

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Волгоградский институт управления – филиал РАНХиГС

Экономический факультет

кафедра информационных систем и математического моделирования

УТВЕРЖДЕНА
учёным советом
Волгоградского института управления –
филиала РАНХиГС
Протокол № 2 от 23.09.2021 г.

ПРОГРАММА БАКАЛАВРИАТА

Финансы и кредит

(наименование образовательной программы)

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ,
реализуемой без применения электронного (онлайн) курса**

Б1.В.04 АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

(индекс, наименование дисциплины (модуля), в соответствии с учебным планом)

38.03.01 Экономика

(код, наименование направления подготовки (специальности))

очная

(форма(ы) обучения)

Год набора – 2022

Волгоград, 2021 г.

Автор(ы)–составитель(и):

к.п.н., доцент, доцент кафедры информационных систем и математического моделирования Ключева И.А.

Заведующий кафедрой информационных систем и математического моделирования,
к.т.н., доцент Астафурова О.А.

РПД Б1.В.03.01 «Экономическая информатика» одобрена на заседании кафедры информационных систем и математического моделирования

Протокол от 31 августа 2021 года № 1

Рабочая программа дисциплины составлена на основе типовой рабочей программы дисциплин Б1.В 04 «Анализ и визуализация данных» для направления подготовки 38.03.01 «Экономика», авторами–составителями которой являются:

старший преподаватель кафедры «Бухгалтерский учет, экономический анализ и аудит» Баркова Е.А.;

старший преподаватель кафедры «Финансы, денежное обращение и кредит» Карпов В.А.;

заведующий кафедрой «Финансы, денежное обращение и кредит» д.э.н., проф. Хандруев А.А.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине , соотнесенных с планируемыми результатами освоения программы.....	4
1.1. Основные компетенции.....	4
1.2. Результаты обучения.....	4
2. Объем и место дисциплины в структуре образовательной программы.....	6
3. Содержание и структура дисциплины.....	7
3.1. Структура дисциплины.....	7
3. 2. Содержание дисциплины (модуля).....	8
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине.....	9
4.1. Текущий контроль успеваемости.....	9
4.2. Типовые материалы текущего контроля успеваемости обучающихся.....	9
5. Оценочные материалы промежуточной аттестации по дисциплине.....	23
5.1. Методы проведения зачета.....	23
5.2. Оценочные материалы промежуточной аттестации.....	23
6. Методические материалы по освоению дисциплины.....	26
7. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине.....	30
7.1. Основная литература.....	30
7.2. Дополнительная литература.....	30
7.3. Учебно-методическое обеспечение самостоятельной работы.....	30
7.4. Нормативные правовые документы.....	30
8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы.....	32
Приложение 1.....	33

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы

1.1. Основные компетенции

Дисциплина Б1.В.04 «Анализ и визуализация данных» обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс ОС II-5	Способность применять информационные технологии для решения прикладных финансово-экономических задач	ПКс ОС II-5.2	Способность анализировать и визуализировать данные о рисках и их потенциальных источниках

1.2. Результаты обучения

В результате освоения дисциплины Б1.В.04 «Анализ и визуализация данных» у студентов должны быть сформированы:

ОТФ/ТФ (при наличии проф-стандарта) трудовые или профессиональные действия	Код этапа освоения компетенции	Результаты обучения
ОТФ В. Разработка отдельных функциональных направлений управления рисками ТФ – В/02.6 Документирование процесса управления рисками и корректировка реестров рисков в рамках отдельных бизнес-процессов и функциональных направлений	ПКс ОС II – 5.2	<p>на уровне знаний:</p> <p>знание структуры и принципов работы современных аналитических платформ; знание моделей, методов и алгоритмов анализа данных; знание способов визуализации данных на разных этапах аналитического процесса; знание методов оценки качества данных и подготовки данных к анализу; знание теоретических основ экономических информационных систем: классификации, принципов построения, основных информационных технологий, используемых в ЭИС, методов обеспечения безопасности ЭИС, тенденций развития ЭИС; знание стандартов MRP, MRP II, ERP; знание структуры, функций и архитектуры ERP-системы; знание примеров западных и российских ERP-систем.</p> <p>на уровне умений:</p> <p>умение осуществлять выбор методов анализа; умение оценить качество данных и подготовить исходные данные для анализа; умение визуализировать данные и интерпретировать полученные результаты; умение решать задачи</p>

		<p>ассоциации, кластеризации, классификации, регрессии, задачи анализа и прогнозирования временных рядов; умение строить аналитическую отчетность; умение регистрировать хозяйственные операции в ERP-системе; умение использовать средства ERP-системы для управления закупками, продажами, кредиторской и дебиторской задолженностью, платежами; умение исправлять ошибки, допущенные при регистрации хозяйственных операций; умение строить производственную и товарную модель предприятия в ERP-системе, планировать потребность в ресурсах; умение использовать аналитические измерения для бизнес-анализа в ERP-системах; умение формировать бюджеты, формировать финансовые и аналитических отчеты в системе и анализировать отклонения.</p> <p>на уровне навыков: навыки решения задач ассоциации, кластеризации, классификации, регрессии, задачи анализа и прогнозирования временных рядов; навыки построения аналитической отчетности; навыки эффективной аналитической работы в среде компьютерных программ Logiном и Power BI; навыки решения практических задач в среде ERP-системы.</p>
--	--	---

2. Объем и место дисциплины в структуре образовательной программы

Учебная дисциплина Б1.В.04 «Анализ и визуализация данных» принадлежит к блоку «Часть, формируемая участниками образовательных отношений». В соответствии с учебным планом по очной и заочной формам обучения дисциплина осваивается в 3 семестре, общая трудоемкость дисциплины в зачетных единицах составляет 2 ЗЕ (72 часа).

Освоение дисциплины Б1.В.04 «Анализ и визуализация данных» происходит после освоения дисциплин Б1.В.03.01 «Экономическая информатика», Б1.О.07 «Алгоритмизация и программирование».

По очной форме обучения количество академических часов, выделенных на контактную работу с преподавателем (по видам учебных занятий) – 38 часов (лекций – 4 часа, практических занятий – 34 часа) и на самостоятельную работу обучающихся – 30 часов, на контроль – 4 часа.

Форма промежуточной аттестации в соответствии с учебным планом:
по очной форме обучения – зачет с оценкой в 3 семестре;

3. Содержание и структура дисциплины

3.1. Структура дисциплины

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.					СР О	Форма текущего контроля успеваемости**, промежуточной аттестации***
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					
			Л/ЭО, ДОТ *	ЛР/ЭО, ДОТ *	ПЗ/ЭО, ДОТ *	КСР		
Очная форма обучения								
3 семестр								
Тема 1	Подготовка данных к анализу	9	1		4		4	О, ПОЗ
Тема 2	Визуализация данных	10			6		4	ПОЗ, ПКР
Тема 3	Классификация и регрессия (Статистические методы)	11	1		6		4	О, ПОЗ
Тема 4	Классификация и регрессия (Машинное обучение)	12			6		6	Т, ПОЗ, ПКР
Тема 5	Кластеризация	13	1		6		6	О, ПОЗ
Тема 6	Анализ и прогнозирование временных рядов	13	1		6		6	О, ПОЗ
Промежуточная аттестация		4						ЗаО
Всего:		72 (2 ЗЕ)	4		34		30	

*Примечание: формы текущего контроля успеваемости: опрос (О), тестирование (Т), контрольная работа (КР), коллоквиум (К), эссе (Э), реферат (Р), диспут (Д), ситуационная задача (СЗ), разбор ситуаций (РС), ** – формы текущего контроля успеваемости: опрос (О), тестирование (Т), онлайн тест (ОТ), практико-ориентированные задания (ПОЗ), практическая контрольная работа (ПКР), *** – форма промежуточной аттестации: зачет с оценкой (ЗаО)*

3. 2. Содержание дисциплины (модуля)

Тема 1. Подготовка данных к анализу

Очистка и предобработка данных.

Оценка качества данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений. Сокращение размерности. Сэмплинг.

Трансформация данных.

Группировка данных. Слияние данных. Квантование. Нормализация и кодирование.

Тема 2. Визуализация данных

Введение в визуализацию. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы для оценки качества моделей. Визуализаторы, применяемые для интерпретации результатов анализа.

Визуализация данных средствами MS Power BI.

Семейство MS Power BI. Конструктор кубов и аналитических представлений. Совместная работа с представлениями – серверные сервисы Power BI. Решение задач объединения

данных из разных источников, вычисляемые поля и Меры, элементы языка DAX, построение различных визуальных представлений по данным из учетных систем и CRM-систем.

