

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гаранин Максим Алексеевич
Должность: Ректор
Дата подписания: 08.10.2025 15:54:59
Уникальный программный ключ:
7708e3a47e66a8ee02711b298d7c78bd1e40bf88

Приложение к рабочей программе дисциплины

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Большие данные

(наименование дисциплины(модуля))

Направление подготовки / специальность

09.03.02 Информационные системы и технологии

(код и наименование)

Направленность (профиль)/специализация

Информационные системы и технологии на транспорте

(наименование)

Форма обучения

Содержание

1. Пояснительная записка.
2. Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций.
3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации.

1. Пояснительная записка

Цель промежуточной аттестации – оценивание промежуточных и окончательных результатов обучения по дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы.

Формы промежуточной аттестации: экзамен- **5 семестр**

Код и наименование компетенции	Код достижения индикатора компетенции
ОПК-8: Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем	ОПК-8.1: Применяет математические модели для проектирования информационных и автоматизированных систем
	ОПК-8.3: Использует средства проектирования информационных и автоматизированных систем

Результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы
ОПК-8.1: Применяет математические модели для проектирования информационных и автоматизированных систем	Обучающийся знает: Математические модели для проектирования информационных и автоматизированных систем для работы с большими данными.	Вопросы тестирования №(1-10)
	Обучающийся умеет: Применять математические модели для проектирования информационных и автоматизированных систем для работы с большими данными.	Задания №(1-2)
	Обучающийся владеет: Работы с математическими моделями для проектирования информационных и автоматизированных систем для работы с большими данными.	Задания №(3-4)
ОПК-8.3: Использует средства проектирования информационных и автоматизированных систем	Обучающийся знает: Средства проектирования информационных и автоматизированных систем для работы с большими данными.	Вопросы тестирования №(11-20)
	Обучающийся умеет: Использовать средства проектирования информационных и автоматизированных систем для работы с большими данными.	Задания №(5-6)
	Обучающийся владеет: Использования средств проектирования информационных и автоматизированных систем для работы с большими данными.	Задания №(7)

5 семестр

Промежуточная аттестация (экзамен) проводится в одной из следующих форм:

- 1) проводится в форме устного ответа на вопросы из перечня
- 2) выполнение заданий в ЭИОС Университета.

2. Типовые¹ контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций

2.1 Типовые вопросы (тестовые задания) для оценки знаниевого образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-8.1: Применяет математические модели для проектирования информационных и автоматизированных систем	Обучающийся знает: Математические модели для проектирования информационных и автоматизированных систем для работы с большими данными.
1.Какой из следующих инструментов является основным для обработки и анализа больших данных в экономике? Hadoop Microsoft Excel Adobe Photoshop Google Chrome 2.Какой термин используется для описания процесса преобразования неструктурированных данных в структурированный формат для анализа? Data Encryption Data Virtualization Data Normalization Data Wrangling 3.Какие из перечисленных характеристик наиболее существенны при работе с экономическими данными в больших объемах? Малый объем данных и простота анализа Медленная скорость обработки данных Высокая скорость обработки данных и масштабируемость Низкая точность данных 4.Какие из перечисленных методов являются типичными для анализа временных рядов в экономике с использованием больших данных? Анализ случайных величин Геостатистика Временные ряды и статистика Анализ текстовых данных 5.Какая роль машинного обучения в обработке больших данных в экономике? Машинное обучение не применимо к экономическим данным Только для создания анимации и графики Используется исключительно для разработки веб-приложений Прогнозирование и выявление закономерностей в данных 6.Какой из следующих форматов обычно используется для хранения больших объемов экономических данных? Текстовые файлы (.txt) Аудиофайлы (.mp3) Базы данных формата NoSQL Графические файлы (.jpg) 7.Какой процесс отвечает за объединение данных из разных источников для создания общего набора	

¹ Приводятся типовые вопросы и задания. Оценочные средства, предназначенные для проведения аттестационного мероприятия, хранятся на кафедре в достаточном для проведения оценочных процедур количестве вариантов. Оценочные средства подлежат актуализации с учетом развития науки, образования, культуры, экономики, техники, технологий и социальной сферы. Ответственность за нераспространение содержания оценочных средств среди обучающихся университета несут заведующий кафедрой и преподаватель – разработчик оценочных средств.

