

Документ подписан электронной подписью
Информация о владельце:
ФИО: Косенок Сергей Михайлович
Должность: ректор
Дата подписания: 16.06.2026 12:46:43
Уникальный идентификатор файла
e3a68f3eaa1a62674b54f4998099d3d6bfdcf836

Оценочные материалы для промежуточной аттестации по дисциплине

Анализ больших данных в финансах

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

Типовые задания для контрольной работы:

Написание эссе по выбранной теме

Общие требования к содержанию эссе:

- ✓ Эссе раскрывает выбранную тему на основе теоретических положений дисциплины и актуальных примеров из финансовой практики.
- ✓ Работа содержит авторскую позицию, аргументированную ссылками на источники (научные статьи, отраслевые отчеты, кейсы финансовых организаций).
- ✓ В эссе отсутствует подмена анализа простым пересказом учебных материалов или популярных статей.
- ✓ Все используемые термины и понятия соответствуют общепринятым определениям в области анализа больших данных и финансов.
- ✓ Эссе не содержит неподтвержденных обобщений и утверждений без указания контекста или источников.

Структура эссе:

Введение (10–15% от объема)

- ✓ Обоснование актуальности выбранной темы для финансовой сферы.
- ✓ Формулировка основного тезиса (ключевой идеи), которую автор будет доказывать.
- ✓ Краткое обозначение логики рассуждения (что будет рассмотрено).

Основная часть (70–80% от объема)

- ✓ Последовательное раскрытие темы с разбивкой на логические блоки (подпункты или абзацы).
- ✓ Использование не менее 3–5 академических или профессиональных источников (статьи из рецензируемых журналов, материалы конференций, технические отчеты крупных финансовых организаций, официальная документация, книги).
- ✓ Приведение конкретных примеров из практики финансовых институтов (банки, страховые компании, финтех-сервисы, биржи).
- ✓ Сравнение различных точек зрения или методов (если применимо к теме).
- ✓ Аргументация авторской позиции с опорой на данные и логические рассуждения.

Заключение (10–15% от объема)

- ✓ Краткое резюме основных выводов.

- ✓ Подтверждение тезиса, заявленного во введении.
- ✓ Указание на возможные ограничения рассмотренных подходов или направления для дальнейших исследований.

Темы для эссе:

1. Трансформация традиционных банковских процессов под влиянием технологий больших данных.
2. Сравнительный анализ архитектур Lambda и Kappa для потоковой обработки финансовых транзакций.
3. Роль неструктурированных данных (тексты новостей, соцсети, аудиозвонки) в кредитном скоринге.
4. Этические дилеммы использования персональных финансовых данных клиентов в системах Big Data.
5. Применение графовых баз данных для выявления мошеннических схем в платежных сетях.
6. Методы обнаружения аномалий в высокочастотных потоках биржевых данных без учителя.
7. Интеграция альтернативных данных (спутниковые снимки, геолокация, телеметрия) в стресс-тестирование портфелей.
8. Проблема «проклятия размерности» при анализе доходности тысяч финансовых инструментов и способы её преодоления.
9. Сравнение эффективности распределенных градиентных бустингов (XGBoost, LightGBM, CatBoost) в задачах антифрода.
10. Особенности построения систем раннего предупреждения финансовых кризисов на основе потоковых операционных данных.
11. Использование техник снижения размерности (PCA, t-SNE, UMAP) для визуализации риск-профилей заемщиков.
12. Роль feature stores в стандартизации и воспроизводимости признаков для моделей анализа финансовых данных.
13. Преимущества и ограничения онлайн-обучения (online learning) в адаптивных системах обнаружения мошенничества.
14. Построение платежного графа: методы поиска транзитных счетов, колец и вееров в межбанковских переводах.
15. Внедрение глубоких нейронных сетей (LSTM, GRU) для прогнозирования оттоков ликвидности в реальном времени.
16. Сопоставление традиционного эконометрического подхода (ARIMA) и методов Big Data (Spark MLlib) для макрофинансового прогнозирования.
17. Регуляторные требования (GDPR, 152-ФЗ, Basel IV) как ограничители и драйверы развития Big Data в финансах.
18. Архитектура дата-лейк (data lake) для хранения сырых логов транзакций в финансовой организации.
19. Разработка скоринговых моделей для небанковских финансовых сервисов (P2P-кредитование, микрозаймы) на основе цифрового следа.
20. Количественная оценка экономического эффекта от внедрения Big Data в систему комплаенс и ПОД/ФТ (противодействие легализации доходов).

