ДОКЛАД

Применение метода деревьев решений к задаче планирования геолого-технических мероприятий

Родион Скобликов

Искусственный интеллект в нефтегазовой отрасли

Тюменский институт нефти и газа

Планирование ГТМ | Деревья решений | Повышение предсказательной способности | Машинное обучение | Повышение достоверности прогноза | Подбор скважин-кандидатов

Актуальность

Геолого-технические мероприятия (ГТМ) - комплекс работ, проводимых на скважинах для регулирования разработки месторождения и поддержания целевых уровней добычи нефти. Другими словами необходимы для поддержания удовлетворительного технического состояния скважин, поддержания или увеличения уровня добычи нефти

Планирование ГТМ

1. Определение скважин-кандидатов
2. Определение вида и характеристик ГТМ
3. Оценка прогнозной эффективности

ТЕХНОЛОГИЧЕСКАЯ ПЛАТФОРМА

АТЛАС

Возможность хранения разноплановой информации

ПЕРСПЕКТИВА ИНТЕГРАЦИИ

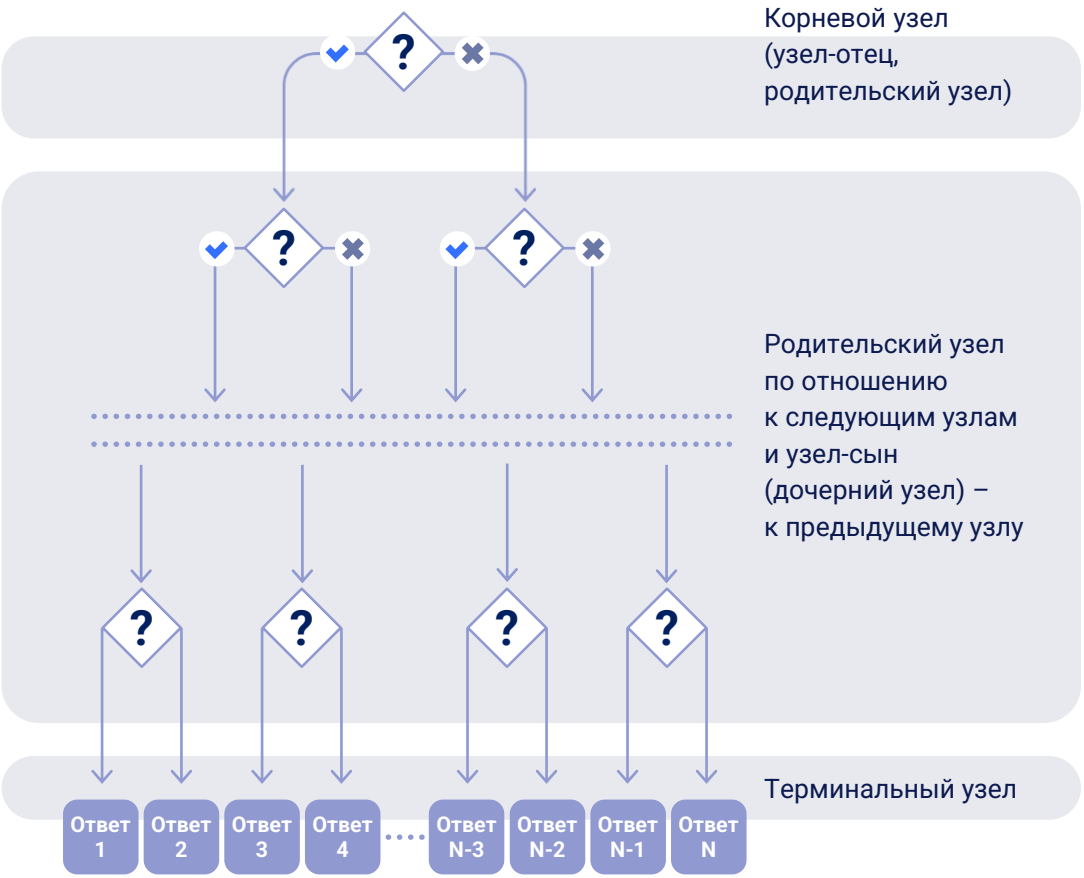
Методы	**Моделирование** Фильтрационные модели Термодинамические модели ...	**Экспертные методы** Решение технолога ...	**Традиционные аналитические методы** Характеристики вытеснения Кривые падения ...	**Машинное обучение** Нейронные сети Деревья решений Линейная регрессия ...
Преимущества	Большой опыт использования Устоявшаяся методология	Большой опыт использования Простота	Большой опыт использования Устоявшаяся методология	Быстрота Автоматизация Нахождение сложных связей
Недостатки	Трудозатратность	Низкая прогнозная способность Нет единой методологии	Не позволяет сформировать план ГТМ Низкая прогнозная способность	Небольшой опыт применения Нет единой методологии

