

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 27.08.2022 17:34:25
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

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

Северо-Западный институт управления - филиал РАНХиГС
«Факультет безопасности и таможен»
«Кафедра таможенного администрирования»

УТВЕРЖДЕНО

Директор

Северо-Западного института управления -
филиала РАНХиГС

Хлутков А.Д.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.14 «Введение в технологии больших данных»

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

ВвТБД

по направлению подготовки (специальности)

38.05.02 «Таможенное дело»

(код и наименование направления подготовки (специальности))

«Информационно-аналитическая деятельность в таможенном деле»

направленность(и) (профиль (и)/специализация(и))

специалист таможенного дела

квалификация

очная, заочная

форма(ы) обучения

Год набора - 2022

Автор–составитель:

Старший преподаватель кафедры бизнес-информатики Е.Н. Томах

Врио Заведующего кафедрой

таможенного администрирования

д с/х.н.

Р.Х.Кочкаров

РПД одобрена на заседании кафедры таможенного администрирования. Протокол от (30.08.2022)
№ 1

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Материалы текущего контроля успеваемости обучающихся и фонд оценочных средств промежуточной аттестации по дисциплине
 - 4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации
 - 4.2. Материалы текущего контроля успеваемости обучающихся
 - 4.3. Оценочные средства для промежуточной аттестации
 - 4.4. Методические материалы
5. Методические указания для обучающихся по освоению дисциплины
6. Учебная литература и ресурсы информационно-телекоммуникационной сети "Интернет", учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине
 - 6.1. Основная литература
 - 6.2. Дополнительная литература
 - 6.3. Учебно-методическое обеспечение самостоятельной работы
 - 6.4. Нормативные правовые документы
 - 6.5. Интернет-ресурсы
 - 6.6. Иные источники
7. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

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

1.1. Дисциплина «Введение в технологии больших данных» обеспечивает овладение следующими компетенциями:

Таблица 1.1

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс-3	Владение навыками управления и анализа больших объемов данных с применением передовых инструментов и методов автоматической обработки структурированной и неструктурированной информации в целях обеспечения безопасности цепей поставок товаров и транспортных средств.	ПКс-3.1	Владеет навыками анализа и администрирования больших объемов данных в целях обеспечения безопасности цепей поставок товаров и транспортных средств

В результате освоения дисциплины у студентов должны быть сформированы:

Таблица 1.2

ОТФ/ТФ (при наличии профстандарта)/ профессиональные действия	Код этапа освоения компетенции	Результаты обучения
Проверка документов и сведений, предоставляемых декларантом и иными лицами, имеющими отношение к перемещению товаров через таможенную границу ЕАЭС, на предмет соблюдения законодательства ЕАЭС и РФ	ПКс-3.1	<p>На уровне знаний:</p> <ul style="list-style-type: none"> -основные цифровые технологии аналитики больших данных; -особенности работы с большими неструктурированными и слабоструктурированными данными; -системы и методы работы с большими данными; - средства бизнес-аналитики, современные языки статистической обработки (R, Python), статистические приложения (SPSS) и графические платформы; <p>На уровне умений:</p> <ul style="list-style-type: none"> - решать задачи аналитики на основе использования технологии больших данных; - использовать средства бизнес-аналитики, современные языки статистической обработки (R, Python), статистические приложения (SPSS) и графические платформы при решении задач аналитики данных; - применять технологии больших данных при решении задач профессиональной деятельности;

2. Объем и место дисциплины в структуре ОП ВО

Объем дисциплины

Общая трудоемкость дисциплины составляет 2 зачетных единиц 72 академических часов. Дисциплины читается по очной форме обучения в пятом семестре по заочной форме в одиннадцатом семестре.

Таблица 2

Очная форма обучения

Вид работы	Трудоемкость (акад/астр.часы)
Общая трудоемкость	72/54
Контактная работа с преподавателем	54/40,5
Лекции	18/13,5
Практические занятия	36/27
Самостоятельная работа	18/13,5
Контроль	
Формы текущего контроля	Задания, контрольная работа, выполнение расчетного задания
Форма промежуточной аттестации	Зачет

Таблица 3

Заочная форма обучения

Вид работы	Трудоемкость (акад/астр.часы)
Общая трудоемкость	72/54
Контактная работа с преподавателем	12/9
Лекции	4/3
Практические занятия	8/6
Самостоятельная работа	56/42
Контроль	4/3
Формы текущего контроля	Задания, контрольная работа, выполнение расчетного задания
Форма промежуточной аттестации	Зачет

Место дисциплины в структуре ОП ВО

Дисциплина Б1.В14 «Введение в технологии больших данных» относится к вариативной части учебного плана по специальности 38.05.2 «Таможенное дело». Является дисциплиной вариативной части. Преподавание дисциплины «Введение в технологии больших данных» основано на дисциплинах – Б1.Б.09 «Математика». В свою очередь она создаёт необходимые предпосылки для освоения программ таких дисциплин, как Б1.В.08 «Таможенная аналитика», Б1.В.05 «Информационно-аналитическая деятельность в таможенном деле», а также Б1.В.13 – «Системы управления базами данных и анализ данных», «Анализ данных в системе контроля таможенной стоимости» изучаемых с дисциплиной одновременно.

Формой промежуточной аттестации в соответствии с учебным планом является зачет.

Дисциплина реализуется с применением дистанционных образовательных технологий

3. Содержание и структура дисциплины

Таблица 4

Очная форма обучения

№ п/п	Наименование тем	Объем дисциплины, час.					Форма текущего контроля успеваемости, промежуточной аттестации	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					СР
			Л	ЛР	ПЗ	КСР		
Тема 1	Общая характеристика сквозных цифровых технологий	12	4		4		4	РГЗ
Тема 2	Большие данные и хранилища данных	12	4		4		4	ДЗ
Тема 3	Общая характеристика языка Python	20	4		12		4	ДЗ, РГЗ
Тема 4.	Машинное обучение на Python	20	4		12		4	ДЗ
Тема 5	Технологии и системы работы с большими данными	8	2		4		2	ДЗ, К
Промежуточная аттестация						2*		Зачет
Всего:		72/54	18/13,5		36/27		18/13,5	

