Приложение 7 ОП ВО

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

ОМСКИЙ ФИЛИАЛ

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 6664D3EAA7DADC690EE8BAC024716D1B
Владелец: Бахта Андрей Сергеевич
Действителен: с 13.06.2024 до 06.09.2025

УТВЕРЖДЕНО

Директор филиала А.С. Бахта

(электронная подпись)

ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА ВЫСШЕГО ОБРАЗОВАНИЯ

ФИНАНСЫ И КРЕДИТ
(наименование образовательной программы)

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

Б1.В.05 Анализ и визуализация данных
(индекс, наименование дисциплины (модуля), в соответствии с учебным планом)

Высшее, бакалавриат
(уровень образования)

38.03.01 Экономика
(код, наименование направления подготовки (специальности)

Бакалавр
(квалификация)

Очная, очно-заочная
(форма(ы) обучения)

Год набора: 2022

Омск, 2024 г.

Авторы-составители:
преподаватель кафедры экономики и управления Е.А. Трушакова

Заведующий кафедрой экономики и управления, к.э.н., доцент
С.С. Стаурский

Начальник учебного отдела
Е.В. Малахова

Рабочая программа дисциплины «Анализ и визуализация данных» рассмотрена и утверждена на заседании Педагогического совета, протокол от 26 января 2022 г. № 4.

Рабочая программа дисциплины «Анализ и визуализация данных» пересмотрена и одобрена на заседании кафедры экономики и управления, протокол от 30 августа 2023 г. № 1, протокол от 30 августа 2024 г. № 1.

Содержание

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы....4
2. Объем и место дисциплины в структуре ОП ВО 5
3. Содержание и структура дисциплины....6
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине....7
5. Методические указания для обучающихся по освоению дисциплины....27
6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине....28
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы....30

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

1.1. Дисциплина Б1.В.05 Анализ и визуализация данных обеспечивает овладение следующими компетенциями:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс ОС II – 6	Способен применять информационные технологии для решения прикладных финансово-экономических задач	ПКс ОС II – 6.2	Способен применять современные информационные технологии для практической работы в экономических информационных системах и решения аналитических задач

1.2. В результате освоения дисциплины у студентов должны быть сформированы:

ОТФ/ТФ (при наличии профстандарта)	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 – 6.2	**На уровне знаний:** 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 обеспечение 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.
		На уровне умений: 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 ассоциации, кластеризации, классификации, регрессии, задачи анализа и прогнозирования временных рядов; умение строить аналитическую отчетность; умение

ОТФ/ТФ (при наличии профстандарта)	Код этапа освоения компетенции	Результаты обучения
		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.
		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 Loginom и Power BI; навыки решения практических задач в среде ERP–системы[1].

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

Дисциплина Б1.В.05 Анализ и визуализация данных составляет 4 зачетные единицы, т.е. 144 академических часа.

На контактную работу с преподавателем выделено 64 часа, из 16 часов лекций и 32 часа практических занятий, на самостоятельную работу обучающихся выделено 92 часов, на промежуточную аттестацию – 4 часа для очной ф/о.

На контактную работу с преподавателем выделено 24 часа, из них 16 часов практических занятий и 8 часов лекций, на самостоятельную работу обучающихся выделено 116 часов для очно-заочной ф/о.

Место дисциплины в структуре ОП ВО

Дисциплина Б1.В.05 Анализ и визуализация данных изучается на 2 курсе, в 4 семестре для студентов очной, очно-заочной форм обучения.

Освоение дисциплины Б1.В.05 Анализ и визуализация данных происходит после освоения дисциплин Б1.В.04 Экономическая информатика, Б1.О.06 Алгоритмизация и программирование. Изучение дисциплины заканчивается зачетом с оценкой.

[1] Свободное программное обеспечение:
1. Orange (свободное ПО) https://orangedatamining.com
2. Loginom Community (свободное ПО) https://loginom.ru
3. MS Power BI Desktop (свободное ПО) https://powerbi.microsoft.com/ru-ru/
4. А также прикладных языков программирования, например, Python (свободное ПО, https://www.online-python.com/; https://playground.programiz.com/), R.