Применение машинного обучения

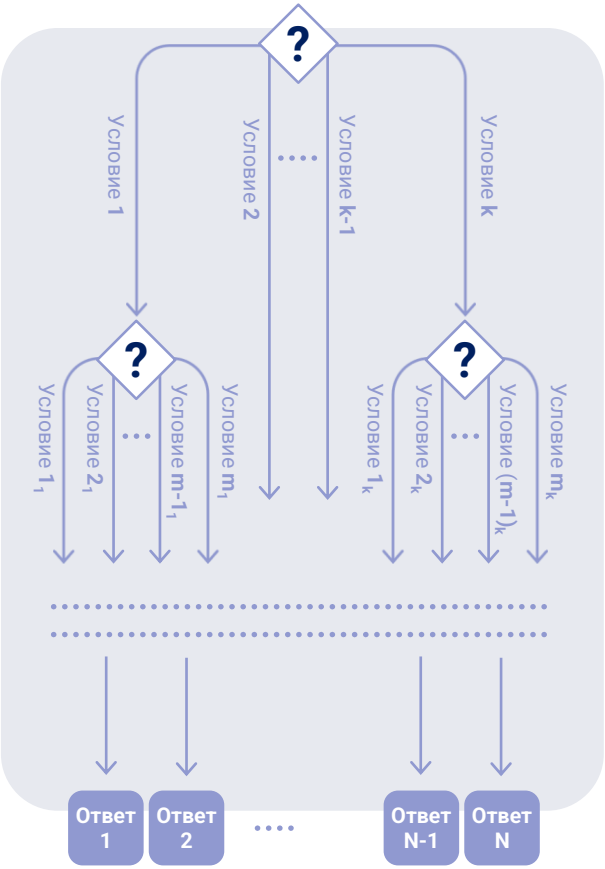
МЕТОД	ПРЕИМУЩЕСТВА	НЕДОСТАТКИ	МЕТОДОЛОГИЯ
Кластерный анализ	Скважины-кандидаты могут быть ранжированы по приоритету	Оценка эффективности количественно не производится Основан на экспертной оценке	Ахмедов К.С.
Эмпирические зависимости	Количественная оценка эффективности Функциональная зависимость	Зависимости построены по данным одного газового месторождения Зависимости построены не по всем ГТМ	Гасумов Р.А., Ахмедов К.С., Толпаев В.А., Филиппов А.Г.
«Ретроспективный» анализ	Функциональная зависимость Скважины-кандидаты могут быть ранжированы по приоритету	Построение функции для каждой скважины Зависимость линейная Результаты не интерпретируемы Длительное обучение Оценка эффективности количественно не производится	Синицына Т.И., Галеев А.А.
Нейронные сети	Полностью автоматизированный метод Количественная оценка эффективности Работа с любым типом данных и большим массивом данных Выявляет как линейные так и нелинейные зависимости	Результаты не интерпретируемые Проведенные исследования по данному методу показали неудовлетворительные результаты прогноза	Курганов Д.В., Кочнев А.А., Козырев Н.Д., Кочнева О.Е., Галкин С.В., Пичугин О.Н., (Совокупность работ)
Линейный дискриминантный анализ	Количественная оценка эффективности Функциональная зависимость	Зависимость линейная Зависимость для одного месторождения Результаты не интерпретируемые	Кочнев А.А., Козырев Н.Д., Кочнева О.Е., Галкин С.В.
Байесовские сети	Основан на принципах Марковского движения	Оценка эффективности количественно не производится Результаты носят исключительно вероятностный характер интерпретации Результаты трудноинтерпретируемые Связи между параметрами сложные	Денисов О.В.
Деревья решений	Более объективные и интерпретируемые результаты Выявляет как линейные так и нелинейные зависимости Малое время обучения Работа с любым типом данных и большим массивом данных Может работать с минимальным набором данных Интеграция других методов	Нестабильность Результаты носят исключительно вероятностный характер интерпретации	Курганов Д.В., Monalisa Jena, Satchidananda Dehuri, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Shamy S, Dr. J Dheeba, Yang, Y. Yi, F. Deng, C. Sun, G

