МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Направление и направленность (профиль)
41.04.05 Международные отношения. Международные отношения и приграничное сотрудничество

Год набора на ОПОП
2026

Форма обучения
очно-заочная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Технологии искусственного интеллекта» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 41.04.05 Международные отношения (утв. приказом Минобрнауки России от 12.07.2017г. №649) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Вишневский А.А.

Кригер А.Б.

Утверждена на заседании научно-образовательный центр "искусственный интеллект" от 27.05.2026 , протокол № 5

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кригер А.Б.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат	1582918206
Номер транзакции	0000000000F8108D
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Цель освоения дисциплины заключается в формировании у обучающихся компетенций в области компьютерных технологий и анализа данных, необходимых для решения научных и практических задач, включая сбор, обработку, визуализацию и интерпретацию данных с использованием современных инструментов и методов. Основные задачи освоения дисциплины:

1. Изучить основы компьютерных технологий как инструмента для хранения, обработки и представления данных.
2. Освоить методы анализа данных, включая работу с большими данными, системами бизнес-аналитики и направлениями Data Science.
3. Научиться применять системы хранения и визуализации данных, включая российские решения.
4. Развить навыки подготовки данных для анализа: работа с табличными процессорами (MS Excel, Open Office Calc), форматами данных (xlsx, csv, txt), выявление и исправление ошибок.
5. Освоить методы численного и нечисленного анализа данных, включая визуализацию показателей и зависимостей.
6. Изучить основы корреляционно-регрессионного анализа, методы классификации (Байесов классификатор, деревья решений) и оценку их качества.
7. Приобрести практический опыт работы с low-code и no-code системами для решения задач бизнес-аналитики.
8. Научиться интерпретировать результаты анализа и применять их для прогнозирования и принятия решений.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине		
			Код результата	Формулировка результата	
41.04.05 «Международные отношения» (М-МО)	ОПК-2 : Способен осуществлять поиск и применять перспективные информационно-коммуникационные технологии и программные средства для комплексной постановки и решения задач профессиональной деятельности	ОПК-2.1к : Оценивает возможности информационно-коммуникационных технологий и программных средств для решения стандартных задач профессиональной деятельности на основе информационной и библиографической культуры и требований информационной безопасности	РД1	Знание	современных направлений развития компьютерных технологий

		ОПК-2.2к : Выбирает информационно-коммуникационные технологии и программные средства для решения стандартных задач профессиональной деятельности	РД1	Умение	применять современные направления компьютерных технологий для хранения, обработки и представления данных
		ОПК-2.3в : Использует информационно-коммуникационные технологии и программные средства для решения стандартных задач профессиональной деятельности	РД1	Навык	работы с инструментами автоматизированной обработки информации

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	**Формирование ценностей**	**Целевые ориентиры**
Формирование гражданской позиции и патриотизма		
Развитие патриотизма и гражданской ответственности	Гуманизм	Внимательность к деталям Гуманность
Формирование духовно-нравственных ценностей		
Формирование ответственного отношения к труду	Созидательный труд	Дисциплинированность Самообучение
Формирование научного мировоззрения и культуры мышления		
Формирование осознания ценности научного мировоззрения и критического мышления	Гуманизм	Системное мышление
Формирование коммуникативных навыков и культуры общения		
Формирование навыков публичного выступления и презентации своих идей	Взаимопомощь и взаимоуважение	Умение работать в команде и взаимопомощь

2 Место дисциплины (модуля) в структуре ОПОП

Дисциплина «Технологии Искуственного интеллекта» относится к Блоку 1 Дисциплины (модули)

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обуче-ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо-емкость	Объем контактной работы (час)						СРС	Форма аттес-тации
				(З.Е.)	Всего	Аудиторная			Внеауди-торная			
						лек.	прак.	лаб.	ПА	КСР		
41.04.05 Международные отношения	ОЗФО	М01.Б	1	2	17	4	12	0	1	0	55	З

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОЗФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОЗФО

№	Название темы	Код ре-зультата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных и информации.	РД1	2	0	0	10	практическое задание
2	Анализ данных, большие данные, направления Data science, системы бизнес-аналитики	РД1	2	4	0	10	практическое задание
3	Анализа и визуализация данных	РД1	0	4	0	10	практическое задание
5	Методы машинного обучения	РД1	0	4	0	10	практическое задание
6	Системы хранения и визуализации данных. Системы класса low-code, no-code.	РД1	0	0	0	15	практическое задание
Итого по таблице			**4**	**12**	**0**	**55**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОЗФО

Тема 1 Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных и информации.

Содержание темы: Современные направления развития компьютерных технологий. Методы и средства хранения, обработки, анализа и представления данных и информации.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция.

Виды самостоятельной подготовки студентов по теме: теоретическая подготовка, компьютерное моделирование.

Тема 2 Анализ данных, большие данные, направления Data science, системы бизнес-аналитики.

Содержание темы: Основные методы анализа данных, различия направлений Data science, их практическое применение. Формирование датафреймов в табличном процессоре (MS Excel | Open Office Calc). Формат xlsx, csv, txt. Поиск ошибок (числовые данные), категории названы по-разному, объединение категорий. Дискретные, непрерывные данные. В чем отличие. Способы визуализации отдельных показателей и зависимостей.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: теоретическая подготовка, компьютерное моделирование.

Тема 3 Анализа и визуализация данных.

Содержание темы: Статистическая оценка числовых данных. Оценка частот и распределения вероятностей. Массивы данных (датафреймы). Оценка качества данных. Формирование «срезов». .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: теоретическая подготовка, компьютерное моделирование.

Тема 5 Методы машинного обучения.

Содержание темы: Методы кластеризации. Методы классификации (наивный Байесов классификатор). Практическое применения методов.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: теоретическая подготовка, компьютерное моделирование.

Тема 6 Системы хранения и визуализации данных. Системы класса low-code, no-code.

Содержание темы: Применение систем класса low-code, no-code для решения практических задач анализа бизнес-данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: теоретическая подготовка, компьютерное моделирование.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

Методические рекомендации по организации самостоятельной работы

В ходе изучения дисциплины студенты должны посещать аудиторные занятия (лекции, практические занятия, консультации). Особое место в овладении частью тем данной дисциплины отводится самостоятельной работе, при этом во время аудиторных

занятий могут быть рассмотрены и проработаны наиболее важные и трудные вопросы по той или иной теме дисциплины, а применение уже освоенные навыков в смежных технологиях вынесены на самостоятельное обучение.

В соответствии с учебным планом направления подготовки процесс изучения дисциплины предусматривает проведение лекций, практических занятий, консультаций, а также самостоятельную работу студентов.

Ниже перечислены предназначенные для самостоятельного изучения студентами те вопросы, которые во время проведения аудиторных занятий изучаются недостаточно или изучение которых носит обзорный характер.

Перечень и тематика самостоятельных работ студентов по дисциплине

1. Теория вероятностей
2. Прикладная статистика

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2025. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/560311 (дата обращения: 14.01.2025).

2. Бессмертный, И. А. Искусственный интеллект. Введение в многоагентные системы : учебник для вузов / И. А. Бессмертный. — Москва : Издательство Юрайт, 2025. — 148 с. — (Высшее образование). — ISBN 978-5-534-20348-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/569279 (дата обращения: 01.09.2025).

