

**Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»**

Направление подготовки/специальность: 15.03.04 Автоматизация технологических процессов и производств

Наименование образовательной программы: Автоматизация технологических процессов и производств

Уровень образования: высшее образование - бакалавриат

Форма обучения: Очная

**Оценочные материалы
по дисциплине
Прикладные методы анализа данных**

**Москва
2024**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ РАЗРАБОТАЛ:

Разработчик

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Криницкий Е.В.
	Идентификатор	Rc6f46e52-KrinitiskyYV-272e3978

Е.В.
Криницкий

СОГЛАСОВАНО:

Руководитель
образовательной
программы

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Щербатов И.А.
	Идентификатор	R6b2590a8-ShcherbatovIA-d91ec17

И.А.
Щербатов

Заведующий
выпускающей кафедрой

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Щербатов И.А.
	Идентификатор	R6b2590a8-ShcherbatovIA-d91ec17

И.А.
Щербатов

ОБЩАЯ ЧАСТЬ

Оценочные материалы по дисциплине предназначены для оценки достижения обучающимися запланированных результатов обучения по дисциплине, этапа формирования запланированных компетенций и уровня освоения дисциплины.

Оценочные материалы по дисциплине включают оценочные средства для проведения мероприятий текущего контроля успеваемости и промежуточной аттестации.

Формируемые у обучающегося компетенции:

1. ПК-2 Способен применять информационные технологии для разработки автоматизированных систем управления технологическими процессами и производствами в области профессиональной деятельности

ИД-6 Применяет современное программное обеспечение для анализа данных

и включает:

для текущего контроля успеваемости:

Форма реализации: Компьютерное задание

1. Контрольная работа № 1 (Индивидуальный проект)
2. Контрольная работа № 2 (Индивидуальный проект)
3. Контрольная работа № 3 (Индивидуальный проект)

Форма реализации: Проверка задания

1. Устный опрос № 1 (Дискуссия)

БРС дисциплины

3 семестр

Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:

- КМ-1 Контрольная работа № 1 (Индивидуальный проект)
КМ-2 Устный опрос № 1 (Дискуссия)
КМ-3 Контрольная работа № 2 (Индивидуальный проект)
КМ-4 Контрольная работа № 3 (Индивидуальный проект)

Вид промежуточной аттестации – Экзамен.

Раздел дисциплины	Веса контрольных мероприятий, %				
	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4
	Срок КМ:	4	8	12	14
Введение в аналитику данных					
Анализ данных как процесс		+	+		
Структурированные данные		+	+		

Технологии и инструменты аналитики данных				
Технологии и прикладные задачи аналитики данных	+	+		
Инструменты аналитики данных. Большие данные. Наука о данных	+	+		
Методы подготовки данных к анализу				
Процесс ETL	+	+		
Трансформация данных	+	+		
Визуализация данных				
Цели, задачи, этапы и методы визуализации			+	+
Стандартизация и очистка нормативно-справочной информации				
Понятия очистки и стандартизации данных	+	+		
Модели Data Mining				
Описательные модели Data Mining: кластеризация, ассоциативные правила	+	+	+	+
Предсказательные модели Data Mining: классификация, регрессия	+	+	+	+
Вес КМ:	25	25	25	25

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

I. Оценочные средства для оценки запланированных результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Индекс компетенции	Индикатор	Запланированные результаты обучения по дисциплине	Контрольная точка
ПК-2	ИД-6ПК-2 Применяет современное программное обеспечение для анализа данных	Знать: Основные принципы, технологии и стандарты бизнес-анализа Уметь: Разрабатывать сценарии обработки данных в аналитической платформе Проектировать сценарии анализа данных сверху вниз для работы с любыми данными, соответствующей заложенной структуре Использовать на практике современные low-code аналитические платформы, востребованные в российских проектных компаниях	КМ-1 Контрольная работа № 1 (Индивидуальный проект) КМ-2 Устный опрос № 1 (Дискуссия) КМ-3 Контрольная работа № 2 (Индивидуальный проект) КМ-4 Контрольная работа № 3 (Индивидуальный проект)

II. Содержание оценочных средств. Шкала и критерии оценивания

КМ-1. Контрольная работа № 1

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Индивидуальный проект

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Студент выполняет типовое индивидуальное задание в аналитической платформе Loginom.