3. Содержание и структура дисциплины

Очная форма обучения

№ п/п	Наименование тем (разделов)	Объем дисциплины, ак. час./ час.						Форма текущего контроля успеваемости**, промежуточной аттестации***
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Тема 1	Подготовка данных к анализу	**18**	2		4		12	*О, ПОЗ*
Тема 2	Визуализация данных	**50**	6		12		32	*ПОЗ, ПКР*
Тема 3	Классификация и регрессия (Статистические методы)	**18**	2		4		12	*О, ПОЗ*
Тема 4	Классификация и регрессия (Машинное обучение)	**18**	2		4		12	*Т, ПОЗ, ПКР*
Тема 5	Кластеризация	**18**	2		4		12	*О, ПОЗ*
Тема 6	Анализ и прогнозирование временных рядов	**18**	2		4		12	*О, ПОЗ*
Промежуточная аттестация		**4**						*ЗаО*
Всего		**144**	**16**		**32**		**92**	

** – формы текущего контроля успеваемости: *опрос (О), тестирование (Т), практико-ориентированные задания (ПОЗ), практическая контрольная работа (ПКР).*

***- форма промежуточной аттестации*: зачет с оценкой (ЗаО)*

Очно-заочная форма обучения

№ п/п	Наименование тем (разделов)	Объем дисциплины, ак. час./ час.						Форма текущего контроля успеваемости**, промежуточной аттестации***
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий				СР	
			Л	ЛР	ПЗ	КСР		
Тема 1	Подготовка данных к анализу	18	-		2		16	*О, ПОЗ*
Тема 2	Визуализация данных	50	2		6		42	*ПОЗ, ПКР*
Тема 3	Классификация и регрессия (Статистические методы)	18	2		2		14	*О, ПОЗ*
Тема 4	Классификация и регрессия (Машинное обучение)	18	2		2		14	*Т, ПОЗ, ПКР*
Тема 5	Кластеризация	18	2		2		14	*О, ПОЗ*
Тема 6	Анализ и прогнозирование временных рядов	18	-		2		16	*О, ПОЗ*
Промежуточная аттестация		4						*ЗаО*
Всего		**144**	**8**		**16**		**116**	

** – формы текущего контроля успеваемости: *опрос (О), тестирование (Т), практико-ориентированные задания (ПОЗ), практическая контрольная работа (ПКР).*

***- форма промежуточной аттестации*: зачет с оценкой (ЗаО)*

Содержание дисциплины

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

4.1.1. Формы текущего контроля успеваемости

Тема (раздел)	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)

4.1.2. Материалы текущего контроля успеваемости

Типовые оценочные материалы по теме 1 «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. 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 неэффективно.

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 значение признаков должны быть приближенно стандартно распределены.

1. Соединить данные: new_v alue.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 2 «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

Отметьте поля Товарная группа (таблица Товары) и Сумма (таблица Продажи). Отсортируйте товарные группы по сумме выручки.

6. На панели Визуализации выберите Круговую диаграмму.

Отметьте поля Область (таблица География), Отдел и Сумма (таблица Продажи).

7. 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.

8. 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.

9. На панели Визуализации выберите Срез: поле Товарная группа (таблица Товары).

10. 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 3 «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 выдачи 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

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

C_{FN} — издержки, связанные с ложноотрицательными наблюдениями;
C_{FP} — издержки, связанные с ложноположительными наблюдениями.
Отношение издержек ${C_{FN}}/{C_{FP}}$ 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 4 «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 Apriori
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.

N 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.

7. Используя визуализатор Таблица сопряженности, определите Чувствительность и Специфичность полученной модели.

8. Оцените качество модели. Сравните ее с моделью, полученной на предыдущем этапе.

9. 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?

10. Сравните качество моделей между собой и с моделью, построенной на шаге Дерево решений.

11. 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 исполнительный лист (судебное решение)	0 / 1
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 –	Целый

	работа велась, и последовало возобновление платежей, «пусто» – работа не велась	
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 остановки 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 A	Число приобретенных за год клиентом книг из категории A
Книги категории B	Число приобретенных за год клиентом книг из категории B
Книги категории C	Число приобретенных за год клиентом книг из категории C
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 A	Деловая, научная, техническая и учебная литература
Книги категории B	Художественная литература
Книги категории C	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 6 «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»).