3. Гуриков, С. Р. Основы алгоритмизации и программирования на Python : учебное пособие / С.Р. Гуриков. — Москва : ИНФРА-М, 2025. — 343 с. — (Высшее образование). - ISBN 978-5-16-020255-6. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2166199 (дата обращения: 31.05.2026)

7.2 Дополнительная литература

1. Лапидус, Л. В. Цифровая экономика, экономика данных и прикладной искусственный интеллект : учебное пособие / Л. В. Лапидус. — Москва : ИНФРА-М, 2026. — 544 с. — (Высшее образование). - ISBN 978-5-16-021561-7. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2230806 (дата обращения: 31.05.2026)

2. Машинное обучение с использованием Python : учебно-методическое пособие / составители А. В. Осин,К. А. Хализев. — Москва :МТУСИ, 2025. — 20 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/501209 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

3. Протодьяконов А.В., Пылов П.А., Садовников В.Е. Алгоритмы Data Science и их практическая реализация на Python : Учебное пособие [Электронный ресурс] : Инфра-Инженерия , 2022 - 392 - Режим доступа: https://znanium.com/catalog/document?id=417222

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Образовательная платформа "ЮРАЙТ"
2. Электронная библиотечная система ZNANIUM.COM - Режим доступа: https://znanium.com/
3. Электронно-библиотечная система "ZNANIUM.COM"
4. Электронно-библиотечная система "ЛАНЬ"
5. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/
6. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/
7. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Коммутатор SuperStack 3 (16*10/100 19")
- Монитор облачный 23" LG23CAV42K/мышь Geniu
- Мультимедийный проектор №1 Casio XJ-V2
- Облачный монитор 23" LG CAV42K
- Облачный монитор LG Electronics черный +клавиатура+мышь

- П/К DNS Office T300, мышь Genius NetScroll 100, клавиатура Genius KB-06X, монитор AOC919 19"
- Проектор Casio XJ-V1
- Уст-во бесп.питания UPS-3000

Программное обеспечение:

- Microsoft OfficeProffessionalPlus 2019 Russian
- Python
- Windows

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

ТЕХНОЛОГИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Направление и направленность (профиль)
41.04.05 Международные отношения. Международные отношения и приграничное сотрудничество

Год набора на ОПОП
2026

Форма обучения
очно-заочная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
41.04.05 «Международные отношения» (М-МО)	ОПК-2 : Способен осуществлять поиск и применять перспективные информационно-коммуникационные технологии и программные средства для комплексной постановки и решения задач профессиональной деятельности	ОПК-2.1к : Оценивает возможности информационно-коммуникационных технологий и программных средств для решения стандартных задач профессиональной деятельности на основе информационной и библиографической культуры и требований информационной безопасности
		ОПК-2.2к : Выбирает информационно-коммуникационные технологии и программные средства для решения стандартных задач профессиональной деятельности
		ОПК-2.3в : Использует информационно-коммуникационные технологии и программные средства для решения стандартных задач профессиональной деятельности

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Компетенция ОПК-2 «Способен осуществлять поиск и применять перспективные информационно-коммуникационные технологии и программные средства для комплексной постановки и решения задач профессиональной деятельности»

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ОПК-2.1к : Оценивает возможности информационно-коммуникационных технологий и программных средств для решения стандартных задач профессиональной деятельности на основе информационной и библиографической культуры и требований информационной безопасности	РД 1	Знание	современных направлений развития компьютерных технологий	понимание разницы направлений развития компьютерных технологий
ОПК-2.2к : Выбирает информационно-коммуникационные технологии и программные средства для решения стандартных задач профессиональной деятельности	РД 1	Умение	применять современные направления компьютерных технологий для хранения, обработки и представления данных	понимание основных методов работы с данными

ОПК-2.3в : Использует информационно-коммуникационные технологии и программные средства для решения стандартных задач профессиональной деятельности	РД 1	Навык	работы с инструментами автоматизированной обработки информации	скорость и точность выполнения типовых операций обработки данных в специализированном ПО

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очно-заочная форма обучения				
РД1	Знание : современных направлений развития компьютерных технологий	1.1. Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных и информации.	Тест	Тест
		1.2. Анализ данных, большие данные, направления Data science, системы бизнес-аналитики	Тест	Тест
РД1	Умение : применять современные направления компьютерных технологий для хранения, обработки и представления данных	1.5. Методы машинного обучения	Практическая работа	Практическая работа
		1.6. Системы хранения и визуализации данных. Системы класса low-code, no-code.	Практическая работа	Практическая работа
РД1	Навык : работы с инструментами автоматизированной обработки информации	1.3. Анализа и визуализация данных	Практическая работа	Практическая работа

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Вид учебной деятельности	Оценочное средство		
	Практическая работа	Тест	Итого
Лекция		10	10
Промежуточная аттестация		30	30

Практические занятия	60		60
Самостоятельная работа			
Итого	60	40	100

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» /	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

	«неудовлетворительно»	

5 Примерные оценочные средства

5.1 Примеры заданий для выполнения практических работ

Задание 1: Написание программного кода, позволяющего подготовить данные для анализа и визуализации (работа с пропусками, приведение данных к стандартному табличному виду и т.д.)

Задание 2: Написание программного кода, позволяющего построить правильные диаграммы распределения для каждого типа данных, объяснение полученных результатов.

Задание 3: Написание программного кода для проведения описательной статистики данных, визуализация и объяснение полученных результатов.

Задание 4: Написание программного кода для проведения расширенного частотного анализа, анализа взаимосвязей между категориальными переменными, визуализация и объяснение полученных результатов.

Задание 5: Написание программного кода, позволяющего создавать датафреймы и срезы, проводить базовые операции над датафреймами с использованием стандартных методов объединения данных, проводить оценку качества данных.

Задание 6: Загрузка данных с помощью систем класса low-code и no-code, проведение подготовки данных, их предварительного анализа и визуализации. Объяснение полученных результатов.

Краткие методические указания

После выполнения каждой практической работы студент должен представить отчет о ее выполнении, а также ответить на сопутствующие вопросы по теме.

Шкала оценки

№	Баллы	Описание
5	10–11	Студент демонстрирует умения на итоговом уровне: умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.
4	8–9	Студент демонстрирует умения на среднем уровне: освоил основные умения, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.
3	6–7	Студент демонстрирует умения и навыки на базовом уровне: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных умений, навыков по дисциплинарной компетенции, испытываются значительные затруднения при оперировании умениями и при их переносе на новые ситуации.
2	3–5	Студент демонстрирует умения и навыки на уровне ниже базового: проявляется недостаточность умений и навыков.
1	0–2	Студентом проявляется полное или практически полное отсутствие умений и навыков.

5.2 Контрольный тест

Тест 1: Анализ данных, большие данные, направления Data Science, системы бизнес-аналитики

1. Какой из перечисленных этапов НЕ входит в стандартный процесс анализа данных (CRISP-DM)?

a) Понимание бизнес-задачи
b) Сбор данных
c) Удаление всех исходных данных
d) Построение моделей

2. Какая из характеристик НЕ относится к "3V" больших данных?

a) Volume (объем)
b) Velocity (скорость)
c) Variety (разнообразие)
d) Validity (валидность)

3. Какое направление Data Science занимается прогнозированием числовых значений?

a) Классификация
b) Кластеризация
c) Регрессионный анализ
d) Ассоциативные правила

4. Какой инструмент НЕ является системой бизнес-аналитики (BI)?

a) Power BI
b) Tableau
c) Qlik Sense
d) Apache Kafka

5. Какой алгоритм машинного обучения относится к обучению без учителя?

a) Линейная регрессия
b) Метод k-ближайших соседей
c) Метод k-средних
d) Дерево решений

6. Какая технология используется для потоковой обработки данных?

a) Apache Hadoop
b) Apache Spark
c) Apache Kafka
d) Microsoft Excel

7. Что означает термин "ETL" в контексте анализа данных?

a) Extract, Transform, Load
b) Encrypt, Transfer, Lock
c) Evaluate, Test, Learn
d) Export, Tag, Label

8. Какой показатель используется для оценки качества классификации?

a) Коэффициент детерминации (R^2)
b) Среднеквадратичная ошибка (MSE)
c) Матрица ошибок
d) Дисперсия

9. Какой язык программирования чаще всего используется в Data Science?

a) Java
b) Python
c) C++
d) PHP

10. Какой метод используется для снижения размерности данных?

a) Линейная регрессия
b) Метод главных компонент (PCA)
c) Дерево решений
d) Логистическая регрессия

Тест 2: Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных

1. Какая из перечисленных баз данных относится к реляционным?

a) MongoDB
b) PostgreSQL
c) Redis
d) Cassandra

2. Какой язык используется для запросов к реляционным базам данных?

a) Python
b) SQL
c) HTML
d) XML

3. Какой формат данных НЕ является структурированным?

a) CSV
b) JSON
c) XML
d) Текстовый документ без разметки

4. Какая технология используется для распределенного хранения больших данных?

a) Microsoft Word
b) Hadoop HDFS
c) PowerPoint
d) Adobe Photoshop

5. Какой инструмент используется для интерактивной аналитики данных?

a) Jupyter Notebook
b) Блокнот
c) Microsoft Paint
d) Adobe Illustrator

6. Какая платформа позволяет выполнять параллельные вычисления на кластерах?

a) Apache Spark
b) Microsoft Excel
c) Adobe Acrobat
d) WinRAR

7. Какой формат данных чаще всего используется в веб-API?

a) JSON
b) MP3
c) AVI
d) EXE

8. Какая СУБД относится к NoSQL?