Тема 3. Классификация и регрессия (Статистические методы)

Введение в классификацию и регрессию. Множественная линейная регрессия. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Нелинейные регрессионные модели. Основы логистической регрессии. Множественная логистическая регрессия. Оценка эффективности и сравнение моделей. ROC-анализ. Оценка ошибки модели. Издержки ошибочной классификации. LIFT- и PROFIT-кривые.

Тема 4. Классификация и регрессия (Машинное обучение)

Введение в деревья решений. Алгоритмы построения деревьев решений. Упрощение деревьев решений. Введение в нейронные сети. Принципы построения нейронных сетей. Процесс обучения нейронной сети. Обучение в условиях несбалансированности классов.

Тема 5. Кластеризация

Введение в кластеризацию. Меры близости, используемые в алгоритмах кластеризации. Методы кластерного анализа. Иерархическая кластеризация. Алгоритм кластеризации k-means. Сети Кохонена. Карты Кохонена.

Тема 6. Анализ и прогнозирование временных рядов

Введение в прогнозирование. Временной ряд и его компоненты. Модели прогнозирования. Исследование наборов данных и выбор метода прогнозирования. Ошибки прогноза. Оценка адекватности выбранного метода прогнозирования. Методы сглаживания и скользящие средние. Декомпозиция временного ряда. Регрессионный анализ временных рядов.

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине

4.1. Текущий контроль успеваемости

В ходе реализации дисциплины Б1.В.04 «Анализ и визуализация данных» используются следующие методы текущего контроля успеваемости обучающихся:

Тема	Методы текущего контроля успеваемости
Подготовка данных к анализу	Опрос Практико-ориентированное задание (компьютерный практикум)
Визуализация данных	Практико-ориентированное задание (компьютерный практикум) Практическая контрольная работа
Классификация и регрессия (Статистические методы)	Опрос Практико-ориентированное задание (компьютерный практикум)
Классификация и регрессия (Машинное обучение)	Тестирование Практико-ориентированное задание (компьютерный практикум) Практическая контрольная работа
Кластеризация	Опрос Практико-ориентированное задание (компьютерный практикум)
Анализ и прогнозирование временных рядов	Опрос Практико-ориентированное задание (компьютерный практикум)
Подготовка данных к анализу	Опрос Практико-ориентированное задание (компьютерный практикум)

4.2. Типовые материалы текущего контроля успеваемости обучающихся

Вопросы для опроса и решения задач

Типовые оценочные материалы по теме 2 «Подготовка данных к анализу»

Опрос

1. Для чего нужны очистка и предобработка данных.
2. Оценка качества данных.
3. Фильтрация данных.
4. Обработка дубликатов и противоречий.
5. Выявление аномальных значений.
6. Восстановление пропущенных значений.
7. Сокращение размерности. Сэмплинг.
8. Трансформация данных.
9. Группировка данных.
10. Слияние данных.
11. Квантование.

12. Нормализация и кодирование.

Практико-ориентированное задание (компьютерный практикум) «Решение задачи мониторинга корректности консолидированных данных истории продаж сети предприятий розничной торговли»

Для принятия управленческих решений о частоте проведения сезонных распродаж, которые должны увеличить объем выручки по итогам года, требуется анализ деятельности магазинов с возможностью прогнозирования выручки на следующий год. Для этого предоставляются данные о продажах 45 магазинов, расположенных в разных регионах - в каждом магазине есть несколько отделов. Компания также проводит несколько рекламных мероприятий по уценке в течение года, которые предшествуют знаменательным праздникам.

Табличные данные представлены в текстовых файлах в формате .csv:

1. stores.csv (магазины);
2. train.csv (продажи);
3. features.csv (признаки).

Магазины (stores.csv) — представлена анонимная информация о 45 магазинах (Store), с указанием типа (Type) и размера торговых площадей в sf (Size) магазина.

Продажи (train.csv) — представлены реальные данные о продажах в сети Walmart, которые охватывают период с 2010 по 2012 года.

Признаки (features.csv) — представлены дополнительные данные, относящиеся к деятельности магазина, отдела и региона на указанные даты.

Выявление аномалий (также обнаружение выбросов) — это опознавание редких данных, событий или наблюдений, которые вызывают подозрения ввиду существенного отличия от большей части данных. Аномалии в данных могут возникать по разным причинам от неправильного ввода до ошибок оцифровки.

Анализ данных Stores

1. Преобразуйте категориальный признак Type в порядковый, так как в данном признаке есть скрытый порядок;
2. Преобразуйте признак Size в кв.м.;
3. Выявите аномалии (выбросы) в признаке Type и определите их в подходящую по значению к одному из типов магазинов, тем самым ослабив эффект выброса;
4. Сохраните полученные данные в файл: new _ stores . csv ;
5. Оцените качество очистки данных по каждому типу магазинов по признаку Size: соотнесите межквартильный размах до очистки данных к межквартильному размаху после очистки.

Внешнее обогащение предполагает привлечение дополнительной информации из внешних источников, что позволит повысить ценность и значимость данных с точки зрения их анализа. Под повышением значимости данных подразумевается, что на основе их анализа можно будет принимать управленческие решения принципиально нового уровня.

Анализ данных Train.

1. Выявите аномалии (выбросы) в признаке Weekly_Sales и преобразуйте их в подходящее значение, тем самым ослабив эффект выброса;
2. Сгруппируйте данные таким образом, чтобы получить новые данные следующего формата: сумма выручки по каждому магазину за определенный день;
3. Сформировать новую таблицу со следующими признаками: «Store», «Value», «Day», «Month», «Year», «IsHoliday»;
4. Сохраните полученные данные в файл: new_train.csv;
Зачастую данные, которые нам нужно использовать, имеют сложный вид; они не всегда бывают цельными, то есть разрозненный набор данных, который в процессе решения поставленной задачи требуется слить в один набор данных.
5. Соедините данные new_stores.csv и new_train.csv;

6. Сохраните полученные данные: value.csv;
7. 7. Рассчитайте долю (%) магазинов сети Walmart каждого типа.
Внутреннее обогащение не предполагает привлечения какой-либо внешней информации.

Оно обычно связано с получением и включением в набор данных полезной информации, которая отсутствует в явном виде, но может быть тем или иным способом получена с помощью манипуляций с имеющимися данными.

Анализ данных Value.

Эффективность использования торговых площадей может оцениваться как по отделам внутри магазина, так и по магазину в целом.

1. Создайте новый признак Effic_space и рассчитайте по каждой позиции в данных эффективность использования одного квадратного метра торговой площади по каждому магазину (сколько выручки приходится на один квадратный метр торговой площади);
2. Сохраните полученные данные: new_value.csv;
3. Определите какой магазин эффективней всех использует торговую площадь, а какой неэффективен.

Шкалирование – это общепринятая задача предобработки данных в машинном обучении. Многие алгоритмы машинного обучения – градиентные методы, которые сильно чувствительны к шкалированию. Они отлично работают на признаках, которые находятся на одной шкале, как правило, от 0 до 1 или от -1 до 1.

Анализ данных Features

Распространенный метод шкалирования данных является стандартизация, при котором значения признаков должны быть приближенно стандартно распределены.

1. Соединить данные: new_value.csv и features.csv;
2. Сделайте преобразования данных (аномалии, выбросы, категориальные признаки и т.п.);
3. Удалить признаки, которые дублируют друг друга или которые информативно включают в себя другие признаки.

После преобразования должны остаться следующие признаки: «Store», «Value», «IsHoliday», «Size», «Effic_space», «Temperature», «Fuel_Price», «Markdown 2», «Markdown 3», «Markdown 4», «Markdown 5», «CPI», «Unemployment» ;

4. Преобразуйте признаки, чтобы они имели среднее значение 0 и стандартное отклонение 1;
5. Определите, какой признак не требуется шкалировать;
6. Определите, какой признак имеет максимальное значение после шкалирования, а какой минимальный.

Типовые оценочные материалы по теме 3 «Визуализация данных»

Опрос

1. Каковы цели и задачи визуализации данных в аналитических технологиях?
2. Зачем нужен визуальный контроль данных после их загрузки в аналитическое приложение?
3. Чем отличаются средства визуализации общего назначения от специализированных средств визуализации?
4. Какие средства визуализации данных относят к визуализаторам общего назначения и почему?
5. В чем заключается OLAP-анализ и каковы его цели?
6. Какова структура OLAP-куба? Какие манипуляции с измерениями можно производить, чтобы сделать представление куба более информативным?

7. Семейство MS Power BI. Конструктор кубов и аналитических представлений.
8. Совместная работа с представлениями – серверные сервисы Power BI.
9. Решение задач объединения данных из разных источников.
10. Вычисляемые поля и Меры.
11. Элементы языка DAX

Практико-ориентированное задание (компьютерный практикум) «Получение аналитической отчетности в MS Power BI»

Легенда: аптечная сеть ООО «ФАРМА – Центр» имеет три торговых точки в разных городах России. Вы получили задание проанализировать работу фирмы за прошедший период и подготовить наглядный отчет.