Критерии оценивания текущего контроля

Текущий контроль знаний студентов обеспечивает оперативное управление учебной деятельностью студента и проводится с целью:

– соответствия уровня и качества подготовки студентов образовательному стандарту Академии;

– проверки качества теоретических знаний по дисциплине;

– проверки наличия умений применять полученные теоретические знания при решении практических задач;

– проверки наличия умений самостоятельной работы с учебной литературой.

Текущий контроль знаний студентов предусматривает систематическую проверку качества полученных студентами знаний, умений и навыков по изучаемой дисциплине.

Оценивание знаний, умений и навыков по учебной дисциплине «Анализ и визуализация данных» осуществляется посредством использования следующих видов оценочных средств:

– устное собеседование (в рамках проведения опросов);

– тестирование;

– выполнение практико-ориентированных практических заданий с использованием компьютерных программ анализа и визуализации данных;

– контрольная работа.

Устное собеседование.

Устные собеседования проводятся во время практических занятий. Тематика обсуждаемых вопросов собеседования не должна выходить за рамки объявленной для данного занятия темы.

При оценке устного собеседования анализу подлежит точность формулировок, связность изложения материала, обоснованность суждений, опора на нормативные документы.

Тестирование.

Тестирование — учебная технология, позволяющая измерять знания, умения и навыки студентов, состоящая из тестовых заданий и формализованных процедур

проведения, обработки и анализа результатов; тесты предоставляются обучающимся в качестве раздаточного материала на бумажном или электронном носителе. При проведении тестирования обучающемуся запрещается пользоваться дополнительной литературой. На выполнение теста обучающемуся дается 30 минут.

Практические задания.

Практические задания осуществляются в форме выполнения практико-ориентированных заданий на компьютере с использованием компьютерных программ анализа и визуализации данных. После выполнения практического задания преподаватель проводит опрос по теме занятия.

Контрольная работа.

В течение семестра студент выполняет две практических контрольных работы с использованием компьютерных программ анализа и визуализации данных. Данный вид работы проверяет усвоение полученных в ходе обучения умений и навыков. Время выполнения контрольной работы составляет 2 часа. Успешное выполнение контрольной работы является обязательным для допуска к промежуточной аттестации.

Шкалы оценивания результатов текущего контроля

Шкала оценивания результатов устных собеседований на практических занятиях

Обозначения		Формулировка требований к степени освоения дисциплины
Цифр.	Оценка	
2	Неудовлетворительно	Студент показывает недостаточные знания программного материала, не способен аргументировано и последовательно его излагать, допускается грубые ошибки в ответах, неправильно отвечает на поставленный вопрос или затрудняется с ответом.
3	Удовлетворительно	Студент показывает достаточные, но не глубокие знания программного материала; при ответе не допускает грубых ошибок или противоречий, однако в формулировании ответа отсутствует должная связь между анализом, аргументацией и выводами. Для получения правильного ответа требуются уточняющие вопросы.
4	Хорошо	Студент показывает глубокие знания материала, грамотно его излагает, достаточно полно отвечает на поставленный вопрос и дополнительные вопросы, умело формулирует выводы. В то же время при ответе допускает несущественные погрешности.
5	Отлично	Студент показывает полные и глубокие знания программного материала, логично и аргументировано отвечает на поставленный вопрос, а также дополнительные вопросы, показывает высокий уровень теоретических знаний.

Шкала оценивания тестирования

Обозначения		Формулировка требований к степени освоения дисциплины
Цифр.	Оценка	
2	Неудовлетворительно	50% и меньше правильных ответов.
3	Удовлетворительно	51–69% правильных ответов.
4	Хорошо	70–84% правильных ответов
5	Отлично	85–100% правильных ответов.

Шкала оценивания результатов практических заданий на практических занятиях

Обозначения		Формулировка требований к степени освоения дисциплины
Цифр.	Оценка	
2	Неудовлетворительно	Студент неправильно решает практическую задачу, не делает выводов по ее результатам, не может объяснить технологию ее решения, показывает полное незнание теоретических аспектов, на дополнительные, уточняющие вопросы не отвечает.
3	Удовлетворительно	Студент допускает несколько незначительных ошибок в решении практической задачи, делает неполные выводы по ее результатам либо недостаточно аргументирует свое решение; отвечает на вопрос о технологии ее решения, но при ответе допускает неточности, что требует дополнительных вопросов.
4	Хорошо	Студент логично и правильно решает практическую задачу, делает грамотные выводы по ее результатам, отвечает на вопрос о технологии ее решения, достаточно аргументирует свое решение, но при ответе допускает погрешности.
5	Отлично	Студент логично и правильно решает практическую задачу, делает грамотные выводы по ее результатам, полно отвечает на вопрос о технологии ее решения, подробно аргументирует свое мнение со ссылками на норму закона, показывает хорошее знание теоретических аспектов.