a) MySQL
b) PostgreSQL
c) MongoDB
d) Oracle

9. Какой инструмент НЕ используется для обработки данных?

a) Pandas
b) NumPy
c) Microsoft PowerPoint
d) SQL

10. Какая технология используется для контейнеризации приложений?

a) Docker
b) Microsoft Word
c) Adobe Photoshop
d) WinZip

Тест 3: Системы хранения и визуализации данных. Российский сегмент рынка

1. Какой российский аналог Tableau существует на рынке?

a) Яндекс.Метрика
b) DataLens (от Яндекса)

c) 1С:Предприятие
d) Тинькофф Аналитика

2. Какая российская СУБД популярна для аналитики больших данных?

a) ClickHouse
b) Oracle
c) MySQL
d) PostgreSQL

3. Какой инструмент визуализации разработан компанией "Точка зрения"?

a) Power BI
b) Tableau
c) DataLens
d) Qlik Sense

4. Какая российская платформа предоставляет облачные решения для хранения данных?

a) Яндекс.Облако
b) Google Drive
c) Dropbox
d) iCloud

5. Какой российский сервис предоставляет аналитику веб-трафика?

a) Яндекс.Метрика
b) Google Analytics
c) Adobe Analytics
d) Facebook Insights

6. Какая российская компания разрабатывает решения для обработки больших данных?

a) СберТех
b) Microsoft
c) IBM
d) Oracle

7. Какой формат визуализации лучше всего подходит для временных рядов?

a) Круговая диаграмма
b) Линейный график
c) Диаграмма рассеяния
d) Гистограмма

8. Какая российская платформа предоставляет инструменты для машинного обучения?

a) TensorFlow
b) PyTorch
c) CatBoost (разработан Яндексом)
d) Scikit-learn

9. Какой инструмент НЕ является системой хранения данных?

a) Hadoop HDFS
b) Amazon S3
c) Яндекс.Облако
d) Microsoft PowerPoint

10. Какой российский сервис предоставляет API для геоаналитики?

a) Яндекс.Карты
b) Google Maps
c) Apple Maps
d) OpenStreetMap

Тест 4: Анализ данных, большие данные, направления Data Science, системы бизнес-аналитики

1. Какой метод анализа данных используется для выявления скрытых закономерностей в больших массивах информации?

a) Описательная статистика
b) Data Mining
c) Линейная регрессия
d) Визуализация

2. Какая характеристика Big Data описывает разнообразие форматов данных?

a) Volume
b) Velocity
c) Variety
d) Veracity

3. Какой тип машинного обучения использует размеченные данные для обучения?

a) Обучение с учителем
b) Обучение без учителя
c) Смешанное обучение
d) Глубокое обучение

4. Какой инструмент позволяет создавать интерактивные дашборды без написания кода?

a) Jupyter Notebook
b) Tableau
c) Apache Spark
d) TensorFlow

5. Какой алгоритм используется для разделения данных на группы по схожести?

a) Линейная регрессия
b) Метод k-средних
c) Дерево решений
d) SVM

6. Какой процесс преобразует сырые данные в пригодный для анализа формат?

a) Data Cleaning
b) Data Aggregation
c) Data Wrangling
d) Data Visualization

7. Какой показатель оценивает точность регрессионной модели?

a) Accuracy
b) F1-score
c) R-квадрат
d) Precision

8. Какая библиотека Python чаще всего используется для работы с табличными данными?

a) NumPy
b) Pandas
c) Matplotlib
d) Scikit-learn

9. Какой метод НЕ используется для обработки пропущенных значений?

a) Удаление строк с пропусками
b) Замена средним значением
c) Замена нулями
d) Шифрование данных

10. Какой инструмент используется для управления workflow в Data Science проектах?

a) Apache Airflow
b) Microsoft Word
c) Adobe Photoshop
d) WinRAR

Тест 5: Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных

1. Какой тип базы данных оптимален для хранения JSON-документов?

a) Реляционная
b) Документоориентированная
c) Графовая
d) Ключ-значение

2. Какой язык используется для работы с Apache Spark?

a) SQL
b) Python (PySpark)
c) HTML
d) CSS

3. Какой формат данных обеспечивает схему для структурированного хранения?

a) CSV
b) JSON
c) Parquet
d) TXT

4. Какая технология используется для быстрого поиска по большим данным?

a) Elasticsearch
b) Microsoft Excel
c) Adobe Acrobat
d) WinZip

5. Какой инструмент НЕ относится к системам управления базами данных?

a) MySQL
b) MongoDB
c) PostgreSQL
d) Microsoft PowerPoint

6. Какой протокол используется для передачи потоковых данных?

a) HTTP
b) FTP
c) MQTT
d) SMTP

7. Какая технология позволяет выполнять SQL-запросы к большим данным?

a) Apache Hive
b) Microsoft Word
c) Adobe Photoshop
d) WinRAR

8. Какой формат данных используется для хранения графовых структур?

a) CSV
b) XML
c) RDF
d) JSON

9. Какой инструмент используется для оркестрации контейнеров?

a) Docker Compose
b) Microsoft Excel
c) Adobe Illustrator
d) WinZip

10. Какая технология используется для хранения временных рядов?

a) Redis
b) InfluxDB
c) MySQL
d) PostgreSQL

Тест 6: Системы хранения и визуализации данных. Российский сегмент рынка

1. Какой российский аналог Power BI предлагает облачные решения?

a) Яндекс.Метрика
b) DataLens
c) 1С:Аналитика
d) Тинькофф Insights

2. Какая российская СУБД разработана для аналитических запросов?

a) Tarantool
b) ClickHouse
c) Postgres Pro
d) Яндекс.База

3. Какой российский сервис предоставляет инструменты для геоаналитики?

a) Яндекс.Карты API
b) Google Maps API
c) 2GIS API
d) OpenStreetMap

4. Какая российская платформа специализируется на обработке потоковых данных?

a) Apache Kafka
b) Яндекс.Потоки
c) Сбер.Аналитика
d) Mail.ru Cloud

5. Какой инструмент визуализации разработан в России?

a) Tableau
b) Qlik Sense
c) DataLens
d) Power BI

6. Какая российская компания разрабатывает решения для компьютерного зрения?

a) Яндекс
b) СберТех
c) Mail.ru Group
d) Все перечисленные

7. Какой формат визуализации лучше всего подходит для сравнения долей?

a) Линейный график
b) Круговая диаграмма
c) Гистограмма
d) Диаграмма рассеяния

8. Какой российский сервис предоставляет API для анализа текстов?

a) Яндекс.SpeechKit
b) CatBoost
c) CloudPayments
d) Tinkoff API

9. Какая российская платформа предоставляет облачные GPU для ML?

a) Яндекс.Облако
b) Selectel

c) Mail.ru Cloud Solutions
d) Все перечисленные

10. Какой инструмент НЕ является российским продуктом для работы с данными?

a) ClickHouse
b) Tarantool
c) Oracle Database
d) DataLens

Тест 7: Анализ данных, большие данные, направления Data Science, системы бизнес-аналитики

1. Какой метод используется для анализа временных рядов?

a) ARIMA
b) K-means
c) SVM
d) Random Forest

2. Что означает термин "Feature Engineering" в машинном обучении?

a) Удаление всех признаков
b) Создание и преобразование признаков
c) Визуализация данных
d) Шифрование данных

3. Какой алгоритм используется для обнаружения аномалий?

a) Линейная регрессия
b) Isolation Forest
c) Логистическая регрессия
d) Метод главных компонент

4. Какой показатель оценивает качество бинарной классификации?

a) MSE
b) RMSE
c) ROC-AUC
d) R-squared

5. Какой инструмент используется для автоматического машинного обучения (AutoML)?

a) TensorFlow
b) PyTorch
c) H2O.ai
d) Scikit-learn

6. Какой метод используется для обработки категориальных признаков?

a) One-Hot Encoding
b) Нормализация
c) Стандартизация
d) PCA

7. Что такое "кросс-валидация" в машинном обучении?

a) Удаление данных
b) Разделение данных на обучающую и тестовую выборки
c) Многократное разбиение данных для оценки модели
d) Визуализация результатов

8. Какой алгоритм НЕ относится к ансамблевым методам?

a) Random Forest
b) Gradient Boosting
c) K-nearest Neighbors
d) XGBoost

9. Какой инструмент используется для версионирования данных?

