



ИНЖИНИРИНГОВЫЙ ЦЕНТР

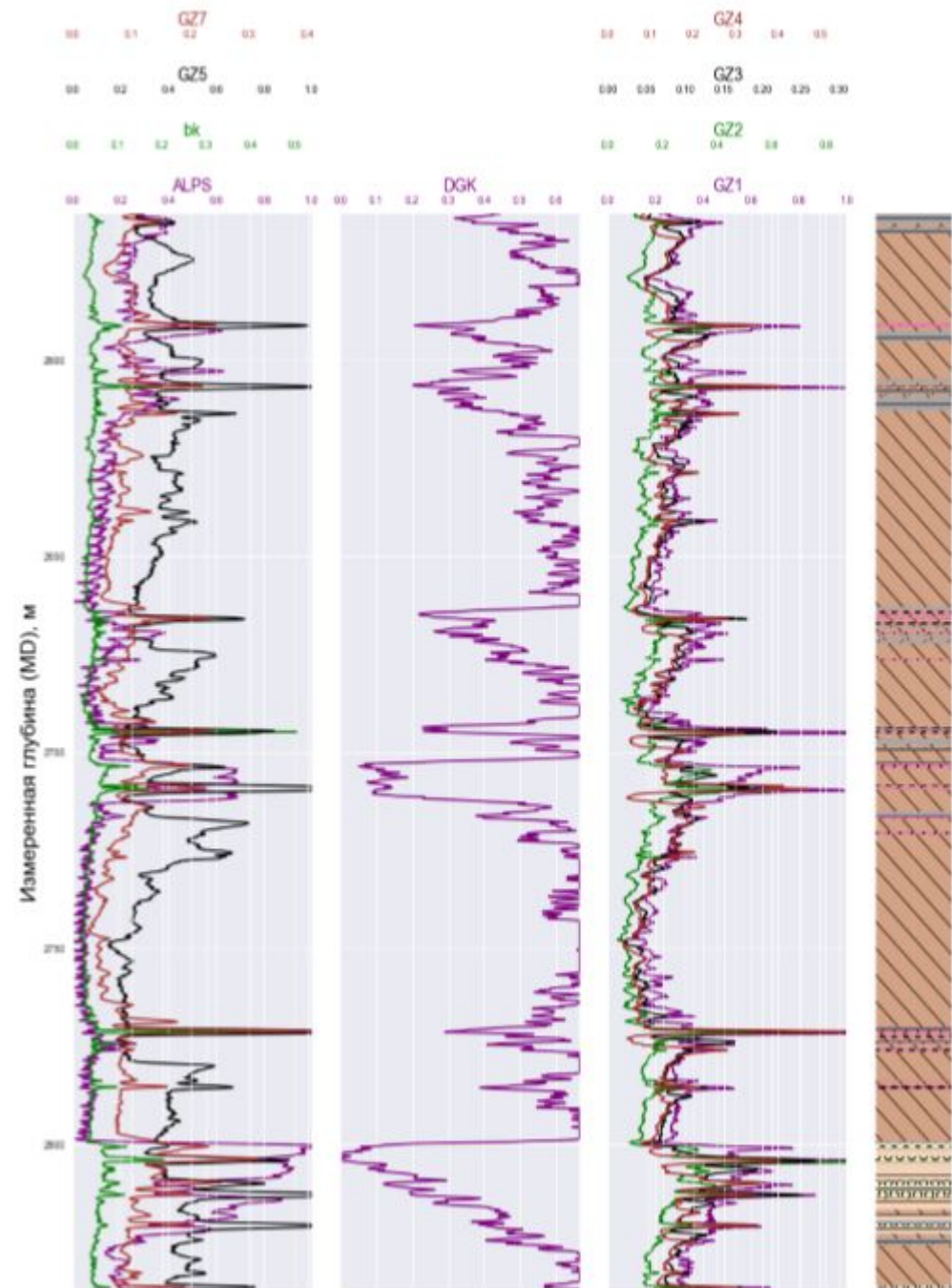
МФТИ

**Тестовое задание на позицию
«Data Scientist»**

Бабушкина Татьяна

Постановка задачи

1. Реализовать модель оценки запасов нефти на основе исторических данных
2. Создать экономическую модель, обосновывающую достаточный комплекс исследований скважины