Краткое содержание задания:

1. Создайте новый пакет и импортируйте в него текстовый файл CreditSample.txt.
2. Создайте новое поле Дата обработки, значения в котором равны текущей дате.
3. Создайте новое поле Размер ссуды у.е., который рассчитывается делением на 70 поля Размер ссуды, руб. Все значения в новом поле должны быть округлены до второго знака.
4. Создайте новое поле Флаг, значение в котором истинно, если выполняется условие: Среднемесячный доход > 2000 и Наличие недвижимости = Да.
5. Создайте еще один столбец, значение в котором равно 1, если выполняется условие: Флаг = TRUE и Давать кредит = FALSE.
6. Создайте новое поле RATE, в котором хранится значение в поле Срок ссуды, возведенное в степень 0,6.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Уметь: Проектировать сценарии анализа данных сверху вниз для работы с любыми данными, соответствующей заложенной структуре	1.Создайте новое поле Сегмент, которое делит всех заемщиков на сегменты по следующим правилам (используйте функцию IF/IFF): а) ЕСЛИ Возраст >= 50 и Среднемесячный доход < 6000 ТО Сегмент = Сегмент 1 б) ЕСЛИ Возраст < 30 и ТО Сегмент = Сегмент 2 Сегмент = Сегмент 3 во всех остальных случаях, не удовлетворяющим п. 1) и 2).

Описание шкалы оценивания:

Оценка: 5 («отлично»)

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно

Оценка: 4 («хорошо»)

Нижний порог выполнения задания в процентах: 60

Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач

Оценка: 3 («удовлетворительно»)

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено

Оценка: 2 («неудовлетворительно»)

Описание характеристики выполнения знания: Оценка "неудовлетворительно" выставляется если задание выполнено неверно или преимущественно не выполнено

КМ-2. Устный опрос № 1

Формы реализации: Проверка задания

Тип контрольного мероприятия: Дискуссия

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Студент дает развернутые ответы на вопросы.

Краткое содержание задания:

1. Назовите интерфейсы аналитической системы Loginom
2. Охарактеризуйте логику работы компонентов аналитической системы Loginom:
 1. Фильтр строк
 2. Калькулятор
 3. Группировка данных
3. Назовите основные особенности обработчика Объединения, включая использование префиксов.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Знать: Основные принципы, технологии и стандарты бизнес-анализа	1. Детально опишите основные области интерфейса Loginom. 2. Опишите способы импорта и экспорта данных, включая подключения из баз данных. 3. Какое расширение имеют Loginom Data файлы. 4. Опишите назначение обработчика Группировка данных в Loginom и его основные особенности 5. Какие области для ввода данных можно выделить в мастере настройки Калькуляторе Loginom. Охарактеризуйте их. 6. Какие способы расчета выражений существует в калькуляторе 7. Назовите назначение обработчика Объединение и его основные особенности 8. С помощью каких элементов управления осуществляется ручная настройка в мастере настройки обработчика Объединение 9. Назовите способы запуска сценария

Описание шкалы оценивания:

Оценка: 5 («отлично»)

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно

Оценка: 4 («хорошо»)

Нижний порог выполнения задания в процентах: 60
Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач

Оценка: 3 («удовлетворительно»)

Нижний порог выполнения задания в процентах: 50
Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено

Оценка: 2 («неудовлетворительно»)

Описание характеристики выполнения знания: Оценка "неудовлетворительно" выставляется если задание выполнено неверно или преимущественно не выполнено

КМ-3. Контрольная работа № 2

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Индивидуальный проект

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Студент выполняет типовое индивидуальное задание в аналитической платформе Loginom.

Краткое содержание задания:

1. Создайте новый проект и импортируйте в него файл *Товары.txt*.
 2. Исключите из набора записи, в которых для товара отсутствует артикул.
 3. С помощью функций компонента **Калькулятор** рассчитайте, сколько месяцев прошло от даты последней продажи каждого товара до 01.04.2018.
 4. Отсортируйте набор данных по количеству месяцев по убыванию.
- Добавьте в набор поле логического типа **Вывод из продажи** и установите значение **true** для товаров, у которых от даты продажи до 01.04.2018 прошло более 10 месяцев.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Уметь: Разрабатывать сценарии обработки данных в аналитической платформе	1. У какого количества товаров отсутствует артикул? 2. Сколько месяцев прошло с продажи отдельных товаров? 3. Какие товары не продавались дольше всего? Как долго? 4. Какое количество товаров пора выводить из продажи?