a) Git
b) DVC (Data Version Control)
c) Docker
d) Kubernetes

10. Какой метод используется для балансировки классов?

a) Удаление примеров мажоритарного класса
b) SMOTE
c) Увеличение веса миноритарного класса
d) Все перечисленные

Тест 8: Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных

1. Какой тип индекса ускоряет поиск в столбцах с текстовыми данными?

a) B-дерево
b) Хэш-индекс
c) Обратный индекс
d) R-дерево

2. Какой формат данных оптимален для хранения вложенных структур?

a) CSV
b) JSON
c) XML
d) Parquet

3. Какой язык используется для запросов в ClickHouse?

a) SQL
b) NoSQL
c) GraphQL
d) Python

4. Какой инструмент используется для оркестрации ETL-процессов?

a) Apache NiFi
b) Apache Kafka
c) Apache Spark
d) Apache Hadoop

5. Какой протокол используется для передачи данных между микросервисами?

a) HTTP
b) gRPC
c) FTP
d) SMTP

6. Какой тип базы данных используется для хранения графов?

a) MongoDB
b) Neo4j
c) Redis
d) Cassandra

7. Какой инструмент используется для мониторинга данных?

a) Grafana
b) Tableau
c) Power BI
d) Excel

8. Какой формат данных используется в Apache Kafka?

a) CSV
b) JSON
c) Avro
d) XML

9. Какой инструмент используется для управления метаданными?

a) Apache Atlas
b) Apache Spark
c) Apache Hadoop
d) Apache Kafka

10. Какой тип хранилища используется для аналитических запросов?

a) OLTP
b) OLAP
c) Key-Value
d) Document

Тест 9: Системы хранения и визуализации данных. Российский сегмент рынка

1. Какой российский инструмент используется для визуализации геоданных?

a) Яндекс.Карты
b) Google Maps
c) 2GIS
d) OpenStreetMap

2. Какая российская платформа предоставляет сервис распознавания лиц?

a) VisionLabs
b) Яндекс.Облако
c) СберТех
d) Mail.ru Group

3. Какой российский продукт используется для управления данными?

a) DataSphere от Яндекса
b) Google Data Studio
c) Tableau
d) Power BI

4. Какой российский сервис предоставляет API для обработки естественного языка?

a) Яндекс.Облако
b) ChatGPT
c) Google NLP
d) IBM Watson

5. Какой российский инструмент используется для анализа логов?

a) ELK Stack
b) ClickHouse
c) Grafana Loki
d) Zabbix

6. Какой российский продукт используется для хранения временных рядов?

a) InfluxDB
b) TimescaleDB
c) VictoriaMetrics
d) Prometheus

7. Какой российский сервис предоставляет облачные GPU для ML?

a) Selectel
b) Яндекс.Облако
c) Mail.ru Cloud Solutions
d) Все перечисленные

8. Какой российский инструмент используется для управления метаданными?

a) Apache Atlas
b) DataLens
c) Amundsen
d) Яндекс.Метрика

9. Какой российский продукт используется для потоковой обработки данных?

a) Apache Kafka
b) Яндекс.Потоки
c) Apache Flink
d) Apache Spark

10. Какой российский сервис предоставляет инструменты для компьютерного зрения?

a) Яндекс.Облако
b) СберТех
c) VisionLabs
d) Все перечисленные

Тест 10: Анализ данных и Data Science

1. **Какой метод используется для уменьшения переобучения в нейронных сетях?**
   a) Увеличение количества слоев
   b) Dropout
   c) Увеличение learning rate
   d) Уменьшение количества эпох

1. **Что измеряет метрика F1-score?**
   a) Точность и полноту одновременно
   b) Только точность
   c) Только полноту
   d) Среднее арифметическое всех метрик

1. **Какой алгоритм НЕ является алгоритмом кластеризации?**
   a) DBSCAN
   b) K-means
   c) Random Forest
   d) Hierarchical clustering

2. **Что такое "холодный старт" в рекомендательных системах?**
   a) Проблема рекомендаций для новых пользователей/объектов
   b) Слишком быстрое обучение модели
   c) Отсутствие данных в системе
   d) Устаревание модели

3. **Какой метод НЕ используется для обработки дисбаланса классов?**
   a) Undersampling
   b) Oversampling
   c) Увеличение learning rate
   d) Использование весов классов

4. **Что такое "transfer learning"?**
   a) Использование предобученной модели для новой задачи
   b) Перенос данных между серверами
   c) Обучение без учителя
   d) Автоматическое машинное обучение

5. **Какой тип нейронных сетей лучше всего подходит для обработки изображений?**
   a) Полносвязные сети
   b) Рекуррентные сети

c) Сверточные сети
d) Автокодировщики

6. **Что такое "batch normalization"?**
   a) Нормализация входных данных
   b) Нормализация активаций между слоями
   c) Уменьшение размера батча
   d) Увеличение скорости обучения

7. **Какой алгоритм оптимизации чаще всего используется в глубоком обучении?**
   a) Градиентный спуск
   b) Стохастический градиентный спуск
   c) Adam
   d) K-means

8. **Что такое "attention mechanism" в нейронных сетях?**
   a) Механизм выделения важных частей входных данных
   b) Метод регуляризации
   c) Тип функции активации
   d) Способ инициализации весов

Тест 11: Большие данные и системы хранения

1. Какой принцип лежит в основе технологии блокчейн?
   a) Репликация данных
   b) Децентрализованное хранение
   c) Распределенный реестр
   d) Все вышеперечисленное

2. Что такое "data lake"?
   a) Хранилище неструктурированных данных
   b) Реляционная база данных
   c) Система визуализации
   d) Инструмент ETL

3. Какой инструмент используется для потоковой обработки данных в реальном времени?
   a) Apache Flink
   b) Apache Hadoop
   c) Apache Hive
   d) Apache Spark

4. Что такое "sharding" в базах данных?
   a) Горизонтальное разделение данных
   b) Вертикальное разделение данных
   c) Сжатие данных
   d) Шифрование данных

5. Какой тип базы данных оптимален для хранения временных рядов?
   a) InfluxDB
   b) MongoDB
   c) PostgreSQL
   d) Redis

6. Что такое "CAP-теорема"?
   a) Теорема о согласованности, доступности и устойчивости к разделению
   b) Теорема о скорости обработки данных
   c) Теорема о безопасности данных
   d) Теорема о масштабируемости

7. Какой формат данных обеспечивает схему для структурированного хранения?
   a) CSV
   b) JSON
   c) Parquet
   d) XML

8. Что такое "lambda-архитектура"?
   a) Подход к обработке больших данных, сочетающий batch и stream processing
   b) Архитектура микросервисов
   c) Модель глубокого обучения
   d) Способ хранения данных

9. Какой инструмент используется для управления workflow в data pipeline?
   a) Apache Airflow
   b) Apache Kafka
   c) Apache Spark
   d) Apache Hadoop

10. Что такое "polyglot persistence"?
    a) Использование разных СУБД для разных типов данных
    b) Хранение данных в одном формате
    c) Метод сжатия данных
    d) Способ репликации данных

Тест 12: Российский сегмент рынка и визуализация данных

1. Какой российский продукт является аналогом Tableau?
   a) DataLens
   b) Яндекс.Метрика
   c) 1С:Аналитика
   d) Тинькофф Business

2. Какая российская компания разрабатывает платформу для компьютерного зрения?
   a) VisionLabs
   b) Яндекс
   c) СберТех
   d) Все вышеперечисленные

3. Какой российский инструмент используется для анализа логов?
   a) ClickHouse
   b) ELK Stack
   c) Grafana
   d) Zabbix

4. Какой российский сервис предоставляет API для обработки естественного языка?
   a) Яндекс.Облако
   b) Сбер AI

c) Mail.ru Cloud
d) Все вышеперечисленные

5. Какой тип визуализации лучше всего подходит для отображения корреляции?
   a) Линейный график
   b) Круговая диаграмма
   c) Диаграмма рассеяния
   d) Гистограмма

6. Какой российский продукт используется для хранения временных рядов?
   a) VictoriaMetrics
   b) InfluxDB
   c) TimescaleDB
   d) Prometheus

7. Какой инструмент позволяет создавать интерактивные веб-дашборды?
   a) Plotly Dash
   b) Matplotlib
   c) Seaborn
   d) Pandas

8. Какой российский сервис предоставляет облачные GPU для ML?
   a) Яндекс.Облако
   b) Selectel
   c) Mail.ru Cloud
   d) Все вышеперечисленные