Параметры входных данных

Обозначение	Расшифровка	Данные
bk	Боковой каротаж	+
GZ1-GZ5, GZ7	Градиент-зонд БКЗ	+
DGK	Гамма-каротаж	+
NKTD, NKTM, NKTR	Нейтрон-нейтронные каротажи	+
ALPS	Каротаж самопроизвольной поляризации	+
lith	Тип породы	+
goal	Наличие нефтяной фазы	+

Полученные значения коэффициента детерминации

Обучение на lith

Алгоритм

Logistic Regression	67,3
SVM with linear kernel	70,8
SVM with polynomial kernel	71,6
SVM with rbf kernel	78,0
kNN 1 neighbour	77,6
kNN 5 neighbours	77,5
kNN 11 neighbours	72,2
Random Forest 10 trees	70,1
Random Forest 100 trees	72,9
Random Forest 200 trees	73,6
	73,8

Обучение на goal

Алгоритм

Logistic Regression	67,3
SVM with linear kernel	70,8
SVM with polynomial kernel	71,6
SVM with rbf kernel	78,0
kNN 1 neighbour	77,6
kNN 5 neighbours	77,5
kNN 11 neighbours	72,2
Gradient Boosting 500 trees	69,6
Random Forest 10 trees	72,6
Random Forest 100 trees	73,6
Random Forest 200 trees	73,8

Выученные уроки: как определяли индекс сложности разработки месторождений RCI?

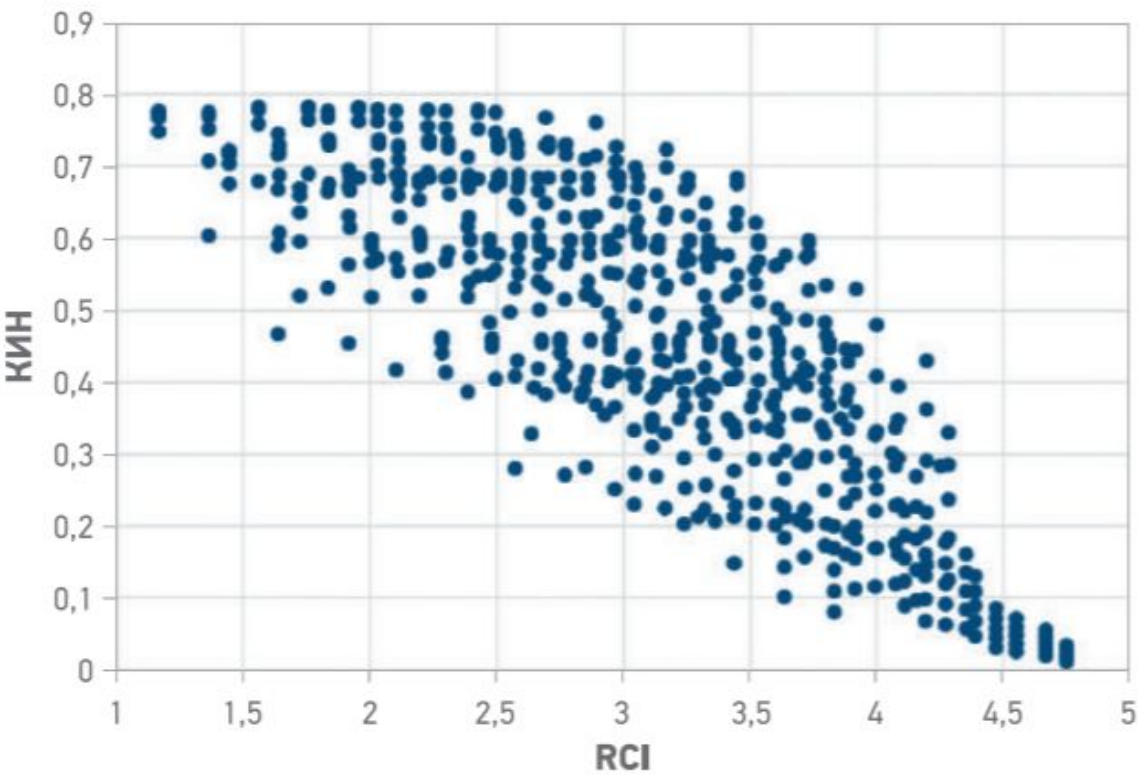
1. Класс сложности каждого параметра определяется по таблице

