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

УТВЕРЖДЕНА

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.06 АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

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

Высшее, бакалавриат

(уровень образования)

38.03.01 Экономика

(код и наименование направления подготовки (специальности))

Финансы и кредит

направленность (профиль/специализация)

бакалавр

квалификация

очная, очно-заочная

форма(ы) обучения

год набора - 2024

Омск, 2024 г.

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

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

Начальник учебного отдела
Е.В. Малахова

Рабочая программа дисциплины «Анализ и визуализация данных» рассмотрена и одобрена на заседании кафедры экономики и управления, протокол от 30 января 2024 г. № 5.

Рабочая программа дисциплины «Анализ и визуализация данных» пересмотрена и одобрена на заседании кафедры экономики и управления, 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....5

4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине....7

5. Методические указания для обучающихся по освоению дисциплины....20

6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", включая перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине....22

7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы....24

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

1.1. Дисциплина Б1.В.06 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-6.1	Демонстрирует навыки использования информационных технологий для решения аналитических и управленческих задач

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

ОТФ/ТФ	Код компонента компетенции	Результаты обучения (дескрипторы)
ОТФ B. Разработка отдельных функциональных направлений управления рисками **ТФ** – B/02.6 Документирование процесса управления рисками и корректировка реестров рисков в рамках отдельных бизнес-процессов и функциональных направлений	ПКс ОС II-6.1.2 Применяет современные информационные технологии для практической работы в экономических информационных системах и решения аналитических задач	**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.
		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 кластеризации, классификации, регрессии; умение строить аналитическую отчетность.
		На уровне навыков: навыки решения задач кластеризации, классификации, регрессии; навыки построения аналитической отчетности и дашбордов; навыки эффективной аналитической работы в прикладном рекомендованном программном обеспечении[1].

[1] Рекомендованным программным обеспечение является:
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.

Данная дисциплина не предусматривает использование ПО офисными офисного типа (MS Excel, MS Access и т. п.) для анализа и визуализации данных.

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

Дисциплина Б1.В.06 Анализ и визуализация данных составляет 4 зачетные единицы, т. е. 144 академических часа (108 астрономических часов).

На контактную работу с преподавателем выделено 48 академических часов (36 астрономических часов), из них 16 академических часов (12 астрономических часов) лекций и 32 академических часа (24 астрономических часа) практических занятий, на самостоятельную работу обучающихся выделено 92 академических часа (69 астрономических часов) для очной и очно-заочной ф/о.

Место дисциплины в структуре ОП ВО

Дисциплина Б1.В.06 Анализ и визуализация данных изучается на 2 курсе в 4 семестре для студентов очной ф/о, на 3 курсе в 5 семестре для студентов очно-заочной ф/о.

Освоение дисциплины Б1.В.06 Анализ и визуализация данных происходит после освоения дисциплин Б1.В.01 Экономическая информатика, Б1.В.02 Дифференциальные и разностные уравнения.

Изучение дисциплины заканчивается зачетом с оценкой.

3. Содержание и структура дисциплины

Очная, 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	**26**	2		4		20	*О, ПОЗ*
Тема 2	Визуализация данных	**34**	2		12		20	*О, ПОЗ, К*
Тема 3	Машинное обучение с учителем	**36**	8		8		20	*О, ПОЗ*
Тема 4	Машинное обучение без учителя	**26**	2		4		20	*О, ПОЗ*
Тема 5	Нейронные сети	**18**	2		4		12	*О, Т, ПОЗ*
Промежуточная аттестация		**4**						*ЗаО*
Всего		**144**	**16**		**32**		**92**	

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

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

Содержание дисциплины

Тема 1. Подготовка данных к анализу

Оценка качества данных. Фильтрация данных. Обработка дубликатов и противоречий. Группировка данных. Слияние данных. Квантование. Кодирование категориальных переменных. Восстановление пропущенных значений. Выявление аномальных значений статистическими методами (удаление и ограничение).

Практические занятия проводятся с использованием рекомендованного прикладного программного обеспечения[2].

[2] Рекомендованным программным обеспечение является:

Тема 2. Визуализация данных

Зачем визуализировать данные? Основные принципы визуализации. Способы визуализации данных.

Визуализация данных с использованием рекомендованного прикладного программного обеспечения[3].