9. Какой метод визуализации лучше всего подходит для иерархических данных?
   a) Дендрограмма
   b) Гистограмма
   c) Box plot
   d) Линейный график

10. Какой российский продукт является аналогом Power BI?
    a) DataLens
    b) Яндекс.Метрика
    c) 1С:Аналитика
    d) Тинькофф Analytics

Тест 13: Продвинутые методы анализа данных и машинного обучения

1. Какой метод используется для интерпретации результатов работы сложных ML-моделей?
   a) LIME
   b) PCA
   c) K-means
   d) ARIMA

2. Что такое "дифференциальная приватность" в контексте анализа данных?
   a) Метод защиты персональных данных
   b) Алгоритм кластеризации
   c) Техника увеличения данных
   d) Метод визуализации

3. Какой алгоритм используется для генерации новых данных, похожих на обучающую выборку?
   a) GAN (Generative Adversarial Networks)
   b) SVM
   c) Random Forest
   d) KNN

4. Что измеряет метрика "точность в top-k"?
   a) Точность среди первых k рекомендаций
   b) Полноту предсказаний
   c) Среднюю точность
   d) Время выполнения запроса

5. Какой метод используется для обработки естественного языка (NLP)?
   a) Word2Vec
   b) K-means
   c) Linear Regression
   d) Decision Trees

6. Что такое "многорукий бандит" (multi-armed bandit) в контексте анализа данных?
   a) Алгоритм для тестирования гипотез
   b) Метод кластеризации
   c) Техника визуализации
   d) Способ хранения данных

7. Какой тип нейронных сетей используется для обработки последовательностей?
   a) RNN (Recurrent Neural Networks)
   b) CNN
   c) GAN
   d) Autoencoders

8. Что такое "обучение с подкреплением" (reinforcement learning)?
   a) Обучение через взаимодействие со средой
   b) Обучение на размеченных данных
   c) Обучение без учителя
   d) Метод кластеризации

9. Какой алгоритм используется для уменьшения размерности в нелинейных случаях?
   a) t-SNE
   b) PCA
   c) LDA
   d) SVD

10. Что такое "функция потерь Хьюбера" (Huber loss)?
    a) Комбинация MSE и MAE
    b) Метод кластеризации
    c) Техника визуализации
    d) Алгоритм рекомендаций

Тест 14: Современные технологии обработки больших данных

1. Что такое "data mesh" архитектура?
   a) Децентрализованный подход к управлению данными

b) Централизованное хранилище данных
c) Метод визуализации
d) Алгоритм машинного обучения

2. Какой инструмент используется для обработки графовых данных в реальном времени?
   a) Apache Flink
   b) Neo4j
   c) Apache Kafka
   d) Apache Spark

3. Что такое "feature store" в ML?
   a) Централизованное хранилище признаков
   b) База данных для логов
   c) Инструмент визуализации
   d) Система мониторинга

4. Какой формат данных используется для эффективного хранения вложенных структур?
   a) Avro
   b) CSV
   c) XML
   d) TXT

5. Что такое "delta lake"?
   a) Открытый формат хранения поверх data lake
   b) Реляционная БД
   c) Графовая БД
   d) Инструмент визуализации

6. Какой инструмент используется для оркестрации ML workflows?
   a) MLflow
   b) Apache Kafka
   c) Apache Hadoop
   d) Apache Spark

7. Что такое "data lineage"?
   a) Отслеживание происхождения и преобразований данных
   b) Метод кластеризации
   c) Алгоритм рекомендаций
   d) Техника визуализации

8. Какой подход используется для обработки потоковых данных с задержкой менее 1 мс?
   a) Event-driven architecture
   b) Batch processing
   c) Microservices
   d) Monolithic architecture

9. Что такое "data fabric"?
   a) Единый уровень доступа к данным
   b) Инструмент визуализации

c) Алгоритм машинного обучения
d) Метод хранения данных

10. Какой инструмент используется для управления ML моделями в production?
a) Kubeflow
b) Apache Kafka
c) Apache Spark
d) Apache Hadoop

Тест 15: Российские технологии в области данных

1. Какой российский фреймворк для ML разработан Яндексом?
a) CatBoost
b) TensorFlow
c) PyTorch
d) Scikit-learn

2. Какой российский сервис предоставляет аналитику в реальном времени?
a) Яндекс.Метрика
b) Google Analytics
c) Adobe Analytics
d) Mixpanel

3. Какой российский продукт является аналогом Snowflake?
a) Яндекс.Облако Data Platform
b) 1С:Предприятие
c) Тинькофф Data Warehouse
d) Сбер.Аналитика

4. Какой российский инструмент используется для управления данными?
a) DataLens
b) Tableau
c) Power BI
d) QlikView

5. Какой российский сервис предоставляет NLP API?
a) Яндекс.Облако SpeechKit
b) Google Cloud NLP
c) AWS Comprehend
d) IBM Watson

6. Какой российский продукт используется для обработки потоковых данных?
a) Яндекс.Потоки
b) Apache Kafka
c) Apache Flink
d) Apache Spark

7. Какой российский инструмент для визуализации геоданных?
a) Яндекс.Карты API
b) Google Maps API
c) Mapbox
d) OpenLayers

8. Какой российский фреймворк для глубокого обучения?
   a) DeepPavlov
   b) TensorFlow
   c) PyTorch
   d) Keras

9. Какой российский сервис предоставляет аналитику для бизнеса?
   a) Яндекс.Метрика
   b) Google Analytics 360
   c) Adobe Analytics
   d) Amplitude

10. Какой российский продукт используется для хранения и обработки больших данных?
    a) ClickHouse
    b) Apache Hadoop
    c) Apache Spark
    d) Elasticsearch

Краткие методические указания

После прохождения теоретической части, студенты должны закрепить материал при помощи контрольного теста.

Шкала оценки

Оценка	Баллы	Описание
5	9-10	
4	6-8	
3	3-5	
2	0-2	

Практическая работа 1

```
import pandas as pd
import numpy as np
def read_american_csv(file_path):
   df = pd.read_csv(file_path, dtype=str)

   for col in df.columns:
      try:
         df[col] = pd.to_numeric(df[col].str.replace(',', ''))
      except ValueError:
         pass

   return df

dataCL=read_american_csv('/content/Covid Live.csv')
print('\n',"Визуализация датасета")
display(dataCL)
print('\n', "Информация о стурктуре данных")
dataCL.info()
print('\n',"ОПИСАТЕЛЬНАЯ СТАТИСТИКА")
dataCL.describe(include='all')
dataCL = dataCL.dropna(thresh=int(len(dataCL)*0.8), axis=1)
dataCL = dataCL.dropna(thresh=int(len(dataCL.loc[1])*0.8), axis=0)
print('\n', "Информация о стурктуре данных")
dataCL.info()
for i in dataCL.select_dtypes(include=['number']).columns:
   dataCL[i] = dataCL[i].fillna(dataCL[i].median())
print('\n', "Информация о стурктуре данных")
dataCL.info()
```

Практическая работа 2

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

age = np.random.normal(35, 10, 100).round()
data = pd.DataFrame({
   'Возраст': age,
   'Доход': 10*age + np.random.lognormal(4, 0.5, 100).round(2),
   'Количество покупок': np.random.randint(1, 10, 100),
   'Пол': np.random.choice(['М', 'Ж'], 100, p=[0.4, 0.6,]),
   'Удовлетворенность': np.random.choice(['Низкая', 'Средняя', 'Высокая'], 100, p=[0.2, 0.5, 0.3]),
})

print("Первые 5 строк данных:")
print(data.head())
print("\nИнформация о типах данных:")
print(data.info())
plt.figure(figsize=(15, 10))

plt.subplot(3, 1, 1)
```

```
sns.countplot(x=data['Количество покупок'], palette='viridis', hue=data['Количество покупок'],
legend=False)
plt.title('Количество покупок (дискретные данные)')

plt.subplot(3, 1, 2)
sns.histplot(data['Возраст'], bins=15, color='skyblue')
plt.title('Распределение возраста (непрерывные данные)')

