



# Олимпиадный тест

Loginom Хакатон 2019

## Раздел 1. Мировой рынок аналитических платформ

<b>№ вопроса</b>	<b>1.1</b>
Тип	Множественный выбор
Условие	В лидерах магического квадрата от Garthner для Data Science and Machine Learning Platforms (2019) значатся:
Варианты	KNIME
	RapidMiner
	SAP
	Microsoft

<b>№ вопроса</b>	<b>1.2</b>
Тип	Одиночный выбор
Условие	Оценки (исследование компаний <b>TAdviser</b> и <b>Инфосистемы Джет</b> ) говорят об объеме российского рынка ИИ по итогам за 2017 год в размере...
Варианты	700 млн. руб.
	28 млрд. руб.
	50 млрд. руб.
	4 млрд. руб.

<b>№ вопроса</b>	<b>1.3</b>
Тип	Одиночный выбор
Условие	Сколько игроков на рынке, представленных в квадрате <b>Data Science and Machine Learning Platforms</b> от Gartner?
Варианты	до 10
	до 20
	около 30

<b>№ вопроса</b>	<b>1.4</b>
Тип	Одиночный выбор
Условие	Согласно оценкам J.P. Morgan и McKinsey, мировой рынок AI (искусственный интеллект) и ML (машинное обучение) в 2018 году составил величину...
Варианты	30 млрд. \$
	300 млрд. \$
	90 млрд. \$
	750 млрд. \$

## Раздел 2. Терминология науки о данных и смежных областей

<b>№ вопроса</b>	<b>2.1</b>
Тип	Одиночный выбор
Условие	Выберите из списка английский термин, означающий <i>сплошное исследование имеющегося массива данных на наличие статистически значимых зависимостей без какого-либо предварительного построения гипотез, что еще называют "слепой анализ данных"</i> .
Варианты	data bagging
	data booting
	data merging
	data mining
	data dredging

<b>№ вопроса</b>	<b>2.2</b>
Тип	Одиночный выбор
Условие	Какая из следующих характеристик Big Data ("4V-модель") больше подходит для Data Science?
Варианты	Объем (Volume)
	Разнообразие (Variety)
	Изменчивость (Variability)
	Скорость (Velocity)

<b>№ вопроса</b>	<b>2.3</b>
Тип	Одиночный выбор
Условие	Какой термин из следующего списка отвечает за обнаружение в данных ранее неизвестных свойств объектов, процессов, явлений?
Варианты	Машинное обучение (Machine Learning)
	Data Mining
	Большие данные (Big Data)
	Разведочный анализ данных (Exploratory Data Analysis)
	Обработка данных (Data Wrangling)

<b>№ вопроса</b>	<b>2.4</b>
Тип	Одиночный выбор
Условие	Какой тип моделей являются "золотым стандартом" в анализе данных?
Варианты	логические
	описательные
	каузальные
	когнитивные
	нет верного ответа

<b>№ вопроса</b>	<b>2.5</b>
Тип	Одиночный выбор
Условие	Определение "Инструменты и приложения для поиска, анализа, моделирования и доставки информации, необходимой для принятия решений" подходят для термина:
Варианты	машинное обучение
	Data Science
	бизнес-аналитика
	Data Mining

<b>№ вопроса</b>	<b>2.6</b>
Тип	Одиночный выбор
Условие	Что является самым важным в Data Science?
Варианты	вопрос, гипотеза
	ответ, результат
	данные

### Раздел 3. Алгоритмы машинного обучения и Data Mining

<b>№ вопроса</b>	<b>3.1</b>
Тип	Поле ввода
Условие	Как называется алгоритм машинного обучения, разработанный и представленный мировому сообществу в 2017 году компанией Яндекс, и который активно используется ей для улучшения результатов поисковой системы?

<b>№ вопроса</b>	<b>3.2</b>
Тип	Одиночный выбор
Условие	Какая из следующих характеристик определяет лучший метод машинного обучения?
Варианты	скорость
	точность
	масштабируемость
	все перечисленные

<b>№ вопроса</b>	<b>3.3</b>
Тип	Одиночный выбор
Условие	<i>Коллаборативные рекомендательные системы</i> в качестве алгоритмической основы чаще всего используют:
Варианты	методы матричной факторизации
	ассоциативные правила
	деревья решающих правил
	сверточные нейронные сети