Задача:

- Сформировать многомерный отчет, отражающий динамику сумм продаж по месяцам года в разрезе товарных групп.
 - Сформировать многомерный отчет, отражающий динамику сумм продаж по месяцам года в разрезе аптек.
 - Сформировать многомерный отчет и график продаж в регионах.
1. Откройте Power BI Desktop. Загрузите в Power BI данные для построения отчета. Тип данных — файлы Excel. Данные содержатся в файлах Продажи, Товары, Курс, География.
 2. Объедините таблицы в единую модель данных.
 3. Перейдите в представление Отчет. Используйте визуализатор «Карточка» для получения общей суммы продаж за период.
 4. Постройте визуализатор «Гистограмма с группировкой», отметив поля Дата (таблица Курс) и Сумма (таблица Продажи). Получите представление по кварталам.
 5. На панели Визуализации выберите Линейчатую диаграмму с группировкой
 6. Отметьте поля Товарная группа (таблица Товары) и Сумма (таблица Продажи). Отсортируйте товарные группы по сумме выручки.
 7. На панели Визуализации выберите Круговую диаграмму.
 8. Отметьте поля Область (таблица География), Отдел и Сумма (таблица Продажи).
 9. На панели Визуализации выберите диаграмму Карта. Отметьте поля Город (таблица География) и Сумма (таблица Продажи). В области форматирования удалите название диаграммы. Параметр «Пузырьковая» установите — 10%. Измените цвета данных и стили карт.
 10. Добавьте в отчет новую страницу. Новую страницу переименуйте в Тренд. На странице Тренд постройте визуализатор График. Отметьте поля Дата (таблица Курс) и Сумма (таблица Продажи). Добавьте Линию тренда.
 11. На панели Визуализации выберите Срез: поле Товарная группа (таблица Товары).
 12. Добавьте срез по городам.

Типовой вариант практической контрольной работы

1. Из нескольких открытых источников подобрать данные, которые позволят
 - дать одного всестороннюю характеристику из регионов РФ;
 - проанализировать динамику основных показателей социально-экономического развития региона.
2. Загрузить данные в MS Power BI, построить модель данных, провести необходимые преобразования данных, вычислить меры.
3. Подобрать средства визуализации данных. Построить отчет, который позволит проанализировать демографическое, социальное, экономическое положение региона РФ и динамику основных показателей.
4. Подготовить презентацию с использованием MS Sway для иллюстрации полученных результатов.

Типовые оценочные материалы к семинару по теме 4 «Классификация и регрессия (статистические методы)»

Опрос

1. В чем заключается задача классификации?
2. Какие алгоритмы используются для решения задач классификации?
3. По какому принципу строится таблица сопряженности и каковы цели ее использования?
4. Как рассчитываются показатели чувствительности и специфичности модели, в чем состоит их смысл?
5. Как строится ROC-кривая?
6. Как можно определить оптимальный порог отсечения?
7. Как строится LIFT-кривая?
8. Какие достоинства и недостатки имеют модели логистической регрессии применительно к задаче кредитного скоринга?

Практико-ориентированное задание (компьютерный практикум) «Скоринговые модели для оценки кредитоспособности заемщиков»

Постановка задачи.

В коммерческом банке имеется продукт «Нецелевой потребительский кредит»: кредиты предоставляются на любые цели с принятием решения в течение нескольких часов. За это время проверяются минимальные сведения о клиенте, в основном, такие, как отсутствие криминального прошлого и кредитная история в других банках.

В банке накоплена статистическая информация о заемщиках и качестве обслуживания ими долга за несколько месяцев. Руководство банка, понимая, что отсутствие адекватных математических инструментов, позволяющих оптимизировать риски, не способствует расширению розничного бизнеса в области потребительского кредитования, поставило перед отделом розничных рисков задачу разработать скоринговые модели с различными стратегиями кредитования, которые позволили бы управлять рисками и минимизировать число «безнадежных» заемщиков.

Решение задачи.

Импортировать файл с кредитными историями Loans.txt.

В задаче представлено 2709 кредитов с известными исходами платежей. Набор данных уже разбит на два множества — обучающее (80%) и тестовое (20%) — таким образом, чтобы в каждом множестве доля «плохих» кредитов была примерно одинаковой.

В структуре данных:

Поле	Описание
Код	Служебный код заявки
Дата	Дата выдачи кредита
О/Д, %	Коэффициент Обязательства/Доход в процентах (Вещественный)
Возраст	Возраст заемщика на момент принятия решения о выдаче кредита
Проживание	Основание для проживания: собственник, муниципальное жилье, аренда
Срок проживания в регионе	Менее 1 года, от 1 года до 5 лет, свыше 5 лет
Семейное положение	Холост/не замужем, женат/замужем, разведен/вдовство, другое
Образование	Среднее, среднее специальное, высшее
Стаж работы на последнем месте	Менее 1 года, от 1 года до 3 лет, свыше 3 лет
Уровень должности	Сотрудник, руководитель среднего звена, руководитель высшего звена
Кредитная история	Информация из бюро кредитных историй. Если есть информация

	о просрочках по прошлым кредитам, то клиенту присваивается категория «отрицательная» (тип данных Строковый);
Просрочки свыше 60 дней	0 – отсутствовали, 1 – имели место (тип данных Целый);
Тестовое множество	Служебный признак, TRUE соответствует тестовому множеству (тип данных Логический).

1. С помощью обработчика «Калькулятор» на основе поля Просрочки более 60 дн. (COL12) получите новое поле Класс заемщика.
2. Постройте модель Логистической регрессии. Выходная переменная — Класс заемщика.
3. Запишите уравнение логистической регрессии.
4. Используя визуализатор «Таблица сопряженности», определите Чувствительность и Специфичность полученной модели.
5. Оцените качество модели.
6. Чему равен оптимальный порог отсечения?
 - Определить оптимальный порог отсечения, используя визуализатор ROC-кривая.
 - Определить оптимальный порог отсечения, используя правило

$$P = \frac{1}{\left(1 + \frac{C_{FN}}{C_{FP}}\right)}$$

C_{FN} — издержки, связанные с ложноотрицательными наблюдениями;

C_{FP} — издержки, связанные с ложноположительными наблюдениями.

Отношение издержек $\frac{C_{FN}}{C_{FP}}$ задать самостоятельно.

7. Как изменяется кредитный рейтинг заемщика с каждым дополнительным иждивенцем?
8. Используя визуализатор «Что-если», определите к какому классу будут отнесены заемщики
 - Мужчина, 43-х лет, состоит в браке, 2 иждивенца, доход — 50000 руб., опыт работы — 20 лет, срок проживания на одном месте — 10 лет., недвижимость 0, месячный платеж 7500 руб.
 - Женщина, 30-ти лет, не замужем, 1 иждивенец, доход — 40000 руб., опыт работы — 6 лет, срок проживания — 10 лет, недвижимость 0, месячный платеж 5000 руб.

Типовые оценочные материалы по теме 5 «Классификация и регрессия (Машинное обучение)»

Тестирование

1. Классификация — это ...
 - a. отнесение объектов к одной из заранее неизвестных групп
 - b. процесс формирования групп и отнесения объектов к одному из них.
 - c. отнесение объектов к одному из заранее известных классов
2. Задачи классификации решаются следующими алгоритмами:
 - a. алгоритмом Argiоi
 - b. нейронные сети
 - c. линейной регрессии
3. Деревья решений относятся к группе (-ам) ...
 - a. кибернетических методов
 - b. методов кросс-табуляции