данных для анализа? Декомпозиция данных Классификация данных Data Integration Хранение данных 8.Какие из следующих языков программирования широко используются для обработки и анализа больших данных в экономике? Visual Basic COBOL h2ython Fortran 9.Что представляет собой понятие «Mah2Reduce» в контексте обработки больших данных? Алгоритм сортировки данных Технология сжатия файлов Функция для удаления дубликатов Модель программирования для обработки параллельных данных 10.Какие из следующих методов используются для обеспечения безопасности больших данных в экономике? Использование открытых сетей для передачи данных Хранение данных в неструктурированном формате Шифрование данных Публичное доступ к данным без ограничений	
ОПК-8.3: Использует средства проектирования информационных и автоматизированных систем	Обучающийся знает: Средства проектирования информационных и автоматизированных систем для работы с большими данными.
11.Какие из следующих терминов относятся к обработке больших данных в экономике? Mah2Reduce Histogram Italicize Sh2ark 12.Что представляет собой «Data Warehousing» в контексте больших данных? Процесс сжатия данных Хранение и управление большим объемом структурированных данных Моделирование данных Централизованное хранилище данных для поддержки анализа и отчетности 13.Какой инструмент чаще всего используется для визуализации больших данных в экономике? Data Encryh2tion Tableau Blockchain Neural Networks 14.Что означает термин «Data Mining» в контексте больших данных? Процесс уничтожения данных Этап развертывания данных Извлечение интересных и ранее неизвестных шаблонов из данных Методология сбора данных 15.Какой вид анализа используется для предсказания будущих тенденций на основе прошлых данных? Static Analysis Descrih2tive Analysis h2redictive Analysis h2rescrih2tive Analysis 16.Что представляет собой «Hadooh2» в экономическом анализе данных? Алгоритм сжатия данных Язык программирования Фреймворк для распределенного хранения и обработки данных Метод структурированного анализа данных 17.Какие из перечисленных терминов относятся к методам машинного обучения? Bar Chart Linear Regression Random Forest	

Scatter h2lot

18.Что означает «Big Data Velocity»?

Способность обрабатывать данные большого объема

Скорость передачи данных

Скорость поступления и обработки потоков данных в реальном времени

Частота анализа данных

19.Какие типы данных включаются в «Big Data Variety»?

Только структурированные данные

Только текстовые данные

Только числовые данные

Разнообразие структурированных и неструктурированных данных

20.Какая особенность «Big Data Volume»?

Большое количество аналитических инструментов

Большой объем хранимых данных

Большая скорость обработки данных

Огромные объемы данных, превышающие возможности традиционных баз данных

2.2 Типовые задания для оценки навыкового образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-8.1: Применяет математические модели для проектирования информационных и автоматизированных систем	Обучающийся умеет: Применять математические модели для проектирования информационных и автоматизированных систем для работы с большими данными.
1. Разведочный анализ данных: оценки центрального положения, оценки вариабельности, обследование распределения данных, обследование двоичных и категориальных данных, корреляция, исследование двух или более переменных Вычислить среднее, среднее усеченное и медиану численности населения, используя R или Python, наборы данных взять из электронного купца https://github.com/andrewgbruce/statistics-for-data-scientists 2. Бутстреп, выборки, распределения - Построить гистограммы для трех выборок: с 1000 значениями, с 1000 средними из 5 значений и с 1000 средними из 20 значений. - Построить бутстраповский доверительный интервал для годового дохода ссудозаявителей на основе выборки из 20 значений. - Построить квантиль-квантильный график выборки из 100 значений, извлеченных из нормального распределения. - Построить квантиль-квантильный график ежедневной доходности акций. - Реализовать одно из распределений (нормальное, Пуассона, Стьюдента, длиннохвостое, биномиальное, экспоненциальное и др).	
ОПК-8.1: Применяет математические модели для проектирования информационных и автоматизированных систем	Обучающийся владеет: Работы с математическими моделями для проектирования информационных и автоматизированных систем для работы с большими данными.
3. Проверка статистических гипотез. ANOVA 1. Компания, продающая относительно дорогостоящую услугу, хочет протестировать, какая из двух веб-презентаций справляется с продажей лучше. Ввиду дороговизны продаваемой услуги, продажи случаются нечасто, и цикл продаж продолжительный; аккумуляирование достаточного количества продаж с целью узнать, какая презентация превосходит, заняло бы слишком много времени. Поэтому компания решает измерить результаты при помощи эрзац-переменной, используя подробную внутреннюю страницу веб-сайта, в которой описывается услуга. Одна из потенциальных эрзац-переменных для компании — это число нажатий на подробной посадочной странице. Еще лучше — сколько времени люди проводят на странице. Разумно полагать, что веб-презентация	