Описание шкалы оценивания:

Оценка: 5 («отлично»)

Нижний порог выполнения задания в процентах: 70
Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно

Оценка: 4 («хорошо»)

Нижний порог выполнения задания в процентах: 60
Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач

Оценка: 3 («удовлетворительно»)

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено

Оценка: 2 («неудовлетворительно»)

Описание характеристики выполнения знания: Оценка "неудовлетворительно" выставляется если задание выполнено неверно или преимущественно не выполнено

КМ-4. Контрольная работа № 3

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Индивидуальный проект

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Студент выполняет типовое индивидуальное задание в аналитической платформе Loginom.

Краткое содержание задания:

1. Создайте новый пакет.
2. Скачайте файл *Вероятности визитов.txt*, который имеет следующие поля:
 - **Клиент.код** – уникальный идентификатор клиента;
 - **Отклик** – поле принимает значение **True** или **False**;
 - **Вероятность визита** – поле принимает значения в диапазоне от 0 до 1;
 - **Признак тестового множества** – поле принимает значение **True** или **False**;
 - **Тип клиента** – поле принимает значение **Тип 1** или **Тип 2**.
3. Создайте в Loginom новый пакет и импортируйте описанный набор данных.
4. Рассчитайте количество клиентов, у которых в идентификаторе **Клиент.код** присутствуют два и более нулей (например, 10006, 30600).

Обратите внимание: для решения данной задачи не требуется создавать сложный сценарий, достаточно изучить функции узла **Калькулятор**.

5. Подключите к вашему пакету библиотеку **Loginom Silver Kit**.
 6. Отберите всех клиентов, которые входят в тестовое множество.
 7. Далее используйте внешний компонент **AUC** из библиотеки **Loginom Silver Kit**.
- Вам потребуются следующие настройки в мэппинге полей:

- Отклик – Событие
- Вероятность визита - Оценка

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Уметь: Использовать на практике современные low-code аналитические платформы, востребованные в российских проектных компаниях	1. Сколько клиентов входит в тестовое множество? 2. У какого клиента из тестового множества вероятность визита максимальна (назовите его код)? 3. У какого клиента из тестового множества вероятность визита минимальна (назовите его код)? 4. Чему равны значения различных показателей на выходе узла AUC? 5. Сколько клиентов типа Тип 1 присутствует среди клиентов, у которых в идентификаторе два и более нулей? 6. Назовите значение переменной Z-оценка,

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
	полученной при использовании компонента AUC на клиентах их тестового множества. Ответ округлите до одного знака после запятой

Описание шкалы оценивания:

Оценка: 5 («отлично»)

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Оценка "отлично" выставляется если задание выполнено в полном объеме или выполнено преимущественно верно

Оценка: 4 («хорошо»)

Нижний порог выполнения задания в процентах: 60

Описание характеристики выполнения знания: Оценка "хорошо" выставляется если большинство вопросов раскрыто. выбрано верное направление для решения задач

Оценка: 3 («удовлетворительно»)

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Оценка "удовлетворительно" выставляется если задание преимущественно выполнено

Оценка: 2 («неудовлетворительно»)

Описание характеристики выполнения знания: Оценка "неудовлетворительно" выставляется если задание выполнено неверно или преимущественно не выполнено

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

3 семестр

Форма промежуточной аттестации: Экзамен

Пример билета

Проект сценария анализа данных

Задание

1. Создайте новый пакет.
2. Скачайте файл *Вероятности визитов.txt*, который имеет следующие поля:
 - **Клиент.код** – уникальный идентификатор клиента;
 - **Отклик** – поле принимает значение **True** или **False**;
 - **Вероятность визита** – поле принимает значения в диапазоне от 0 до 1;
 - **Признак тестового множества** – поле принимает значение **True** или **False**;
 - **Тип клиента** – поле принимает значение **Тип 1** или **Тип 2**.
3. Создайте в Loginom новый пакет и импортируйте описанный набор данных.
4. Рассчитайте количество клиентов, у которых в идентификаторе **Клиент.код** присутствуют два и более нулей (например, 10006, 30600).

