

Документ подписан электронной подписью
 Информация о владельце:
 ФИО: Косенок Сергей Михайлович
 Должность: ректор
 Дата подписания: 16.06.2026 12:46:50
 Уникальный идентификатор:
 e3a68f3eaa1a62674b54f4998099d3d6bfdcf836

Тестовое задание для диагностического тестирования по дисциплине:

Анализ больших данных в финансах, 1 и 2 семестры

Код направления подготовки	38.04.08 Финансы и кредит
Направленность (профиль)	Финансовые технологии
Форма обучения	очная
Кафедра-разработчик	Экономики, учета и финансов
Выпускающая кафедра	Экономики, учета и финансов

1 семестр

Проверяемая компетенция	Задание	Варианты ответов	Тип сложности вопроса
ОПК-2.1	Что понимается под «четырьмя V» (4V) больших данных в финансовом контексте?	A) Value, Venture, Vote, Vision B) Volume, Velocity, Variety, Veracity C) Vector, Virtual, Voice, Variable D) Valid, Vague, Vast, Vital	низкий
ОПК-2.1	Какой метод машинного обучения чаще всего применяется для обнаружения мошеннических транзакций в реальном времени?	A) Линейная регрессия B) Метод главных компонент (PCA) C) Градиентный бустинг D) Простое скользящее среднее	низкий
ОПК-2.1	Технология распределенных вычислений, используемая для пакетной обработки петабайт финансовых данных, включает экосистему Apache ...		низкий
ОПК-2.2	Для анализа связей между счетами, транзакциями и контрагентами при противодействии		низкий

	отмыванию денег наиболее эффективны ... базы данных		
ОПК-2.2	Для оценки кредитного риска физического лица на основе больших данных (данные о покупках, платежах, геолокации) чаще всего используется:	<p>A) ARIMA-модель</p> <p>B) Альтернативные кредитные скоринговые модели на основе графовых нейросетей</p> <p>C) Классическая модель Альтмана (Z-score)</p> <p>D) Модель Блэка-Шоулза</p>	низкий
ОПК-2.3	Какую базу данных оптимально использовать для хранения и анализа потоков биржевых сделок в реальном времени (высокая скорость записи и мгновенные запросы)?	<p>A) MySQL</p> <p>B) PostgreSQL</p> <p>C) InfluxDB</p> <p>D) MongoDB</p>	средний
ОПК-2.1	Какие источники данных относят к «альтернативным данным» (alternative data) в финансовом анализе?	<p>A) Данные о спутниковых снимках парковок розничных сетей</p> <p>B) Котировки акций с биржи</p> <p>C) Данные геолокации мобильных устройств</p> <p>D) Годовая финансовая отчетность по МСФО</p> <p>E) Ключевая ставка ЦБ</p>	средний
ОПК-2.1	Установите соответствие между инструментом обработки данных и его основной функцией в финансовой аналитике	<p>Инструмент обработки данных:</p> <p>A) Apache Kafka</p> <p>B) Apache Flink</p> <p>C) Apache Hadoop</p> <p>D) ClickHouse</p> <p>Основная функция:</p> <p>1) Распределённое хранение больших данных (HDFS)</p> <p>2) Очередь сообщений для потоковых финансовых событий</p> <p>3) Поточковая обработка с низкой задержкой (миллисекунды)</p> <p>4) Колоночная база данных</p>	средний

		для быстрых агрегаций по котировкам	
ОПК-2.3	Установите соответствие между типом базы данных и областью её применением в финансах	<p>Тип базы данных:</p> <p>A) Графовая база данных (Neo4j)</p> <p>B) Time-series БД (InfluxDB)</p> <p>C) Колоночная БД (ClickHouse)</p> <p>D) Key-value БД (Redis)</p> <p>Область применения:</p> <p>1) Хранение биржевых тиков с временной меткой</p> <p>2) Анализ цепочек транзакций для AML</p> <p>3) Кэширование часто запрашиваемых курсов валют</p> <p>4) Агрегация миллионов транзакций для отчётности</p>	средний
ОПК-2.2	Какой алгоритм кластеризации часто применяется для сегментации клиентов банка по транзакционным данным (Big Data)?	<p>A) K-means</p> <p>B) Логистическая регрессия</p> <p>C) LSTM</p> <p>D) Дерево решений с одним узлом</p>	средний
ОПК-2.1	Что из перечисленного является примером неструктурированных данных в финансах?	<p>A) Строка в таблице с суммой перевода</p> <p>B) Дата и время транзакции</p> <p>C) Текст новостного сообщения о компании</p> <p>D) Номер счета отправителя</p>	средний
ОПК-2.2	Метод машинного обучения, основанный на ансамбле деревьев решений и часто применяемый для обнаружения фрода в реальном времени, называется градиентный ...		средний
ОПК-2.3	Для потоковой обработки биржевых тиков в реальном времени часто используется		средний