Суть деревьев решений

Структура бинарного дерева решений



Структура небинарного дерева решений



Концепция применения деревьев решений

- **Формирование** обучающей выборки
- **Прогноз** (классификация)
- **Обучение** (выявление более значимых условий, влияющих на результат)

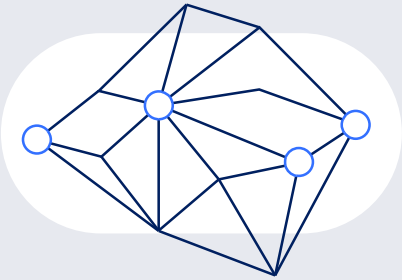
Популярные методы деревьев решений

Метод	Математическая основа	Критерий расщепления	Особенности
ID3	Информационная неоднородность	Информационная энтропия	Нестабилен, больше подходит для классификаций
CART	Информационная неоднородность	Индекс Джини	Нестабилен, больше подходит для классификаций
Random Forest	Бэггинг Информационная неоднородность	Индекс Джини Информационная энтропия	Стабилен, хорошо подходит для прогнозирования Неинтерпретируемые результаты (ансамбль деревьев)
Gradient Boosting Trees	Бэггинг Информационная неоднородность Градиентный бустинг	Индекс Джини Информационная энтропия	Стабилен, хорошо подходит для прогнозирования, Неинтерпретируемые результаты (ансамбль деревьев) Сильно зависит от размерности выборки
QUEST	Вероятностное распределение	Квадратичный дискриминантный анализ	Стабилен, хорошо подходит для прогнозирования, Сильно зависит от размерности выборки Не работает с непрерывными ответами
CHAID	Вероятностное распределение	Хи-квадрат	Стабилен, хорошо подходит для прогнозирования
Treed Polynomial Classifier	Полином	Изменение полинома	Нестабилен, труднорегистрируемые результаты Работает только с непрерывными величинами Не подходит для выборок с большим количеством предикторов
REPTree	Рекуррентная нейронная сеть	Условие нейронной сети	Нестабилен, хорошо подходит для прогнозирования, труднорегистрируемые результаты