- c. статистических методов
 - d. логических методов
4. Внутренний узел дерева решений называют также ...
 - a. узлом проверки
 - b. конечным узлом
 - c. вершиной
 - d. листом
 5. Конечный узел дерева решений называют также ...
 - a. узлом проверки
 - b. узлом решения
 - c. листом
 6. При помощи метода деревьев решений решаются задачи ...
 - a. классификации и кластеризации
 - b. классификации и прогнозирования
 - c. кластеризации, классификации и прогнозирования
 - d. кластеризации и прогнозирования
 7. Процесс отсечения ветвей или замена некоторых ветвей поддеревом ...
 - a. происходит сверху вниз, т.е. является нисходящим
 - b. происходит снизу вверх, т.е. является восходящим
 - c. может быть, как нисходящим, так и восходящим
 8. Алгоритм конструирования дерева решений ...
 - a. не требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых
 - b. требует от пользователя выбора из набора входных атрибутов (независимых переменных), наиболее значимых
 - c. на вход алгоритма можно подавать все существующие атрибуты, алгоритм сам выберет наиболее значимые среди них, и только они будут использованы для построения дерева
 9. Если зависимая переменная принимает непрерывные значения, то дерево решений решает задачу:
 - a. классификации и численного прогнозирования
 - b. численного прогнозирования
 - c. классификации
 10. Если зависимая переменная принимает дискретные значения, при помощи метода дерева решений решается задача:
 - a. классификации и численного прогнозирования
 - b. классификации
 - c. численного прогнозирования
 11. Метод, который делает заключения относительно данной ситуации по результатам поиска аналогий, хранящихся в базе прецедентов относится к категории ...
 - a. «обучение с учителем»
 - b. самообучающейся системы
 - c. «обучение без учителя»
 12. Явление переобучения характеризуется ...
 - a. чрезмерно точным соответствием нейронной сети конкретному набору обучающих примеров, при котором сеть теряет способность к обобщению
 - b. возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры нейронной сети
 - c. возникновением, в случае слишком долгого обучения, слишком большого числа обучающих примеров или слишком сложной структуры нейронной сети
 13. Многослойный персептрон – это сеть ...
 - a. обратного распространения сигнала (с обратными связями)

- b. в которой входной сигнал преобразуется в выходной, проходя последовательно через несколько слоев
 - c. прямого распространения сигнала (без обратных связей)
14. В многослойном персептроне ...
- a. должен быть хотя бы один скрытый слой
 - b. присутствие нескольких скрытых слоев оправдано лишь в случае использования нелинейных функций активации
 - c. может быть какое угодно количество скрытых слоев, они также могут вообще отсутствовать
15. Заполните пропуски в формулировке: "... – выходная связь нейрона, с которой сигнал (возбуждения или торможения) поступает на... следующих нейронов"
- a. синапс, аксоны
 - b. аксон, точку ветвления
 - c. аксон, синапсы
16. Однонаправленные входные связи, соединенные с выходами других нейронов – это ...
- a. синапсы
 - b. слои сети
 - c. аксоны
17. Слой нейронной сети – это ...
- a. выходная связь данного нейрона, с которой сигнал (возбуждения или торможения) поступает на синапсы следующих нейронов
 - b. один или несколько нейронов, на входы которых подается один и тот же общий сигнал
 - c. однонаправленные входные связи, соединенные с выходами других нейронов
18. Ошибкой обучения нейронной сети называется ...
- a. переобучение нейронной сети
 - b. разность между желаемым и полученным на выходе сигналами
 - c. целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети
19. Целевая функция, требующая минимизации в процессе управляемого обучения нейронной сети – это ...
- a. функция ошибок
 - b. функция переобучения
 - c. ошибка обучения

Практико-ориентированное задание (компьютерный практикум) «Прогнозирование ухода к другому оператору клиентов телекоммуникационной сети»

Предлагается решить задачу из области прогнозирования ухода к другому оператору клиентов телекоммуникационной сети.

В двух файлах telcojan.txt (обучающая выборка) и telcofeb.txt (тестовая выборка) имеются статистические данные клиентах телекоммуникационной компании по состоянию на январь и февраль некоторого года.

Структура файлов одинакова и содержит уже отобранные аналитиками наиболее значимые характеристики, влияющие на уход клиента (отказ от пользования услугами): демографическая информация и используемый тариф.

№ п/п	Имя поля	Описание
1	Срок клиента	Количество месяцев, прошедших с момента подключения
2	Возраст	Возраст клиента
3	Адрес	Срок проживания по последнему адресу
4	Стаж	Стаж работы общий

5	Семейное положение	Семейное положение клиента
6	Образование	Уровень образования
7	Пенсионер	–
8	Пол	–
9	Тариф	Текущий тариф клиента
10	Уход	Отказ от пользования услугами (соответствует значению 1)

1. Постройте модель «Дерево решений».
2. Используя визуализатор «Таблица сопряженности», определите Чувствительность и Специфичность полученной модели.
3. Оцените качество модели. Сравните ее с моделью, полученной на предыдущем этапе.
4. С помощью визуализатора «Значимость атрибутов» охарактеризуйте, степень нелинейной зависимости между каждым фактором, включенным в модель, и независимой переменной.
5. Постройте модель «Нейронная сеть». Используйте алгоритм BACKPROP.
6. Используя визуализатор Таблица сопряженности, определите Чувствительность и Специфичность полученной модели.
7. Оцените качество модели. Сравните ее с моделью, полученной на предыдущем этапе.
8. Переобучите нейронную сеть
 - алгоритмом RPOP;
 - изменив количество скрытых слоев и количество нейронов в каждом слое;
 - уменьшив количество входных переменных (в этом случае следует определить новое значение требуемого количества нейронов).
 Удалось ли значительно улучшить общую ошибку классификации?
9. Сравните качество моделей между собой и с моделью, построенной на шаге Дерево решений.
10. Сформулируйте рекомендации по использованию результатов моделирования в работе компании.

Типовой вариант практической контрольной работы:

Предлагается решить задачу из области коллекторского скоринга на тему «Оптимизация выбора стратегии взыскания просроченной задолженности».

В файле debtors.txt имеются статистические данные о заемщиках банка, которые стали должниками, то есть имеют просроченную задолженность. Структура полей файла приведена в следующей таблице.

Имя поля	Описание	Тип данных
Код	Код заемщика	Целый
Сумма кредита	Сумма кредита в рублях	Вещественный
Срок кредита	Срок кредита в мес. (6, 12, 18)	Целый
Ежемесячный платеж	Размер аннуитетного ежемесячного платежа в рублях	Вещественный
Возраст	Возраст, лет	Целый
Пол	Пол клиента	Строковый
Срок просрочки	Срок просрочки в днях	Целый
Число платежей до выхода на просрочку	Сколько сделал заемщик платежей до того, как была впервые зафиксирована просроченная задолженность	Целый
Исполнительный лист	Имеется ли по заемщику исполни-	0 / 1

	тельный лист (судебное решение)	
Остаток долга (тело)	Сумма в рублях остатка «тела» кредита	Вещественный
Остаток долга (проценты)	Сумма в рублях остатка «процентов» по кредиту	Вещественный
Штрафы	Начисленные штрафы и пенни	Вещественный
Просрочка/Долг	Отношение, рассчитанное на основе трех предыдущих полей по формуле: $(12) / (11 + 10)$	Вещественный
Возобновление платежей	Начал ли заемщик снова оплачивать кредит после работы специалиста коллекторского подразделения: 0 – работа велась, и возобновления платежей не последовало, 1 – работа велась, и последовало возобновление платежей, «пусто» – работа не велась	Целый
Тестовое множество	Признак участия примера в тестовом множестве	Логический

Известна следующая дополнительная информация:

Ориентировочная стоимость плана мероприятий по взысканию просроченной задолженности с 1 заемщика оценивается в 1 000 рублей (зарплата персоналу, оплата звонков, выездов и т.д.)

Средняя ожидаемая сумма погашения от 1 должника 12 000 рублей.

Задание:

1. Построить модели, позволяющие классифицировать должников на 2 класса в зависимости от исхода мероприятий, проводимого специалистом коллекторского подразделения банка (логистическая регрессия, дерево решений, многослойный персептрон). Разделение на обучающее и тестовое множество уже проведено – используйте поле Тестовое множество. За положительный исход примите Возобновление платежей = 1.
2. Построить LIFT-кривые для моделей (там, где это возможно) и сравнить модели.
3. Используя дополнительную информацию, выбрать лучшую модель с двух позиций:
 - Максимизация ожидаемых поступлений от должников;
 - Минимизация рабочего времени, которое будет потрачено на мероприятия по взысканию.
4. Ответьте на вопрос: какой из алгоритмов оказался наименее чувствителен к несбалансированности классов? Комментарии впишите к любому узлу-обработчику.
5. Объедините лучшие модели в Ансамбль моделей.
6. «Прогоните» через ансамбль моделей тех должников, к которым еще не применялись мероприятия по взысканию и получите список клиентов, с которыми нужно работать, ранжированный по убыванию вероятности положительного исхода.
7. Все важные визуализаторы вывести на панель отчетов. Свои комментарии впишите к любому узлу-обработчику.

Типовые оценочные материалы по теме 5 «Кластеризация»

Опрос

1. Дайте формальную постановку задачи кластеризации.
2. Каковы цели кластеризации?
3. Что такое метрика расстояния?
4. Каковы основные шаги алгоритма k-means?

5. Как в алгоритме k-means определяется принадлежность точки к тому или иному кластеру?
6. Что такое центроид (центр тяжести кластера) и какова его роль в алгоритме k-means? Что является условием останова алгоритма k-means?
7. Каковы преимущества и недостатки алгоритма k-means?
8. Какова структура сети Кохонена? Какова роль входных и выходных нейронов карты?
9. В чем заключаются процедуры конкуренции, объединения и подстройки весов?
10. По какому правилу производится подстройка весов нейронов карты?
11. Что такое радиус обучения и какую роль он играет в процессе построения карты?
12. Чем определяется требуемое число выходных нейронов сети Кохонена?