(страница), которая задерживает внимание людей дольше, приведет к большему количеству продаж. Следовательно, при сравнении страницы А со страницей В метрическим показателем будет среднее время сеанса. Речь идет о внутренней странице специального назначения, поэтому она не получает огромного числа посетителей. Также стоит отметить, что служба Google Analytics (GA), при помощи которой мы измеряем время сеанса, не может измерить время сеанса последнего посещения страницы клиентом. Вместо того чтобы удалить этот сеанс из данных, GA записывает его как 0, поэтому данные требуют доработки, чтобы удалить эти сеансы. В результате получается в общей сложности 36 сеансов для двух разных презентаций: 21 сеанс для страницы А и 15 сеансов для страницы В. Используя программный пакет ggplot, сравните времена сеансов при помощи парных коробчатых диаграмм.

Времена сеансов страницы В продолжительнее в среднем на 21,4 секунды в отличие от страницы А. Вопрос состоит в том, находится ли эта разница внутри диапазона того, что может породить случайная возможность, или же, напротив, является статистически значимой. Ответьте на этот вопрос, применяя перестановочный тест

Покажите в виде гистограммы перераспределение разниц во временах сеансов

Создайте гистограмму произвольно перестановленных разниц в уровне конверсии

2. Проверка на основе t-статистики

3. Реализуйте процедуру ANOVA:

*Объединить все данные в одной коробке.

*Перетасовать и вынуть четыре повторных выборки с пятью значениями для каждой.

*Записать среднее значение каждой из четырех групп.

*Записать дисперсию среди средних значений четырех групп.

*Повторить шаги 2–4 множество раз (скажем, 1000).

3. Точная проверка Фишера и хи-квадрат

4. Регрессия и предсказание

1. прямая регрессия

2. подогнанные значения и остатки

3. множественная линейная регрессия:

Задача: оценщики округа должны оценивать стоимость домов в целях обложения налогами. Потребители недвижимости и профессионалы в этой области консультируются на популярных веб-сайтах, чтобы удостовериться в справедливости цены. Цель состоит в том, чтобы предсказать продажную цену на основе остальных переменных.

Вычислите метрические показатели регрессионной модели

4. Шаговая регрессия

5. Взвешенная регрессия

6. Представление фиктивных переменных (для задачи 3), диагностика модели

7. Перекрестная проверка

8. Отбор модели и шаговая регрессия

9. Взвешенная регрессия

10. Представление фиктивных переменных

11. Многоуровневые факторные переменные

12. Интерпретация уравнения регрессии

13. Проверка допущений: диагностика регрессии

14. Графики частных остатков и нелинейность

15. Нелинейная регрессия

ОПК-8.3: Использует средства проектирования информационных и автоматизированных систем

Обучающийся умеет:

Использовать средства проектирования информационных и автоматизированных систем для работы с большими данными.