<b>№ вопроса</b>	<b>3.4</b>
Тип	Одиночный выбор
Условие	Укажите неверное утверждение.
Варианты	Кластеризация $k$ -средних это, по сути, метод векторного квантования
	Кластеризация $k$ -средних имеет целью разделить $n$ наблюдений на $k$ кластеров
	Метод "к-ближайших соседей" - аналог метода $k$ -средних
	Все перечисленные верны

<b>№ вопроса</b>	<b>3.5</b>
Тип	Одиночный выбор
Условие	Что из перечисленного не является алгоритмом машинного обучения?
Варианты	SVG
	SVM
	Random Forest
	ни один из упомянутых

<b>№ вопроса</b>	<b>3.6</b>
Тип	Одиночный выбор
Условие	Что из следующего является правильным утверждением для модели случайного леса (Random Forest)?
Варианты	Случайный лес трудно интерпретировать, но часто модель показывает высокую точность
	Случайный лес легко интерпретировать, и часто модель показывает высокую точность
	Случайный лес трудно интерпретировать, и часто модель показывает невысокую точность
	Ни одно из приведенных

## Раздел 4. Технология Big Data, Deep Learning

<b>№ вопроса</b>	<b>4.1</b>
Тип	Одиночный выбор
Условие	Hadoop можно использовать для создания распределенных кластеров на основе обычных серверов, которые обеспечивают недорогую обработку и хранение неструктурированных данных, файлов журналов и других форм больших данных.
Варианты	верно
	неверно

<b>№ вопроса</b>	<b>4.2</b>
Тип	Одиночный выбор
Условие	Какой инструмент предназначен для передачи данных между Hadoop и реляционными базами данных?
Варианты	RDBMS
	Hive
	Groovy
	SQL
	Sqoop

<b>№ вопроса</b>	<b>4.3</b>
Тип	Одиночный выбор
Условие	По мнению аналитиков, на чем могут основываться традиционные ИТ-системы, когда они интегрированы с такими технологиями больших данных, как Hadoop?
Варианты	Управление большими данными
	Хранилища данных и бизнес-аналитика
	Управление кластерами Hadoop
	Сбор и хранение неструктурированных данных

<b>№ вопроса</b>	<b>4.4</b>
Тип	Одиночный выбор
Условие	Что такое HDFS?
Варианты	Инструмент инфраструктуры хранилища данных для обработки структурированных данных в Hadoop
	Инструмент для импорта таблиц и их экспорта из файловой системы Hadoop
	Колоночная СУБД, созданная для работы поверх файловой системы Hadoop
	Основная система хранения, используемая приложениями Hadoop

## Раздел 5. Предметная бизнес-аналитика

<b>№ вопроса</b>	<b>5.1</b>
Тип	Одиночный выбор
Условие	В какой предметной области нашел применение анализ выживаемости (оценки Каплана-Майера, регрессия Кокса и т.п.)?
Варианты	Клиентская аналитика
	Кредитный скоринг
	Управление запасами
	Рекомендательные системы
	Анализ текстов

<b>№ вопроса</b>	<b>5.2</b>
Тип	Одиночный выбор
Условие	Где показали эффективность и широко применяются сверточные нейронные сети?
Варианты	Кредитный скоринг
	Клиентская аналитика
	Распознавание изображений
	Рекомендательные системы
	Интернет-маркетинг

<b>№ вопроса</b>	<b>5.3</b>
Тип	Одиночный выбор
Условие	Из какой предметной области данные термины? IV, Gini индекс, винтаж, KS-статистика, шансы
Варианты	Клиентская аналитика
	Кредитный скоринг
	Управление запасами
	Рекомендательные системы
	Распознавание изображений

<b>№ вопроса</b>	<b>5.4</b>
Тип	Одиночный выбор
Условие	Из какой предметной области данные термины? LTV, Redumption Rate, CTR, Net WoE
Варианты	Клиентская аналитика
	Кредитный скоринг
	Управление запасами
	Рекомендательные системы

<b>№ вопроса</b>	<b>5.5</b>
Тип	Множественный выбор
Условие	Какие метрики используются для оценки качества рекомендательных систем?
Варианты	Полнота
	Точность
	Индекс AUC
	Ошибка RSME
	KS-статистика
	F-мера



<b>№ вопроса</b>	<b>5.6</b>
Тип	Одиночный выбор
Условие	На чем основан современный статистический кредитный скоринг?
Варианты	WoE- и IV-анализ
	РСА - метод главных компонент
	Стресс-тестирование
	Градиентный бустинг