Типовые вопросы к зачету:

1. Определение больших данных (Big Data) и их ключевые характеристики: объем (Volume), скорость (Velocity), разнообразие (Variety), достоверность (Veracity) и ценность (Value).
2. Основные отличия анализа больших данных от классического эконометрического анализа в финансах.

3. Этапы жизненного цикла больших данных в финансовой организации: сбор, хранение, обработка, анализ, визуализация.
4. Роль распределенных вычислительных систем (Hadoop, Spark) в обработке финансовых логов и транзакционных данных.
5. Понятие потоковых данных (streaming data) в финансах: tick-данные, котировки в реальном времени, потоки транзакций.
6. NoSQL-базы данных в финансах: документоориентированные (MongoDB), колоночные (Cassandra) и графовые (Neo4j) для управления рисками и мошенничеством.
7. Проблема неструктурированных данных в финансах: тексты новостей, отчеты эмитентов, посты в соцсетях, аудиозаписи колл-центров.
8. Методы предобработки больших финансовых данных: очистка, нормализация, агрегация по временным окнам, обработка пропусков в высокочастотных рядах.
9. Понятие трех V (Volume, Velocity, Variety) как основа для выбора архитектуры хранения и анализа финансовых данных.
10. Выбросы (outliers) в больших массивах транзакций: причины возникновения и методы обнаружения (Z-оценка, MAD, изолирующий лес).
11. Принципы выборки из больших данных при построении финансовых моделей: стратифицированная, временная и случайная выборки.
12. Использование графовых баз данных для анализа связей между счетами, контрагентами и транзакциями (транзакционные графы).
13. Понятие и назначение data lake («озеро данных») в финансовой организации в сравнении с data warehouse.
14. Высокочастотные финансовые данные (HFTP): тики, стаканы заявок, временные интервалы в миллисекундах и микроструктура рынка.
15. Основные вызовы при работе с временными рядами сверхбольшой размерности (сотни тысяч финансовых инструментов).
16. Инструменты визуализации больших данных в финансах: дашборды с агрегацией в реальном времени (Grafana, Tableau, Power BI).
17. Партиционирование и шардирование финансовых данных как способы ускорения запросов к историческим транзакциям.
18. Использование облачных хранилищ (AWS S3, Google BigQuery, Snowflake) для масштабируемого анализа финансовых логов.
19. Этические и регуляторные ограничения при сборе и анализе персональных финансовых данных (GDPR, 152-ФЗ, PCI DSS).
20. Различие между пакетной обработкой (batch processing) и обработкой в реальном времени (stream processing) в финансовом мониторинге.

2 семестр

Типовые задания для контрольной работы:

Кейс-проект «ФинТех-дашборд»

Входные данные

Необработанный датасет финансового учреждения (транзакции, анкеты клиентов, логи активности, факт оттока).

Цель: Построить интерактивный дашборд, обучить модель прогнозирования оттока ($ROC-AUC \geq 0,75$) и предложить управленческие решения.

Этапы выполнения проекта:

1. ETL и предобработка

- Устранить аномалии (отрицательные суммы, возраст > 120 , выбросы по IQR/Z-оценке).

- Заполнить пропуски: числовые - медианой по сегменту, категориальные - модой/отдельной категорией.
 - Создать признаки: тенор, средний чек, дни без операций, динамика активности.
 - Результат: очищенный датасет `final_churn_dataset.csv`.
- 2. Интерактивный дашборд (Plotly / Tableau Public)**
- Динамика оттока по месяцам с фильтром по сегменту.
 - Сегментация: отток по возрасту, продукту, городу (столбцы/heatmap).
 - KPI: уровень оттока, число активных клиентов.
 - Фильтры: период, сегмент, возраст, тенор.
 - Результат: HTML-дашборд или публичная ссылка.
- 3. Модель прогнозирования оттока**
- Разделение 70/30 (с сохранением временного порядка).
 - Модели: Logistic Regression и Random Forest (минимум две).
 - Балансировка классов.
 - Оценка: ROC-AUC $\geq 0,75$, также Precision-Recall AUC, F1.
 - Анализ важности признаков.
 - Результат: обученная модель + код скоринга.
- 4. Управленческие решения**
- Каждое решение содержит:
- Какой признак модели его обосновывает.
 - Конкретное действие.
 - Ожидаемый эффект.
- 5. Защита проекта**
- Презентация.
 - Демонстрация кода.
 - Демонстрация дашборда в браузере.