Тема 3. Машинное обучение с учителем

Постановка задачи классификации и задачи регрессии.

Классификация

Расстояние в многомерном пространстве: Евклидово расстояние, манхэттенское расстояние, расстояние Чебышёва. Перемасштабирование данных (приведение к единой размерности).

Основные алгоритмы классификации: Метод ближайших соседей. Логистическая регрессия. Метод опорных векторов. Дерево решений, случайный лес, градиентный бустинг на решающих деревьях.

Матрица ошибок. Метрики качества классификатора: Правильность (Accuracy), Точность (Precision), Полнота (Recall) , F_1 - мера, AUC

Валидация моделей. Сравнение моделей.

Уменьшение размерности данных. Метод главных компонент.

Регрессия

Основные алгоритмы построения регрессионных моделей. Линейная регрессия. Регуляризация линейной модели: гребневая (ridge) регрессия, лассо (lasso) регрессия. Метод ближайших соседей. Метод опорных векторов. Случайный лес, градиентный бустинг на решающих деревьях.

Метрики качества регрессионной модели. MAE, MAPE, MSE, RMSE, R^2

Валидация моделей. Сравнение моделей.

Тема 4. Машинное обучение без учителя

Постановка задачи кластеризации.

Метод k-means. Иерархическая агломеративная кластеризация. Расстояния в иерархической кластеризации. Дендрограмма. Метод DBSCAN.

Оценка качества кластеризации: среднее внутрикластерное расстояние, среднее межкластерное расстояние, коэффициент силуэта.

Проектирование многомерных данных на плоскость с использованием t-SNE.

Тема 5. Нейронные сети

Персептрон Розенблатта. Многослойный персептрон. Обучение НС на основе алгоритма обратного распространения ошибки. Глубокое обучение.

Использование нейронной сети для решения задач регрессии и классификации. Проблемы решения задач анализа данных с использованием НС (количество и качество данных, переобучение). Использование нейронной сети для решения задачи кластеризации. Сеть Кохоннена. Архитектура НС для анализа данных различного типа (тексты, изображения): сверточные и рекуррентные нейросети.

---

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, R.

Данная дисциплина не предусматривает использование ПО офисными офисного типа (MS Excel, MS Access и т. п.) для анализа и визуализации данных.

[3] Рекомендуется ознакомить студентов со следующими визуализациями: http://blog.atkcg.ru/wp-content/uploads/2015/07/1-й-слайд1.jpg

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
Тема 1. Подготовка данных к анализу	Опрос Практико-ориентированное задание (компьютерный практикум)
Тема 2. Визуализация данных	Опрос Практико-ориентированное задание (компьютерный практикум) Контрольная работа
Тема 3. Машинное обучение с учителем	Опрос Практико-ориентированное задание (компьютерный практикум)
Тема 4. Машинное обучение без учителя	Опрос Практико-ориентированное задание (компьютерный практикум) Контрольная работа
Тема 5. Нейронные сети	Опрос Тестирование Практико-ориентированное задание (компьютерный практикум)

4.1.2. Материалы текущего контроля успеваемости

Типовые оценочные материалы по теме 1 «Подготовка данных к анализу»

Примерные темы для опроса

1. Оценка качества данных.
2. Фильтрация данных.
3. Обработка дубликатов и противоречий.
4. Группировка данных.
5. Слияние данных.
6. Квантование.
7. Кодирование категориальных переменных.
8. Восстановление пропущенных значений.
9. Стандартизация данных.
10. Метод главных компонент.
11. Выявление аномальных значений статистическими методами (удаление и ограничение).

Примерное 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.

Табличные данные представлены в текстовых файлах в формате .csv: stores.csv (магазины), train.csv (продажи), features.csv (признаки).

I. 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.

II. 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.

III. 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.

IV. 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»

Примерные темы для опроса (на примере MS Power BI)

1. Цели и задачи визуализации данных?
2. Зачем нужен визуальный контроль данных после их загрузки в аналитическое приложение?
3. Основные принципы визуализации.
4. Способы визуализации данных.
5. Что такое дашборд?
6. Правила разработки эффективного дашборда.
7. Семейство MS Power BI.
8. Подготовка данных с помощью Power Query.
9. Модель данных в Power BI.
10. Вычисляемые поля и меры.
11. Формулы DAX.
12. Построение различных визуальных представлений в Power BI.
13. Публикация отчета Power BI в сети Интернет.

