

Ошибка SQL:!error!

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)
40.04.01 Юриспруденция. Теория и практика правоприменения

Год набора на ОПОП
2025

Форма обучения
заочная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции и	Код и формулировка индикатора достижения компетенции
40.04.01 «Юриспруденция» (М-ЮП)	ОПК-7 : Способен применять информационные технологии и использовать правовые базы данных для решения задач профессиональной деятельности с учетом требований информационной безопасности	ОПК-7.3в : Применяет программные инструментальные 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-7 «Способен применять информационные технологии и использовать правовые базы данных для решения задач профессиональной деятельности с учетом требований информационной безопасности»

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код	Тип	Результат	
ОПК-7.3в : Применяет программные инструментальные средства для решения стандартных задач профессиональной деятельности	РД1	Знание	современных направлений развития компьютерных технологий	понимание разницы направлений развития компьютерных технологий
	РД1	Умение	применять современные направления компьютерных технологий для хранения, обработки и представления данных	понимание основных методов работы с данными
	РД1	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
Заочная форма обучения				
РД1	Знание : современных направлений развития компьютерных технологий	1.2. Анализ данных, большие данные, направления Data science, системы бизнес-аналитики	Тест	Тест
		1.9. Методы кластеризации (сегментация)	Тест	Тест
		1.10. Методы классификации	Тест	Тест
РД1	Умение : применять современные направления компьютерных технологий для хранения, обработки и представления данных	1.1. Компьютерные технологии как инструмент хранения, обработки, анализа и представления данных и информации.	Тест	Тест
РД1	Навык : работы с инструментами автоматизированной обработки информации	1.11. Системы класса low-code, no-code	Практическая работа	Практическая работа
РД2	Знание : основных концепций Data Science и бизнес-аналитики	1.3. Системы хранения и визуализации данных. Российский сегмент рынка	Тест	Тест
		1.11. Системы класса low-code, no-code	Тест	Тест
РД2	Умение : анализировать большие данные	1.8. Массивы данных (датафреймы)	Практическая работа	Практическая работа
		1.9. Методы кластеризации (сегментация)	Практическая работа	Практическая работа
		1.10. Методы классификации	Практическая работа	Практическая работа
РД2	Навык : использования информационных технологий для решения практических задач.	1.5. Числовые и нечисловые данные (выборки, векторы)	Практическая работа	Практическая работа
РД3	Знание : форматов хранения данных (xlsx, csv, txt)	1.4. Подготовка данных для анализа и визуализации. Начальные навыки. Табличные процессоры, форматы хранения данных	Практическая работа	Практическая работа
РД3	Умение : подготавливать данные для анализа	1.7. Нечисловые данные	Практическая работа	Практическая работа
РД3	Навык : очистки и преобразования данных в табличных процессорах (Excel, Calc)	1.6. Числовые характеристики выборок	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» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

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 (от Яндекса)
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) 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) СберTex
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	