Шкала оценивания выполнения контрольной работы

Обозначения		Формулировка требований к степени освоения дисциплины
Цифр.	Оценка	
2	Неудовлетворительно	Выполнение менее 50 % объема задания контрольной работы, наличие существенных замечаний, студент делает неправильные выводы по ее результатам либо не аргументирует свое решение.
3	Удовлетворительно	Выполнение не менее 50 % объема задания контрольной работы (возможно наличие замечаний), студент делает неполные выводы по ее результатам либо недостаточно аргументирует свое решение.
4	Хорошо	Выполнение не менее 70 % объема задания контрольной работы (возможно наличие несущественных замечаний), наличие в работе аргументированных выводов на основе проведенного анализа
5	Отлично	Выполнение не менее 85% объема задания контрольной работы без замечаний, наличие в работе аргументированных выводов на основе проведенного анализа.

4.2. Промежуточная аттестация

4.2.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

Показатели и критерии оценивания компетенций с учетом этапа их формирования

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс ОС II – 6	Способен применять информационные технологии для решения прикладных финансово-экономических задач	ПКс ОС II – 6.2	Способен применять современные информационные технологии для практической работы в экономических информационных системах и решения аналитических задач

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
ПКс ОС II – 6.2 Способен применять современные информационные технологии для практической работы в экономических информационных системах и решения аналитических задач	Умеет использовать для решения аналитических и управленческих задач информационные технологии	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

4.2.2. Форма и средства (методы) проведения промежуточной аттестации

Форма промежуточной аттестации в соответствии с учебным планом - зачет с оценкой.

4.2.3. Типовые оценочные средства

Зачет с оценкой по дисциплине «Анализ и визуализация данных» служит для оценки работы студента в течение всего времени обучения по данной дисциплине.

Зачет с оценкой проводится в конце семестра и призван выявить уровень, прочность и систематичность полученных студентом теоретических и практических знаний, приобретения навыков самостоятельной работы, развития творческого мышления, умение синтезировать полученные знания и применять их в решении практических задач.

Зачет осуществляется в форме устного ответа и решения практической задачи на компьютере. Задание билета включает в себя один теоретический вопрос и одно практическое задание для решения в компьютерных программах визуализации и анализа данных.

Пример типового билета к зачету с оценкой
по дисциплине «Анализ и визуализация данных»

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ
Омский филиал

Дисциплина «Анализ и визуализация данных»

Билет №1

1. Основные этапы построения моделей. KDD — методика извлечения знаний.

2. Задача.

Импортировать файл Loans.txt.

С помощью обработки «Нейронная сеть» построить модель, которая могла бы предсказывать наличие в будущем у заемщика просрочек по обслуживанию долга.

1. Оценить качество модели.
2. К какому классу будут отнесен заемщик

Мужчина, 45-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 8000 руб.

Заведующий кафедрой /_____/

Примеры задач для решения в компьютерных программах анализа и визуализации данных для промежуточной аттестации (практическая часть)

Вариант 1.

На основе данных хранилища данных «Банк» получить следующие отчеты и вынести их на Панель Отчетов.

1. Количество одобренных/отклоненных заявок по каждому филиалу и отделению на основе всех имеющихся данных.
2. 10 «наиболее злостных» неплательщиков по состоянию на 10.11.2019
3. Среднюю интенсивность поступления заявок по дням недели каждого из отделений банка на основании всех имеющихся данных. Отчет дополните соответствующим графиком.

Вариант 2.

Импортировать текстовый файл Loans_demo.txt.

С помощью обработки «Логистическая регрессия» построить модель, которая могла бы предсказывать наличие в будущем у заемщика просрочек по обслуживанию долга.