Практико-ориентированное задание (компьютерный практикум) «Сегментация клиентов компании»

Легенда: руководство отдела маркетинга сети книжных магазинов по продаже книг различной направленности (от деловой литературы до художественной и развлекательной) поставило задачу сегментации своих постоянных покупателей, выявления их предпочтений. Предполагается, что это будет использоваться в будущем для оптимизации ассортимента книжных магазинов, проведения рекламных акций и для других целей.

В качестве исходных данных для сегментации выступают данные о постоянных покупателях книжной сети, у которых имеются дисконтные карты. Благодаря ним, о них известна следующая информация: пол, возраст и осуществленные покупки.

Задание. Вы выступаете в роли бизнес-аналитика. В вашем распоряжении имеются данные клиентах и их покупках за последний год (файл bookstore.txt).

Название поля	Описание
Код клиента	Уникальный код клиента
Возраст	Возраст клиента
Пол	Пол клиента
Посещений за год	Число посещений клиентом книжных магазинов, закончившихся покупкой товаров
Книги категории А	Число приобретенных за год клиентом книг из категории А
Книги категории В	Число приобретенных за год клиентом книг из категории В
Книги категории С	Число приобретенных за год клиентом книг из категории С
Книги категории D	Число приобретенных за год клиентом книг из категории D
Книги категории E	Число приобретенных за год клиентом книг из категории E
Среднее число книг в чеке	Среднее число книг в чеке (за год)
Средняя сумма чека	Средняя сумма чека (за год)

Для удобства проведения анализа все книги разделены на пять основных категорий.

Категория книжной продукции	Описание
Книги категории А	Деловая, научная, техническая и учебная литература
Книги категории В	Художественная литература
Книги категории С	Книги по литературе и искусству и культуре
Книги категории D	Народная медицина и здоровье
Книги категории E	Детективы, любовные романы, фантастика

Перед вами стоит задача сегментации клиентской базы при помощи самоорганизующихся карт Кохонена.

1. Разбейте всех клиентов на сегменты. Перед построением аналитической модели рекомендуется ответить на следующие вопросы:
 - Какие признаки включить в алгоритм кластеризации?
 - Какие их признаков более значимые, какие – нет?
 - Какой размер карты Кохонена выбрать?

2. Проанализировав карты, определите, какое количество кластеров дает самую наглядную картину?
3. Оцените емкость каждого сегмента (мощность кластера).
4. Охарактеризуйте каждый сегмент и дайте ему краткое название.
5. Постарайтесь описать социальный портрет покупателей каждого сегмента. Дайте название каждому сегменту. При интерпретации кластеров вы можете привлекать любую статистическую и (или) графическую информацию по кластеру: количество объектов, среднее, минимум, максимум и т.д.

Типовые оценочные материалы по теме 7 «Анализ и прогнозирование временных рядов»

Опрос

1. Детерминированная и случайная составляющие временного ряда.
2. Модели временных рядов.
3. Компоненты временного ряда.
4. Исследование временных рядов и автокорреляция.
5. В чем заключается принцип построения ретропрогноза?
6. Оценка адекватности выбранного метода прогнозирования.
7. Методы сглаживания.
8. Прогнозирование методом скользящего среднего.
9. Декомпозиция временного ряда.
10. Регрессионный анализ временных рядов.

Практико-ориентированное задание (компьютерный практикум) «Прогнозирование объема продаж»

Легенда: оптовая компания занимается сбытом строительных материалов, ассортимент насчитывает несколько тысяч торговых позиций, объединенных в группы (сухие смеси, плитка, напольные покрытия, грунтовка и т.п.)

Перед вами поставлена задача: автоматизировать ежемесячный расчет потоварного прогноза на следующие три периода.

1. Подключить хранилище данных materials.gdb
2. Для импорта выбрать только процесс Продажи
3. Выбрать для импорта все измерения, атрибуты и факты процесса Продажи, для фактов агрегирующей функцией является Сумма.
4. Определить срез для выбранных измерений. Типовой сценарий будет построен на примере товарной группы «Грунтовка». Установить фильтр «Кроме последнего периода», т.к. данные текущего, еще не завершившегося месяца, в построении модели участия не принимают.
5. Для последующей группировки по месяцам выполнить Преобразование даты к виду «Дата (Год+Месяц)».
6. Выполнить Группировку по измерению «Дата (Год+Месяц)» и факту «Сумма» и/или «Количество».
7. Построить график продаж. Оценить, содержит ли Кривая продаж шумы и выбросы, которые необходимо удалить для получения более качественного прогноза. Для сглаживания данных воспользоваться «Парциальной обработкой».

Парциальная обработка может включать в себя удаление аномалий и сглаживание (фильтрацию) данных (если требуется). При настройке парциальной обработки указать для поля Количество:

- Редактирование аномальных значений с о Степенью подавления (Малая),
- Сглаживание данных с Полосой пропускания — 80.

Отобразить результат преобразования данных на диаграмме. Оценить качество сглаживания.

8. С помощью обработчика «Автокорреляция» определить сезонность продаж данной группы товара. Какие значения АКФ говорят о наличии Тренда и наличии Сезонности?
9. Для построения модели прогноза воспользоваться инструментом «Скользящее окно». Скользящее окно — преобразование временного ряда к таблице смежных значений.
10. Построить несколько моделей временных рядов.
В качестве моделей использовать:
 - «Наивную» модель скользящего среднего за предыдущие три месяца.
 - Линейную регрессию с периодами (1,2,3) и (1,2,12).
 - Нейронную сеть периодами (1,2,12) — Рекомендуется перебрать 2-3 модели нейронной сети, изменяя число нейронов в скрытом слое.
 В качестве визуализаторов выбрать Диаграмму рассеивания и Диаграмму, на которой отобразить Количество и Количество_OUT (ретропрогноз).
11. По диаграммам оценить качество моделирования.
12. Вычислить среднюю процентную ошибку для каждой модели. Оценить качество моделирования.
13. К узлу лучшей модели применить обработчик «Прогнозирование». Обработчик Прогнозирование — позволяет получать предсказание значений временного ряда на число отсчетов, соответствующее заданному горизонту прогнозирования.
14. К узлу Прогнозирование применить Фильтр: Шаг прогноза = Не пустой.
15. После получения прогнозных цифр по группе товаров произвести разгруппировку прогнозных значений для получения прогноза для каждой товарной позиции (обработчик «Разгруппировка»).

Шкала оценивания

Уровень знаний, умений и навыков обучающегося при устном ответе во время проведения текущего контроля определяется баллами в диапазоне 0-100 %. Критериями оценивания при проведении устного опроса является демонстрация основных теоретических положений, в рамках осваиваемой компетенции, умение применять полученные знания на практике, овладение навыками анализа и систематизации информации в области финансов.

Тестирование

Уровень знаний, умений и навыков обучающегося при устном ответе во время проведения текущего контроля определяется баллами в диапазоне 0 – 100%. Критерием оценивания при проведении тестирования, является количество верных ответов, которые дал студент на вопросы теста. При расчете количества баллов, полученных студентом по итогам тестирования, используется следующая формула:

$$B = \frac{B}{O} \times 100\%$$

где B – количество баллов, полученных студентом по итогам тестирования;

B – количество верных ответов, данных студентом на вопросы теста;

O – общее количество вопросов в тесте.

Решение задач

Уровень знаний, умений и навыков обучающегося при устном ответе во время проведения текущего контроля определяется баллами в диапазоне 0 – 100 %. Критерием оценивания при проведении тестирования, является количество верных ответов, которые дал студент на вопросы теста. При расчете количества баллов, полученных студентом по итогам тестирования, используется следующая формула:

$$B = \frac{B}{O} \times 100\%$$

где Б – количество баллов, полученных студентом по итогам тестирования;

В – количество верных ответов, данных студентом на вопросы теста;

О – общее количество вопросов в тесте.

5. Оценочные материалы промежуточной аттестации по дисциплине

5.1. Методы проведения зачета

Зачет с оценкой проводится с применением следующих методов: метод устного опроса по вопросам из перечня примерных вопросов из п. 5.2. и выполнение практических заданий.

5.2. Оценочные материалы промежуточной аттестации

Компонент компетенции	Промежуточный / ключевой индикатор оценивания	Критерий оценивания
ПКс ОС П-5.2 Способность анализировать и визуализировать данные о рисках и их потенциальных источниках	- представляет анализ и визуализацию данных о рисках и их потенциальных источниках; - делает выводы на основе проведенного анализа.	Знает тенденции развития современных технических средств и информационных технологий, решает задачи ассоциации, кластеризации, классификации, регрессии, задачи анализа и прогнозирования временных рядов в среде компьютерных программ.