2* - не входит в общий объем дисциплины

ДЗ – кейс,

РГЗ- расчетно- графическая работа

К – практическое контрольное задание

Таблица 5

Заочная форма обучения

№ п/п	Наименование тем	Объем дисциплины, час.					Форма текущего контроля успеваемости, промежуточной аттестации	
		Всего	Контактная работа обучающихся с преподавателем по видам учебных занятий					СР
			Л	ЛР	ПЗ	КСР		
Тема 1	Общая характеристика сквозных цифровых технологий	14	2				12	РГЗ
Тема 2	Большие данные и хранилища данных	12			2		10	ДЗ
Тема 3	Общая характеристика языка Python	14			2		12	ДЗ, РГЗ
Тема 4.	Машинное обучение на Python	16	2		2		12	ДЗ
Тема 5	Технологии и системы работы с большими данными	12			2		10	ДЗ, К
Контроль		4				4		
Промежуточная аттестация								Зачет
Всего:		72	4		8	4	56	

2* - не входит в общий объем дисциплины

ДЗ – кейс,

РГЗ- расчетно- графическая работа

К – практическое контрольное задание

Содержание дисциплины

Тема 1. Общая характеристика сквозных цифровых технологий

Введение. Предмет и задачи дисциплины «Введение в технологии больших данных». Место и роль дисциплины «Введение в технологии больших данных» в системе учебных дисциплин. Основа развития цифровой экономики России. Экономическая роль цифровой экономики. Генезис теории постиндустриального (информационного) общества. Доклад о мировом развитии «Цифровые дивиденды»

Определение больших данных. Жизненный цикл проекта по аналитике больших данных. Типовая архитектура проекта в области больших данных.

Классификация современных цифровых технологий. Жизненный цикл процесса анализа данных. Классификация методов Data Mining. Модели Data Mining. Понятие данные и знания. Процесс обнаружения знаний. Классификация задач Data Mining. Методы анализа данных. Разведочный Цифровые технологии. Очистка и фильтрация данных. Статистические диаграммы. «Ящичные» диаграммы. Диаграммы «ствол-листья». Задачи классификации и регрессии. Использование статистических пакетов для интеллектуального анализа данных. Понятие бизнес-аналитики. Средства бизнес-аналитики. Средства легкой бизнес-аналитики. Qlik View, Qlik Sence, Power BI. Распределенный реестр и криптовалюта

Общая характеристика языка R. Графические средства языка.

Тема 2. Большие данные и хранилища данных

Задачи систем поддержки принятия решений. OLTP и OLAP-системы. Принципы построения информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Правила Кодда. Размерностные модели. MOLAP, ROLAP, HOLAP- системы. Витрины данных. ETL (Extract Transform Load) – средство извлечения, обработки и загрузки данных. Добыча данных. Добыча данных в управлении качеством. Data Mining. Стандарты Data Mining. Стандарт CWM, CRISP, PMML. Методология KDD. Задачи предобработки данных. Технология ETL. Просмотр данных. Очистка данных. Оценка качества данных. Заполнение пропущенных данных. Аномальные и предельные данные. Использование ящичной диаграммы. Выявление дубликатов и противоречий. Корреляционный анализ. Использование факторного анализа при предобработке данных. Трансформация данных. Квантование. Сэмплинг. Группировка данных. Решение задач предобработки и очистки данных в R (Python).

Тема 3. Общая характеристика языка Python

Сравнительный анализ Python, R. Общая характеристика языка Python. Среда разработки. Платформа Anaconda. Основы синтаксиса языка. Переменные, ключевые слова. Основы программирования на языке Python. Типы данных. Функциональность для работы с данными. Установка пакетов научных вычислений на Python. Установка Keras. Решение задач анализа данных. Понятие тензора. Скаляры, векторы, матрицы, тензоры третьего и высшего рангов. Временные ряды или последовательности. Изображения

Тема 4. Машинное обучение на Python

Постановка задач кластерного анализа. Определение кластера. Параметры кластера. Меры близости. Метрики кластерного анализа. Базовые алгоритмы кластеризации. Иерархическая кластеризация. Дендрограммы. Метод K-средних. Понятие центроида. Профили кластеров. Взаимосвязь кластерного и регрессионного анализа. Кластерный анализ на Python.

Формулировка задачи классификации. Классификационный анализ с обучением. Деревья решений. Алгоритмы построения деревьев решений. Методы и алгоритмы

построения деревьев. Алгоритм CART. Определение прекращения построения дерева классификации. Логистическая регрессия. Сравнение результатов классификации различными методами. Метод k-ближайших соседей.

Оценка качества задач классификации. Таблица сопряженности. Понятие чувствительности и специфичности. ROC-кривая. Ошибки первого и второго рода при решении задач классификации.

Решение задач анализа текстов. Решение задач анализа изображений.

Общая характеристика навигатора Anaconda. Виджеты Orange. Использование виджетов для решения задач машинного обучения.

Тема 5. Технологии и системы работы с большими данными

Фреймворк для обработки больших данных. Парадигма MapReduce. Основная идея, формальное описание. Обзор реализаций. Утилиты Hadoop. Схема его работы, роли серверов в Hadoop-кластере. Подходы, направленные на реализацию хранилищ данных. NoSQL. Библиотеки Python Pandas, SciKit-Learn, Pyspark, Keras. TensorFlow.

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

4.1. Формы и методы текущего контроля успеваемости обучающихся и промежуточной аттестации

В ходе реализации дисциплины «Введение в технологии больших данных» используются следующие методы текущего контроля успеваемости обучающихся:

Таблица 4.1

Тема (раздел)	Формы (методы) текущего контроля успеваемости
Тема 1. Общая характеристика сквозных цифровых технологий	Защита расчетно-графического задания
Тема 2. Большие данные и хранилища данных	Защита задания
Тема 3. Общая характеристика языка Python	Защита расчетно-графического задания
Тема 4. Машинное обучение на Python	Защита задания, Тестирование
Тема 5. Технологии и системы работы с большими данными	Защита задания, контрольная работа

В дисциплине используются следующие активные и интерактивные методы обучения:

- выполнение и защита заданий и контрольной работы;
- интерактивная работа по решению практических задач на компьютерах в компьютерном классе с текущим обсуждением хода и результатов решения задачи, использованию современных программных средств аналитики, data mining;
- выполнение тестирования;
- тренинги в решении практических задач, направленных на формирование универсальных и общепрофессиональных компетенций;

Признаками данных методов являются:

- активизация мышления студентов, причем учащийся вынужден быть активным;
- длительное время активности — студент работает не эпизодически, а в течение всего учебного процесса. Поэтому данные методы в основном реализуются на занятиях семинарского типа;
- самостоятельность в выработке и поиске решений поставленных задач;
- мотивированность к обучению путем использовать балльно-рейтинговой системы оценивания.