Типовые вопросы к экзамену:

1. Методы машинного обучения для классификации транзакций при обнаружении мошенничества (fraud detection) в реальном времени.
2. Кластеризация клиентов финансовой организации на основе транзакционных данных (сегментация по паттернам расходов, риск-профилям).
3. Построение кредитного скоринга с использованием больших данных: альтернативные данные (платежи ЖКХ, телеком-данные, цифровой след).
4. Анализ социальных сетей для прогнозирования поведения финансовых рынков (sentiment analysis новостей и постов трейдеров).
5. Методы снижения размерности (PCA, t-SNE, UMAP) при анализе доходности тысяч финансовых активов.
6. Применение распределенных градиентных бустингов (XGBoost, LightGBM, CatBoost) к большим транзакционным массивам в задачах антифрода.
7. Обнаружение аномальных паттернов в потоках платежей с использованием методов без учителя (autoencoders, isolation forest).
8. Понятие и расчет Value-at-Risk (VaR) на основе моделирования больших массивов исторических доходностей (исторический метод).
9. Стресс-тестирование портфеля с использованием big data: сценарии на основе альтернативных данных (спутниковые снимки, погода, трафик).
10. Использование NLP (распознавание именованных сущностей, тематическое моделирование) для анализа финансовых отчетов (10-K, 10-Q).
11. Архитектура Lambda и Карра для обработки потоковых финансовых данных: преимущества и недостатки.
12. Обнаружение инсайдерской торговли с помощью анализа временных меток транзакций и сетевых связей.

13. Понятие «цифрового следа» клиента банка и его использование для прогнозирования оттока (churn prediction).
14. Регрессионные модели на больших данных для предсказания доходности облигаций с учетом макро- и микроструктурных факторов.
15. Оптимизация кредитного конвейера с использованием big data: автоматическая проверка благонадежности заемщика.
16. Использование Apache Kafka как шины для потоковой передачи транзакций в системы риск-мониторинга.
17. Проблема асимметрии классов в финансовых данных (например, редкое мошенничество) и методы борьбы (SMOTE, взвешивание классов).
18. Построение когортного анализа (cohort analysis) для оценки LTV клиента на основе больших массивов транзакций.
19. Методы онлайн-обучения (online learning) для адаптации моделей к меняющемуся поведению финансовых мошенников.
20. Анализ платежного графа: поиск колец, вееров и подозрительных транзитных структур (transit accounts).
21. Временные ряды высокой размерности в управлении ликвидностью банка: прогнозирование оттоков с использованием LSTM и GRU.
22. Сравнение эффективности традиционных ARIMA-моделей и глубоких нейронных сетей на больших финансовых временных рядах.
23. Понятие и применение feature stores в финансовом big data для управления признаками (стандартизация, версионирование, кэширование).
24. Применение технологии потоковой обработки (Apache Flink, Spark Streaming) для обнаружения мошеннических транзакций за миллисекунды.
25. Верификация и валидация моделей на больших данных: метод скользящего окна (rolling window) и расширяющегося окна (expanding window).
26. Использование графовых нейронных сетей (GNN) для выявления сложных схем финансовых махинаций в платежных сетях.
27. Количественная оценка риска контрагента на основе нетрадиционных источников: данные GPS, история звонков, транзакции в маркетплейсах.
28. Проблема мультиколлинеарности и переобучения в моделях с тысячами признаков (регуляризация Lasso, Ridge, ElasticNet в распределенной среде).
29. Автоматизированная генерация финансовых отчетов на основе сырых логов транзакций с применением big data ETL-процессов.
30. Построение системы раннего предупреждения (Early Warning System) финансового кризиса компании на основе потоковых операционных данных.