# Box-plot для выбросов
plt.subplot(3, 1, 3)
sns.boxplot(data['Доход'], color='lightgreen')
plt.title('Распределение дохода (правостороннее)')

plt.tight_layout()
plt.show()
plt.figure(figsize=(15, 5))

plt.subplot(1, 2, 1)
data['Пол'].value_counts().plot.pie(autopct='%1.1f%%', colors=['lightcoral', 'lightblue', 'lightgreen'])
plt.title('Распределение по полу')

plt.subplot(1, 2, 2)
sns.countplot(data=data, x='Удовлетворенность', order=['Низкая', 'Средняя', 'Высокая'],
palette='Blues_r', hue='Удовлетворенность', legend=False)
plt.title('Уровень удовлетворенности')

from statsmodels.graphics.mosaicplot import mosaic
mosaic(data, ['Пол', 'Удовлетворенность'], title='Пол vs Удовлетворенность')
plt.show()
```

Практическая работа 3

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

np.random.seed(42)

data = pd.DataFrame({
   'Возраст': np.random.normal(loc=35, scale=10, size=1000).round(),
   'Доход': np.random.lognormal(mean=4, sigma=0.7, size=1000).round(2),
   'Количество_покупок': np.random.poisson(lam=3, size=1000),
   'Сумма_покупок': np.random.gamma(shape=2, scale=20, size=1000).round(2),
   'Пол': np.random.choice(['М', 'Ж'], 1000, p=[0.45, 0.55]),
   'Образование': np.random.choice(['Среднее', 'Среднее спец.', 'Высшее', 'Ученая степень'],
                        1000, p=[0.3, 0.4, 0.25, 0.05])
})

data.loc[np.random.choice(1000, 20, replace=False), 'Доход'] *= 3
def print_central_tendency(df, columns):
   for col in columns:
      print(f"\nАнализ столбца: {col}")
```

```
        print(f"Среднее значение: {df[col].mean():.2f}")
        print(f"Медиана: {df[col].median():.2f}")
        mode = df[col].mode()
        print(f"Мода: {', '.join(map(str, mode.values))} (встречается {df[col].value_counts().max()} раз)")

        diff = abs(df[col].mean() - df[col].median())
        if diff > 0.1 * df[col].std():
            print("Заметная разница между средним и медианой указывает на асимметрию
распределения")


        plt.figure(figsize=(10, 5))
        sns.histplot(df[col], kde=True)
        plt.axvline(df[col].mean(), color='r', linestyle='--', label=f'Среднее: {df[col].mean():.2f}')
        plt.axvline(df[col].median(), color='g', linestyle='-', label=f'Медиана: {df[col].median():.2f}')
        plt.title(f'Распределение {col} с отмеченными мерами центра')
        plt.legend()
        plt.show()

numeric_cols = ['Возраст', 'Доход', 'Количество_покупок', 'Сумма_покупок']
print_central_tendency(data, numeric_cols)

def print_variability_measures(df, columns):
    variability = pd.DataFrame(index=columns,
                               columns=['Размах', 'Дисперсия', 'Станд. отклонение', 'IQR', 'Коэф. вариации'])

    for col in columns:
        variability.loc[col, 'Размах'] = df[col].max() - df[col].min()
        variability.loc[col, 'Дисперсия'] = df[col].var()
        variability.loc[col, 'Станд. отклонение'] = df[col].std()
        variability.loc[col, 'IQR'] = df[col].quantile(0.75) - df[col].quantile(0.25)
        variability.loc[col, 'Коэф. вариации'] = df[col].std() / df[col].mean()

    print("\nМеры изменчивости:")
    print(variability)

    plt.figure(figsize=(12, 6))
    sns.boxplot(data=df[columns], orient='h')
    plt.title('Boxplot для анализа разброса данных')
    plt.show()

print_variability_measures(data, numeric_cols)
```

Практическая работа 4

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats

np.random.seed(42)

data = pd.DataFrame({
```

```
    'Пол': np.random.choice(['Мужчина', 'Женщина'], 1500, p=[0.47, 0.53]),
    'Возрастная_группа': np.random.choice(['18-25', '26-45', '46+'],
                                1500, p=[0.2, 0.55, 0.25]),
    'Образование': np.random.choice(
        ['Колледж', 'Бакалавр', 'Магистр'],
        1500,
        p=[0.45, 0.35, 0.2]
    )
})


age_satisfaction = {
    '18-25': [0.6, 0.25, 0.15],  # Молодежь менее удовлетворена
    '26-45': [0.3, 0.55, 0.15],
    '46+': [0.05, 0.40, 0.55]     # Пожилые более удовлетворены
}

data['Удовлетворенность'] = data['Возрастная_группа'].apply(
    lambda x: np.random.choice(
        ['Низкая', 'Средняя', 'Высокая'],
        p=age_satisfaction[x]
    )
)


data.loc[data.sample(frac=0.03).index, 'Образование'] = np.nan
data.loc[data.sample(frac=0.02).index, 'Удовлетворенность'] = np.nan

print(f"\n=== Основные статистические характеристики ===")
print(data.describe(),'\n')
def extended_frequency_analysis(df, column):
    print(f"\n=== Расширенный анализ для '{column}' ===")

    counts = df[column].value_counts(dropna=False)
    rel_freq = df[column].value_counts(normalize=True, dropna=False)

    freq_df = pd.DataFrame({
        'Количество': counts,
        'Доля': rel_freq,
        'Накопленная доля': rel_freq.cumsum()
    })
    print(freq_df)

    probs = rel_freq[rel_freq > 0]
    entropy = -np.sum(probs * np.log2(probs))
    diversity = 1 - np.sum(probs**2)

    print(f"\nЭнтропия: {entropy:.3f} бит")
    print(f"Индекс разнообразия Симпсона: {diversity:.3f}")

    plt.figure(figsize=(12, 5))
    plt.subplot(1, 2, 1)
    sns.barplot(x=counts.index, y=counts.values)
    plt.title(f'Абсолютные частоты ({column})')
```

```
    plt.xticks(rotation=45)

    plt.subplot(1, 2, 2)
    plt.pie(rel_freq, labels=rel_freq.index, autopct='%1.1f%%')
    plt.title(f'Относительные частоты ({column})')

    plt.tight_layout()
    plt.show()

extended_frequency_analysis(data, 'Образование')
def ordinal_analysis(df, column, order):
    print(f"\n=== Анализ порядковой переменной '{column}' ===")

    df[column] = pd.Categorical(df[column], categories=order, ordered=True)

    cum_freq = df[column].value_counts(normalize=True).sort_index().cumsum()
    median_cat = cum_freq[cum_freq >= 0.5].index[0]
    print(f"Медианная категория: {median_cat}")

    plt.figure(figsize=(10, 5))
    ax = sns.countplot(data=df, x=column, order=order)
    plt.title(f'Распределение {column} с учетом порядка')
    plt.xticks(rotation=45)

    ymax = ax.get_ylim()[1]
    plt.plot([median_cat, median_cat], [0, ymax], 'r--', alpha=0.5)
    plt.text(median_cat, ymax*0.9, ' Медиана', color='red')
    plt.show()

satisfaction_order = ['Очень низкая', 'Низкая', 'Средняя', 'Высокая', 'Очень высокая']
ordinal_analysis(data, 'Удовлетворенность', satisfaction_order)
```

Практическая работа 5

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.DataFrame({
    'Возраст': np.random.normal(35, 10, 100).round(),
    'Доход': np.random.lognormal(4, 0.5, 100).round(2),
    'Количество покупок': np.random.randint(1, 10, 100),
    'Пол': np.random.choice(['М', 'Ж'], 100, p=[0.4, 0.6]),
    'Удовлетворенность': np.random.choice(['Низкая', 'Средняя', 'Высокая'], 100, p=[0.2, 0.5, 0.3])
})

data.loc[data.sample(frac=0.1).index, 'Количество покупок'] = np.nan
data.loc[data.sample(frac=0.15).index, 'Удовлетворенность'] = np.nan

print("Первые 5 строк датафрейма:")
print(data.head())

print("\nИнформация о датафрейме:")
print(data.info())
```

```
print("\nОписательная статистика числовых данных:")
print(data.describe())

print("\nОписательная статистика категориальных данных:")
print(data.describe(include=['object']))

print("\nКоличество пропущенных значений в каждом столбце:")
print(data.isnull().sum())

print("\nКоличество полных дубликатов строк:")
print(data.duplicated().sum())

print("\nАнализ выбросов (количественные показатели):")
numeric_cols = data.select_dtypes(include=['int64', 'float64']).columns
for col in numeric_cols:
   q1 = data[col].quantile(0.25)
   q3 = data[col].quantile(0.75)
   iqr = q3 - q1
   lower_bound = q1 - 1.5 * iqr
   upper_bound = q3 + 1.5 * iqr
   outliers = data[(data[col] < lower_bound) | (data[col] > upper_bound)][col]
   print(f"{col}: {len(outliers)} выбросов")

print("\nЛюди с доходом выше среднего:")
high_income = data[data['Доход'] > data['Доход'].mean()]
print(high_income.head())

print("\nЖенщины старше 30 лет с высоким уровнем удовлетворенности:")
filtered_data = data[(data['Пол'] == 'Ж') &
                     (data['Возраст'] > 30) &
                     (data['Удовлетворенность'] == 'Высокая')]
print(filtered_data.head())

print("\nЛюди с количеством покупок больше 5 и доходом в интервале [50, 100]:")
query_result = data.query('`Количество покупок` > 5 and 50 <= Доход <= 100')
print(query_result.head())

print("\nСредний доход по полу и уровню удовлетворенности:")
grouped_data = data.groupby(['Пол', 'Удовлетворенность'])['Доход'].mean().unstack()
print(grouped_data)
# Настройка стиля графиков
sns.set(style="whitegrid")

plt.figure(figsize=(10, 6))
sns.boxplot(x='Пол', y='Возраст', data=data)
plt.title('Распределение возраста по полу')
plt.show()

plt.figure(figsize=(10, 6))
sns.violinplot(x='Удовлетворенность', y='Доход', data=data)
plt.title('Распределение дохода по уровню удовлетворенности')
plt.show()
```