4.1.2. Зачет проводится с применением следующих методов (средств):

Зачет проводится в компьютерном классе в устной форме. Во время зачета проверяется уровень знаний по дисциплине «Введение в технологии больших данных», а также уровень умений решать учебные задачи анализа данных с использованием программных приложений. К зачету студенты должны решить задания по всем темам учебной дисциплины. Результаты решения задач могут быть использованы при решении практической задачи в соответствии с имеемым перечнем задач. Пример задач приведен в программе. При ответе на вопросы студент показывает умение решать практические задачи на языках аналитики и статистических пакетах.

4. 2. Материалы текущего контроля успеваемости обучающихся

4.2.1 Кейсы (задания)

Задание 1. Измеряются 13 характеристик химического состава вина. Необходимо по значениям имеющихся переменных определить тип вина.

Имеются данные о трех сортах вина. Сорт вина указан в трех столбцах класс_1, класс_2 и класс_3. Если в первом столбце (класс_3) стоит единица, то наблюдение соответствует виду вина третьего типа, если во втором столбце (класс_2) стоит единица, то наблюдение соответствует виду вина второго типа, если в третьем столбце (класс_1) стоит единица, то наблюдение соответствует виду вина первого типа. Таким образом, столбцы дублируют друг друга.

Список переменных:

- 1) Alcohol (содержание алкоголя)
- 2) Malic acid (яблочная кислота)
- 3) Ash (зола)
- 4) Alcalinity of ash
- 5) Magnesium (магний)
- 6) Total phenols (общее содержание фенола (карболовой кислоты))
- 7) Flavanoids (ароматические вещества)
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity (интенсивность цвета)
- 11) Hue (окраска? красители?)
- 12) OD280/OD315 of diluted wines
- 13) Proline (пролин)

Число наблюдений – 178

Число переменных – 13, все измерены в количественной (непрерывной) шкале.

Источник задачи: UCI Machine learning Database

<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/wine>

Обсуждение задачи:

<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/wine>

Кейс 2. Имеются данные рейтинга Global Firepower, который основан на более чем 50 факторах, чтобы определить оценку PowerIndex (PwrIdx) данной страны. Данные взяты из наборов данных kaggle

Приведенная при расчете рейтинга формула позволяет более мелким, но более технологически развитым странам конкурировать с более крупными, менее развитыми. Модификаторы (в виде бонусов и штрафов) добавляются для дальнейшего уточнения списка. Некоторые пункты, которые соблюдены в отношении окончательного рейтинга:

- ранжирование не просто зависит от общего количества оружия, доступного какой-либо одной стране, а скорее сосредоточено на разнообразии оружия в пределах количества,

чтобы обеспечить лучший баланс доступных огневых мощностей (т. е., например, 100 тральщиков не соответствует стратегической и тактической ценности 10 авианосцам);

- ядерные запасы не принимаются во внимание, но признанные/подозреваемые ядерные державы получают бонус;

- географические факторы, логистическая гибкость, природные ресурсы и местная промышленность влияют на окончательный рейтинг;

- доступные трудовые ресурсы являются ключевым фактором; Страны с большим населением, как правило, выше;

- страны, не имеющие выхода к морю, «не наказаны» за отсутствие военно-морского флота; морские державы «наказываются» при расчете рейтинга из-за отсутствия разнообразия в имеющихся морских активах;

- альянсы НАТО получают небольшой бонус за теоретический обмен ресурсами, возможную организацию коалиций;

- нынешнее политическое/военное руководство, их политика, роль не принимается во внимание.

На 2017 год в базу данных GFP входит в общей сложности 133 страны. (<http://www.globalfirepower.com/countries-listing.asp>).

Решить задачу кластерного анализа методом иерархической кластеризации и методом k-средних. При решении задачи пропущенные данные заменять медианой. Для выявления пропущенных данных использовать библиотеку `misc`.

В анализируемом наборе имеется 47 атрибутов, первые два из которых символьные. Третий атрибут – ранг страны получается путем анализа всех остальных. Поэтому данный рейтинг не следует учитывать при решении задачи.

Кейс 3. Даны данные учебного набора «Титаник» решить задачу классификации различными методами. Сравнить результаты классификации. Для решения задачи использовать файл `train.csv`. При решении задачи использовать следующие атрибуты

`Pclass + Sex + Age + SibSp + Parch + Fare`.

Для выделения нужных признаков использовать операцию конкатенации, например `dat[,c(2,3,4,6,9)]`

Для проверки качества классификации использовать тестовую и обучающую выборки. Размер выборок сделать равным. При построении выборки использовать функцию `sample`.

Построить таблицу сопряженности по результатам проверки качества работы классификатора на тестовой выборке.

Кейс 4. Построить диаграммы объема импорта (экспорта) из стран мира. <https://ru-stat.com/>

4.2.2. Задачи на домашние работы

Задача 1. Решить задачу построения дерева решений по данным

	Соперник	Играем	Лидеры	Дождь	Победа
Выше	Дома	На месте	Да	Нет	
Выше	Дома	На месте	Нет	Да	
Выше	Дома	Пропускают	Нет	Нет	
Ниже	Дома	Пропускают	Нет	Да	
Ниже	В гостях	Пропускают	Нет	Нет	Нет
Ниже	Дома	Пропускают	Да	Да	
Выше	В гостях	На месте	Да	Нет	
Ниже	В гостях	На месте	Нет	Нет	

Задача 2. С помощью какой функции можно решить задачу кластерного анализа методом k-средних?

- С помощью функции `k-mean`
- С помощью функции `mean`
- С помощью функции `cluster`

- С помощью функции `kn`

Задача 3. С помощью функции `rngm(100,0,0.1)` генерируется последовательность случайных чисел:

- с математическим ожиданием равным 100;
- с математическим ожиданием равным 0.1;
- с математическим ожиданием равным 0;
- размер последовательности равен 1.

Тесты

Задание 1. Сколько Петабайт в Зеттабайте?

Задание 2. Укажите фактор, способствовавший появлению тренда больших данных

- Маркетинговые кампании крупных корпораций
- Снижение издержек на хранение данных
- Появление новых технологий обработки потоковых данных
- Выпуск баз данных с обработкой данных в памяти

Задание 3. Определение больших данных: 1. Выберите верный ответ

- Большие данные – это обработка или хранение более 1 Тб информации.
- Проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна.
- Большие данные – это огромная PR-акция крупных вендоров и не более того.
- Большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект.