	платформа Apache ...		
ОПК-2.1	При использовании градиентного бустинга (XGBoost) для обнаружения фрода типичное минимальное количество деревьев (n_estimators) в базовой модели составляет ...		средний
ОПК-2.1	Доля сохраняемой дисперсии при использовании метода главных компонент (PCA) для финансовых данных обычно составляет ... %		средний
ОПК-2.2	Упорядочьте этапы типового пайплайна обработки больших финансовых данных в хронологическом порядке	<ol style="list-style-type: none"> 1. Визуализация результатов (дашборд) 2. Приём потоковых данных из Kafka 3. Очистка и валидация транзакций 4. Хранение в колоночной БД (ClickHouse) 	высокий
ОПК-2.3	Упорядочьте шаги по построению модели градиентного бустинга (XGBoost) для обнаружения фрода в реальном времени	<ol style="list-style-type: none"> 1. Обучение модели на исторических транзакциях 2. Сбор и разметка транзакций (фрод / не фрод) 3. Применение модели к новому потоку транзакций 4. Отбор признаков и обработка дисбаланса классов 	высокий
ОПК-2.3	Какие технологии распределенной обработки данных подходят для потокового анализа финансовых транзакций?	<ol style="list-style-type: none"> A) Apache Flink B) Apache Hadoop (MapReduce) C) Apache Kafka Streams D) ClickHouse E) Apache Oozie 	высокий
ОПК-2.2	Какие методы снижения размерности применяются при анализе больших	<ol style="list-style-type: none"> A) Метод главных компонент (PCA) B) T-SNE C) Удаление строк с 	высокий

	финансовых данных?	пропусками D) Автоэнкодеры (нейросетевые) E) Нормировка признаков	
ОПК-2.2	Какие задачи решает анализ больших данных в управлении рисками банка?	A) Стресс-тестирование на исторических данных B) Расчет VAR (Value at Risk) на распределенных вычислениях C) Автоматическая генерация текста договора D) Мониторинг санкционных списков в реальном времени E) Создание 3D-графиков	высокий

2 семестр

Проверяемая компетенция	Задание	Варианты ответов	Тип сложности вопроса
ОПК-2.1	Какая технология распределенных вычислений является стандартом де-факто для обработки петабайт финансовых данных в пакетном режиме?	A) Apache Kafka B) Apache Hadoop (MapReduce) / Spark C) Redis D) SQLite	низкий
ОПК-2.2	При анализе больших данных для построения «стресс-тестов» банка (например, при резком изменении курсов валют) используется:	A) Модель Монте-Карло на распределенных вычислительных кластерах B) Только линейная интерполяция C) Визуализация в Power BI без моделирования D) Метод Фибоначчи	низкий
ОПК-2.1	Метод анонимизации клиентских данных, добавляющий контролируемый шум в ответы на запросы, называется ... приватность		низкий
ОПК-2.1	Метод обнаружения аномалий в		низкий

	финансовых транзакциях, который изолирует выбросы через случайные разбиения пространства признаков, называется изолирующий ...		
ОПК-2.1	Что означает понятие «латеральное чтение данных» при аудите финансовых больших данных?	<p>A) Шифрование данных</p> <p>B) Отслеживание происхождения и трансформации данных от источника до отчета</p> <p>C) Удаление дубликатов</p> <p>D) Копирование данных на ленту</p>	низкий
ОПК-2.2	Какой инструмент экосистемы Apache чаще всего используют для потоковой обработки финансовых событий (например, обнаружение аномалий в чековых транзакциях в реальном времени)?	<p>A) Apache Hive</p> <p>B) Apache Flink или Kafka Streams</p> <p>C) Apache Pig</p> <p>D) Apache Oozie</p>	средний
ОПК-2.3	Какие проблемы решает графовый анализ данных в финансах?	<p>A) Поиск цепочек транзакций между счетами</p> <p>B) Выявление связанных сторон в кредитном скоринге</p> <p>C) Хранение больших таблиц с временными рядами</p> <p>D) Обнаружение колец поручительства в банковских рисках</p> <p>E) Замена реляционных баз данных для бухгалтерии</p>	средний
ОПК-2,2	Установите соответствие между этапом пайплайна больших данных и его действием в финансах	<p>Этап:</p> <p>A) Ingestion</p> <p>B) Storage</p> <p>C) Processing</p> <p>D) Visualization</p> <p>Действие:</p> <p>1) Удаление дубликатов</p>	средний

