



Технологии создания и практического внедрения прогнозных моделей машинного обучения.

Андрейченко А.Е., PhD, к.ф.-м.н.

руководитель направления искусственного интеллекта компании "К-Скай",
г. Петрозаводск

09 февраля 2023, Москва



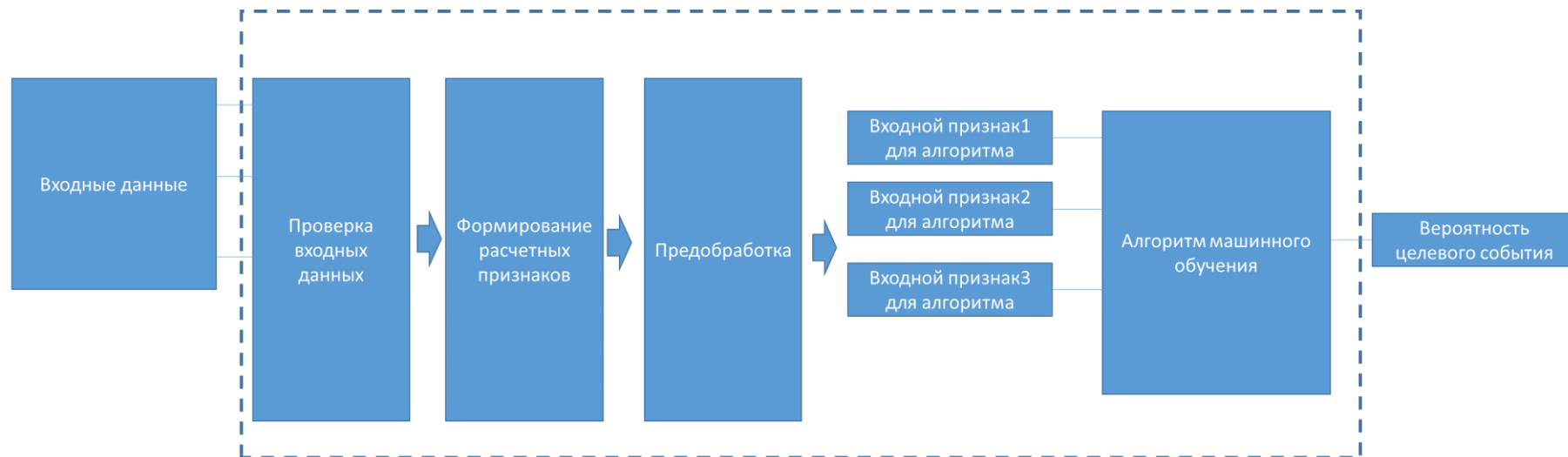
Система машинного обучения



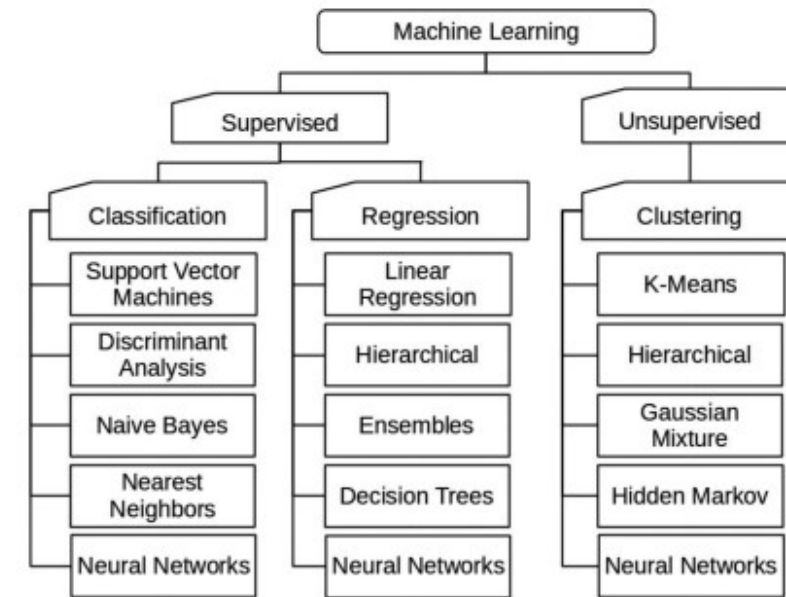
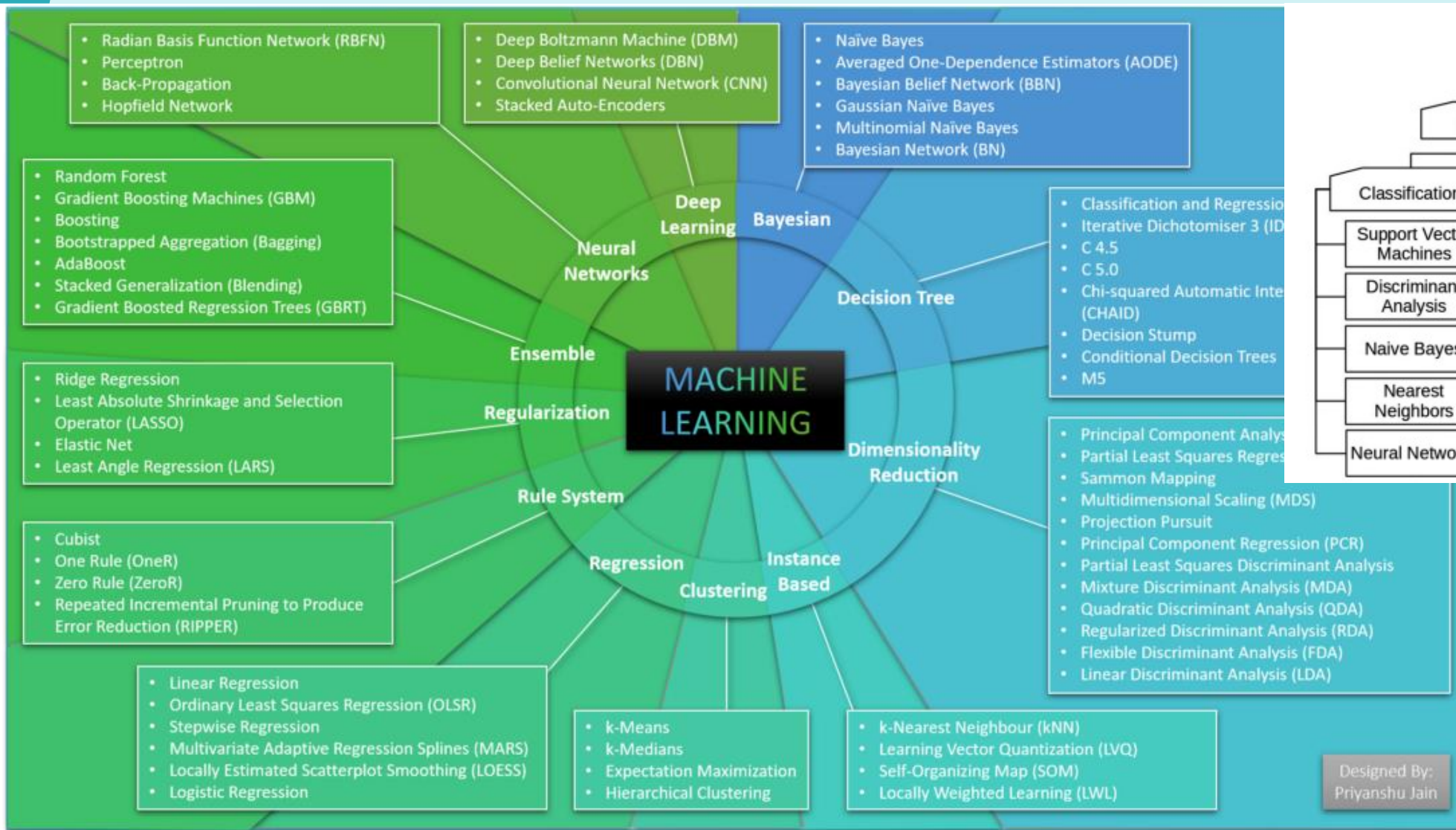
Моделирование