Задание 4. Выберите неверный ответ:

- Большие данные – это данные объёма свыше 1 Тб
- Проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна.
- Большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров.
- Большие данные как правило не структурированы.

Задание 5. Отметьте те из вариантов, в которых данные структурированы:

- Данные о продажах компании, представленные в виде помесечных отчётов в формате MS Word. ii. Таблица с ежедневными показаниями температуры помещения за год в файле формата csv.
- Текст педагогической поэмы А.С. Макаренко, представленный в формате PDF.
- Библиотека фильмов, представленных в формате `mpeg4` на одном жестком диске.

Задание 6. Выберите неверное высказывание:

- Большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных.
- Увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации.
- Удешевление систем хранения на единицу информации привело к росту рынка больших данных.
- Большое разнообразие источников данных

Задание 7. Отметьте неверное понимание Variety в контексте характеристик Big Data:

- Высокая скорость генерирования данных.
- Разные типы данных в колонках таблиц реляционных СУБД.
- Разнообразие отраслей, являющихся источниками данных.
- Разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные.

ЗАДАНИЕ 8. Коэффициент парной корреляции характеризует тесноту _____ связи между _____ переменными.

ВАРИАНТЫ ОТВЕТОВ:

- | | |
|-----------------------------|-------------------------------|
| 1) линейной ... несколькими | 2) нелинейной ... несколькими |
| 3) линейной ... двумя | 4) нелинейной ... двумя |

ЗАДАНИЕ 9. Установите соответствие между наименованиями элементов уравнения $Y=b_0+b_1X+e$ и их буквенными обозначениями:

1. параметры регрессии
2. объясняющая переменная
3. объясняемая переменная
4. случайные отклонения

ВАРИАНТЫ ОТВЕТОВ:

- | | |
|--------|---------------|
| A) Y | B) b_0, b_1 |
| C) X | D) e |

ЗАДАНИЕ 10. Для линейного уравнения регрессии $y = a + bx + \varepsilon$ метод наименьших квадратов используется при оценивании параметров...

ВАРИАНТЫ ОТВЕТОВ:

- | | |
|--------|--------|
| 1) B | 2) y |
| 3) X | 4) a |

ЗАДАНИЕ 11. Сколько параметров содержит парное линейное уравнение регрессии?

ВАРИАНТЫ ОТВЕТОВ:

- A) 1
- Б) 2
- В) 3
- Г) 4

ЗАДАНИЕ 12. При выполнении предпосылок МНК оценки параметров регрессии обладают свойствами:

ВАРИАНТЫ ОТВЕТОВ:

- | | |
|------------------|----------------------|
| 1) достоверность | 2) эффективность |
| 3) несмещенность | 4) несостоятельность |

ЗАДАНИЕ 13. Как влияет увеличение объема выборки на величину остаточной дисперсии случайной величины?

ВАРИАНТЫ ОТВЕТОВ:

- А) Никак.
- Б) Остаточная дисперсия увеличивается.
- В) Остаточная дисперсия уменьшается.
- Г) Результат зависит от конкретного вида случайной величины.

ЗАДАНИЕ 14. В кластерном анализе используются методы объединения ...

ВАРИАНТЫ ОТВЕТОВ:

- 1) Ближнего соседа
- 2) Дальнего соседа
- 3) Среднего соседа
- 4) центроидный метод

ЗАДАНИЕ 15. В кластерном анализе для определения близости между кластерами используются метрики ...

ВАРИАНТЫ ОТВЕТОВ:

- 1) Эвклидово расстояние
- 2) Куб Эвклидова расстояния
- 3) Взвешенное эвклидово расстояние
- 4) Квадрат Эвклидова расстояния

ЗАДАНИЕ 16. В дискриминантном анализе обучающая выборка используется для ...

ВАРИАНТЫ ОТВЕТОВ:

- 1) Выявления значимых признаков
- 2) Выявления аномального измерения
- 3) Разделения объектов на классы
- 4) Выбора вида модели

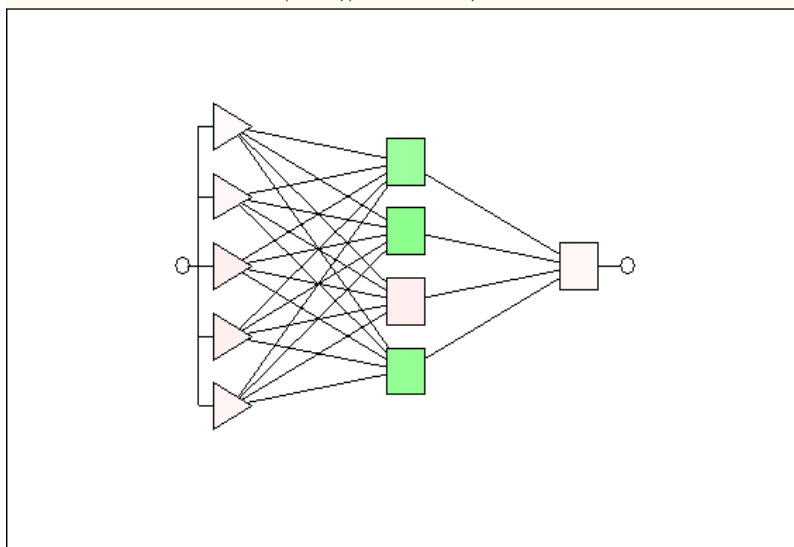
ЗАДАНИЕ 17. Сигмоидальная активизационная функция искусственного нейрона имеет вид...

ВАРИАНТЫ ОТВЕТОВ:

- 1) $y = \begin{cases} 1, & \text{если } S \geq T \\ 0, & \text{если } S < T \end{cases}$
- 2) $y = \frac{1}{1 + e^{-s}}$
- 3) $y = \begin{cases} 1, & \text{если } S > 0 \\ -1, & \text{если } S \leq 0 \end{cases}$
- 4) $y = \begin{cases} S, & \text{если } S \geq 0 \\ 0, & \text{если } S < 0 \end{cases}$

ЗАДАНИЕ 18. На рисунке приведена архитектура многослойного персептрона. Определить число рецепторных, реагирующих и ассоциативных элементов