Группа вероятности	$\mu_{\text{н}}/\mu_{\text{в}}$		$K_{\text{выт}}$		$k, 10^{-3} \text{ мкм}^2$		$V_{\text{др}}$	
	Интервал	Класс	Интервал	Класс	Интервал	Класс	Интервал	Класс
[P1..20]	[0,67...0,81]	1	[0,21...0,4]	5	[0,16...9,17]	5	[0...0,1]	5
[P20...P40]	[0,81...0,83]	2	[0,4...0,44]	4	[9,17...42,39]	4	[0,1...0,3]	4
[P40...P60]	[0,83...0,84]	3	[0,44...0,47]	3	[42,39...78,64]	3	[0,3...0,5]	3
[P60...P80]	[0,84...0,85]	4	[0,47...0,52]	2	[78,64...129,74]	2	[0,5...0,7]	2
[P80...P100]	[0,85...0,93]	5	[0,52...0,65]	1	[129,74...1085,58]	1	[0,7...1]	1

2. RCI вычисляется по следующей формуле, где λ – весовой коэффициент, показывающий степень влияния рассматриваемого параметра на результат; K – класс сложности параметра при кластеризации

$$RCI = \sum_{i=1}^N \lambda_i K_i$$

количество слагаемых(новых исследований) мы выбираем самостоятельно в зависимости от уже полученного значения RCI



Облако значений КИН – RCI

Чем выше сложность месторождения, тем меньше нефти удастся добыть

Источники: Е.В. Белоногов, Д.С. Перец «Газпромнефть НТЦ»
“Совершенствование методики расчета коэффициента сложности разработки месторождения”

Какой из параметров самый важный?

Исследования проводим в порядке возрастания влияния полученной информации на goal, то есть чем больше корреляция этого признака и goal, тем раньше это исследование делаем(bk - ALPS - GZ1 - GZ5 - NKTD - NKTM).

	well_id	depth_m	bk	GZ1	GZ2	GZ3	GZ4	GZ5	GZ7	DGK	NKTD	NKTM	NKTR	ALPS
well_id	1.000000	0.077138	0.010641	-0.066570	-0.033818	0.073035	-0.008449	-0.016586	0.044055	0.048030	0.204927	0.289328	-0.267935	-0.035179
depth_m	0.077138	1.000000	0.225927	-0.033161	-0.129006	0.198673	-0.038665	0.108160	0.187192	-0.013077	0.190254	0.274496	-0.332965	-0.031809
bk	0.010641	0.225927	1.000000	0.339005	-0.297446	0.110476	-0.216892	0.355722	0.059959	-0.309604	0.346504	0.297684	-0.385680	0.091254
GZ1	-0.066570	-0.033161	0.339005	1.000000	-0.145835	-0.093114	-0.206725	0.367165	-0.108582	-0.180689	0.247468	0.128326	-0.217430	0.076823
GZ2	-0.033818	-0.129006	-0.297446	-0.145835	1.000000	-0.305916	0.291211	-0.180571	-0.256954	0.175620	-0.111651	-0.114944	0.147920	0.031923
GZ3	0.073035	0.198673	0.110476	-0.093114	-0.305916	1.000000	-0.025064	0.068071	0.403175	-0.012289	0.117413	0.156402	-0.181295	-0.079826
GZ4	-0.008449	-0.038665	-0.216892	-0.206725	0.291211	-0.025064	1.000000	-0.284129	-0.012114	0.161874	-0.107980	-0.045566	0.099679	-0.001219
GZ5	-0.016586	0.108160	0.355722	0.367165	-0.180571	0.068071	-0.284129	1.000000	-0.004907	-0.190391	0.343596	0.240168	-0.322409	0.040603
GZ7	0.044055	0.187192	0.059959	-0.108582	-0.256954	0.403175	-0.012114	-0.004907	1.000000	-0.010658	0.056502	0.106141	-0.133538	-0.074331
DGK	0.048030	-0.013077	-0.309604	-0.180689	0.175620	-0.012289	0.161874	-0.190391	-0.010658	1.000000	-0.080532	-0.048167	0.119764	0.006296
NKTD	0.204927	0.190254	0.346504	0.247468	-0.111651	0.117413	-0.107980	0.343596	0.056502	-0.080532	1.000000	0.803551	-0.581461	0.058207
NKTM	0.289328	0.274496	0.297684	0.128326	-0.114944	0.156402	-0.045566	0.240168	0.106141	-0.048167	0.803551	1.000000	-0.470499	0.082250
NKTR	-0.267935	-0.332965	-0.385680	-0.217430	0.147920	-0.181295	0.099679	-0.322409	-0.133538	0.119764	-0.581461	-0.470499	1.000000	-0.000714
ALPS	-0.035179	-0.031809	0.091254	0.076823	0.031923	-0.079826	-0.001219	0.040603	-0.074331	0.006296	0.058207	0.082250	-0.000714	1.000000
lith	-0.107453	-0.092688	0.449937	0.256789	-0.173477	-0.082366	-0.159424	0.180178	-0.065279	-0.322571	0.074683	0.045146	-0.095533	0.167164
goal	-0.110368	-0.125804	0.401875	0.221030	-0.112740	-0.128354	-0.134134	0.142911	-0.107349	-0.259724	0.011401	0.002383	-0.013842	0.241537