1. Выбор признаков
2. Обучение и тестирование алгоритмов
3. Выбор алгоритма и упаковка модели для эксплуатации



Алгоритмы машинного обучения



<https://dl.acm.org/doi/10.1145/3341145>

Designed By:
Priyanshu Jain

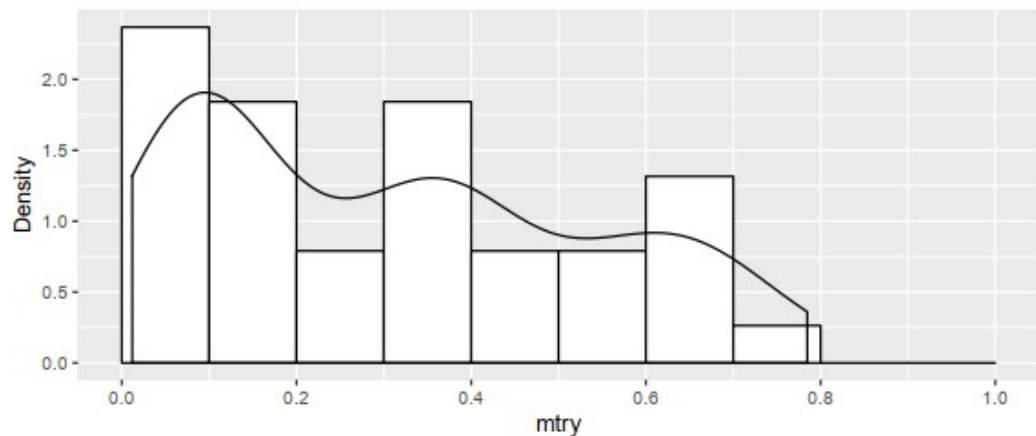
Гиперпараметры и обучение

Вопросы при разработке моделей и создании алгоритмов:

1. Какие гиперпараметры необходимо настраивать? В каких пределах?
2. Какие надежные настройки гиперпараметров по умолчанию?

Универсальный подход отсутствует, ответы зависят как от алгоритма, так и от набора данных.

Распределение оптимального значения гиперпараметра vs. набор данных (Random Forest)



Algorithm	Hyperparameter	Type	Lower	Upper	Trafo
glmnet					
(Elastic net)	alpha	numeric	0	1	-
	lambda	numeric	-10	10	2^x
rpart					
(Decision tree)	cp	numeric	0	1	-
	maxdepth	integer	1	30	-
	minbucket	integer	1	60	-
	minsplit	integer	1	60	-
kkn					
(k-nearest neighbor)	k	integer	1	30	-
svm					
(Support vector machine)	kernel	discrete	-	-	-
	cost	numeric	-10	10	2^x
	gamma	numeric	-10	10	2^x
	degree	integer	2	5	-
ranger					
(Random forest)	num.trees	integer	1	2000	-
	replace	logical	-	-	-
	sample.fraction	numeric	0.1	1	-
	mtry	numeric	0	1	$x \cdot p$
	respect.unordered.factors	logical	-	-	-
	min.node.size	numeric	0	1	n^x
xgboost					
(Gradient boosting)	nrounds	integer	1	5000	-
	eta	numeric	-10	0	2^x
	subsample	numeric	0.1	1	-
	booster	discrete	-	-	-
	max_depth	integer	1	15	-
	min_child_weight	numeric	0	7	2^x
	colsample_bytree	numeric	0	1	-
	colsample_bylevel	numeric	0	1	-
	lambda	numeric	-10	10	2^x
	alpha	numeric	-10	10	2^x

Гиперпараметры и обучение

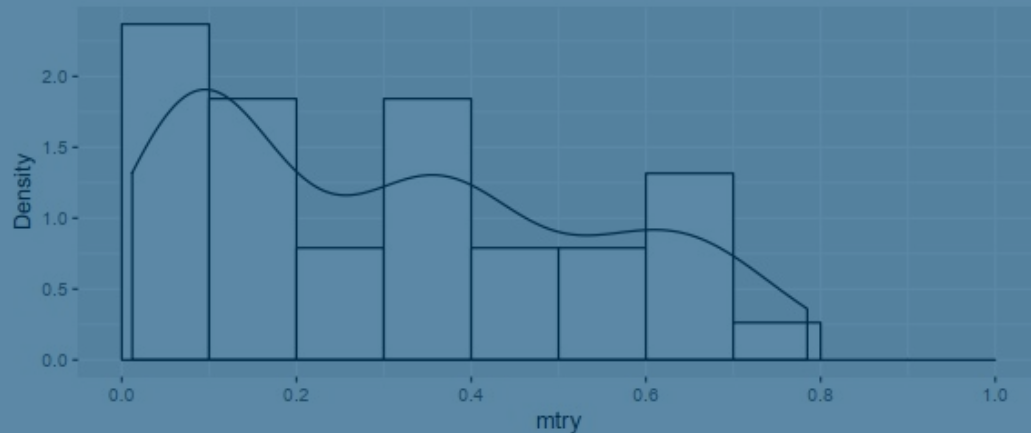
Вопросы при разработке моделей и создании алгоритмов:

1. Какие гиперпараметры необходимо настраивать? В каких пределах?
2. Какие надежные настройки гиперпараметров по умолчанию?