Примерное практико-ориентированное задание (компьютерный практикум) «Получение аналитической отчетности (на примере 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. 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.

Примерный вариант 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 дашборд, который позволит проанализировать демографическое, социальное, экономическое положение региона РФ и динамику основных показателей.

Типовые оценочные материалы по теме 3 «Машинное обучение с учителем»:

Примерные темы для опроса

1. В чем заключается задача классификации?
2. Расстояние в многомерном пространстве: Евклидово расстояние, манхэттенское расстояние, расстояние Чебышёва.
3. Основные алгоритмы классификации: Метод ближайших соседей.
4. Основные алгоритмы классификации: Логистическая регрессия.
5. Основные алгоритмы классификации: Метод опорных векторов.
6. Дерево решений и случайный лес.
7. Матрица ошибок.
8. Метрики качества классификатора: Правильность (Accuracy), Точность (Precision), Полнота (Recall), F_1 – мера.
9. Как строится ROC-кривая?
10. Метрика качества классификатора AUC.
11. Валидация моделей.
12. Сравнение моделей.
13. В чем заключается задача регрессии?
14. Основные алгоритмы построения регрессионных моделей: метод ближайших соседей.
15. Основные алгоритмы построения регрессионных моделей: Линейная регрессия.
16. Регуляризация линейной модели: гребневая (ridge) регрессия,
17. Регуляризация линейной модели: Лассо (lasso) регрессия.
18. Метрики качества регрессионной модели: MAE, MAPE, MSE, RMSE, R2

Примерное 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.

Руководство банка 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 наборе представлено 2709 кредитов с известными исходами платежей.

1. С помощью обработчика «Калькулятор» на основе поля Просрочки более 60 дн. (COL12) получите новое поле Класс заемщика.

2. Разделите набор на тестовое и обучающее множества.

3. Постройте модель Логистической регрессии. Выходная переменная — Класс заемщика.

4. Запишите уравнение логистической регрессии.

5. Оцените качество модели, используя различные метрики.

6. Чему может быть равно значение параметра «порог отсечения»?

— Выберите значение параметра «порог отсечения», используя визуализатор ROC-кривая.

— Выберите значение параметра «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. 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 «Машинное обучение без учителя»:

Примерные темы для опроса

1. В чем заключается задача кластеризации.
2. Кластеризация методом k-means.
3. Расстояние в многомерном пространстве: Евклидово расстояние, манхэттенское расстояние, расстояние Чебышёва.
4. Что такое центроид?
5. Как в алгоритме k-means определяется принадлежность точки к тому или иному кластеру?
6. Что является условием остановки алгоритма k-means?
7. Иерархическая агломеративная кластеризация.
8. Расстояния в иерархической кластеризации.
9. Дендрограмма.
10. Кластеризация методом DBSCAN
11. Оценка качества кластеризации: среднее внутрикластерное расстояние, среднее межкластерное расстояние.
12. Оценка качества кластеризации: Коэффициент силуэта.

13. Проектирование многомерных данных на плоскость с использованием t-SNE

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 этому, о покупателях 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).

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

Категория	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. Используйте различные алгоритмы для проведения кластеризации.

1. Разбейте всех клиентов на сегменты. Перед построением аналитической модели рекомендуется ответить на следующие вопросы:

– Какие признаки включить в алгоритм кластеризации?

– Какое количество кластеров выбрать?

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.

6. Какой сегмент приносит компании самый большой доход?

7. Какой сегмент вы считаете наиболее перспективным?

Типовые оценочные материалы по теме 5 «Нейронные сети»:

Примерные темы для опроса

1. Персептрон Розенблатта.
2. Многослойный персептрон.
3. Обучение НС на основе алгоритма обратного распространения ошибки.
4. Глубокое обучение.
5. Использование нейронной сети для решения задачи регрессии
6. Использование нейронной сети для решения задачи классификации.
7. Проблема переобучения НС.
8. Использование нейронной сети для решения задачи кластеризации. Сеть Кохоннена.
9. В чем заключаются процедуры конкуренции, объединения и подстройки весов?

10. Что такое радиус обучения и какую роль он играет в процессе построения карты?