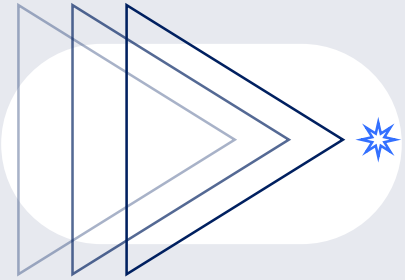
Цель и задачи работы

Разработка формализованного алгоритма оценки эффективности и планирования ГТМ

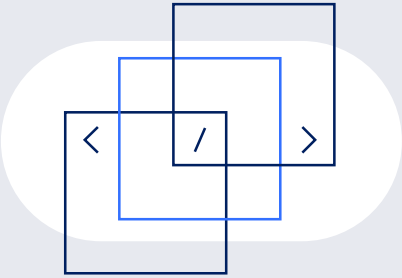
01 **Формирование** модели данных



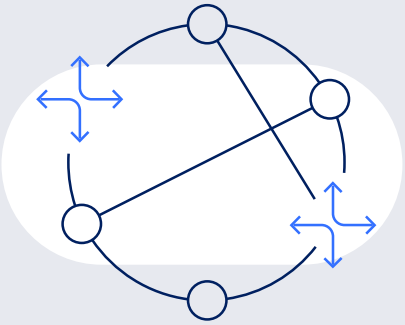
02 **Выбор** оптимального метода



03 **Разработка** критериев применения



04 **Апробация** разработанного алгоритма



Концепция применения деревьев решений

Модель данных

↑

Фактическая информация выполненных ГТМ

Построение

↑

Математическая модель

Результат

1. Интерпретируемость
2. Формализация
3. Автоматизация

Формирование модели данных

Категории входных параметров

1	**Общие сведения**	Номер скважины, пласт, месторождение, дата и т.д.
2	**Геологические**	Коэффициент расчленённости, коэффициент песчанистости, НГП, тип коллектора, мощность и т.д.
3	**Сведения об эксплуатации скважин**	Конструкция скважины, дебит нефти, дебит жидкости, обводненность, забойное давление и т.д.
4	**Параметры технологии ГТМ**	Давление разрыва пласта, расход агента ОПЗ, масса пропанта, полудлина трещины ГРП, столб жидкости над ГНО и т.д.
5	**Параметры эффективности**	Прирост дебита нефти, доп. добыча, обводненность и т.д.

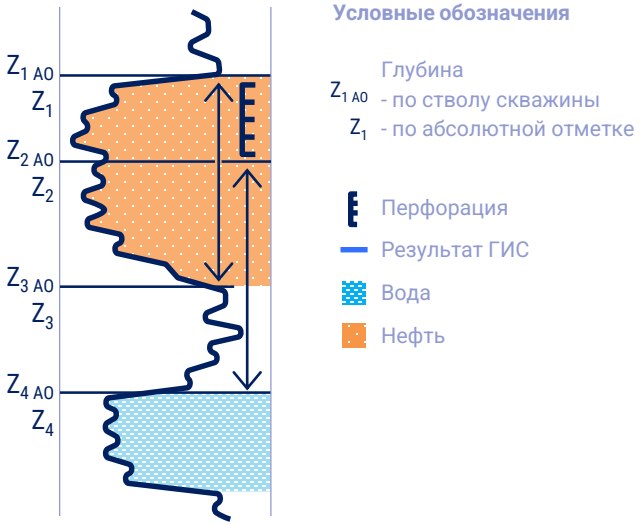
Всего **56** параметров

39 количественных

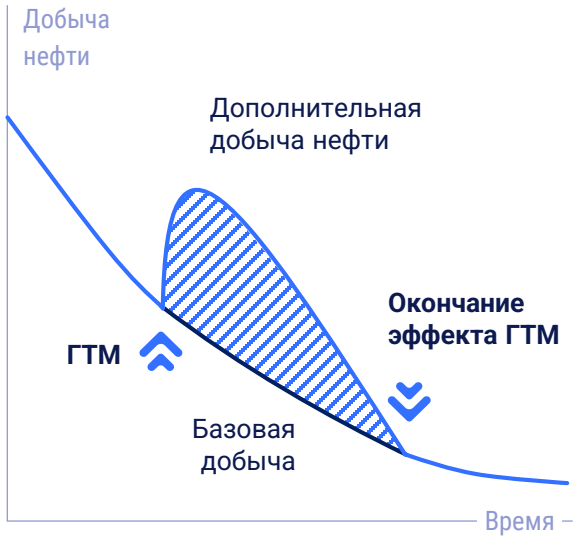
17 номинальных

- Учет параметров описывающие геолого-технические особенности
- Учет особенностей технологии ГТМ

Корреляционная диаграмма



Оценка технологического эффекта



Модель данных

641 наблюдение (фактические ГТМ)

ОПЗ, ГРП, Оптимизация, ПВР, РИР

3 НГП (Волго-Уральская, Тимано-Печорская, Западно-Сибирская)

8 месторождений

Выбор оптимального метода

Определение точности

Одно дерево

Разделение на обучающую и тестовую выборки случайным образом, и формирование разных классов

№ скв	Параметры		Эффект от ГТМ	Класс
157	X_1	Y_1		Обучение
354	X_2	Y_2		Тест

Сопоставление результата дерева решений с фактическим и данными

Прогноз = Факт

Оценка качества построения дерева

№	Прогноз	Факт	Соответствие
1	Класс 1	Класс 1	+
2	Класс 1	Класс 2	–
...	...	...	...
	Процент правильных прогнозов:		***N %***

Класс - категория в которую модель пытается отнести входные данные. Например доп. добыча >1 (класс 1) и <1 (класс 2) тыс. тонн

Генерация множества деревьев (>800)

№ построения	Процент правильных прогнозов
1	N_1 %
2	N_2 %
...	...
Средний процент правильных прогнозов	***P %***

Точность

Среднее значение % верно спрогнозированных исходов по множеству уникальных деревьев