Универсальный алгоритм

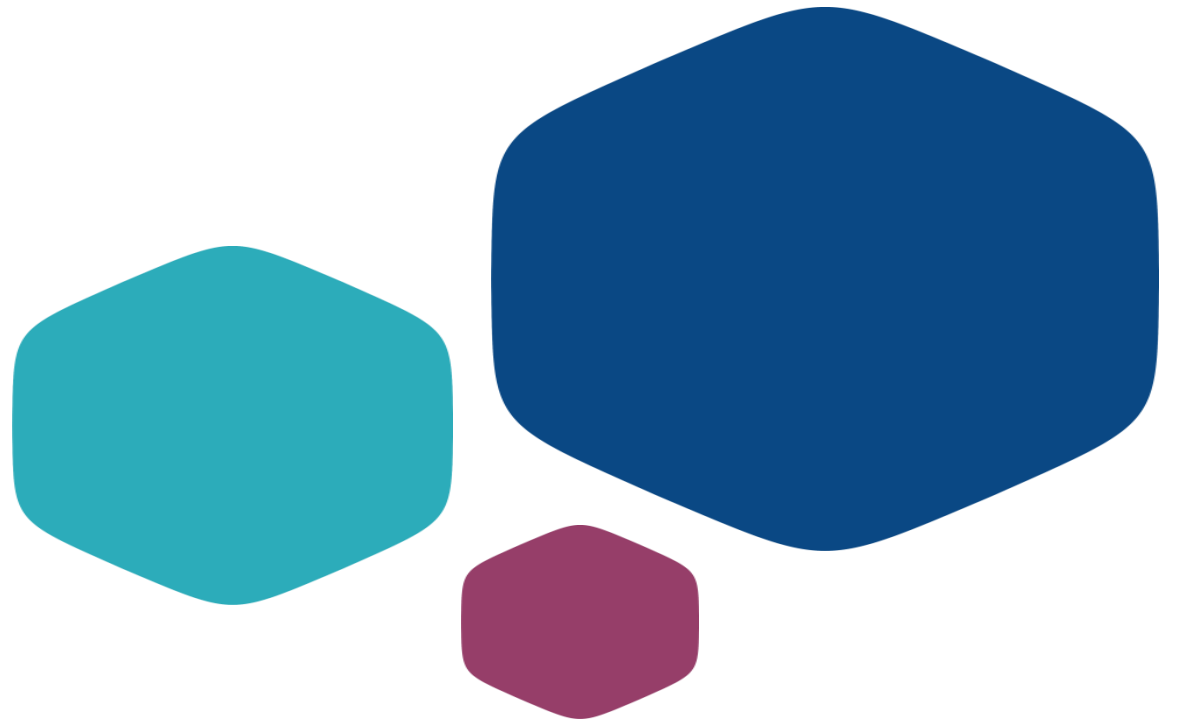
+ Предобработка входных параметров...
 + Выбор целевой функции при оптимизации ...

Распределение оптимального значения гиперпараметра vs. набор данных (Random Forest)



Algorithm	Hyperparameter	Type	Lower	Upper	Trafo
glmnet (Elastic net)	alpha	numeric	0	1	-
	lambda	numeric	-10	10	2^x
rpart (Decision tree)	cp	numeric	0	1	-
	maxdepth	integer	1	30	-
	minbucket	integer	1	60	-
	minsplit	integer	1	60	-
ranger (Random forest)	num.trees	integer	1	2000	-
	replace	logical	-	-	-
xgboost (Gradient boosting)	sample.fraction	numeric	0.1	1	-
	mtry	numeric	0	1	$x \cdot p$
	respect.unordered.factors	logical	-	-	-
	min.node.size	numeric	0	1	n^x
	nrounds	integer	1	5000	-
xgboost (Gradient boosting)	eta	numeric	-10	0	2^x
	subsample	numeric	0.1	1	-
	booster	discrete	-	-	-
	max_depth	integer	1	15	-
	min_child_weight	numeric	0	7	2^x
	colsample_bytree	numeric	0	1	-
	colsample_bylevel	numeric	0	1	-
	lambda	numeric	-10	10	2^x
	alpha	numeric	-10	10	2^x
	degree	integer	2	5	-

AutoML

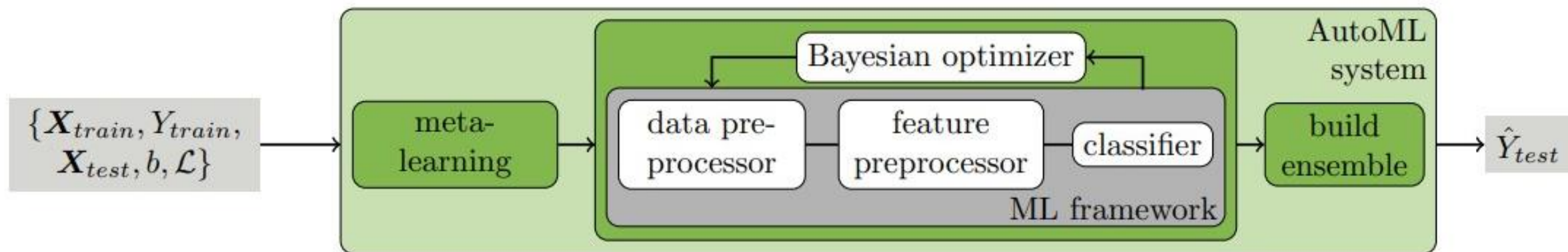


AutoML_scikit-learn:

- 15 классификаторов
- 14 методов предобработки параметров
- 4 метода обработки данных

AutoML_scikit-learn позволяет автоматизировать:

- выбор алгоритма/ансамбля алгоритмов классификации
- настройку гиперпараметров
- предобработку данных и параметров



Цель: сравнить производительность ручной разработки моделей с автоматической для дихотомных задач прогнозирования в медицине

Методы:

- 4 набора медицинских данных (всего/отрицательный класс/положительный класс);
- ручная разработка 5 алгоритмов классификации;
- применение AutoML для тех же данных;
- сравнение по основным метрикам точности (порог по методу Юдена) и длительности разработки.

Наборы данных:

1. Возникновение преэклампсии во время беременности (37255/36294/961)
2. Наличие атеросклеротической бляшки БЦА (357/200/157)
3. Выживаемость в течение года пациентов с ЗНО легкого (26165/21165/5000)
4. Выживаемость в течение года пациентов с ЗНО молочной железы (80484/78093/2391)

Результаты: набор данных №1

	Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0	Logistic Regression	0.65 [0.44 0.8]	0.74 [0.72 0.77]	0.06 [0.03 0.09]	0.99 [0.98 0.99]	0.11 [0.06 0.16]	0.85 [0.83 0.87]	0.72 [0.6 0.81]	0.74 [0.71 0.77]	0.50	[[2156, 747], [27, 50]]
1	LightGBM	0.78 [0.61 0.92]	0.76 [0.74 0.79]	0.08 [0.05 0.11]	0.99 [0.99 1.]	0.15 [0.09 0.2]	0.86 [0.85 0.88]	0.81 [0.71 0.89]	0.76 [0.74 0.79]	0.50	[[2218, 685], [17, 60]]
2	XGBoost	0.77 [0.6 0.94]	0.78 [0.75 0.8]	0.08 [0.05 0.12]	0.99 [0.99 1.]	0.15 [0.09 0.21]	0.87 [0.86 0.89]	0.79 [0.68 0.88]	0.78 [0.75 0.8]	0.46	[[2262, 641], [18, 59]]
3	Random Forest	0.79 [0.62 0.95]	0.7 [0.67 0.73]	0.06 [0.04 0.1]	0.99 [0.99 1.]	0.12 [0.08 0.17]	0.82 [0.8 0.84]	0.79 [0.7 0.87]	0.7 [0.67 0.73]	0.44	[[2024, 879], [16, 61]]
4	Multi Layer Perceptron	0.81 [0.62 0.95]	0.63 [0.6 0.66]	0.05 [0.03 0.08]	0.99 [0.98 1.]	0.1 [0.06 0.14]	0.77 [0.75 0.79]	0.74 [0.63 0.83]	0.63 [0.6 0.66]	0.44	[[1824, 1079], [15, 62]]
	Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0	AutoML	0.17 [0.04 0.32]	0.99 [0.99 1.]	0.34 [0.07 0.62]	0.98 [0.97 0.99]	0.23 [0.05 0.4]	0.98 [0.98 0.99]	0.79 [0.7 0.87]	0.97 [0.96 0.98]	0.41	[[2878, 25], [64, 13]]