Алгоритм определения стоимости исследований месторождения

1. Проводить исследования в порядке возрастания важности полученной информации для обучения модели.
2. На первом шаге необходимо получить параметры: b_k - ALPS - GZ1 - GZ5 - NKTD - NKTM, обучить модель (на этих признаках точность предсказания не более 74,8%)
3. В зависимости от результатов, оценить есть ли смысл вкладывать в дальнейшую разведку.
4. С получением каждого нового признака обучать модель и смотреть, получается ли найти необходимое количество нефтенасыщенных пропластков с достаточной точностью.

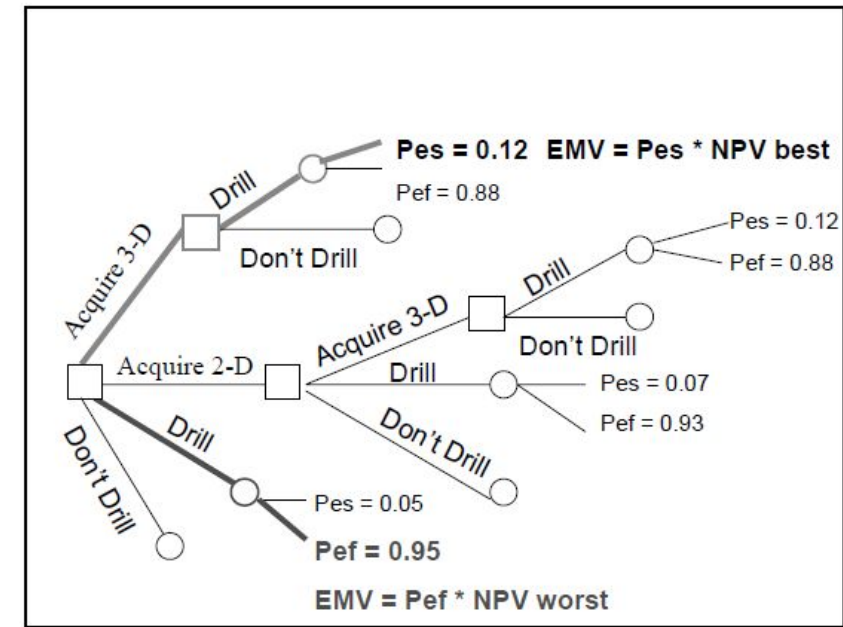
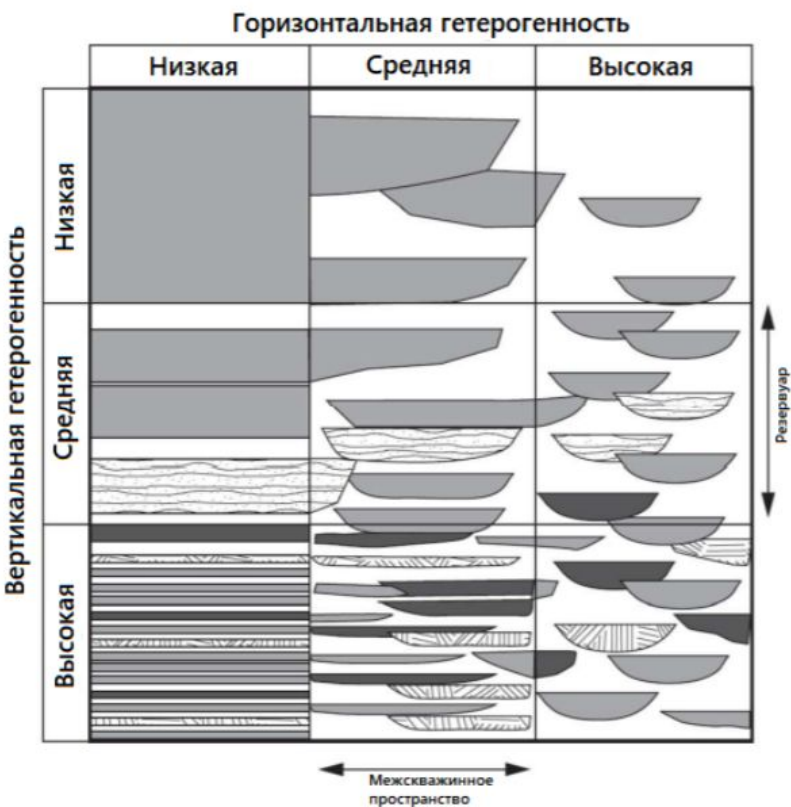