Типовые оценочные средства промежуточной аттестации

Вопросы к зачету с оценкой по дисциплине «Анализ и визуализация данных»

1. Аналитический и информационный подход к моделированию
2. Основные этапы построения моделей
3. Формы представления, типы и виды данных
4. Классификация задач Data Mining
5. Виды источников данных
6. Хранилища данных
7. Многомерная модель данных
8. Оценка качества данных
9. Предобработка данных
10. Фильтрация данных
11. Обработка дубликатов и противоречий
12. Выявление аномальных значений
13. Восстановление пропущенных значений
14. Сокращение размерности
15. Трансформация данных
16. Группировка данных
17. Слияние данных
18. Квантование
19. Нормализация и кодирование
20. Цели и задачи визуализации данных в аналитических технологиях
21. Виды визуализаторов данных
22. OLAP-анализ
23. Задача поиска ассоциаций
24. Значимость и полезность ассоциативных правил
25. Алгоритм apriori

26. Задача кластеризации
27. Алгоритмы решения задачи кластеризации
28. Алгоритм k-means
29. Сети и карты Кохонена
30. Задача классификации
31. Статистические методы решения задачи классификации
32. Деревья решений
33. Нейронные сети
34. Эффект переобучения
35. Оценка качества классификационных моделей
36. Модели временных рядов
37. Методы сглаживания временных рядов
38. Выбор метода прогнозирования
39. Оценка адекватности выбранного метода прогнозирования
40. Ансамбли моделей

Шкала оценивания

Уровень знаний, умений и навыков обучающегося при устном ответе во время промежуточной аттестации определяется оценками «зачтено», «не зачтено», «отлично», «хорошо», «удовлетворительно» и «неудовлетворительно». Критериями оценивания на зачете и зачете с оценкой является демонстрация основных теоретических положений в рамках осваиваемой компетенции, умение применять полученные знания на практике, овладение навыками решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

Для дисциплин, формой промежуточной аттестации которых является зачет с оценкой, приняты следующие соответствия:

- 90-100% – «отлично» (5);
- 75-89% – «хорошо» (4);
- 60-74% – «удовлетворительно» (3);
- менее 60% – «неудовлетворительно» (2).

При оценивании результатов обучения используется следующая шкала оценок:

100% – 90% (отлично)	Этапы компетенции, предусмотренные образовательной программой, сформированы на высоком уровне. Свободное владение материалом, выявление межпредметных связей. Уверенное владение понятийным аппаратом дисциплины. Практические навыки профессиональной деятельности сформированы на высоком уровне. Способность к самостоятельному нестандартному решению практических задач
89% – 75% (хорошо)	Этапы компетенции, предусмотренные образовательной программой, сформированы достаточно. Детальное воспроизведение учебного материала. Практические навыки профессиональной деятельности в значительной мере сформированы. Присутствуют навыки самостоятельного решения практических задач с отдельными элементами творчества.
74% – 60% (удовлетворительно)	Этапы компетенции, предусмотренные образовательной программой, сформированы на минимальном уровне. Наличие минимально допустимого уровня в усвоении учебного материала, в т.ч. в самостоятельном решении практических задач. Практические навыки профессиональной деятельности сформированы не в

	полной мере.
менее 60% (неудовлетворительно)	Этапы компетенции, предусмотренные образовательной программой, не сформированы. Недостаточный уровень усвоения понятийного аппарата и наличие фрагментарных знаний по дисциплине. Отсутствие минимально допустимого уровня в самостоятельном решении практических задач. Практические навыки профессиональной деятельности не сформированы.

Фонды оценочных средств промежуточной аттестации по дисциплине представлены в приложении 1.

6. Методические материалы по освоению дисциплины
Рекомендации по планированию и организации времени, необходимого на изучение дисциплины

Структура времени, необходимого на изучение дисциплины

Форма изучения дисциплины	Время, затрачиваемое на изучение дисциплины, %
Изучение литературы, рекомендованной в учебной программе	40
Решение задач, практических упражнений и ситуационных примеров	40
Изучение тем, выносимых на самостоятельное рассмотрение	20
Итого	100

Методические рекомендации по подготовке к практическому (семинарскому) занятию

Практическое (семинарское) занятие – одна из основных форм организации учебного процесса, представляющая собой коллективное обсуждение студентами теоретических и практических вопросов, решение практических задач под руководством преподавателя. Основной целью практического (семинарского) занятия является проверка глубины понимания студентом изучаемой темы, учебного материала и умения изложить его содержание ясным и четким языком, развитие самостоятельного мышления и творческой активности у студента. На практических (семинарских) занятиях предполагается рассматривать наиболее важные, существенные, сложные вопросы которые, наиболее трудно усваиваются студентами. При этом готовиться к практическому (семинарскому) занятию всегда нужно заранее. Подготовка к практическому (семинарскому) занятию включает в себя следующее:

- обязательное ознакомление с планом занятия, в котором содержатся основные вопросы, выносимые на обсуждение;
- изучение конспектов лекций, соответствующих разделов учебника, учебного пособия, содержания рекомендованных нормативных правовых актов;
- работа с основными терминами (рекомендуется их выучить);
- изучение дополнительной литературы по теме занятия, делая при этом необходимые выписки, которые понадобятся при обсуждении на семинаре;
- формулирование своего мнения по каждому вопросу и аргументированное его обоснование;
- запись возникших во время самостоятельной работы с учебниками и научной литературы вопросов, чтобы затем на семинаре получить на них ответы;
- обращение за консультацией к преподавателю.

Практические (семинарские) занятия включают в себя и специально подготовленные рефераты, выступления по какой-либо сложной или особо актуальной проблеме, решение задач. На практическом (семинарском) занятии студент проявляет свое знание предмета, корректирует информацию, полученную в процессе лекционных и внеаудиторных занятий, формирует определенный образ в глазах преподавателя, получает навыки устной речи и культуры дискуссии, навыки практического решения задач.

Рекомендации по изучению методических материалов

Методические материалы по дисциплине позволяют студенту оптимальным образом организовать процесс изучения данной дисциплины. Методические материалы по дисциплине призваны помочь студенту понять специфику изучаемого материала, а в конеч-

ном итоге – максимально полно и качественно его освоить. В первую очередь студент должен осознать предназначение методических материалов: структуру, цели и задачи. Для этого он знакомится с преамбулой, оглавлением методических материалов, говоря иначе, осуществляет первичное знакомство с ним. В разделе, посвященном методическим рекомендациям по изучению дисциплины, приводятся советы по планированию и организации необходимого для изучения дисциплины времени, описание последовательности действий студента («сценарий изучения дисциплины»), рекомендации по работе с литературой, советы по подготовке к зачету с оценкой и разъяснения по поводу работы с тестовой системой курса и над домашними заданиями. В целом данные методические рекомендации способны облегчить изучение студентами дисциплины и помочь успешно сдать зачет с оценкой. В разделе, содержащем учебно-методические материалы дисциплины, содержание практических занятий по дисциплине.

Рекомендации по самостоятельной работе студентов

Неотъемлемым элементом учебного процесса является самостоятельная работа студента. При самостоятельной работе достигается конкретное усвоение учебного материала, развиваются теоретические способности, столь важные для современной подготовки специалистов. Формы самостоятельной работы студентов по дисциплине: написание конспектов, подготовка ответов к вопросам, написание рефератов, решение задач, исследовательская работа, выполнение контрольной работы.

Задания для самостоятельной работы включают в себя комплекс аналитических заданий, выполнение которых предполагает тщательное изучение научной и учебной литературы, периодических изданий, а также законодательных и нормативных документов предлагаемых в п. 7 «Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине». Задания предоставляются на проверку в печатном виде.

Рекомендации по работе с литературой

При изучении курса учебной дисциплины особое внимание следует обратить на рекомендуемую основную и дополнительную литературу.

Важным элементом подготовки к семинару является глубокое изучение основной и дополнительной литературы, рекомендованной по теме занятия, а также первоисточников. При этом полезно прочитанную литературу законспектировать. Конспект должен отвечать трем требованиям: быть содержательным, по возможности кратким и правильно оформленным.

Содержательным его следует считать в том случае, если он передает все основные мысли авторов в целостном виде. Изложить текст кратко – это значит передать содержание книги, статьи в значительной мере своими словами. При этом следует придерживаться правила - записывать мысль автора работы лишь после того, как она хорошо понята. В таком случае поставленная цель будет достигнута. Цитировать авторов изучаемых работ (с обязательной ссылкой на источник) следует в тех случаях, если надо записывать очень важное определение или положение, обобщающий вывод.