11. Сверточная НС

12. Рекуррентная НС

Примерное тестирование

1. Выберите верное утверждение:

a. Нейросети появились недавно, в 2000-х годах

b. Название “глубокое обучение” произошло от того, что нужно всегда очень глубоко погружаться в процесс чтения статей, чтобы понять алгоритм

c. Название “глубокое обучение” произошло от того, что в современных и сильных нейросетях обычно много слоёв (больше, чем, скажем, 5)

d. Deep Learning нужно использовать везде, где это возможно, даже там, где уже очень хорошо работают обычные алгоритмы (без машинного обучения)

e. Для хорошей работы нейросети её обычно нужно совсем мало данных, которые легко и быстро собираются

2. Задачи классификации решаются следующими алгоритмами:

a. алгоритмом Apriori

b. нейронные сети

c. линейной регрессии

3. 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"

4. Как много данных нужно для обучения современных нейронных сетей?

a. Нисколько: при обучении нейросетей данные не используются

b. Мало: нейросети успешно обучаются на выборках всего из десятков объектов

c. Среднее количество: на выборках из сотен или тысяч объектов можно обучить нейронную сеть, которая будет работать хорошо в широком множестве случаев

d. Много: для обучения современных нейронных сетей необходимы сотни тысяч или даже миллионы объектов

5. Что происходит во время обучения нейронной сети?

a. Настраивается число слоев и число нейронов в каждом слое

b. Настраивается длина шага градиентного спуска

c. Настраивается входной объект

d. Настраиваются параметры слоев (веса)

6. Что такое градиентный спуск?

a. Алгоритм постепенного обновления весов нейронной сети во время ее обучения

b. Алгоритм случайного выбора весов нейронной сети

c. Алгоритм выполнения предсказания с помощью нейронной сети

d. Алгоритм выбора объектов для обучения нейронной сети

7. 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

8. М**ногослойный 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)

9. 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

10. 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

11. Однонаправленные входные связи, соединенные с выходами других нейронов – это …

a. синапсы

b. слои сети

c. аксоны

12. 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

13. 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

14. 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.

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.

1. Постройте модель «Нейронная сеть».

3. Оцените качество модели.

4. Переобучите нейронную сеть

– 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?

5. Сравните качество моделей.

6. Сформулируйте рекомендации по использованию результатов моделирования в работе компании.

Шкалы оценивания результатов текущего контроля

Шкала оценивания результатов устных собеседований на практических занятиях

Оценка	Формулировка требований к степени освоения дисциплины
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.
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.
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.
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

Оценка	Формулировка требований к степени освоения дисциплины
Неудовлетворительно	50% и меньше правильных ответов.
Удовлетворительно	51–69% правильных ответов.
Хорошо	70–84% правильных ответов
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-6.1	Демонстрирует навыки использования информационных технологий для решения аналитических и управленческих задач

Компонент компетенции	Промежуточный индикатор оценивания	Критерий оценивания
ПКс ОС II-6.1.2 Применяет современные информационные технологии для практической работы в экономических информационных системах и решения аналитических задач	ПКс ОС II – 6.1 Демонстрирует навыки использования информационных технологий для решения аналитических и управленческих задач	Знает тенденции развития современных технических средств и информационных технологий, решает задачи кластеризации, классификации, регрессии, 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 обучающегося 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.

Примерный перечень вопросов для подготовки к зачету с оценкой

1. Основные этапы построения моделей
2. Формы представления, типы и виды данных
3. Классификация задач Data Mining
4. Оценка качества данных.
5. Фильтрация данных.
6. Обработка дубликатов и противоречий.
7. Группировка данных.
8. Слияние данных.
9. Квантование.
10. Кодирование категориальных переменных.
11. Восстановление пропущенных значений.
12. Стандартизация данных.
13. Метод главных компонент.
14. Выявление аномальных значений статистическими методами (удаление и ограничение).
15. Каковы цели и задачи визуализации данных.
16. Зачем нужен визуальный контроль данных после их загрузки в аналитическое приложение?
17. Способы визуализации данных.
18. Что такое дашборд? Правила разработки эффективного дашборда.
19. Вычисляемые поля и меры в Power BI.
20. В чем заключается задача классификации?
21. Расстояние в многомерном пространстве: Евклидово расстояние, манхэттенское расстояние, расстояние Чебышёва.
22. Основные алгоритмы классификации: Метод ближайших соседей.