Результаты: набор данных №1

	Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0	Logistic Regression	0.65 [0.44 0.8]	0.74 [0.72 0.77]	0.06 [0.03 0.09]	0.99 [0.98 0.99]	0.11 [0.06 0.16]	0.85 [0.83 0.87]	0.72 [0.6 0.81]	0.74 [0.71 0.77]	0.50	[[2156, 747], [27, 50]]
1	LightGBM	0.78 [0.61 0.95]	0.74 [0.72 0.77]	0.06 [0.03 0.09]	0.99 [0.98 0.99]	0.11 [0.06 0.16]	0.86 [0.85 0.88]	0.81 [0.71 0.89]	0.76 [0.74 0.79]	0.50	[[2218, 685], [17, 60]]
2	XGBoost	0.77 [0.6 0.94]	0.74 [0.72 0.77]	0.06 [0.03 0.09]	0.99 [0.98 0.99]	0.11 [0.06 0.16]	0.87 [0.86 0.89]	0.79 [0.68 0.88]	0.78 [0.75 0.8]	0.46	[[2262, 641], [18, 59]]
3	Random Forest	0.79 [0.62 0.96]	0.74 [0.72 0.77]	0.06 [0.03 0.09]	0.99 [0.98 0.99]	0.11 [0.06 0.16]	0.82 [0.8 0.84]	0.79 [0.7 0.87]	0.7 [0.67 0.73]	0.44	[[2024, 879], [16, 61]]
4	Multi Layer Perceptron	0.81 [0.62 0.95]	0.63 [0.6 0.66]	0.05 [0.03 0.08]	0.99 [0.98 1.]	0.1 [0.06 0.14]	0.77 [0.75 0.79]	0.74 [0.63 0.83]	0.63 [0.6 0.66]	0.44	[[1824, 1079], [15, 62]]
0	AutoML	0.17 [0.04 0.3]	0.74 [0.72 0.77]	0.06 [0.03 0.09]	0.99 [0.98 0.99]	0.11 [0.06 0.16]	0.98 [0.98 0.99]	0.79 [0.7 0.87]	0.97 [0.96 0.98]	0.41	[[2878, 25], [64, 13]]

~ 1 неделя

~ 10 минут

ROC-AUC, Точность

Результаты: набор данных №2

14

	Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0	Logistic Regression	0.75 [0.38 1.]	0.68 [0.61 0.75]	0.09 [0.03 0.17]	0.98 [0.96 1.]	0.16 [0.06 0.28]	0.8 [0.75 0.85]	0.67 [0.49 0.84]	0.68 [0.62 0.74]	0.04	[[127, 60], [2, 6]]
1	LightGBM	0.62 [0.25 1.]	0.72 [0.65 0.78]	0.09 [0.02 0.16]	0.98 [0.95 1.]	0.15 [0.04 0.26]	0.83 [0.78 0.87]	0.64 [0.37 0.86]	0.71 [0.65 0.78]	0.04	[[134, 53], [3, 5]]
2	XGBoost	0.88 [0.6 1.]	0.52 [0.46 0.59]	0.07 [0.03 0.13]	0.99 [0.97 1.]	0.13 [0.05 0.24]	0.69 [0.62 0.74]	0.65 [0.44 0.81]	0.54 [0.48 0.6]	0.00	[[98, 89], [1, 7]]
3	Random Forest	0.62 [0.25 1.]	0.93 [0.89 0.96]	0.26 [0.08 0.5]	0.98 [0.96 1.]	0.37 [0.12 0.59]	0.95 [0.93 0.97]	0.78 [0.6 0.95]	0.91 [0.88 0.95]	0.05	[[173, 14], [3, 5]]
4	Multi Layer Perceptron	0.75 [0.33 1.]	0.76 [0.69 0.82]	0.12 [0.03 0.2]	0.99 [0.97 1.]	0.2 [0.05 0.33]	0.86 [0.81 0.9]	0.67 [0.4 0.88]	0.76 [0.69 0.82]	0.05	[[142, 45], [2, 6]]
5	AutoML	0.88 [0.5 1.]	0.59 [0.52 0.66]	0.08 [0.03 0.15]	0.99 [0.97 1.]	0.15 [0.05 0.25]	0.74 [0.69 0.8]	0.67 [0.43 0.82]	0.61 [0.54 0.67]	0.30	[[111, 76], [1, 7]]

Результаты: набор данных №2

15

Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок	
0 Logistic Regression	0.75 [0.38 1.]	0.68 [0.61 0.75]	0.09 [0.03 0.17]	0.98 [0.96 1.]	0.16 [0.06 0.28]	0.8 [0.75 0.85]	0.67 [0.49 0.84]	0.68 [0.62 0.74]	0.04	[[127, 60], [2, 6]]	
1 LightGBM	0.62 [0.25 1.]	~ 1 неделя				0.4 [0.25 0.26]	0.83 [0.78 0.87]	0.64 [0.37 0.86]	0.71 [0.65 0.78]	0.04	[[134, 53], [3, 5]]
2 XGBoost	0.88 [0.6 1.]					0.65 [0.44 0.81]	0.5 [0.24 0.24]	0.69 [0.62 0.74]	0.54 [0.48 0.6]	0.00	[[98, 89], [1, 7]]
3 Random Forest	0.62 [0.25 1.]	0.93 [0.89 0.96]	0.26 [0.08 0.5]	0.98 [0.96 1.]	0.37 [0.12 0.59]	0.95 [0.93 0.97]	0.78 [0.6 0.95]	0.91 [0.88 0.95]	0.05	[[173, 14], [3, 5]]	
4 Multi Layer Perceptron	0.75 [0.33 1.]	0.76 [0.69 0.82]	0.12 [0.03 0.2]	0.99 [0.97 1.]	0.2 [0.05 0.33]	0.86 [0.81 0.9]	0.67 [0.4 0.88]	0.76 [0.69 0.82]	0.05	[[142, 45], [2, 6]]	
5 AutoML	0.88 [0.5 1.]	~ 1 минута				0.5 [0.25 0.25]	0.74 [0.69 0.8]	0.67 [0.43 0.82]	0.61 [0.54 0.67]	0.30	[[111, 76], [1, 7]]