3. Оценить качество модели.
4. Чему равен оптимальный порог отсечения?
5. Как изменяется кредитный рейтинг заемщика с каждым дополнительным иждивенцем?
6. К какому классу будут отнесен заемщик

Мужчина, 43-х лет, состоит в браке, 3 иждивенца, 20500 доход, опыт работы 20, срок проживания 10, недвижимость 0, месячный платеж 6725.

Вариант 3.

Импортировать текстовый файл responses.txt.

С помощью обработок «Дерево решений» и «Нейронная» сеть построить модели, позволяющие предсказать отклик клиента на рассылку.

1. Сравнить качество моделей.
2. Какую модель для отбора респондентов из рассмотренных вы выберете, если известно, что бюджет рассылки составляет 95 000 руб.?
3. Скольким клиентам пришлось бы отправить письма, если бы мы проводили рассылку случайно? Какую экономию дает наша модель?

Вариант 4.

Импортировать текстовый файл mobile.txt.

1. Провести сегментацию абонентов оператора мобильной связи.
2. Построить профили абонентов путем выявления их схожего поведения в плане частоты, длительности и времени звонков, а также ежемесячных расходов;
3. Провести оценку наиболее и наименее доходных сегментов.

Вариант 5.

Исходные данные представлены в файле Чеки.txt.

Проанализировать данные и решить следующие задачи:

1. Предсказать то, какие товары покупатели могут выбрать в зависимости от того, что уже есть в их корзинах;
2. Выявить наиболее популярные товарные наборы, состоящие из более чем 1 предмета;
3. Предложить рекламные акции типа «каждому купившему A и B товар C в подарок».

Вариант 6.

На основе данных хранилища данных Materials

1. Построить модель, позволяющую, автоматизировать ежемесячный расчет потоварного прогноза на следующие три периода. Типовой сценарий построить на примере товарной группы «Грунтовка».
2. Оценить качество предложенной модели.

Список вопросов для подготовки к зачету с оценкой

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

4.3. Методические материалы

Процедура проведения зачета с оценкой

Зачет с оценкой осуществляется в форме устного ответа и решения практической задачи на компьютере. Задание билета включает в себя один теоретический вопрос и одно практическое задание для решения в компьютерных программах визуализации и анализа данных. При проведении зачета с оценкой билет выбирает сам обучающийся в случайном порядке.

Оценка результатов устного аттестационного испытания объявляется обучающимся в день его проведения. При проведении зачета с оценкой в аудитории могут одновременно находиться не более шести студентов. По окончании ответа на вопросы билета преподаватель может задать обучающемуся дополнительные и уточняющие вопросы в пределах учебного материала, вынесенного на зачет с оценкой.

При оценке знаний студента преподаватель должен руководствоваться приведенной в рабочей программе шкалой оценивания результатов, а также следующими критериями:

- правильность ответа на вопросы билета;
- полнота ответа, знание терминологии;
- логика и аргументированность ответа;
- способность делать выводы.
-

Критерии и шкала оценивания результатов промежуточной аттестации (зачет с оценкой)

Обозначения		Формулировка требований к степени освоения дисциплины
Цифр.	Оценка	
2	Неудовлетворительно	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине. Учебные достижения в семестровый период и результаты рубежного контроля демонстрировали низкую степень овладения программным материалом.
3	Удовлетворительно	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ. Учебные достижения в семестровый период и результаты рубежного контроля демонстрируют достаточную (удовлетворительную) степень овладения программным материалом.
4	Хорошо	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на

		вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения. Учебные достижения в семестровый период и результаты рубежного контроля демонстрируют хорошую степень овладения программным материалом.
5	Отлично	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно излагает его на зачете, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал монографической литературы, правильно обосновывает принятое решение. Учебные достижения в семестровый период и результаты рубежного контроля демонстрируют высокую степень овладения программным материалом.

5. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины «Анализ и визуализация данных» предполагает, как аудиторную, так и самостоятельную работу студентов.

Аудиторная работа проводится в форме практических занятий. Подготовка к занятиям должна носить систематический характер. Это позволит обучающемуся в полном объеме выполнить все требования преподавателя.

Самостоятельная работа является обязательным компонентом процесса подготовки бакалавров, она формирует самостоятельность, познавательную активность, вырабатывает практические навыки работы с учебной литературой.

Общий объем аудиторной и самостоятельной работы определяется учебно-тематическим планом. Изучение дисциплины завершается зачетом с оценкой.