71,3%

Метод	Учет номинальных параметров	Интерпретируемость	Возможность регрессионного подхода	Возможность небинарного расщепления	Гибкость	Точность
CHAID	+	+	+	+	+	+
QUEST	–	+	+	+	–	+
CART	+	+	–	–	+	–
RF	+	–	–	–	+	+

Подход

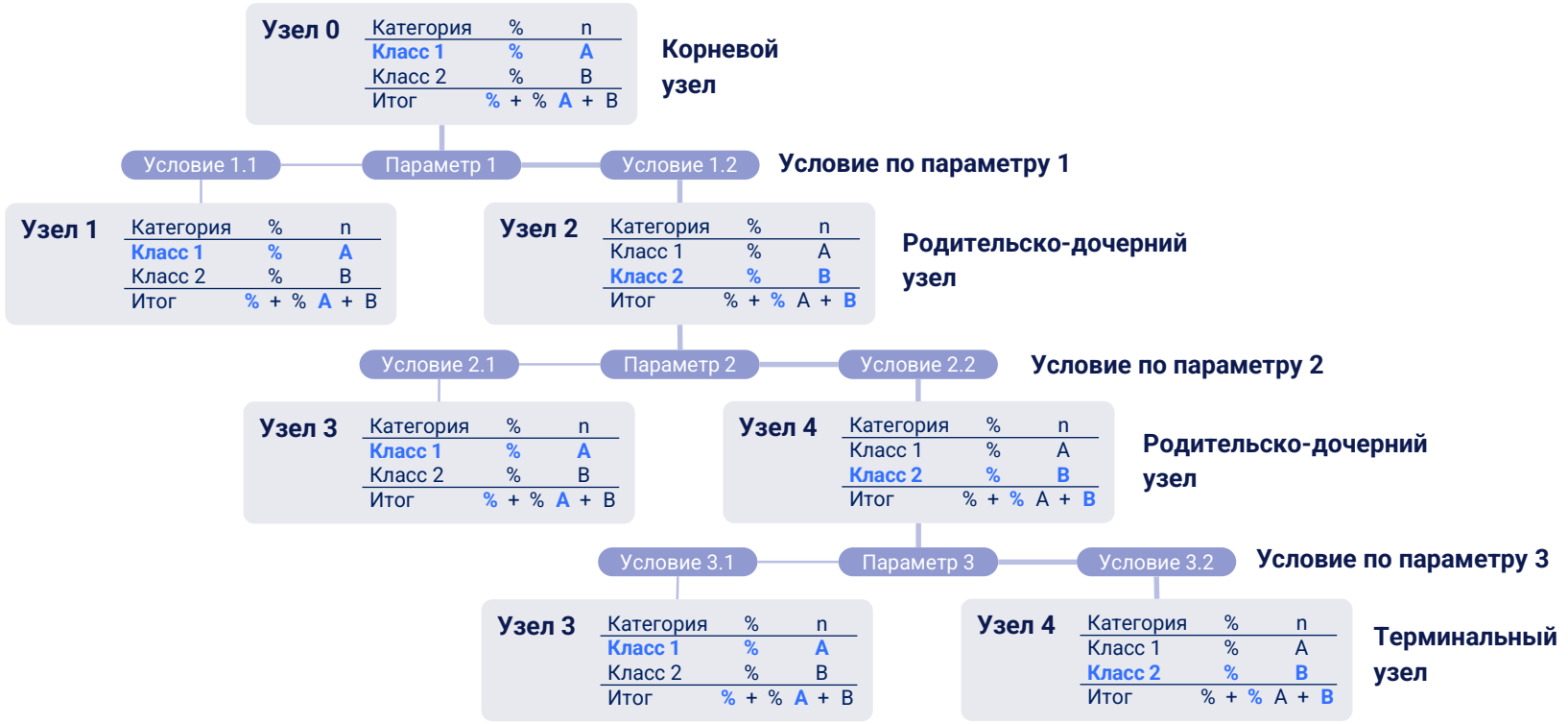
Регрессионный — Дает количественный ответ, но имеет большую погрешность в рассматриваемой задаче