5. Классификация

1. Реализовать наивный байесовский алгоритм

2. Логистическая регрессия и обобщенная линейная модель

3. Предсказанные значения в логистической регрессии

4. Диагностика модели

5. Оценивание моделей классификации

6. ROC-кривая

7. Стратегии в отношении несбалансированных данных

6. К ближайших соседей. Древоидные модели. Бэггинг и случайный лес. Бустинг

1. Задача: предсказание невозврата ссуды

2. Создайте признак, который представляет кредитоспособность заемщика

3. Сгенерируйте структурно распечатанную версию дерева

4. Алгоритм рекурсивного сегментирования

5. Постройте график предсказанных исходов из случайного леса применительно к данным о невозвратных ссудах

6. Выполните подгонку модели к данным о невозвратных ссудах

7. Постройте график предсказанных исходов из XGBoost применительно к данным о невозвратных ссудах

8.Выполните подгонку xgboost к данным о ссудах для тренировочного набора со всеми включенными в модель переменными 9.Гиперпараметры и перекрестная проверка	
ОПК-8.3: Использует средства проектирования информационных и автоматизированных систем	Обучающийся владеет: Использования средств проектирования информационных и автоматизированных систем для работы с большими данными.
7. Анализ главных компонент. Кластеризация на основе K средних. Иерархическая кластеризация 1. Выполнить PCA на доходности курса акций Chevron (CVX) и ExxonMobil (XOM) 2.Отобразите на графике главные компоненты вместе с данными 3. Визуализируйте относительную важность главных компонент (покажите нагрузки для верхних пяти главных компонент доходности курса акций) 4. Найдите четыре кластера на основе двух переменных — доходности акций ExxonMobil (XOM) и Chevron (CVX) 5. Визуализируйте кластеры 6. Покажите средние значения переменных в каждом кластере ("центроиды") 7.Постройте иерархическую кластеризацию к доходностям акций для ряда компаний 8. Сравнение мер различия применительно к данным об акциях (на графике) 9. Примените модельно-ориентированную кластеризацию к данным доходности акций 10. ВИС-оценки для данных о доходностях акций для разных количеств кластеров (компонент) 11. Постройте график каменистой осыпи 12. Постройте дендограмму hclust применительно к выборке данных о невозвратных ссудах с типами смешанных переменных	

2.3. Перечень вопросов для подготовки обучающихся к промежуточной аттестации

1. Назначение и основные компоненты системы больших данных.
- 2.Обзор современных систем управления большими данными.
- 3.Уровни представления больших данных.
- 4.Понятие схемы и подсхемы.
- 5.Модели данных (ER, семантическая объектная модель, логическая, физическая).
- 6.Иерархическая модель данных.
- 7.Сетевая модель данных.
- 8.Реляционная модель данных.
- 9.Схема отношения.
- 10.Язык манипулирования данными для реляционной модели.
- 11.Реляционная алгебра и язык SQL.
- 12.Проектирование баз данных для больших данных.
- 13.Функциональные зависимости.
- 14.Декомпозиция отношений.
- 15.Транзитивные зависимости.
- 16.Проектирование с использованием метода сущность-связь.
- 17.Создание и модификация больших данных.
- 18.Поиск, сортировка, индексирование больших данных.
- 19.Разработка форм и отчетов.
- 20.Физическая организация больших данных.
- 21.Хешированные, индексированные файлы.
- 22.Защита больших данных.
- 23.Целостность и сохранность больших данных.
- 24.Нормализация отношений
- 25.ER-проектирование больших данных.
- 26.Инфологическое моделирование
- 27.Даталогическое моделирование
- 28.Семантическая модель данных
- 29.Понятие о технологии, информации, данных
- 30.Скалярные типы переменных
- 31.Векторные типы переменных

- 32.Сложный тип переменных. Вложенность
- 33.Управление пользователями больших данных.
- 34.Аудит базы данных
- 35.Обеспечение целостности базы данных
- 36.Создание базы данных. (файлы параметров)
- 37.Запуск и останов базы данных
- 38.Различные режимы работы базы данных
- 39.Резервное копирование базы данных
- 40.Динамический SQL
- 41.Объектно-ориентированные БД
- 42.Иерархическая, сетевая и реляционная модели данных
- 43.Схемы и объекты схемы
- 44.Блоки данных, экстенды и сегменты.
- 45.Структуры памяти и процессы
- 46.Журнал Повторений
- 47.Транзакция
- 48.Этапы концептуального моделирования