Успешное изучение дисциплины требует посещения и активной работы на практических занятиях, выполнения учебных заданий преподавателя, ознакомления с основной и дополнительной литературой.

5.2. Методические указания для обучающихся по подготовке к практическим занятиям.

Практические занятия служат для закрепления изученного материала, развития умений и навыков работы в компьютерном классе в среде аналитического программного обеспечения, а также для контроля преподавателем степени подготовленности студентов по изучаемой дисциплине.

При подготовке к практическому занятию, обучающемуся необходимо:

– до очередного практического занятия по рекомендованным литературным источникам проработать теоретический материал, соответствующий теме занятия;

– в начале занятия задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения;

– в ходе занятия давать конкретные, четкие ответы по существу вопросов;

– на практическом занятии доводить каждую задачу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

Студентам, пропустившим практическое занятие, либо не подготовившимся к нему, рекомендуется не позже чем в 2-недельный срок явиться на консультацию к преподавателю и отчитаться по теме занятия.

5.3. Методические указания по организации самостоятельной работы обучающихся

Основной целью самостоятельной работы студентов является улучшение профессиональной подготовки обучающихся, направленное на формирование действенной системы фундаментальных и профессиональных знаний, умений и навыков, которые они могли бы свободно и самостоятельно применять в практической деятельности.

Самостоятельная работа студентов по дисциплине «Анализ и визуализация данных» способствует более глубокому усвоению изучаемого курса и проводится в следующих видах:

- подготовка к занятиям в соответствии с заданиями на самостоятельную работу с использованием материалов практических занятий и приведенных ниже источников литературы;
- выполнение домашних заданий;
- подготовка к промежуточной аттестации.

Приступая к изучению той или иной темы, выделяемой по предметно-систематизированному принципу, необходимо по отдельности и последовательно рассмотреть каждую из частей, из которых состоит тема. При изучении курса, обучающиеся должны уметь пользоваться учебной литературой для самостоятельной подготовки к занятиям.

5.4. Методические указания по подготовке обучающихся к зачету с оценкой

Подготовка к зачету с оценкой осуществляется студентом самостоятельно с использованием перечня вопросов к зачету с оценкой, конспекта лекций по дисциплине и рекомендованных литературных источников.

В период подготовки к зачету с оценкой студенты вновь обращаются к пройденному учебному материалу. При этом они не только укрепляют полученные знания, но и получают новые.

Подготовка студента к зачету с оценкой включает в себя два этапа:

- самостоятельная работа в течение семестра;
- непосредственная подготовка в дни, предшествующие зачету с оценкой по темам курса.

Основным источником подготовки к зачету с оценкой является конспект занятий, где учебный материал дается в систематизированном виде, ключевые его положения детализируются, подкрепляются современными фактами и информацией, которые в силу новизны не вошли в опубликованные печатные источники.

В ходе подготовки к зачету с оценкой студентам необходимо обращать внимание не только на уровень запоминания, но и на степень понимания излагаемых проблем.

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Основная литература

1. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2019. — 490 с. —

(Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/432178

2. Билл, Фрэнкс Революция в аналитике: Как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Фрэнкс Билл ; перевод И. Евстигнеева ; под редакцией В. Мылова. — Москва : Альпина Паблишер, 2020. — 320 с. — ISBN 978-5-9614-5302-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: https://www.iprbookshop.ru/93032.html

6.2. Дополнительная литература

1. 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

2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2019. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://www.biblio-online.ru/bcode/432851

3. Подкорытова, О. А. Анализ временных рядов : учебное пособие для бакалавриата и магистратуры / О. А. Подкорытова, М. В. Соколов. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2019. — 267 с. — (Бакалавр и магистр. Модуль). — ISBN 978-5-534-02556-9. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://biblio-online.ru/bcode/433180

4. Маккинли, У. Python и анализ данных / У. Маккинли; перевод А. Слинкина. - Саратов : Профобразование, 2019. - 482 с. - ISBN 978-5-4488-0046-7. - Текст : электронный // ЭБС "IPRBooks" : [сайт]. - URL: https://www.iprbookshop.ru/88752.html

6.3. Учебно-методическое обеспечение самостоятельной работы

Не предусмотрено.