Результаты: набор данных №3

16

	Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0	Logistic Regression	0.76 [0.7 0.82]	0.55 [0.51 0.59]	0.29 [0.25 0.33]	0.91 [0.88 0.93]	0.42 [0.37 0.46]	0.68 [0.65 0.71]	0.72 [0.68 0.76]	0.59 [0.56 0.62]	0.46	[[9280, 7573], [969, 3094]]
1	Multi Layer Perceptron	1.0 [1. 1.]	0.0 [0. 0.]	0.19 [0.17 0.22]	nan [nan nan]	0.33 [0.29 0.36]	nan [nan nan]	0.63 [0.59 0.67]	0.19 [0.17 0.22]	0.00	[[0, 16853], [0, 4063]]
2	Random Forest	0.66 [0.59 0.73]	0.79 [0.77 0.82]	0.43 [0.37 0.49]	0.91 [0.88 0.93]	0.52 [0.47 0.57]	0.85 [0.83 0.87]	0.81 [0.78 0.85]	0.77 [0.74 0.8]	0.80	[[13380, 3473], [1391, 2672]]
3	Extra Trees classifier	0.73 [0.68 0.8]	0.66 [0.62 0.69]	0.34 [0.3 0.39]	0.91 [0.89 0.93]	0.46 [0.42 0.51]	0.76 [0.74 0.79]	0.77 [0.73 0.8]	0.67 [0.65 0.7]	0.49	[[11068, 5785], [1086, 2977]]
4	XGBoost	0.73 [0.67 0.79]	0.8 [0.77 0.83]	0.47 [0.41 0.52]	0.92 [0.91 0.94]	0.57 [0.51 0.62]	0.86 [0.84 0.88]	0.86 [0.82 0.88]	0.79 [0.76 0.81]	0.22	[[13488, 3365], [1107, 2956]]
5	AutoML	0.97 [0.93 0.99]	0.87 [0.85 0.9]	0.65 [0.59 0.7]	0.99 [0.98 1.]	0.78 [0.73 0.81]	0.93 [0.91 0.94]	0.98 [0.97 0.99]	0.89 [0.87 0.91]	0.32	[[14728, 2125], [138, 3925]]

Результаты: набор данных №3

Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок	
0 Logistic Regression	0.76 [0.7 0.82]	0.55 [0.51 0.59]	0.29 [0.25 0.33]	0.91 [0.88 0.93]	0.42 [0.37 0.46]	0.68 [0.65 0.71]	0.72 [0.68 0.76]	0.59 [0.56 0.62]	0.46	[[9280, 7573], [969, 3094]]	
1 Multi Layer Perceptron	1.0 [1. 1.]	~ 1 неделя				0.36	nan [nan nan]	0.63 [0.59 0.67]	0.19 [0.17 0.22]	0.00	[[0, 16853], [0, 4063]]
2 Random Forest	0.66 [0.59 0.73]					0.77 [0.74 0.8]	0.85 [0.83 0.87]	0.80	[[13380, 3473], [1391, 2672]]		
3 Extra Trees classifier	0.73 [0.68 0.8]	0.66 [0.62 0.69]	0.34 [0.3 0.39]	0.91 [0.89 0.93]	0.46 [0.42 0.51]	0.76 [0.74 0.79]	0.77 [0.73 0.8]	0.67 [0.65 0.7]	0.49	[[11068, 5785], [1086, 2977]]	
4 XGBoost	0.73 [0.67 0.79]	0.8 [0.77 0.83]	0.47 [0.41 0.52]	0.92 [0.91 0.94]	0.57 [0.51 0.62]	0.86 [0.84 0.88]	0.86 [0.82 0.88]	0.79 [0.76 0.81]	0.22	[[13488, 3365], [1107, 2956]]	
5 AutoML	0.97 [0.93 0.99]	~ 10 минут				0.81	0.93 [0.91 0.94]	0.98 [0.97 0.99]	0.89 [0.87 0.91]	0.32	[[14728, 2125], [138, 3925]]

Результаты: набор данных №4

18

Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0 Logistic Regression	0.0 [0. 0.]	1.0 [1. 1.]	0.0 [0. 0.]	0.97 [0.96 0.98]	0.0 [0. 0.]	0.99 [0.98 0.99]	0.72 [0.62 0.82]	0.97 [0.96 0.98]	0.97	[[15850, 20], [439, 0]]
1 Multi Layer Perceptron	0.0 [0. 0.]	1.0 [1. 1.]	0.0 [0. 0.]	0.97 [0.96 0.98]	0.0 [0. 0.]	0.99 [0.98 0.99]	0.5 [0.5 0.5]	0.97 [0.96 0.98]	0.50	[[15870, 0], [439, 0]]
2 Random Forest	0.13 [0.03 0.29]	0.98 [0.98 0.99]	0.18 [0.05 0.36]	0.98 [0.97 0.98]	0.15 [0.04 0.31]	0.98 [0.97 0.99]	0.74 [0.66 0.83]	0.96 [0.95 0.97]	0.63	[[15608, 262], [382, 57]]
3 Extra Trees classifier	0.12 [0. 0.26]	0.97 [0.96 0.98]	0.11 [0. 0.26]	0.98 [0.97 0.98]	0.11 [0. 0.25]	0.97 [0.97 0.98]	0.7 [0.61 0.78]	0.95 [0.94 0.96]	0.68	[[15457, 413], [388, 51]]
4 XGBoost	0.31 [0.12 0.52]	0.93 [0.92 0.95]	0.12 [0.04 0.21]	0.98 [0.97 0.99]	0.17 [0.06 0.28]	0.96 [0.95 0.97]	0.78 [0.68 0.86]	0.92 [0.9 0.94]	0.67	[[14833, 1037], [303, 136]]
5 K-nearest neighbors	0.09 [0. 0.21]	0.95 [0.94 0.96]	0.05 [0. 0.11]	0.97 [0.96 0.98]	0.06 [0. 0.15]	0.96 [0.95 0.97]	0.52 [0.47 0.58]	0.93 [0.91 0.94]	0.10	[[15110, 760], [400, 39]]
6 AutoML	0.0 [0. 0.03]	1.0 [1. 1.]	0.07 [0. 0.33]	0.97 [0.96 0.98]	0.0 [0. 0.06]	0.99 [0.98 0.99]	0.71 [0.63 0.81]	0.97 [0.96 0.98]	0.51	[[15857, 13], [438, 1]]