Важно и внешнее оформление конспекта. В его начале надо указать тему семинара, дату написания, названия литературных источников, которые будут законспектированы. Глубокая самостоятельная работа над ними обеспечит успешное усвоение изучаемой дисциплины.

Одним из важнейших средств серьезного овладения теорией является **конспектирование первоисточников.**

Для составления конспекта рекомендуется сначала прочитать работу целиком, чтобы уяснить ее общий смысл и содержание. При этом можно сделать пометки о ее структуре, об основных положениях, выводах, надо стараться отличать в тексте основное от второстепенного, выводы от аргументов и доказательств. Если есть непонятные слова, надо в энциклопедическом словаре найти, что это слово обозначает. Закончив чтение (параграфа, главы, статьи) надо задать себе вопросы такого рода: В чем главная мысль? Каковы основные звенья доказательства ее? Что вытекает из утверждений автора? Как это согласуется с тем, что уже знаете о прочитанном из других источников?

Ясность и отчетливость восприятия текста зависит от многого: от сосредоточенности студента, от техники чтения, от настойчивости, от яркости воображения, от техники фиксирования прочитанного, наконец, от эрудиции – общей и в конкретно рассматриваемой проблеме.

Результатом первоначального чтения должен быть простой **план текста и четкое представление о неясных местах**, отмеченных в книге. После предварительного ознакомления, при повторном чтении следует **выделить основные мысли автора** и их развитие в произведении, обратить внимание на обоснование отдельных положений, на методы и формы доказательства, наиболее яркие примеры. В ходе этой работы окончательно отбирается материал для записи и определяется ее вид: **план, тезисы, конспект**.

План это краткий, последовательный перечень основных мыслей автора. Запись прочитанного в виде тезисов – это выявление и запись опорных мыслей текста. Разница между планом и тезисами заключается в следующем: в плане мысль называется (ставь всегда вопрос: о чем говорится?), в тезисах – формулируется – (что именно об этом говорится?). Запись опорных мыслей текста важна, но полного представления о прочитанном на основании подобной записи не составишь. Важно осмыслить, как автор доказывает свою мысль, как убеждает в истинности своих выводов. Так возникает конспект. Форма записи, как мы уже отметили, усложняется в зависимости от целей работы: план – о чем?; тезисы – о чем? что именно?; конспект – о чем? что именно? как?

Конспект – это краткое последовательное изложение содержания. Основу его составляет план, тезисы и выписки. Недостатки конспектирования: многословие, цитирование не основных, а связующих мыслей, стремление сохранить стилистическую связанность текста в ущерб его логической стройности. Приступать к конспектированию необходимо тогда, когда сложились навыки составления записи в виде развернутого подробного плана.

Форма записи при конспектировании требует особого внимания: важно, чтобы собственные утверждения, размышления над прочитанным, четко отделялись при записи. Разумнее выносить свои пометки на широкие поля, записывать на них дополнительные справочные данные, помогающие усвоению текста (дата события, упомянутого авторами; сведения о лице, названном в книге; точное содержание термина). Если конспектируется текст внушительного объема, необходимо указывать страницы книги, которые охватывает та или иная часть конспекта.

Для удобства пользования своими записями важно озаглавить крупные части конспекта, подчеркивая **заголовки**. Следует помнить о назначении красной строки, стремиться к четкой графике записей – уступами, колонками. Излагать главные мысли автора и их систему аргументов необходимо преимущественно своими словами, перерабатывая таким образом информацию, – так проходит уяснение ее сути. Мысль, фразы, понятые в контексте, могут приобрести более пространное изложение в записи. Но текст оригинала свертывается, и студент, отрабатывая логическое мышление, учится выделять главное и обобщать однотипные суждения, однородные факты. Кроме того, делая записи своими словами, обобщая, студент учится письменной речи.

Знание общей стратегии чтения, техники составления плана и тезисов определяет и технологию конспектирования:

- внимательно читать текст, попутно отмечая непонятные места, незнакомые термины и понятия. **Выписать на поля** значение отмеченных понятий.
- при первом чтении текста необходимо составить его **простой план**, последовательный перечень основных мыслей автора.
- при повторном чтении текста выделять **систему доказательств** основных положений работы автора.
- заключительный этап работы с текстом состоит в осмыслении ранее отмеченных мест и их краткой последовательной записи.
- при конспектировании нужно стремиться **выразить мысль автора своими словами**, это помогает более глубокому усвоению текста.
- в рамках работы над первоисточником важен умелый **отбор цитат**. Необходимо учитывать, насколько ярко, оригинально, сжато изложена мысль. Цитировать необходимо те суждения, на которые впоследствии возможна ссылка как на авторитетное изложение мнения, вывода по тому или иному вопросу.

Конспектировать целесообразно не на отдельном листе, а в общей тетради на одной странице листа. Обратная сторона листа может быть использована для дополнений, необходимость которых выяснится в дальнейшем. При конспектировании литературы следует оставить широкие поля, чтобы записать на них план конспекта. Поля могут быть использованы также для записи своих замечаний, дополнений, вопросов. При выступлении на семинаре студент может пользоваться своим конспектом для цитирования первоисточника. Все обучающиеся внимательно слушают выступления одногруппников, отмечают спорные или ошибочные положения в них, вносят поправки, представляют свои решения и обоснования обсуждаемых проблем.

В конце семинара, когда преподаватель подводит итоги занятия, студенты с учетом рекомендаций преподавателя и выступлений сокурсников дополняют или исправляют свои конспекты.

Рекомендации по изучению методических материалов

Методические материалы по дисциплине позволяют студенту оптимальным образом организовать процесс изучения данной дисциплины. Методические материалы по дисциплине призваны помочь студенту понять специфику изучаемого материала, а в конечном итоге – максимально полно и качественно его освоить. В первую очередь студент должен осознать предназначение методических материалов: структуру, цели и задачи. Для этого он знакомится с преамбулой, оглавлением методических материалов, говоря иначе, осуществляет первичное знакомство с ним. В разделе, посвященном методическим рекомендациям по изучению дисциплины, приводятся советы по планированию и организации необходимого для изучения дисциплины времени, описание последовательности действий студента («сценарий изучения дисциплины»), рекомендации по работе с литературой, советы по подготовке к зачету с оценкой и разъяснения по поводу работы с тестовой системой курса и над домашними заданиями. В целом данные методические рекомендации способны облегчить изучение студентами дисциплины и помочь успешно сдать зачет с оценкой. В разделе, содержащем учебно-методические материалы дисциплины, содержание практических занятий по дисциплине, словарь основных терминов дисциплины.

Рекомендации для подготовки к зачету с оценкой

При подготовке к зачету с оценкой студент внимательно просматривает вопросы, предусмотренные рабочей программой, и знакомится с рекомендованной основной литературой. Основой для сдачи зачета с оценкой студентом является изучение конспектов

лекций, прослушанных в течение семестра, информация, полученная в результате самостоятельной работы в течение семестра.

7. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет», учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине

7.1. Основная литература.

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511020>
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511121>

7.2. Дополнительная литература.

1. Подкорытова, О. А. Анализ временных рядов : учебное пособие для вузов / О. А. Подкорытова, М. В. Соколов. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2020. — 267 с. — (Высшее образование). — ISBN 978-5-534-02556-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/450587>
2. Чубукова, И. А. Data Mining: учебное пособие / И. А. Чубукова. — 3-е изд. — Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 469 с. — ISBN 978-5-4497-0289-0. — Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/89404.html>

7.3. Учебно-методическое обеспечение самостоятельной работы.

Не предусмотрено.

7.4. Нормативные правовые документы.

1. Федеральный закон Российской Федерации от 27 июля 2006 г. N 149-ФЗ Об информации, информационных технологиях и о защите информации, Собрание Законодательства РФ. Официальное издание. 2006 г.
2. Федеральный закон «Об информации, информатизации и защите информации», от 20 февраля 1995 Г. N 24-ФЗ (с изменениями от 10 января 2003 г.)

7.5. Интернет-ресурсы.