Практическая работа 6

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
import statsmodels.api as sm

numeric_cols = data.select_dtypes(include=['number']).columns
cat_cols = data.select_dtypes(include=['object', 'category']).columns

if len(numeric_cols) > 0:
    print("\nГИСТОГРАММЫ ЧИСЛОВЫХ ПЕРЕМЕННЫХ:")
    data[numeric_cols].hist(bins=15, figsize=(15, 10))
    plt.tight_layout()
    plt.show()

    print("\nБОКСПЛОТЫ ЧИСЛОВЫХ ПЕРЕМЕННЫХ:")
    plt.figure(figsize=(15, 5))
    for i, col in enumerate(numeric_cols, 1):
        plt.subplot(1, len(numeric_cols), i)
        sns.boxplot(y=data[col])
        plt.title(col)
    plt.tight_layout()
    plt.show()
numeric_data = data.select_dtypes(include=['number'])

if len(numeric_data.columns) > 1:
    print("\n" + "="*50)
    print("АНАЛИЗ МУЛЬТИКОЛЛИНЕАРНОСТИ")
    print("="*50)

    print("\nМАТРИЦА КОРРЕЛЯЦИЙ:")
    corr = numeric_data.corr()
    sns.heatmap(corr, annot=True, fmt=".2f", cmap='coolwarm', center=0)
    plt.show()

    threshold = 0.7  # Пороговое значение корреляции
    strong_correlations = set()  # Множество для хранения пар переменных

