективная/неэффективная терапия

II Этап – прогнозирование среди неэффективных избыточной/неэффективной терапии

Feature importances



Всего выявлено: 57 факторов

Из них 6 факторов имеют гораздо большую значимость, чем остальные

Примеры:

- Социально-демографические признаки пациента
- Наличие у пациента хронических вирусных заболеваний
- Длительность применения назначенных препаратов
- Длительность нахождения на диализном лечении



Нейросети — не подходят для применения в медицине.

Причина — поведение обученной нейросети не всегда может быть однозначно предсказуемо, что увеличивает риск применения.



**Наш выбор метода для оценка эффективности лечения —
Модифицированный бустинг.**

Причина модификации: повышение эффективности вычислительных ресурсов и, как следствие, уменьшение времени выполнения.



Использовали библиотеку `xgboost.sklearn` (XGBClassifier)

Матрица неточностей

Модель	Фактически	
	Положительно	Отрицательно
Положительно	TP	FP
Отрицательно	FN	TN

- **TP** (True Positives) – верно классифицированные положительные примеры
- **TN** (True Negatives) – верно классифицированные отрицательные примеры
- **FN** (False Negatives) – положительные примеры, классифицированные как отрицательные
- **FP** (False Positives) – отрицательные примеры, классифицированные как положительные

Метрики качества определялись на основе матрицы неточностей

Чувствительность и специфичность алгоритма:

$$Se = TPR = \frac{TP}{TP + FN} \cdot 100\%$$
$$Sp = \frac{TN}{TN + FP} \cdot 100\% = 100 - FPR$$

F-мера:

$$F = 2 \frac{TPR \cdot FPR}{TPR + FPR}$$

Accuracy:

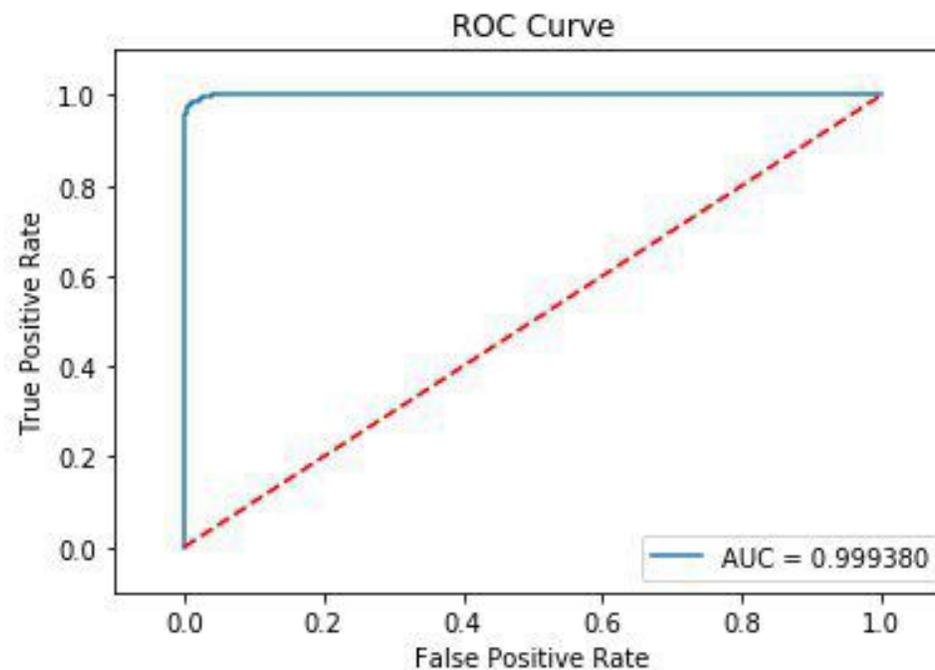
$$\frac{TP + TN}{TP + FN + FP + TN}$$

Чувствительность — **98%**

Специфичность — **98%**

F = **0,98**

AUC = **0,99**





Использовался алгоритм мэтчинга.

Критерием качества в мэтчинге считалось соответствие терапии профилю пациента не менее, чем на 80%.

Поиск соответствия проводился по базе прецедентов. Алгоритм мэтчинга был адаптирован за счет учета всех параметров по поиску наибольшего соответствия с весами.

Веса были предоставлены:

- в ходе экспертного опроса врачей-нефрологов
- в результате многократного тестирования предложенных методов
- в результате долевого распределения 10 ранжированных по значимости влияния признаков (для этого использовалась логистическая регрессия, построенная над всеми размеченными данными)



Алгоритм SequenceMatcher : сравнения пар последовательностей

Библиотека в Python: difflib. SequenceMatcher

Процент эффективных назначений, полученных из ERP-системы по оценке ИИ:

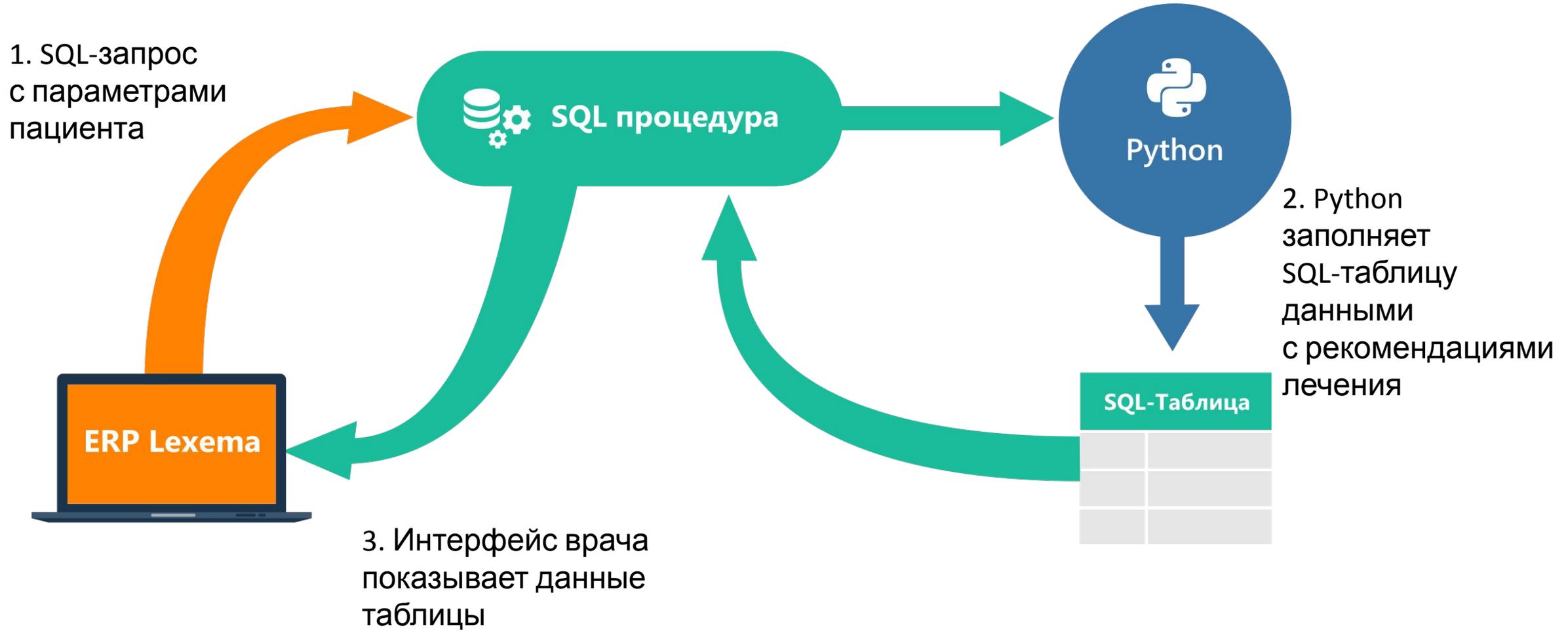
49%

Целевой показатель эффективности для работа-врача: 98%

Эффективности назначений, т.е. улучшений показателей крови

КАК РЕШЕНИЕ/МОДЕЛЬ ВСТРОЕНО В ERP-СИСТЕМУ?





Назначения КСГ X

Рекомендации | Показатели

Пациент Период 01.09.2019 - 30.09.2019

Лечение X

Выбор	Лечение	Тип	Медикамент	Ед.изм.	Способ исполнения	Периодичность	Доза назначения	Кол-во процедур	Цена	Анализы
▼ Лечение: Вариант 1 совпадение 88.2%. Стоимость лечения: 5993.53 руб.										
<input type="checkbox"/>	Вариант 1 сов...	Анемия	Эральфон 2000 ЕД р-р для и...	ЕД	Через экстрако...	Ежепроцедурно	2 000.00	13	5 100.33	Гемоглобин: 103; Ферр...
<input type="checkbox"/>	Вариант 1 сов...	ФКО	Альфа Д3-Тева 0,25 мкг . таб...	мкг.	Внутри	Ежедневно	0.50	35	602.00	Кальций: 2.1; Фосфор н...
<input type="checkbox"/>	Вариант 1 сов...	ФКО	Кальция карбонат 2,0	гр.	Внутри	Ежедневно	2.00	35	291.20	Кальций: 2.1; Фосфор н...
▼ Лечение: Вариант 2 совпадение 88.2%. Стоимость лечения: 5993.53 руб.										
<input type="checkbox"/>	Вариант 2 сов...	Анемия	Эральфон 2000 ЕД р-р для и...	ЕД	Через экстрако...	Ежепроцедурно	2 000.00	13	5 100.33	Гемоглобин: 103; Ферр...
<input type="checkbox"/>	Вариант 2 сов...	ФКО	Альфа Д3-Тева 0,25 мкг . таб...	мкг.	Внутри	Ежедневно	0.50	35	602.00	Кальций: 2.1; Фосфор н...
<input type="checkbox"/>	Вариант 2 сов...	ФКО	Кальция карбонат 2,0	гр.	Внутри	Ежедневно	2.00	35	291.20	Кальций: 2.1; Фосфор н...
▼ Лечение: Вариант 3 совпадение 82.4%. Стоимость лечения: 18437.83 руб.										
<input type="checkbox"/>	Вариант 3 сов...	Анемия	Эральфон 2000 ЕД р-р для и...	ЕД	Через экстрако...	Ежепроцедурно	2 000.00	13	5 100.33	Гемоглобин: 103; Ферр...
<input type="checkbox"/>	Вариант 3 сов...	ФКО	Мимпара 30мг. №28	мг	Внутри	Ежедневно	30.00	35	13 337...	Кальций: 2.3; Фосфор н...

	 Anemia Control Management Германия	 Strategic Anemia Advisor США	 Lexema-Medicine Россия
Эффективность	70-83%	65%	98%
Чувствительность	92%		98%
Специфичность	75%		98%



Лакман Ирина

Руководитель направления
ИИ

в Lexema

Тел: 8 800 222 55 13

С УДОВОЛЬСТВИЕМ

ОТВЕЧУ НА ВСЕ ВАШИ ВОПРОСЫ