3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации

Критерии формирования оценок по ответам на вопросы, выполнению тестовых заданий

- оценка **«отлично»** выставляется обучающемуся, если количество правильных ответов на вопросы составляет 100 – 90 % от общего объёма заданных вопросов;
- оценка **«хорошо»** выставляется обучающемуся, если количество правильных ответов на вопросы – 89 – 76 % от общего объёма заданных вопросов;
- оценка **«удовлетворительно»** выставляется обучающемуся, если количество правильных ответов на тестовые вопросы – 75–60 % от общего объёма заданных вопросов;
- оценка **«неудовлетворительно»** выставляется обучающемуся, если количество правильных ответов – менее 60 % от общего объёма заданных вопросов.

Критерии формирования оценок по результатам выполнения заданий

«Зачтено» – ставится за работу, выполненную полностью без ошибок и недочетов в соответствии с заданием. Обучающийся полностью владеет информацией по теме работы, решил все поставленные в задании задачи.

«Не зачтено» - ставится за работу, если обучающийся правильно выполнил менее 2/3 всего задания, использовал при выполнении неправильные алгоритмы, допустил грубые ошибки при программировании, сформулировал неверные выводы по результатам работы.

Виды ошибок:

- *грубые ошибки: незнание основных понятий, правил, норм; незнание приемов решения задач; ошибки, показывающие неправильное понимание условия предложенного задания.*
- *негрубые ошибки: неточности формулировок, определений; нерациональный выбор хода решения.*
- *недочеты: нерациональные приемы выполнения задания; отдельные погрешности в формулировке выводов; небрежное выполнение задания.*

Критерии формирования оценок по результатам выполнения практических работ

«Зачтено» – ставится за работу, выполненную полностью без ошибок и недочетов в соответствии с заданием. Обучающийся полностью владеет информацией по теме работы, решил все поставленные в задании задачи.

«Не зачтено» - ставится за работу, если обучающийся правильно выполнил менее 2/3 всей работы, использовал при выполнении работы неправильные алгоритмы, допустил грубые ошибки при расчетах, сформулировал неверные выводы по результатам работы.

Критерии формирования оценок по экзамену

«Отлично» (5 баллов) – обучающийся демонстрирует знание всех разделов изучаемой дисциплины: содержание базовых понятий и фундаментальных проблем; умение излагать программный материал с демонстрацией конкретных примеров. Свободное владение материалом должно характеризоваться логической ясностью и четким видением путей применения полученных знаний в практической деятельности, умением связать материал с другими отраслями знания.

«Хорошо» (4 балла) – обучающийся демонстрирует знания всех разделов изучаемой дисциплины: содержание базовых понятий и фундаментальных проблем; приобрел необходимые умения и навыки, освоил вопросы практического применения полученных знаний, не допустил фактических ошибок при ответе, достаточно последовательно и логично излагает теоретический материал, допуская лишь незначительные нарушения последовательности изложения и некоторые неточности. Таким образом, данная оценка выставляется за правильный, но недостаточно полный ответ.

«Удовлетворительно» (3 балла) – обучающийся демонстрирует знание основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем. Однако знание основных проблем курса не подкрепляется конкретными практическими примерами, не полностью раскрыта сущность вопросов, ответ недостаточно логичен и не всегда последователен, допущены ошибки и неточности.

«Неудовлетворительно» (0 баллов) – выставляется в том случае, когда обучающийся демонстрирует фрагментарные знания основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем. У экзаменуемого слабо выражена способность к самостоятельному аналитическому мышлению, имеются затруднения в изложении материала, отсутствуют необходимые умения и навыки, допущены грубые ошибки и незнание терминологии, отказ отвечать на дополнительные вопросы, знание которых необходимо для получения положительной оценки.