Обратите внимание: для решения данной задачи не требуется создавать сложный сценарий, достаточно изучить функции узла **Калькулятор**.

5. Подключите к вашему пакету библиотеку **Loginom Silver Kit**.
6. Отберите всех клиентов, которые входят в тестовое множество.
7. Далее используйте внешний компонент **AUC** из библиотеки **Loginom Silver Kit**.

Вам потребуются следующие настройки в мэппинге полей:

- Отклик – Событие
- Вероятность визита - Оценка

После настройки входного порта достаточно запустить узел на выполнение.

8. Ответьте на вопросы:
 - Сколько клиентов входит в тестовое множество?
 - У какого клиента из тестового множества вероятность визита максимальна (назовите его код)?

Процедура проведения

Экзамен проводится в устной форме по билетам в виде подготовки и изложения развернутого ответа на теоретические вопросы и пояснение выполнения практического задания. Время на выполнение задания/подготовку ответа – 50 минут

1. Перечень компетенций/индикаторов и контрольных вопросов проверки результатов освоения дисциплины

1. Компетенция/Индикатор: ИД-6ПК-2 Применяет современное программное обеспечение для анализа данных

Вопросы, задания

1.

1.	Что из себя представляет аналитическая платформа?
2.	Дайте определение понятию кластеризация данных.
3.	Поясните особенности структуры самоорганизующихся карт.
4.	Для чего предназначено хранилище данных?
5.	Для чего используют нейронные сети?
6.	Дайте определение понятию искусственный интеллект?

7.	С какой целью выполняют процесс очистки данных?
8.	В чем состоит процесс нормализации данных?
9.	В чем состоит моделирование системы?
10.	Назовите ключевые особенности метода дерева решений?
11.	Дайте определение понятию структурированные данные.
12.	На основе какого способа машинного обучения построены карты Кохонена и какие задачи она решает?
13.	Охарактеризуйте процесс трансформации данных.
14.	Что из себя представляет процесс консолидации данных?
15.	Дайте определение понятию Data Mining?

Материалы для проверки остаточных знаний

1. Аналитическая платформа — ...

Ответы:

- а) специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) подразделение искусственного интеллекта, изучающее методы построения алгоритмов, способных обучаться на данных

Верный ответ: а

2. Машинное обучение - ...

Ответы:

- а) специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных
- б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.
- г) подразделение искусственного интеллекта, изучающее методы построения алгоритмов, способных обучаться на данных.

Верный ответ: г

3. Очистка данных — ...

Ответы:

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

Верный ответ: а

4. Обогащение данных — ...

Ответы:

- а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.
- б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязки аналитических задач
- в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязки аналитической задачи
- г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

Верный ответ: б

5.Предобработка данных

Ответы:

- а) определение информации о некотором атрибуте (поле) источника данных и проверяется ее соответствие заданным ограничениям
- б) комплекс методов и алгоритмов, применяемых в аналитическом приложении с целью подготовки данных к решению конкретной задачи и приведения их в соответствие с требованиями, определяемыми спецификой задачи и способами ее решения
- в) выявление дубликатов
- г) выявление противоречий

Верный ответ: б

6.Транзакция —

Ответы:

- а) некоторый набор операций над базой данных, который рассматривается как единственное завершённое, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных
- б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов
- в) высокоуровневые средства отражения информационной модели и описания структуры данных
- г) это установление зависимости дискретной выходной переменной от входных переменных

Верный ответ: а

II. Описание шкалы оценивания

Оценка: 5 («отлично»)

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Работа выполнена в рамках "продвинутого" уровня. Ответы даны верно, четко сформулированные особенности практических решений

Оценка: 4 («хорошо»)

Нижний порог выполнения задания в процентах: 60

Описание характеристики выполнения знания: Работа выполнена в рамках "базового" уровня. Большинство ответов даны верно. В части материала есть незначительные недостатки

Оценка: 3 («удовлетворительно»)

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Работа выполнена в рамках "порогового" уровня. Основная часть задания выполнена верно.

Оценка: 2 («неудовлетворительно»)

Описание характеристики выполнения знания: Работа не выполнена или выполнена преимущественно неправильно

III. Правила выставления итоговой оценки по курсу