		<p>транзакций и исправление валют</p> <p>2) Визуализация тепловой карты рисков по отделениям банка</p> <p>3) Приём потока биржевых тиков через Kafka</p> <p>4) Запись очищенных данных в ClickHouse</p>	
ОПК-2.2	<p>Установите соответствие между методом снижения размерности и его особенностью для финансовых данных</p>	<p>Метод:</p> <p>A) PCA</p> <p>B) t-SNE</p> <p>C) Автоэнкодер</p> <p>D) Отбор признаков по важности (XGBoost)</p> <p>Особенность:</p> <p>1) Линейное преобразование, сохраняющее максимальную дисперсию</p> <p>2) Нелинейное отображение, хорошо для визуализации кластеров клиентов</p> <p>3) Нейросетевой метод, может обучаться на размеченных транзакциях</p> <p>4) Использует встроенную важность для кредитного скоринга</p>	средний
ОПК-2.1	<p>В чем основное преимущество использования графовых баз данных (Neo4j, Amazon Neptune) в финансовом анализе больших данных?</p>	<p>A) Высокая скорость агрегации больших чисел</p> <p>B) Эффективный анализ связей между счетами, транзакциями и контрагентами</p> <p>C) Хранение исторических котировок по времени</p> <p>D) Замена Excel</p>	средний
ОПК-2.3	<p>Какая метрика в первую очередь используется для оценки модели бинарной классификации (например, мошенник / не мошенник) в</p>	<p>A) Accuracy</p> <p>B) Precision, Recall и F1-score</p> <p>C) Mean Absolute Error</p> <p>D) R-squared</p>	средний

	условиях сильного дисбаланса классов (99% нормальных транзакций, 1% фродовых)?		
ОПК-2.2	Технология распределенной потоковой обработки, часто используемая вместе с Kafka для построения пайплайнов финансовых событий, называется Kafka ...		средний
ОПК-2.2	Анализ тональности новостей и сообщений в соцсетях для прогнозирования движения акций называется ... анализом		средний
ОПК-2.3	Для предотвращения переобучения LSTM при прогнозировании доходности акций рекомендуемое минимальное соотношение числа обучающих примеров к числу параметров сети составляет ... к одному		средний
ОПК-2.3	В задаче кредитного скоринга на больших данных стандартный порог отсечения по вероятности дефолта составляет ... %		средний
ОПК-2.3	Упорядочьте действия при применении метода главных компонент (РСА) к многомерным финансовым временным рядам	<ol style="list-style-type: none"> 1. Проекция данных на выбранные главные компоненты 2. Вычисление ковариационной матрицы 3. Выбор топ-k компонент по объяснённой дисперсии 4. Расчёт собственных векторов и собственных чисел 	высокий

ОПК-2.2	Упорядочьте шаги расчёта Value at Risk (VaR) методом исторической симуляции на больших данных	<ol style="list-style-type: none"> 1. Расчёт квантиля распределения убытков (например, 5%) 2. Расчёт дневных изменений стоимости портфеля по историческим данным (5 лет) 3. Сортировка полученных изменений по убыванию 4. Загрузка исторических котировок всех активов в распределённую таблицу 	высокий
ОПК-2.2	Какие методы машинного обучения подходят для прогнозирования временных рядов в финансах (например, доходность акций) на больших данных?	<ol style="list-style-type: none"> A) LSTM B) GRU C) ARIMA D) Prophet E) Линейная регрессия с одним признаком 	высокий
ОПК-2.1	Какие подходы используются для обеспечения конфиденциальности клиентских данных при анализе больших данных в финансах?	<ol style="list-style-type: none"> A) Анонимизация B) Псевдонимизация C) Дифференциальная приватность D) Хранение данных в открытом виде E) Отказ от анализа данных 	высокий
ОПК-2.1	Какие утверждения верны для использования градиентного бустинга (CatBoost, XGBoost) в финансовом скоринге?	<ol style="list-style-type: none"> A) Хорошо работает с пропусками в данных B) Поддерживает категориальные признаки C) Требуется масштабирования признаков D) Эффективен на табличных больших данных E) Работает только на GPU 	высокий