    print(f"\nПАРЫ ПЕРЕМЕННЫХ С КОРРЕЛЯЦИЕЙ > {threshold}:")
    for i in range(len(corr.columns)):
        for j in range(i):
            if abs(corr.iloc[i, j]) > threshold:
                col1 = corr.columns[i]
                col2 = corr.columns[j]
                strong_correlations.add((col1, col2))
                print(f"{col1} и {col2}: {corr.iloc[i, j]:.2f}")
```

Тест 1: Анализ данных, большие данные, направления Data Science, системы бизнес-аналитики

1. Какой из перечисленных этапов НЕ входит в стандартный процесс анализа данных (CRISP-DM)?
   a) Понимание бизнес-задачи
   b) Сбор данных
   c) Удаление всех исходных данных
   d) Построение моделей
   Правильный ответ: c) Удаление всех исходных данных
2. Какая из характеристик НЕ относится к "3V" больших данных?
   a) Volume (объем)
   b) Velocity (скорость)
   c) Variety (разнообразие)
   d) Validity (валидность)
   Правильный ответ: d) Validity (валидность)
3. Какое направление Data Science занимается прогнозированием числовых значений?
   a) Классификация
   b) Кластеризация
   c) Регрессионный анализ
   d) Ассоциативные правила
   Правильный ответ: c) Регрессионный анализ
4. Какой инструмент НЕ является системой бизнес-аналитики (BI)?
   a) Power BI
   b) Tableau
   c) Qlik Sense
   d) Apache Kafka
   Правильный ответ: d) Apache Kafka
5. Какой алгоритм машинного обучения относится к обучению без учителя?
   a) Линейная регрессия
   b) Метод k-ближайших соседей
   c) Метод k-средних

d) Дерево решений

Правильный ответ: c) Метод k-средних

6. Какая технология используется для потоковой обработки данных?

   a) Apache Hadoop

   b) Apache Spark

   c) Apache Kafka

   d) Microsoft Excel

   Правильный ответ: c) Apache Kafka

7. Что означает термин "ETL" в контексте анализа данных?

   a) Extract, Transform, Load

   b) Encrypt, Transfer, Lock

   c) Evaluate, Test, Learn

   d) Export, Tag, Label

   Правильный ответ: a) Extract, Transform, Load

8. Какой показатель используется для оценки качества классификации?

   a) Коэффициент детерминации (R^2)

   b) Среднеквадратичная ошибка (MSE)

   c) Матрица ошибок

   d) Дисперсия

   Правильный ответ: c) Матрица ошибок

9. Какой язык программирования чаще всего используется в Data Science?

   a) Java

   b) Python

   c) C++

   d) PHP

   Правильный ответ: b) Python

10. Какой метод используется для снижения размерности данных?

    a) Линейная регрессия

    b) Метод главных компонент (PCA)

    c) Дерево решений

    d) Логистическая регрессия

    Правильный ответ: b) Метод главных компонент (PCA)

Тест 2: Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных

1. Какой тип индекса ускоряет поиск в столбцах с текстовыми данными?
   a) B-дерево
   b) Хэш-индекс
   c) Обратный индекс
   d) R-дерево
   Правильный ответ: c) Обратный индекс
2. Какой формат данных оптимален для хранения вложенных структур?
   a) CSV
   b) JSON
   c) XML
   d) Parquet
   Правильный ответ: b) JSON
3. Какой язык используется для запросов в ClickHouse?
   a) SQL
   b) NoSQL
   c) GraphQL
   d) Python
   Правильный ответ: a) SQL
4. Какой инструмент используется для оркестрации ETL-процессов?
   a) Apache NiFi
   b) Apache Kafka
   c) Apache Spark
   d) Apache Hadoop
   Правильный ответ: a) Apache NiFi
5. Какой протокол используется для передачи данных между микросервисами?
   a) HTTP
   b) gRPC
   c) FTP

d) SMTP

Правильный ответ: b) gRPC

6. Какой тип базы данных используется для хранения графов?

   a) MongoDB

   b) Neo4j

   c) Redis

   d) Cassandra

   Правильный ответ: b) Neo4j

7. Какой инструмент используется для мониторинга данных?

   a) Grafana

   b) Tableau

   c) Power BI

   d) Excel

   Правильный ответ: a) Grafana

8. Какой формат данных используется в Apache Kafka?

   a) CSV

   b) JSON

   c) Avro

   d) XML

   Правильный ответ: c) Avro

9. Какой инструмент используется для управления метаданными?

   a) Apache Atlas

   b) Apache Spark

   c) Apache Hadoop

   d) Apache Kafka

   Правильный ответ: a) Apache Atlas

10. Какой тип хранилища используется для аналитических запросов?

    a) OLTP

    b) OLAP

    c) Key-Value

    d) Document

    Правильный ответ: b) OLAP

Тест 3: Российские технологии в области данных

1. Какой российский фреймворк для ML разработан Яндексом?
   a) CatBoost
   b) TensorFlow
   c) PyTorch
   d) Scikit-learn
   Правильный ответ: a) CatBoost
2. Какой российский сервис предоставляет аналитику в реальном времени?
   a) Яндекс.Метрика
   b) Google Analytics
   c) Adobe Analytics
   d) Mixpanel
   Правильный ответ: a) Яндекс.Метрика
3. Какой российский продукт является аналогом Snowflake?
   a) Яндекс.Облако Data Platform
   b) 1С:Предприятие
   c) Тинькофф Data Warehouse
   d) Сбер.Аналитика
   Правильный ответ: a) Яндекс.Облако Data Platform
4. Какой российский инструмент используется для управления данными?
   a) DataLens
   b) Tableau
   c) Power BI
   d) QlikView
   Правильный ответ: a) DataLens
5. Какой российский сервис предоставляет NLP API?
   a) Яндекс.Облако SpeechKit
   b) Google Cloud NLP
   c) AWS Comprehend
   d) IBM Watson
   Правильный ответ: a) Яндекс.Облако SpeechKit
6. Какой российский продукт используется для обработки потоковых данных?
   a) Яндекс.Потоки

b) Apache Kafka

c) Apache Flink

d) Apache Spark

Правильный ответ: a) Яндекс.Потоки

7. Какой российский инструмент для визуализации геоданных?

   a) Яндекс.Карты API

   b) Google Maps API

   c) Mapbox

   d) OpenLayers

   Правильный ответ: a) Яндекс.Карты API

8. Какой российский фреймворк для глубокого обучения?

   a) DeepPavlov

   b) TensorFlow

   c) PyTorch

   d) Keras

   Правильный ответ: a) DeepPavlov

9. Какой российский сервис предоставляет аналитику для бизнеса?

   a) Яндекс.Метрика

   b) Google Analytics 360

   c) Adobe Analytics

   d) Amplitude

   Правильный ответ: a) Яндекс.Метрика

10. Какой российский продукт используется для хранения и обработки больших данных?

    a) ClickHouse

    b) Apache Hadoop

    c) Apache Spark

    d) Elasticsearch

    Правильный ответ: a) ClickHouse

Тест 4: Анализ данных, большие данные, направления Data Science, системы бизнес-аналитики

1. Какой метод анализа данных используется для выявления скрытых закономерностей в больших массивах информации?
   a) Описательная статистика
   b) Data Mining
   c) Линейная регрессия
   d) Визуализация
   Правильный ответ: b) Data Mining
2. Какая характеристика Big Data описывает разнообразие форматов данных?
   a) Volume
   b) Velocity
   c) Variety
   d) Veracity
   Правильный ответ: c) Variety
3. Какой тип машинного обучения использует размеченные данные для обучения?
   a) Обучение с учителем
   b) Обучение без учителя
   c) Смешанное обучение
   d) Глубокое обучение
   Правильный ответ: a) Обучение с учителем
4. Какой инструмент позволяет создавать интерактивные дашборды без написания кода?
   a) Jupyter Notebook
   b) Tableau
   c) Apache Spark
   d) TensorFlow
   Правильный ответ: b) Tableau
5. Какой алгоритм используется для разделения данных на группы по схожести?
   a) Линейная регрессия
   b) Метод k-средних
   c) Дерево решений

d) SVM

Правильный ответ: b) Метод k-средних

6. Какой процесс преобразует сырые данные в пригодный для анализа формат?

   a) Data Cleaning

   b) Data Aggregation

   c) Data Wrangling

   d) Data Visualization

   Правильный ответ: c) Data Wrangling

7. Какой показатель оценивает точность регрессионной модели?

   a) Accuracy

   b) F1-score

   c) R-квадрат

   d) Precision

   Правильный ответ: c) R-квадрат

8. Какая библиотека Python чаще всего используется для работы с табличными данными?

   a) NumPy

   b) Pandas

   c) Matplotlib

   d) Scikit-learn

   Правильный ответ: b) Pandas

9. Какой метод НЕ используется для обработки пропущенных значений?

   a) Удаление строк с пропусками

   b) Замена средним значением

   c) Замена нулями

   d) Шифрование данных

   Правильный ответ: d) Шифрование данных

10. Какой инструмент используется для управления workflow в Data Science проектах?

    a) Apache Airflow

    b) Microsoft Word

    c) Adobe Photoshop

d) WinRAR

Правильный ответ: a) Apache Airflow

Тест 5: Большие данные и системы хранения

1. Какой принцип лежит в основе технологии блокчейн?
   a) Репликация данных
   b) Децентрализованное хранение
   c) Распределенный реестр
   d) Все вышеперечисленное
   Правильный ответ: d) Все вышеперечисленное
2. Что такое "data lake"?
   a) Хранилище неструктурированных данных
   b) Реляционная база данных
   c) Система визуализации
   d) Инструмент ETL
   Правильный ответ: a) Хранилище неструктурированных данных
3. Какой инструмент используется для потоковой обработки данных в реальном времени?
   a) Apache Flink
   b) Apache Hadoop
   c) Apache Hive
   d) Apache Spark
   Правильный ответ: a) Apache Flink
4. Что такое "sharding" в базах данных?
   a) Горизонтальное разделение данных
   b) Вертикальное разделение данных
   c) Сжатие данных
   d) Шифрование данных
   Правильный ответ: a) Горизонтальное разделение данных
5. Какой тип базы данных оптимален для хранения временных рядов?
   a) InfluxDB

b) MongoDB

c) PostgreSQL

d) Redis

Правильный ответ: a) InfluxDB

6. Что такое "CAP-теорема"?

   a) Теорема о согласованности, доступности и устойчивости к разделению

   b) Теорема о скорости обработки данных

   c) Теорема о безопасности данных

   d) Теорема о масштабируемости

   Правильный ответ: a) Теорема о согласованности, доступности и устойчивости к разделению

7. Какой формат данных обеспечивает схему для структурированного хранения?

   a) CSV

   b) JSON

   c) Parquet

   d) XML

   Правильный ответ: c) Parquet

8. Что такое "lambda-архитектура"?

   a) Подход к обработке больших данных, сочетающий batch и stream processing

   b) Архитектура микросервисов

   c) Модель глубокого обучения

   d) Способ хранения данных

   Правильный ответ: a) Подход к обработке больших данных, сочетающий batch и stream processing

9. Какой инструмент используется для управления workflow в data pipeline?

   a) Apache Airflow

   b) Apache Kafka

   c) Apache Spark

   d) Apache Hadoop

   Правильный ответ: a) Apache Airflow

10. Что такое "polyglot persistence"?
    a) Использование разных СУБД для разных типов данных
    b) Хранение данных в одном формате
    c) Метод сжатия данных
    d) Способ репликации данных
    Правильный ответ: a) Использование разных СУБД для разных типов данных

Тест 6: Системы хранения и визуализации данных. Российский сегмент рынка

1. Какой российский аналог Tableau существует на рынке?
    a) Яндекс.Метрика
    b) DataLens (от Яндекса)
    c) 1С:Предприятие
    d) Тинькофф Аналитика
    Правильный ответ: b) DataLens (от Яндекса)
2. Какая российская СУБД популярна для аналитики больших данных?
    a) ClickHouse
    b) Oracle
    c) MySQL
    d) PostgreSQL
    Правильный ответ: a) ClickHouse
3. Какой инструмент визуализации разработан компанией "Точка зрения"?
    a) Power BI
    b) Tableau
    c) DataLens
    d) Qlik Sense
    Правильный ответ: c) DataLens
4. Какая российская платформа предоставляет облачные решения для хранения данных?
    a) Яндекс.Облако
    b) Google Drive
    c) Dropbox

d) iCloud

Правильный ответ: a) Яндекс.Облако

5. Какой российский сервис предоставляет аналитику веб-трафика?

   a) Яндекс.Метрика

   b) Google Analytics

   c) Adobe Analytics

   d) Facebook Insights

   Правильный ответ: a) Яндекс.Метрика

6. Какая российская компания разрабатывает решения для обработки больших данных?

   a) СберТех

   b) Microsoft

   c) IBM

   d) Oracle

   Правильный ответ: a) СберТех

7. Какой формат визуализации лучше всего подходит для временных рядов?

   a) Круговая диаграмма

   b) Линейный график

   c) Диаграмма рассеяния

   d) Гистограмма

   Правильный ответ: b) Линейный график

8. Какая российская платформа предоставляет инструменты для машинного обучения?

   a) TensorFlow

   b) PyTorch

   c) CatBoost (разработан Яндексом)

   d) Scikit-learn

   Правильный ответ: c) CatBoost (разработан Яндексом)

9. Какой инструмент НЕ является системой хранения данных?

   a) Hadoop HDFS

   b) Amazon S3

   c) Яндекс.Облако

d) Microsoft PowerPoint

Правильный ответ: d) Microsoft PowerPoint

10. Какой российский сервис предоставляет API для геоаналитики?

    a) Яндекс.Карты

    b) Google Maps

    c) Apple Maps

    d) OpenStreetMap

    Правильный ответ: a) Яндекс.Карты