23. Основные алгоритмы классификации: Логистическая регрессия.
24. Основные алгоритмы классификации: Метод опорных векторов.
25. Дерево решений и случайный лес.
26. Матрица ошибок.
27. Метрики качества классификатора: Правильность (Accuracy), Точность (Precision), Полнота (Recall), F1 – мера.
28. Как строится ROC-кривая?
29. Метрика качества классификатора AUC.
30. Валидация моделей.
31. Сравнение моделей.
32. В чем заключается задача регрессии?
33. Линейная регрессия.
34. Регуляризация линейной модели: гребневая (ridge) регрессия,
35. Регуляризация линейной модели: Лассо (lasso) регрессия.
36. Метрики качества регрессионной модели: MAE, MAPE, MSE, RMSE, R2
37. В чем заключается задача кластеризации.
38. Кластеризация методом k-means.
39. Иерархическая агломеративная кластеризация.
40. Кластеризация методом DBSCAN
41. Оценка качества кластеризации: среднее внутрикластерное расстояние, среднее межкластерное расстояние.
42. Оценка качества кластеризации: Коэффициент силуэта.
43. Персептрон Розенблатта.
44. Многослойный персептрон.
45. Обучение НС на основе алгоритма обратного распространения ошибки.
46. Глубокое обучение.
47. Проблема переобучения НС.
48. Использование нейронной сети для решения задачи кластеризации. Сеть Кохоннена.
49. Сверточная НС
50. Рекуррентная НС

Примеры практических задач для решения в компьютерных программах

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.

1. Оценить качество модели.
2. Чему равен оптимальный порог отсечения?
3. Как изменяется кредитный рейтинг заемщика с каждым дополнительным иждивенцем?

4. 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 файлах Заявки.scv, Договоры.csv, Обслуживание.csv

Используя MS Power BI, постройте набор визуализаторов, которые помогут дать ответ на следующие вопросы:

1. Кто чаще обращался за кредитом: мужчины или женщины? Верно ли это для всех типов кредитов?
2. Кредиты какого типа чаще всего выдавал банк за отчетный период? Справедливо ли это для всех филиалов? Для всех отделений?
3. Зависит ли задержка выплат по кредиту от семейного положения?

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 обучающегося 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)

Оценка	Формулировка требований к степени освоения дисциплины

Неудовлетворительно	Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится обучающимся, 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.
Удовлетворительно	Оценка «удовлетворительно» выставляется обучающемуся, 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.
Хорошо	Оценка «хорошо» выставляется обучающемуся, 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.
Отлично	Оценка «отлично» выставляется обучающемуся, 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.

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.

Обучающимся, 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.

Методические рекомендации по проведению опроса

Устные опросы 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.

Методические рекомендации по проведению тестирования

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.

Онлайн тестирование проводится в среде СДО РАНХ и ГС.

Методические рекомендации по проведению компьютерного практикума (практико-ориентированные задание)

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.

Методические рекомендации по проведению контрольной работы

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.

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.

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 обучающегося 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 обучающимся 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. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/520544

6.2. Дополнительная литература

1. 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

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

3. 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

4. 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

5. 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. http://www.intuit.ru — Национальный Открытый Университет «ИНТУИТ»;
2. http://www.planetaexcel.ru — сайт о возможностях MS Excel (приемы, видео-уроки, книги);
3. http://www.clouderp.ru/tags/BUSINESS_INTELLIGENCE/ Что такое Business Intelligence? Обзор BI систем
4. http://netpeak.net/ru/blog/kak-rabotat-s-microsoft-power-bi-podrobnoe-rukovodstvo/ Как работать с Microsoft Power BI — подробное руководство
5. http://community.powerbi.com/ Форум (англ. яз.) по Microsoft Power BI. Примеры применения. Ответы на вопросы. Использование на мобильных устройствах и др.

6. https://www.microsoft.com/ru-RU/download/details.aspx?id=45331 – ссылка для скачивания Microsoft Power BI;

7. https://docs.microsoft.com/ru-ru/power-bi/desktop-getting-started – техническая документация Microsoft, документация Power BI;

8. 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;

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

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

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

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

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

14. ПО прикладных языков программирования Python https://www.online-python.com/ ; https://playground.programiz.com/

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 с. — 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; 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/; ПО прикладных языков программирования Python https://www.online-python.com/ (свободное ПО, онлайн).

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».