1. www.biblio-online.ru — Электронно-библиотечная система [ЭБС] Юрайт;
2. <http://www.iprbookshop.ru> — Электронно-библиотечная система [ЭБС] «Iprbooks»
3. <https://e.lanbook.com> — Электронно-библиотечная система [ЭБС] «Лань».
4. <https://dlib.eastview.com> — East View Information Services, Inc. Коллекции

- электронных научных и практических журналов.
5. <http://elibrary.ru/> — Научная электронная библиотека Elibrary.ru.
 6. <http://www.intuit.ru> — Национальный Открытый Университет «ИНТУИТ»;
 7. <http://www.planetaexcel.ru> — сайт о возможностях MS Excel (приемы, видео-уроки, книги);
 8. http://www.clouderp.ru/tags/BUSINESS_INTELLIGENCE/ Что такое Business Intelligence? Обзор BI систем
 9. <http://netpeak.net/ru/blog/kak-rabotat-s-microsoft-power-bi-podrobnoe-rukovodstvo/> Как работать с Microsoft Power BI — подробное руководство
 10. <http://community.powerbi.com/> Форум (англ. яз.) по Microsoft Power BI. Примеры применения. Ответы на вопросы. Использование на мобильных устройствах и др.
 11. <https://www.microsoft.com/ru-RU/download/details.aspx?id=45331> – ссылка для скачивания Microsoft Power BI;
 12. <https://docs.microsoft.com/ru-ru/power-bi/desktop-getting-started> – техническая документация Microsoft, документация Power BI;
 13. <https://support.office.com/ru-ru/article/Вход-в-sway-db30c63b-6548-456d-b0e29fcd68bfa7a0?ui=ru-RU&rs=ru-RU&ad=RU> – руководство по созданию презентации в MS Sway;
 14. <http://www.gks.ru> – официальный сайт Федеральной службы государственной статистики;
 15. <http://moscow.gks.ru> – Территориальный орган Федеральной службы государственной статистики по городу Москве;
 16. <https://data.mos.ru> – портал открытых данных Правительства Москвы

7.6. Иные источники

8. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

Материально-техническое обеспечение дисциплины включает в себя:

- лекционные аудитории, оборудованные видеопроекционным оборудованием для презентаций, средствами звуковоспроизведения, экраном;
- помещения для проведения семинарских и практических занятий, оборудованные учебной мебелью;

Дисциплина должна быть поддержана соответствующими лицензионными программными продуктами.

Программные средства обеспечения учебного процесса должны включать:

- программы презентационной графики;
- текстовые и табличные редакторы.

Вуз обеспечивает каждого обучающегося рабочим местом в компьютерном классе в соответствии с объемом изучаемых дисциплин, обеспечивает выход в сеть Интернет.

Помещения для самостоятельной работы обучающихся включают следующую оснащенность: столы аудиторные, стулья, доски аудиторные, компьютеры с подключением к локальной сети института (включая правовые системы) и Интернет.

Для изучения учебной дисциплины используются автоматизированная библиотечная информационная система и электронные библиотечные системы: «Университетская библиотека ONLINE», «Электронно-библиотечная система издательства ЛАНЬ», «Электронно-библиотечная система издательства «Юрайт», «Электронно-библиотечная система IPRbooks», «Научная электронная библиотека eLIBRARY» и др.

Приложение 1
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

Волгоградский институт управления – филиал РАНХиГС

Экономический факультет

кафедра информационных систем и математического моделирования

УТВЕРЖДЕНА
учёным советом
Волгоградского института управления – филиала
РАНХиГС
Протокол № 2 от 23.09.2021 г.

ПРОГРАММА БАКАЛАВРИАТА

Финансы и кредит

(наименование образовательной программы)

ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

Б1.В.04 АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

(индекс, наименование дисциплины (модуля), в соответствии с учебным планом)

краткое наименование дисциплины (модуля) (при наличии)

38.03.01 Экономика

(код, наименование направления подготовки (специальности))

очная, заочная

(форма(ы) обучения)

Год набора – 2022

Волгоград, 2021 г.

1. Вопросы к зачету с оценкой по дисциплине «Анализ и визуализация данных»

1. Аналитический и информационный подход к моделированию
2. Основные этапы построения моделей
3. Формы представления, типы и виды данных
4. Классификация задач Data Mining
5. Виды источников данных
6. Хранилища данных
7. Многомерная модель данных
8. Оценка качества данных
9. Предобработка данных
10. Фильтрация данных
11. Обработка дубликатов и противоречий
12. Выявление аномальных значений
13. Восстановление пропущенных значений
14. Сокращение размерности
15. Трансформация данных
16. Группировка данных
17. Слияние данных
18. Квантование
19. Нормализация и кодирование
20. Цели и задачи визуализации данных в аналитических технологиях
21. Виды визуализаторов данных
22. OLAP-анализ
23. Задача поиска ассоциаций
24. Значимость и полезность ассоциативных правил
25. Алгоритм apriori
26. Задача кластеризации
27. Алгоритмы решения задачи кластеризации
28. Алгоритм k-means
29. Сети и карты Кохонена
30. Задача классификации
31. Статистические методы решения задачи классификации
32. Деревья решений
33. Нейронные сети
34. Эффект переобучения
35. Оценка качества классификационных моделей
36. Модели временных рядов
37. Методы сглаживания временных рядов
38. Выбор метода прогнозирования
39. Оценка адекватности выбранного метода прогнозирования
40. Ансамбли моделей

2. Тестовые задания

1. Аналитик это ...
 - а) специалист в области анализа и моделирование
 - б) специалист в предметной области;
 - в) человек, решающий определенные задачи;
 - г) человек, который имеет опыт в программировании.

2. Эксперт это ...
 - а) специалист в области анализа и моделирование;
 - б) специалист в предметной области;
 - в) человек, решать определенные задачи;
 - г) человек, который имеет опыт в программировании.
3. Задача классификации сводится к ...
 - а) нахождения частых зависимостей между объектами или событиями;
 - б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекта значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
4. Задача регрессии сводится к ...
 - а) нахождения частых зависимостей между объектами или событиями;
 - б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекта значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
5. Задача кластеризации заключается в ...
 - а) нахождения частых зависимостей между объектами или событиями;
 - б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекта значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
6. Целью поиска ассоциативных правил является ...
 - а) нахождения частых зависимостей между объектами или событиями;
 - б) определения класса объекта по его характеристиками;
 - в) определение по известным характеристиками объекта значение некоторого его параметра;
 - г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
7. До предполагаемых моделей относятся такие модели данных:
 - а) модели классификации и последовательностей;
 - б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - в) классификации, кластеризации, исключений, итоговые и ассоциации;
 - г) модели классификации, последовательностей и исключений.
8. В описательных моделей относятся следующие модели данных:
 - а) модели классификации и последовательностей;
 - б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
 - в) классификации, кластеризации, исключений, итоговые и ассоциации;
 - г) модели классификации, последовательностей и исключений.
9. Модели классификации описывают ...
 - а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 - б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 - в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 - г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.
10. Модели последовательностей описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

11. Регрессивные модели описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

12. Виды лингвистической неопределенности:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами;
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);
- в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость)
- г) неоднозначность смысла фраз (Синтаксическая и семантическая).

13. Модели исключений описывают ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

14. Итоговые модели обнаружат ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

15. Модели ассоциации проявляют ...

- а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;
- б) ограничения на данные анализируемого массива;
- в) закономерности между связанными событиями;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

16. Виды физической неопределенности данных:

- а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью)
- б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);

в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);

г) неоднозначность смысла фраз (Синтаксическая и семантическая).

17. Очистка данных — ...

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.

18. Обогащение — ...

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.

19. Консолидация — ...

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.

20. Транзакция — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, 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.

21. Метаданные — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, 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

22. Классификация – ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, 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.

23. Регрессия – ...

а) это установление зависимости непрерывной выходной переменной от входных переменных;

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;

в) выявление закономерностей между связанными событиями;

г) это установление зависимости дискретной выходной переменной от входных переменных.

24. Кластеризация – ...

а) это установление зависимости непрерывной выходной переменной от входных переменных;

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;

в) выявление закономерностей между связанными событиями;

г) это установление зависимости дискретной выходной переменной от входных переменных.

25. Ассоциация – ...

а) это установление зависимости непрерывной выходной переменной от входных переменных;

б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов;

в) выявление закономерностей между связанными событиями;

г) это установление зависимости дискретной выходной переменной от входных переменных.

26. Машинное обучение – ...

а) специализированный программный решение (или набор решений), который 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 изучающий методы построения алгоритмов, способных обучаться на данных.

27. Аналитическая платформа – ...

- а) специализированный программный решение (или набор решений), который 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.

28. Обучающая выборка – ...

- а) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов;
- б) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат;
- в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат;
- г) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

29. Ошибка обучения – ...

- а) это ошибка, допущенная моделью на учебной множества;
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества;
- в) имена, типы, метки и назначения полей исходной выборки данных;
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат.

30. Ошибка обобщения – ...

- а) это ошибка, допущенная моделью на учебной множества;
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества;
- в) имена, типы, метки и назначения полей исходной выборки данных;
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат.

Ответы

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
А	Б	Б	В	Г	А	А	Б	А	Б	В	Б	А	Б	В
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
А	А	Б	Г	А	В	Г	А	Б	В	Г	А	Б	А	Б