6.4. Нормативные правовые документы

1. Федеральный закон Российской Федерации от 27 июля 2006 г. N 149-ФЗ «Об информации, информационных технологиях и о защите информации» // СПС КонсультантПлюс.

2. Федеральный закон от 20 февраля 1995 Г. N 24-Ф «Об информации, информатизации и защите информации» // СПС КонсультантПлюс.

6.5. Интернет-ресурсы

1. https://urait.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://new.znanium.com Электронно-библиотечная система [ЭБС] «Znanium.com».
5. https://dlib.eastview.com – Информационный сервис «East View». Коллекции электронных научных и практических журналов.
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.microsoft.com/ru-ru/office/%D1%81%D0%BE%D0%B7%D0%B4%D0%B0%D0%BD%D0%B8%D0%B5-%D0%B2-sway-db30c63b-6548-456d-b0e2-9fcd68bfa7a0 – руководство по созданию презентации в MS Sway;

14. https://rosstat.gov.ru/– официальный сайт Федеральной службы государственной статистики;

15. https://omsk.gks.ru/ - Территориальной орган Федеральной службы государственной статистики по Омской области

6.6. Иные источники

1. Паклин Н.Б., Орешков В.И Бизнес-аналитика: от данных к знаниям.: Учебное пособие — СПб.: Питер, 2013. — 704 с.

2. Барский А.Б. Введение в нейронные сети [Электронный ресурс]/ А.Б. Барский— Электрон. текстовые данные. — М.: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. — 358 c. — Режим доступа: http://www.iprbookshop.ru/52144.html. — ЭБС «IPRbooks»

3. Ханк Д.Э., Уичерн Д.У., Райтс А.Дж. Бизнес-прогнозирование. — М.: Вильямс, 2013. — 656 с.

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

Для проведения занятий по дисциплине необходимо материально-техническое обеспечение учебных аудиторий в зависимости от типа занятий: семинарского и лекционного типов, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации: комплект специализированной учебной мебели, доска аудиторная, компьютерная техника, презентационное оборудование. Выход в Интернет и доступ в электронную информационно-образовательную среду организации, в т.ч. с рабочих мест обучающихся.

Программное обеспечение: операционная система для ПК; пакеты прикладных программ, СПС КонсультантПлюс, антивирусное программное обеспечение.

Свободное программное обеспечение:

Orange (свободное ПО) https://orangedatamining.com

Loginom Community (свободное ПО) https://loginom.ru

MS Power BI Desktop (свободное ПО) https://powerbi.microsoft.com/ru-ru/

А также прикладных языков программирования, например, Python (свободное ПО, https://www.online-python.com/; https://playground.programiz.com/).

Помещение для самостоятельной работы обучающихся. Комплект специализированной учебной мебели, компьютерная техника с возможностью подключения к сети «Интернет» и доступом в электронную информационно-образовательную среду организации.

Базы данных, информационно-справочные и поисковые системы

1. https://urait.ru – Электронно-библиотечная система [ЭБС] Юрайт

2. http://www.iprbookshop.ru – Электронно-библиотечная система [ЭБС] «IPR SMART»

3. https://e.lanbook.com - Электронно-библиотечная система [ЭБС] «Лань»

4. https://new.znanium.com - Электронно-библиотечная система [ЭБС] «Znanium.com».

5. https://ibooks.ru - Электронно-библиотечная система система [ЭБС] «Айбукс.ру/ibooks.ru»

6. https://lib.ranepa.ru/ru/informatsionnye-resursy/russkoyazychnye-resursy/58-elibrary-ru - Научная электронная библиотека Elibrary.ru

7. http://www.edu.ru - Федеральный образовательный портал «Российское образование»

8. https://dlib.eastview.com – Информационный сервис «East View»

9. https://grebennikon.ru/ - Электронная библиотека Издательского дома «Гребенников», содержит полные тексты научно-практических журналов по маркетингу, менеджменту, финансам и управлению персоналом.

10. https://lib.ranepa.ru/ru/informatsionnye-resursy/zarubezhnye-resursy/49-ebsco-publishing – информационный ресурс EBSCO, содержащий издания по экономике, бизнесу, менеджменту и др.

11. https://rusneb.ru - Национальная электронная библиотека (НЭБ)

12. Справочно-правовая система «КонсультантПлюс».