Дискретный — Дает номинальный ответ, но инвариантен и более точен

Гибкость

Возможность работать с относительно небольшой размерностью выборки

Разработка критериев применения



Структура и прогнозная способность деревьев **зависят** как от самих **параметров,** так и от их **конфигурации**

Параметры выборки:

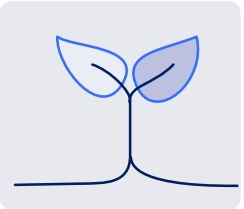
1. Вариативность
2. Размерность
3. Информативность

Параметры настройки дерева:

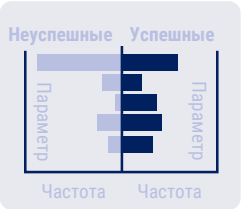
1. Минимальное количество наблюдений в родительском узле
2. Минимальное количество наблюдений в дочернем узле
3. Количество интервалов разбиения количественных переменных
4. Максимальная глубина дерева

Анализ влияния параметров настроек

Деление исходной выборки на **подвыборки-группы исследований** с учетом влияющих параметров построения



Построение деревьев решений, **сбор** статистических **данных**



Анализ результатов построений

- Критерий Манна-Уитни
- Критерий Колмогорова-Смирнова

Успешное дерево – дерево с более 70% правильно прогнозируемых событий

Критерии оптимальной применимости

Параметр	Критерий
Принцип аналогии	Необходимо использовать информацию соответствующую объекту прогноза (вид ГТМ, одновозрастные отложения…)
Количество наблюдений в выборке	Не менее 90
Вариативность выборки	Более 33%
Количество наблюдений в каждом обучающем классе	Не менее 40 При небольшой выборке необходима однородность в классах
Мин. кол-во наблюдений в родительском узле	от 15% до 25% от выборки
Минимальное количество наблюдений в дочернем узле	от 10% до 20% от выборки
Количество интервалов разбиения количественных переменных	от 10 до 20