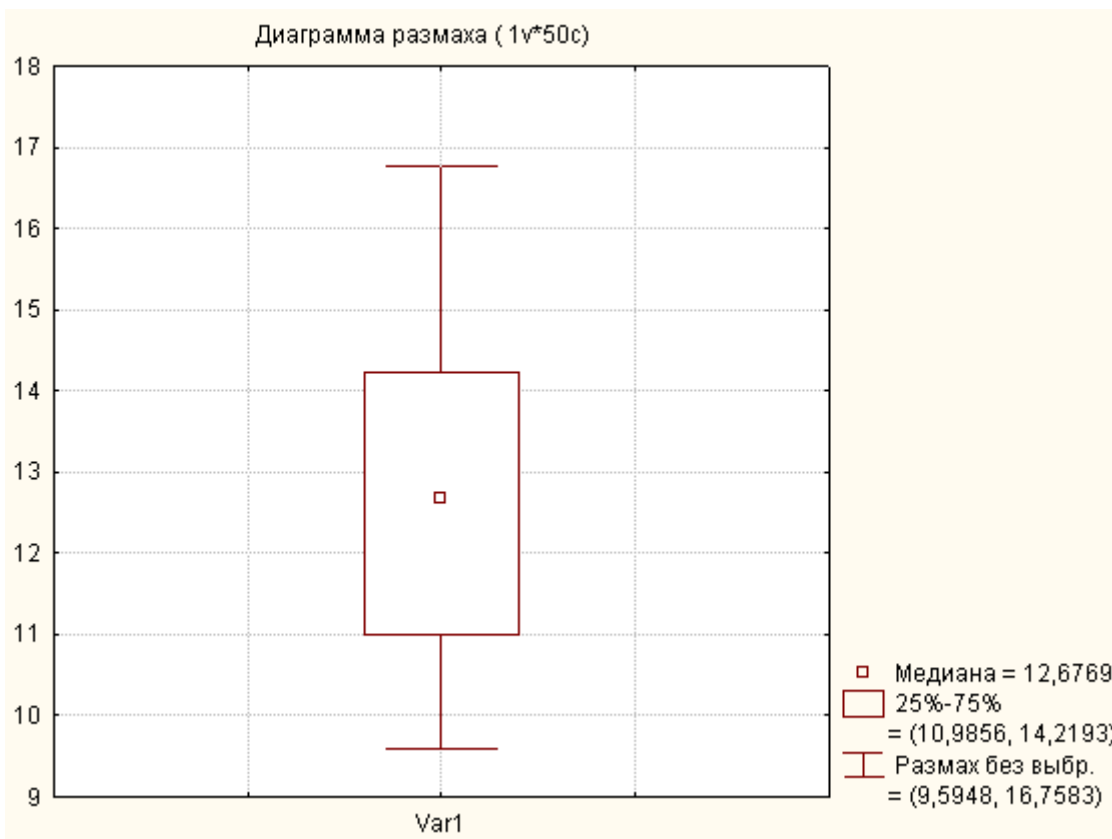
Архитектура : МП s5 1:5-4-1:1 , N = 27
 Производительность обуч. = 0,219250 , Контр. производительность = 0,338934 , Тест.
 производительность = 0,268458



ВАРИАНТЫ ОТВЕТОВ:

- | | |
|---|---|
| <p>1) Рецепторных 4, ассоциативных 5,
 реагирующих -1</p> <p>3) Рецепторных 5, ассоциативных 4,
 реагирующих -1</p> | <p>2) Рецепторных 4,
ассоциативных 1, реагирующих -5</p> <p>4) Рецепторных 1,
ассоциативных 5, реагирующих -4</p> |
|---|---|

ЗАДАНИЕ 19. На рисунке приведена диаграмма размаха («ящик с усами»). Определить интерквартильный размах с точностью до второго знака



ВАРИАНТЫ ОТВЕТОВ:

- | | | |
|----|-------|----------|
| 1) | 3,23 | 2) 7,16 |
| 3) | 12,68 | 4) 16,76 |

Задача 20. Принцип MapReduce состоит в том, чтобы

- Производить вычисления на узлах, где информация изначально была сохранена
- Использовать вычислительные мощности систем хранения
- Использовать функциональное программирование для решения задач массивно-параллельной обработки

4.3. Оценочные средства для промежуточной аттестации.

Таблица 3

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПКс-3	Владение навыками управления и анализа больших объемов данных с применением передовых инструментов и методов автоматической обработки структурированной и неструктурированной информации в целях обеспечения безопасности цепей поставок товаров и транспортных средств.	ПКс-3.1	Владеет навыками анализа и администрирования больших объемов данных в целях обеспечения безопасности цепей поставок товаров и транспортных средств

Этап освоения компетенции	Показатель оценивания	Критерий оценивания
ПКс-3.1	1. Демонстрирует знания основных технологий работы с большими данными. Демонстрирует умение решать задачи информационно-аналитической деятельности с использованием технологии больших данных	1. Полнота и качество ответов. 2. Правильность решения задач оценки качества 3. Активность на занятиях. Своевременность и полнота решения задач

Для оценки сформированности компетенций, знаний и умений, соответствующих данным компетенциям, используются контрольные вопросы, а также задачи, при решении которых необходимо построить имитационные модели, спланировать и провести эксперименты с ними.

Типовые вопросы, выносимые на зачет:

1. Дать определение понятия «большие данные». Перечислить свойства больших

данных. Перечислить основные сквозные цифровые технологии.

2. Указать роль аналитика по данным (Data Scientist). Сформулировать ключевые компетенции аналитика. Выделить отличия BI от Data Science.

3. Сформулировать парадигма Map Reduce. Описать принцип работы. Нарисовать диаграмму. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.

4. Перечислить задачи систем поддержки принятия решений. OLTP и OLAP-системы. Принципы построения информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Правила Кодда.

5. Дать характеристику размерностных моделей, моделей MOLAP, ROLAP, HOLAP-системы.

6. Характеризовать витрины данных.

7. Сделать обзор ETL (Extract Transform Load) технологии.

8. Характеризовать стандарты Data Mining, стандарты CWM, CRISP, PMML.

9. Рассмотреть содержание и этапы методология KDD. Дать характеристику задачи предобработки данных.

10. Рассмотреть модель распределенных вычислений. Map Reduce. Реализацию Map Reduce на Python

11. Сделать обзор языков Python. Описать основы языка Python. Привести примеры среды разработки. Сделать обзор платформы Anaconda.

12. Продемонстрировать организацию использования пакета Anaconda.

13. Характеризовать основы синтаксиса языка, переменных, ключевых слов языка, основ программирования на языке Python.

14. Характеризовать типы данных языка Python, функциональности для работы с данными. Описать организацию установки пакетов научных вычислений на Python, установки Keras.