Результаты: набор данных №4

Модель	Чувствительность (Recall)	Специфичность	Прогностическая ценность положительного класса (Precision)	Прогностическая ценность отрицательного класса	F1 (положительный)	F1 (отрицательный)	ROC-AUC	Точность	Порог классификации	Матрица ошибок
0 Logistic Regression	0.0 [0. 0.]	1.0 [1. 1.]	0.0 [0. 0.]	0.97 [0.96 0.98]	0.0 [0. 0.]	0.99 [0.98 0.99]	0.72 [0.62 0.82]	0.97 [0.96 0.98]	0.97	[[15850, 20], [439, 0]]
1 Multi Layer Perceptron	0.0 [0. 0.]	1.0 [1. 1.]	0.0 [0. 0.]	0.97 [0.96 0.98]	0.0 [0. 0.]	0.99 [0.98 0.99]	0.5 [0.5 0.5]	0.97 [0.96 0.98]	0.50	[[15870, 0], [439, 0]]
2 Random Forest	0.13 [0.03 0.25]	0.93 [0.92 0.95]	0.12 [0.04 0.21]	0.98 [0.97 0.99]	0.04 [0.01 0.31]	0.98 [0.97 0.99]	0.74 [0.66 0.83]	0.96 [0.95 0.97]	0.63	[[15608, 262], [382, 57]]
3 Extra Trees classifier	0.12 [0.02 0.25]	0.93 [0.92 0.95]	0.12 [0.04 0.21]	0.98 [0.97 0.99]	0.04 [0.01 0.25]	0.97 [0.97 0.98]	0.7 [0.61 0.78]	0.95 [0.94 0.96]	0.68	[[15457, 413], [388, 51]]
4 XGBoost	0.31 [0.12 0.52]	0.93 [0.92 0.95]	0.12 [0.04 0.21]	0.98 [0.97 0.99]	0.17 [0.06 0.28]	0.96 [0.95 0.97]	0.78 [0.68 0.86]	0.92 [0.9 0.94]	0.67	[[14833, 1037], [303, 136]]
5 K-nearest neighbors	0.09 [0. 0.21]	0.95 [0.94 0.96]	0.05 [0. 0.11]	0.97 [0.96 0.98]	0.06 [0. 0.15]	0.96 [0.95 0.97]	0.52 [0.47 0.58]	0.93 [0.91 0.94]	0.10	[[15110, 760], [400, 39]]
6 AutoML	0.0 [0. 0.]	1.0 [1. 1.]	0.0 [0. 0.]	0.97 [0.96 0.98]	0.0 [0. 0.06]	0.99 [0.98 0.99]	0.71 [0.63 0.81]	0.97 [0.96 0.98]	0.51	[[15857, 13], [438, 1]]

~ 1 неделя

~ 10 минут

Преимущества:

- Высокая скорость и эффективность подбора гиперпараметров при обучении моделей
- Прозрачность и отсутствие человеческого фактора при создании моделей
- Доступность разработки моделей для специалистов с небольшим опытом в машинном обучении и DataScience

Ограничения:

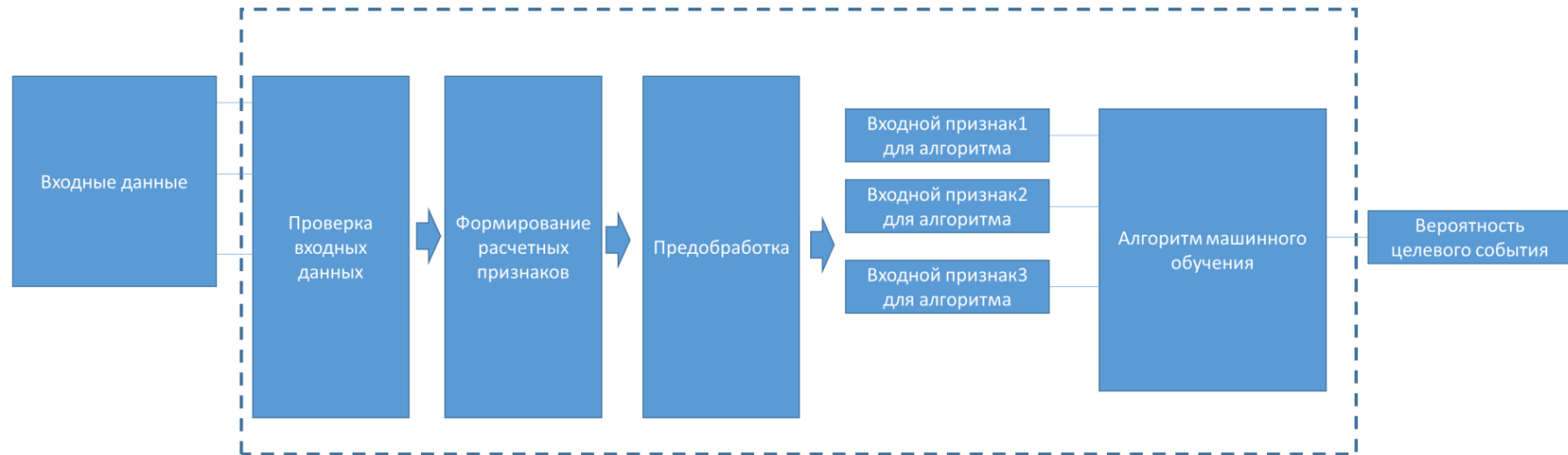
- Невозможно оперативное использование передовых алгоритмов
- Относительная сложность итоговых ансамблей алгоритмов

Выводы:

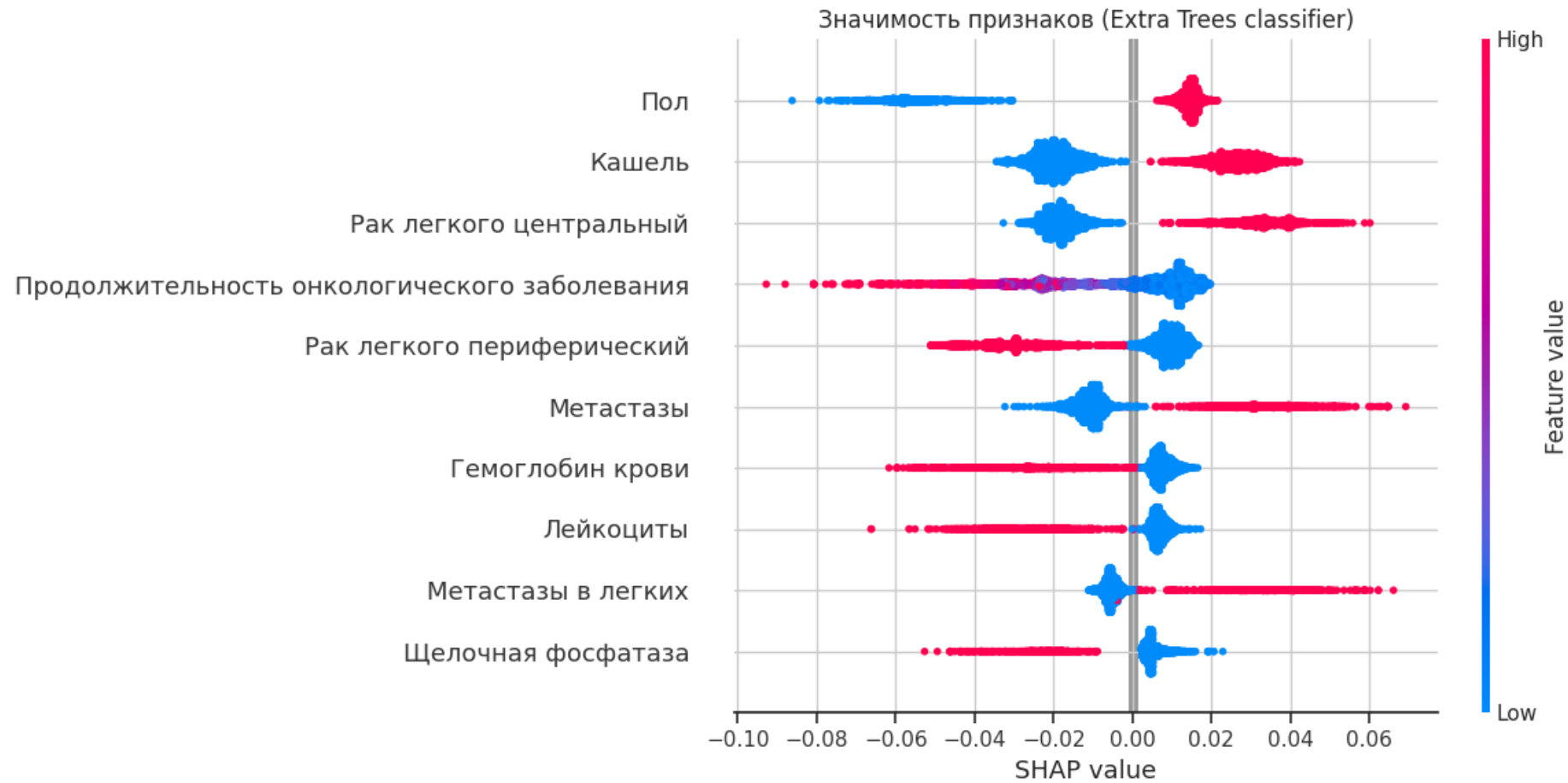
- AutoML демонстрирует сравнимые метрики по сравнению с ручной разработкой
- Возможность разрабатывать модели с допустимыми метриками точности при минимальных навыках в машинном обучении