Требования были сформированы по результату анализа более

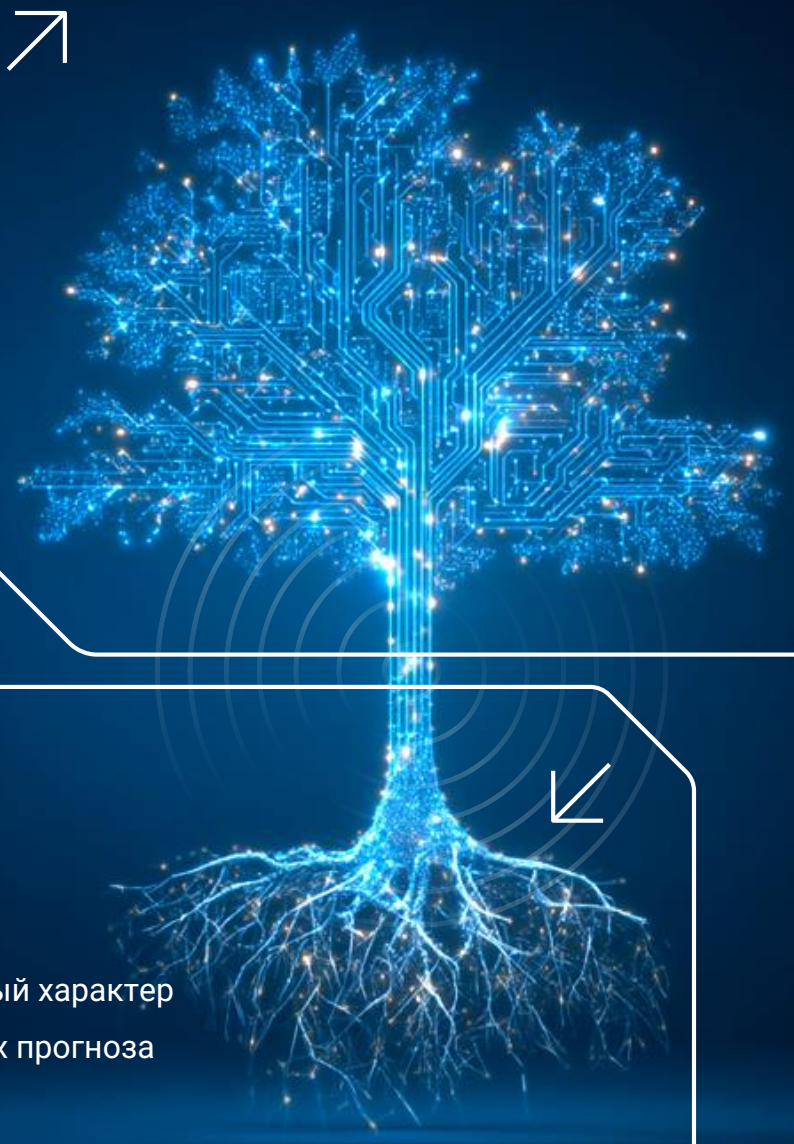
6 000 деревьев

Выполнение указанных критериев обеспечивает **оптимальные условия** применения деревьев решений в задаче планирования ГТМ

Средний процент правильно предсказанных событий деревьями при соблюдении разработанных критериев составляет

75%

Выводы



Разработанный алгоритм позволяет:

1. Получать достоверные прогнозы
2. Получать интерпретированные результаты
3. Формализовать задачу и унифицировать подход
4. Получать прогноз по нескольким сценариям

Минусы:

1. Нет количественного ответа
2. Ответ носит исключительно вероятностный характер
3. Нет дополнительной верификации данных прогноза

Перспективы развития темы

1 Модель данных

Сведения о пластовой системе и насыщаемых флюидах

- Проницаемость пласта
- Толщина пласта
- Расчлененность пласта
- Начальная нефтенасыщенность
- Стратиграфические границы пласта
- Начальное пластовое давление
- Давление насыщения
- Вязкость нефти, воды

Промысловые показатели

- Траектория скважины
- Интервалы вскрытия пласта
- Дебиты нефти/жидкости
- Коэффициент эксплуатации
- Газовый фактор
- Забойное давление
- Карты изобар
- Результаты ПГИ на техническое состояние

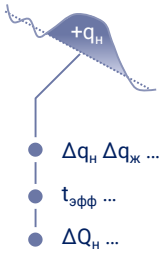
Технологические параметры ГТМ

- ГРП: масса проппанта, количество стадий
- ОПЗ: вид технологии, объем закачки
- ПВР: интервалы вскрытия, тип перфоратора, плотность отверстий

2 Алгоритм

Оценка фактической эффективности ГТМ

- Расчет базовых показателей
- Определение даты завершения эффекта
- Расчет технологических показателей эффективности: амплитудные приросты дебитов (приемистостей), продолжительность эффекта, дополнительная добыча нефти



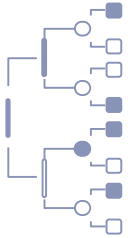
Обучение модели машинного обучения

Методы:

- CHAID/Random Forest…
- LSTM, RNN, Boruto…

Результат

Модель машинного обучения, применимая в задаче подбора скважин-кандидатов для ГТМ и прогнозирования технологического эффекта от проведения работ



3 Результаты работы функции

Добывающий фонд

№ скв.	Вид ГТМ	Технология	Показатели до ГТМ			Ожидаемые показатели после ГТМ			$t_{эфф}$, мес
			$q_н$ т/сут	$q_ж$ т/сут	$f_в$ %	$q_н$ т/сут	$q_ж$ т/сут	$f_в$ %	
5022	ГРП	40 т проппанта, 1 стадия	13	40	69	35	120	71	17
	ОПЗ	СКО	13	40	69	17	54	69	6
	ПВР	Интервал дострела [2783, 2790]	13	40	69	18	55	68	29
	РИР	Мероприятие нецелесообразно							

Нагнетательный фонд

№ скв.	Приемистость, $м^3$/сут		Рекомендация		Расчетная приемистость после ГТМ $м^3$/сут
	Текущая	Оптимальная	Вид ГТМ	Технология	
4087	25	70	ОПЗ	СКО	65
4022	32	85	ПВР	Реперфорация [2678; 2688]	90
4087	75	150	ОПЗ	СКО	145

Спасибо за внимание!

Родион Скобликов

Применение метода деревьев решений к задаче планирования геолого-технических мероприятий

Искусственный интеллект в нефтегазовой отрасли

ТИНГ Тюменский институт нефти и газа

Планирование ГТМ | Деревья решений | Повышение предсказательной способности | Машинное обучение | Повышение достоверности прогноза | Подбор скважин-кандидатов