15. Сделать обзор организации решения задач анализа данных. Дать определение понятия тензора, скаляров, векторов, матриц, тензоров третьего и высшего рангов, временных рядов или последовательностей. Рассмотреть операции над тензорами.

16. Характеризовать фреймворк Keras, библиотеки Google TensorFlow

17. Определить содержание задач кластерного анализа, понятие кластера, параметров кластера. Объяснить меры близости между кластерами, используемые метрики кластерного анализа.

18. Характеризовать базовые алгоритмы кластеризации. Рассмотреть иерархическую кластеризацию, понятие дендрограммы, организацию ее использования.

19. Характеризовать метод К-средних. Раскрыть понятие центроида, профиля кластеров. Связать понятия кластерного и регрессионного анализа. Описать организацию решения задач кластерного анализа на Python.

20. Дать определение задачи классификации. Перечислить методы классификации. Дать общую характеристику классификационного анализа с обучением. Продемонстрировать решение задачи классификации в Orange.

21. Объяснить основное содержание метода деревьев решений. Привести примеры деревьев решений. Характеризовать алгоритмы построения деревьев решений. Показать организацию решения задачи классификации в Python.

22. Характеризовать метод логистической регрессии, организацию ее решения в Python. Привести пример решения задачи различными методами классификации с помощью Orange. Исследовать результаты классификации различными методами.

23. Объяснить содержание метода k-ближайших соседей. Продемонстрировать решение задачи и сравнить с результатами решения другими методами.

24. Сделать обзор инструментов, используемых при оценке качества задач классификации. Таблица сопряженности (матрица ошибок). Дать определение понятий чувствительности и специфичности, ROC-кривой, критериев оценки качества

классификации.

25. Дать определение понятия нейронной сети. Характеризовать архитектуру нейронной сети.

26. Сделать обзор основ машинного обучения. Характеризовать организацию оценки моделей машинного обучения, предназначение тренировочных, проверочных и контрольных данных, понятие функции потерь.

27. Характеризовать механизм нейронных сетей на основе обучения, метод обратного распространения ошибки.

28. Продемонстрировать решение задач классификации, регрессии, прогнозирования с помощью нейронных сетей в Python, Orange.

29. Дать определения понятий поверхностного и глубокого обучения, глубокого обучения в технологиях компьютерного зрения.

30. Характеризовать содержание глубокого обучения для текста и последовательностей. Сделать обзор рекуррентных нейронных сетей, слоев LSTM, GRU.

Типовые контрольные задания на зачет:

Кейс 1. Проверить гипотезу о значимом отличии среднего балла за экзамены в десятом и одиннадцатом классах, используя критерий Стьюдента и критерий Манна-Уитни. Построить диаграммы «ящик с усами» для школьников, имеющих разные хобби. Построить диаграмму «дерево-листья». Данные находятся в файле тесты Школа.txt. Задачу решить в R и в SPSS.

Построить задачу классификации хобби в зависимости от результатов тестирования. Задачу классификации решить с помощью деревьев решений в R.

Кейс 2. Создать случайную последовательность размером в 500 наблюдений с использованием генератора равномерно распределенных чисел в диапазоне от 0 до 10. Проверить статистическую гипотезу о числовых значениях параметров:

$$1 H_0 : a = 0,5; H_1 : a \neq 0,5 .$$

$$2 H_0 : a = 5; H_1 : a > 5 .$$

Построить гистограмму распределения в R. Построить гистограмму частот и гистограмму относительных частот. При построении гистограммы оценить и задать число интервалов. Указать название осей и название гистограммы, а также заливку синего цвета. На диаграмму поместить кривую ядерной плотности, а также аппроксимацию равномерным законом распределения. При построении кривой регулировать ее гладкость.

- Оценить статистические характеристики.
- При проверке гипотезы: использовать одновыборочный T-критерий. Задать уровень значимости 0,05. Использовать одностороннюю и двухстороннюю проверки гипотезы.
- Проверить гипотезу о равномерном законе распределения с помощью критерия Колмогорова-Смирнова.

В R использовать функцию t.test

Кейс 3. В файле ГосСлужба.txt приведены данные по стажу работы, стажу в должности и возрасту в государственной службе.

- Построить гистограммы распределения случайных величин.
- Оценить выборочные характеристики.
- Проверить статистические гипотезы о значимом отличии стажа в должности, стажа работы на гос. службе и возраста для мужчин и женщин с использованием t-критерия и критерия Манна-Уитни.
- Построить диаграммы размаха для случайных величин: возраст, стаж службы.

Задачу решить в SPSS.

Кейс 4. Таксомоторную компанию интересует зависимость между средним пробегом автомашины в расчете на 1 л топлива и возрастом машины. Были взяты 12 автомашин одной

марки. Поскольку водителями были мужчины и женщины, предполагалось, что какая-то часть изменчивости пробега определяется разной техникой вождения у мужчин и женщин. Значения среднего пробега были рассчитаны на основе сведений о расходе горючего после прохождения машиной расстояния 100 км. Данные приведены в таблице.

Пол (мужчины, женщины)	Возраст машины, лет	Расход горючего, км.
мужчина	3	8,92
женщина	4	8,8
женщина	3	9,48
мужчина	2	9,68
женщина	1	10,2
мужчина	5	8,44
мужчина	4	8,24
мужчина	1	9,6
женщина	1	10,4
мужчина	2	9,24
женщина	2	9,92
мужчина	3	8,08

- Определить, значимы ли различия между пробегом для водителей-мужчин и водителей женщин, используя Т-тест для независимых групп (двухсторонний и односторонний). Для проверки гипотезы проверить гипотезу о постоянстве дисперсии. Сравнить результаты проверки гипотезы с результатами проверки по критерию Манна-Уитни. Построить диаграммы размаха.
- Построить ящичные диаграммы для водителей мужчин и водителей-женщин.
- Решить задачу построения описательной статистики в SPSS.

Для проверки гипотезы по критерию Манна-Уитни в R использовать функцию `wilcox.test(y ~ x, data)`

Кейс 5. Создать две случайные последовательности двух случайных величинах, размером в 200 наблюдений, полученных с помощью генераторов нормально распределенных случайных чисел, имеющих одинаковое математическое ожидание, равное 5 и ско, соответственно 1 и 2.

- Проверить гипотезу о равенстве математических ожиданий и дисперсий данных величин.
- Изменить генератор, добавив в первый генератор смещение математического ожидания. Вновь проверить статистическую гипотезу.
- Проверить гипотезы о нормальном законе распределения.
- Найти сумму пяти случайных величин, равномерно распределенных на интервале 0, 2. Проверить гипотезу о нормальном законе распределения суммы.