Входные данные





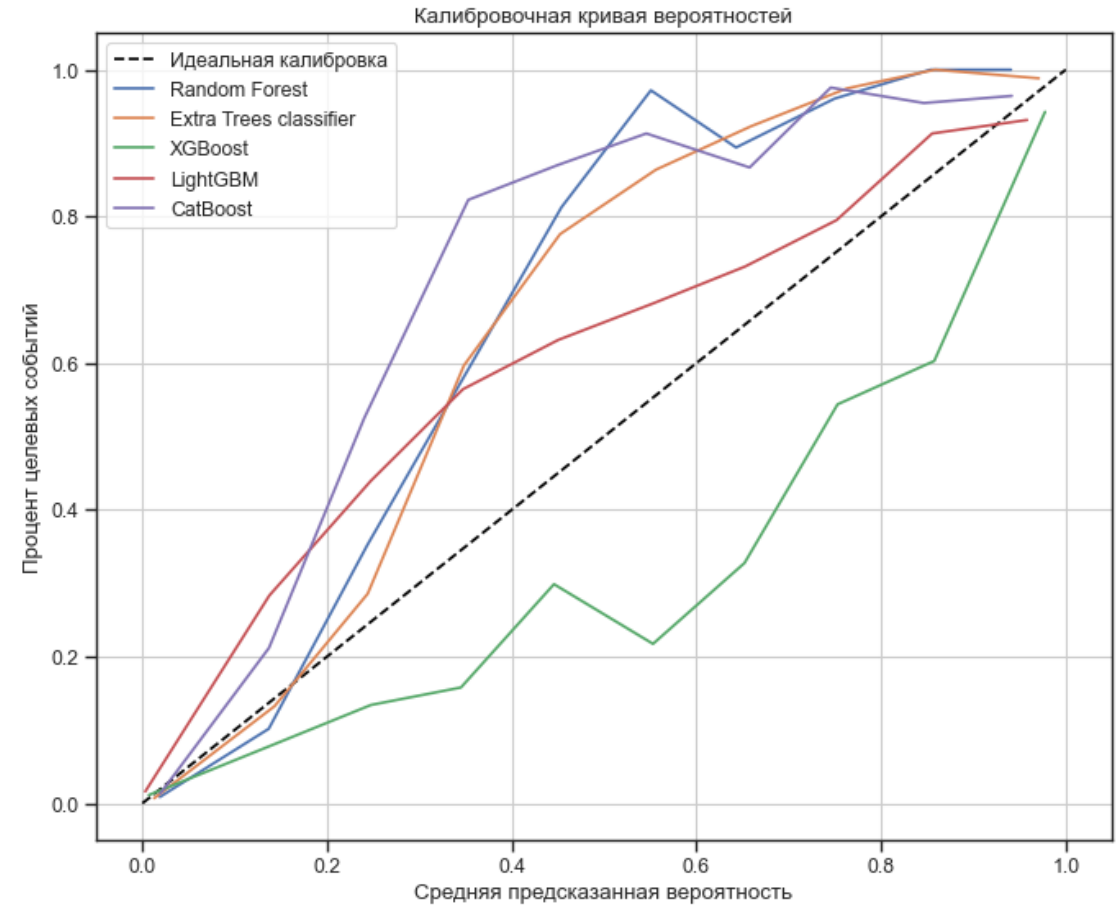
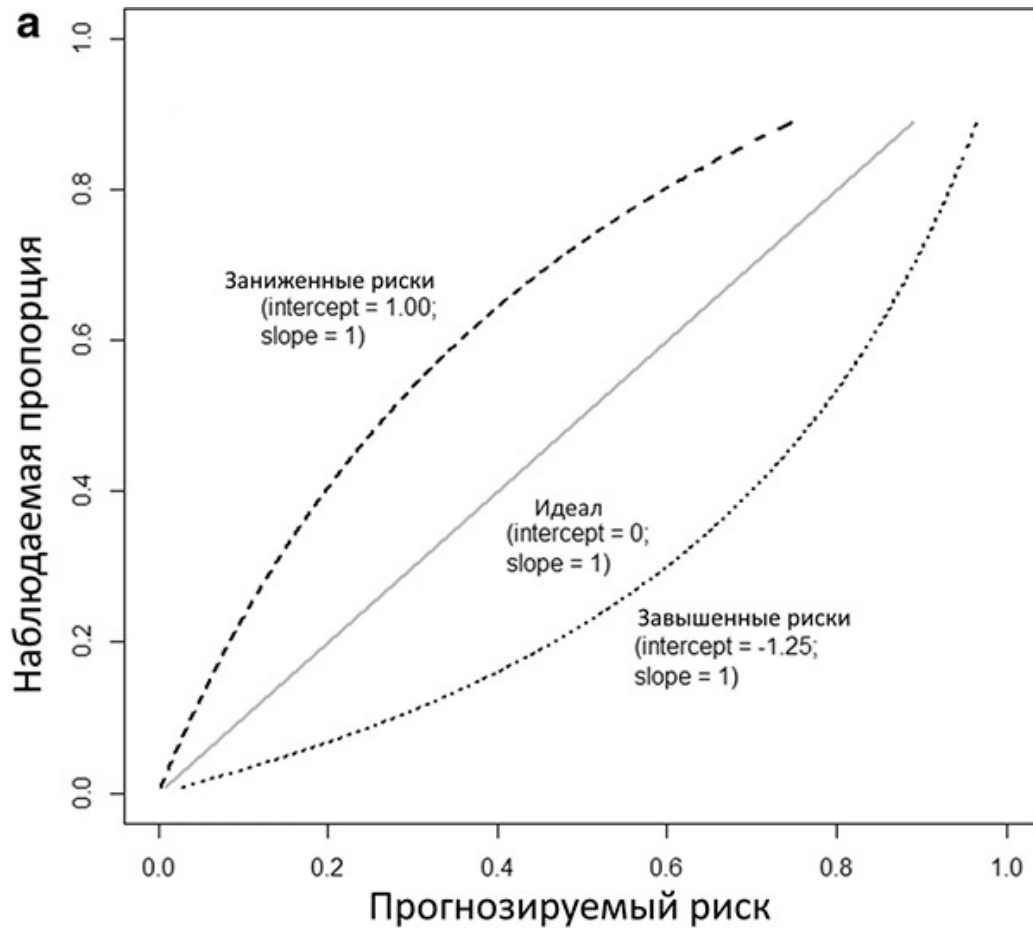
Входные данные \neq входные признаки для алгоритма