Fig. 3: What alternatives exist?
3-D data influences Pes and NPV

Важно учитывать не только толщину нефтенасыщенных пропластков, но и их форму



На коэффициент извлечения нефти(КИН) влияет не только толщина пропластка, но и его архитектурная сложность. Высокая гетерогенность пласта требует больших вложений при добыче нефти, что наша экономическая модель не учитывает

Матрица гетерогенности архитектуры резервуара

		Горизонтальная гетерогенность		
		Низкая	Средняя	Высокая
Вертикальная гетерогенность	Низкая	Дельта волнового типа Барьерный остров Пляж барьерного острова Пляж береговой линии	Приустьевые бары дельты флювиального типа Приливно-отливные отложения	Высокопесчанистые меандрирующие реки Эстуарии Дельты флювиального влияния
	Средняя	Золотые отложения Дистальная зона дельты волнового влияния	Аллювиальные конусы выноса Озерная дельта Проксимальная часть дельты волнового влияния	Дельты приливно-отливного влияния Многорусловые реки
	Высокая	Дистальная зона глубоководного конуса выноса	Грубозернистые меандрирующие реки Многорусловая дельта	Высокопесчанистые дельты речного влияния Высокопесчанистые мелкозернистые меандрирующие реки

Матрица гетерогенности архитектуры обстановок осадконакопления

Расчет рентабельности проведения исследований скважины



Для оценки выручки

PROPERTIES	Параметр	Значение
	Пористость, φ	0.7
	Площадь запасов, приведенный на одну скважину S	100 м ²
	Плотность нефти, ρ	860 кг/м ³
	Цена нефти за 1 тонну, p_{nb}	4150 руб.



Для оценки затрат

EXPENSES	Тип исследования	Стоимость, руб./м.
	Bk	2450
	GZ(1-7)	2050
	DGK	1300
	NKT(D/M/R)	2050
	ALPS	1150

$$VALUE = p_{nb} \varphi h S \rho - Expenses$$

Искомый параметр - суммарная толщина нефтенасыщенных пропластков

Номер скважины:	47	95	126	164	176
h, м	43,759	39,897	647,046	663,106	38,805
H, м	115,989	103,913	1645,368	1500,378	101,583
Value, руб	8 223 967	7 541 098	123 232 159	130 629 945	7 322 690
рентабельность	-	-	+	+	-