Задачу решить с помощью статистических критериев в R. Построить вероятностные и квантиль-квантиль графики.

Кейс 6. Решить задачу кластерного анализа для файла Семейное положение.txt. при решении задачи кластерного анализа:

- определить склонность к кластеризации;
- определить лучшую метрику иерархической кластеризации;
- выполнить иерархическую кластеризацию;
- определить состав и центроиды кластеров;
- Решить задачу кластеризации методом k-средних;
- Выполнить интерпретацию полученных кластеров;
- Визуализировать полученную кластеризацию;
- Задачу решить в RStudio и в SPSS.

Кейс 7. В наборе Animals библиотеки cluster имеются данные о 20 животных. Заданы 6 бинарных признаков: теплокровные/нетеплокровные; летают/не летают; позвоночный/беспозвоночный; находящихся под угрозой вымирания; живущих в группах. Решить задачу кластерного анализа наблюдений в SPSS и в R. Использовать иерархическую кластеризацию и кластеризацию методом k-средних.

Шкала оценивания.

Оценка результатов производится на основе балльно-рейтинговой системы (БРС). Использование БРС осуществляется в соответствии с приказом от 06 сентября 2019 г. №306 «О применении балльно-рейтинговой системы оценки знаний обучающихся».

Схема расчетов сформирована в соответствии с учебным планом направления, согласована с руководителем научно-образовательного направления, утверждена деканом факультета.

Схема расчетов доводится до сведения студентов на первом занятии по данной дисциплине, является составной частью рабочей программы дисциплины и содержит информацию по изучению дисциплины, указанную в Положении о балльно-рейтинговой системе оценки знаний обучающихся в РАНХиГС.

В соответствии с балльно-рейтинговой системой максимально-расчетное количество баллов за семестр составляет 100, из них в рамках дисциплины отводится:

40 баллов - на промежуточную аттестацию

40 баллов - на работу на практических занятиях

20 баллов - на посещаемость занятий

В случае если студент в течение семестра не набирает минимальное число баллов, необходимое для сдачи промежуточной аттестации, то он может заработать дополнительные баллы, отработав соответствующие разделы дисциплины, получив от преподавателя компенсирующие задания.

Критерии оценки ответа на вопросы на зачёте:

«Зачтено» ставится в том случае, если студент продемонстрирует знание основных понятий, относящихся к изучаемой дисциплине, правильно ответить, по крайней мере, на один дополнительный вопрос, в состоянии выполнить практические действия. Ответ должен быть логичным и последовательным, либо студент способен уточнить содержание ответа

«Не зачтено» ставится в том случае, если студент не демонстрирует знание основных понятий, относящихся к изучаемой дисциплине, не отвечает ни на один дополнительный вопрос, и изложение ответа на вопрос не последовательное и не логичное. При этом, студент не в состоянии выполнить практические действия.

51-100 баллов - зачет

0-50 баллов - незачет

5. Методические указания для обучающихся по освоению дисциплины

Рабочей программой дисциплины предусмотрены следующие виды аудиторных занятий: лекции, практические занятия, контрольные работы. На лекциях рассматриваются наиболее сложный материал дисциплины. Лекция сопровождается презентациями, компьютерными текстами лекции, что позволяет студенту самостоятельно работать над повторением и закреплением лекционного материала. Для этого студенту должно быть предоставлено право самостоятельно работать в компьютерных классах в сети Интернет.

Практические занятия предназначены для самостоятельной работы студентов по решению конкретных задач дискретно математики. Ряд практических занятий проводится в компьютерных классах с использованием Excel. Каждое практическое занятие сопровождается домашними заданиями, выдаваемыми студентам для решения внеаудиторное время. Для оказания помощи в решении задач имеются тексты практических

заданий с условиями задач и вариантами их решения.

Большинство тем основано на использовании приложения Deductor. Каждый студент может скачать бесплатную версию приложения (академическая версия), получить доступ к portalу данного приложения для получения актуальной информации о нем. Академическая версия имеет ограниченный функционал. В частности, нет возможности использовать современные хранилища данных. Встроенная бесплатная база данных позволяет построить хранилища. Однако все возможности современных хранилищ данных не реализованы.

Расчетно-графическое кейс выполняется в средах бизнес-аналитики Qlik View, Qlik Sense, MS BI. Отчет представляется в распечатанной виде. В нем должны быть скрины основных окон разработанных платформ.

С целью контроля сформированности компетенций разработан фонд контрольных заданий. Его использование позволяет реализовать балльно-рейтинговую оценку, определенную приказом от 28 августа 2014 г. №168 «О применении балльно-рейтинговой системы оценки знаний студентов».

Для активизации работы студентов во время контактной работы с преподавателем отдельные занятия проводятся в интерактивной форме. В основном, интерактивная форма занятий обеспечивается при проведении занятий в компьютерном классе. Интерактивная форма обеспечивается наличием разработанных файлов с заданиями, наличием контрольных вопросов, возможностью доступа к системе дистанционного обучения, а также к тестеру.

Для работы с печатными и электронными ресурсами СЗИУ имеется возможность доступа к электронным ресурсам. Организация работы студентов с электронной библиотекой указана на сайте института (странице сайта – «Научная библиотека»).

Контрольные вопросы для подготовки к занятиям

Таблица 4.3

№ п/п	Наименование темы дисциплины	Контрольные вопросы для самопроверки
1	Тема 1. Основы языка Python	<ol style="list-style-type: none"> 1. Дайте сравнительный анализ OLAP и OLTP систем. Сферы их применения. 2. В чем отличие информационного хранилища от баз данных? 3. Принципы построения информационных хранилищ. Классификация информационных хранилищ. 4. Модели информационных хранилищ. Многомерная модель данных. Нормальная форма. Денормализация моделей данных. 5. Правила Кодда. Зачем применяется денормализация моделей? 6. Размерностные модели. В чем отличие таблицы фактов от размерностной таблицы? 7. Дайте характеристику стандартам Data Mining.
2	Тема 2. Представление данных на Python	<ol style="list-style-type: none"> 1. Дайте характеристику этапа ETL (Extracting Transforming and Loading). 2. Какие задачи решаются Data Mining? 3. Каково предназначение и средства разведочный АНализ данных и машинное обучение? Дайте характеристику диаграммы «ящик с усами» 4. Назовите какие операции выполняются при агрегировании данных. 5. Приведите примеры использования статистических пакетов для разведочного анализа. 6. Назовите и выполните сравнительный анализ графических средств анализа. Дайте характеристику биржевых диаграмм. 7. Для чего используются диаграммы рассеяния?
3	Тема 3. Машинное обучение на Python	<ol style="list-style-type: none"> 1. Зачем используются ассоциативные правила? Приведите примеры задач использования ассоциативных правил.