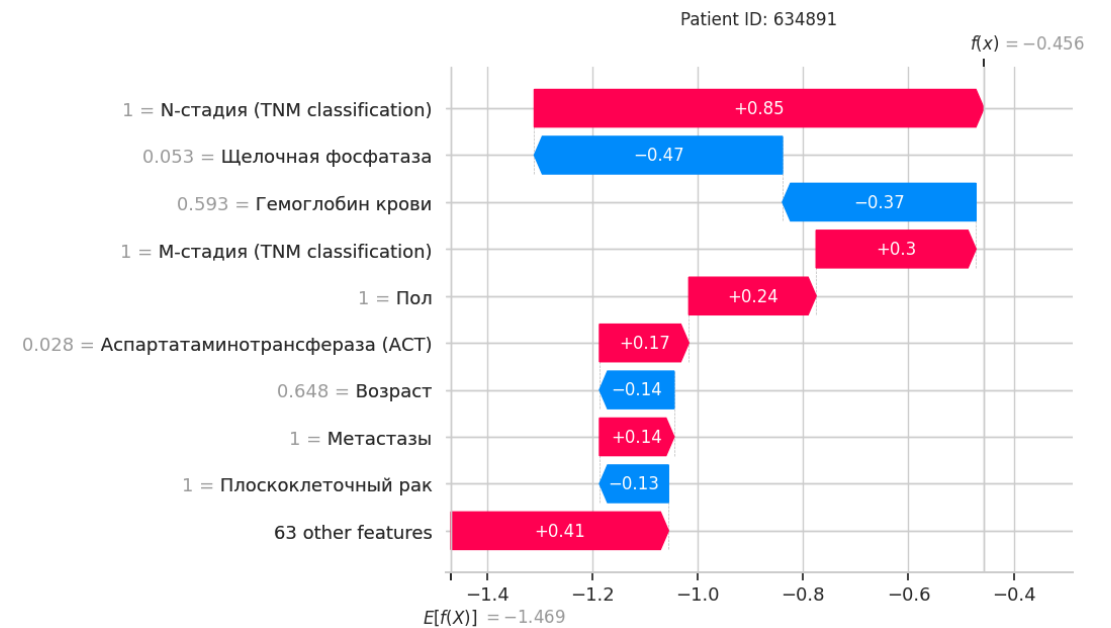
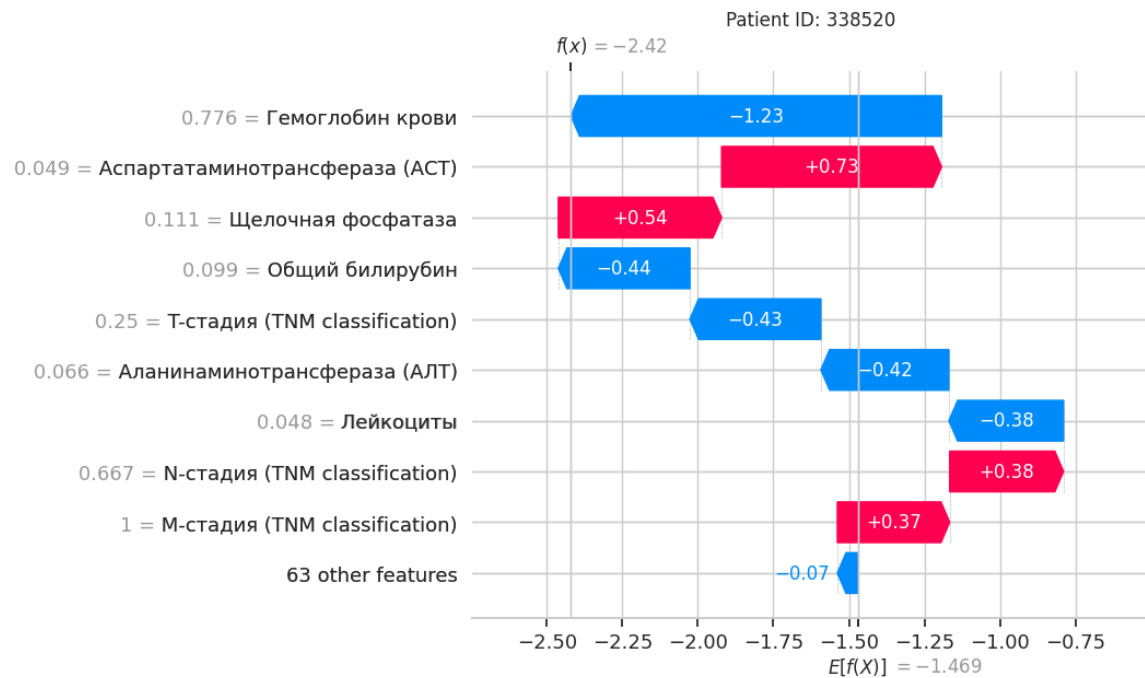
Методика векторов Шепли для выбора обязательных признаков модели



Калибровка и интерпретация результатов



Методика векторов Шепли для визуализации вклада в прогноза каждого признака



- Практическое внедрение моделей машинного обучения требует наличия дополнительных технологических блоков, обеспечивающих ценность и надежность работы систем машинного обучения.
- Автоматизированное моделирование сравнимо с ручным и должно применяться для обеспечения экономической целесообразности создания моделей в условиях необходимости итеративного уточнения клинико-практических задач.
- Необходимо при внедрении уделять внимание интерпретируемости результатов моделей конечными пользователями.

Благодарим за внимание!

Вопросы?

Контакты:

Андрейченко Анна

email: aandreychenko@webiomed.ru

telegram: [@anna_medAI](https://www.t.me/@anna_medAI)



К-ЛАБ