		<p>2. Дайте определение ассоциативного правила. Зачем используются обобщенные правила? Что такое транзакция. Приведите примеры транзакций.</p> <p>3. Какие показатели используются для построения правила?</p> <p>4. Алгоритмы построения ассоциативных правил. Алгоритм apriori.</p> <p>5. Что понимается под кластером? Назовите характеристики кластера. Что такое «центроид» кластера?</p> <p>6. Дайте классификацию методов кластерного анализа. Приведите примеры их применения в практической жизни.</p> <p>7. Зачем используются меры близости? Назовите методы определения близости между кластерами.</p> <p>8. Когда применяется метод ближнего соседа, дальнего соседа? Сравните их.</p> <p>9. Дайте характеристику метрик кластерного анализа.</p> <p>10. Поясните содержание «дендрограммы» и организацию ее применения.</p> <p>11. Что понимается под профилем кластера.</p>
4	Тема 4. Нейронные сети на Python	<p>1. Дайте определение задачи классификации. Какие методы решения задачи классификации Вы знаете?</p> <p>2. Особенности решения задач классификации с обучением.</p> <p>3. Деревья классификации и их свойства.</p> <p>4. Приведите примеры алгоритмов построения деревьев.</p> <p>5. Как определяется правило остановки построения дерева?</p> <p>6. Алгоритм CART? Приведите пример его использования.</p>
5	Тема 5. Технологии и системы работы с большими данными	<p>7. Дайте общую характеристику фреймворка для обработки больших данных. MapReduce.</p> <p>8. Охарактеризовать возможности Hadoop. Схема его работы, роли серверов в Hadoop-кластере.</p> <p>9. Особенности СУБД NoSQL.</p> <p>10. Какие основные объекты библиотеки Python Pandas.</p> <p>11. Какие основные объекты библиотеки Python SciKit-Learn,</p> <p>12. Какие основные объекты библиотеки Python Pyspark,</p> <p>13. Какие основные объекты библиотеки Python Keras.</p> <p>14. Какие основные объекты библиотеки Python TensorFlow.</p>

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. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Цифровые технологии и процессов. 3-е изд. [Электронный ресурс]- СПб. : БХВ-Петербург, 2010, 512 с.-
2. Методы и модели прогнозирования социально- экономических процессов : [учеб. пособие] / Т. С. Клебанова [и др.] ; Федер. гос. бюджетное образовательное учреждение высш. проф. образования, Рос. акад. нар. хоз-ва и гос. службы при Президенте Рос. Федерации, Сев.-Зап. ин-т упр. - СПб. : Изд-во СЗИУ РАНХиГС, 2012. - 564 с.
3. Миркин Б.Г. Введение в Цифровые технологии [Электронный ресурс] : учебник и практикум / Б. Г. Миркин ; Нац. исслед. ун-т Высш. шк. экономики. - Электрон. дан. - М. : Юрайт, 2016. - 174 с.
4. Наследов А.Д. Математические методы психологического исследования : анализ и интерпретация данных : [учебное пособие] / А.Д. Наследова. - СПб. : Речь, 2007. - 390 с.
5. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям : [хранилища данных и OLAP, очистка и предобработка данных, основные алгоритмы Data Mining, сравнение и

ансамбли моделей, решение бизнес задач на аналитической платформе Deductor] : учеб. пособие / Н. Паклин, В. Орешков. - 2-е изд., испр. - СПб.[и др.] : Питер, 2013. - 701 с.

6. Сузи Р.А. Python [Электронный ресурс] – СПб.: БХВ-Петербург, 2020.-350с.

7. Федоров Д.Ю. Программирование на языке высокого уровня python [Электронный ресурс]- М.:Юрайт, 2018.

Все источники основной литературы взаимозаменяемы.

6.2 Дополнительная литература

1. Барсегян А.А, Куприянов М.С., Степаненко В.В., Холод И.И. Технология анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – СПб.: БХВ-Петербург, 2004.

2. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows. – М.: Финансы и статистика, 2000.

3. Бринк Х., Ричардс Дж., Феверолф М. Машинное обучение. – СПб.: Питер, 2018.

4. Винстон, Уэйн Л. Excel 2007 : Цифровые технологии и бизнес- моделирование = Excel 2007: Data Analysis and Business Modelling : [пер. с англ.] / Уэйн Л. Винстон. - М. : Рус. Редакция ; СПб. : БХВ-Петербург, 2008. - 594 с.

5. Кацко И.А., Паклин Н.Б. Практикум по анализу данных на компьютере. – М.: КолосС, 2009. -278 с.

6. Ларсон Б. Разработка Бизнес-аналитики в Microsoft SQL Server 2005. – Москва: «Питер», 2008.

7. Наследов А. SPSS 19. Профессиональный статистический Цифровые технологии. – СПб. : Питер, 2011.

8. Наумов В.Н. Средства бизнес-аналитики. – СПб.: СЗИУ, 2016.

9. Наумов В.Н. Анализ данных и машинное обучение. Методы и инструментальные средства. – СПб.: СЗИУ, 2020.

10. Тихомиров Н.П. Методы эконометрики и многомерного статистического анализа. – М.: Экономика, 2011.

11. Чак Лэм — Nadoor в действии. - Издательство "ДМК Пресс" - 2012 - ISBN: 978-5-94074-785-7 - Текст электронный // ЭБС Лань - URL: <https://e.lanbook.com/book/39997>

12. Шолле Ф. Глубокое обучение на Python. – СПб. :Питер, 2018.

6.3. Учебно-методическое обеспечение самостоятельной работы.

1. Положение об организации самостоятельной работы студентов федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» (в ред. приказа РАНХиГС от 11.05.2016 г. № 01-2211);

2. Положение о курсовой работе (проекте) выполняемой студентами федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» (в ред. приказа РАНХиГС от 11.05.2016 г. № 01-2211)

6.4. Нормативные правовые документы.

Не используются

6.5. Интернет-ресурсы.

СЗИУ располагает доступом через сайт научной библиотеки <http://nwapa.spb.ru/> к следующим подписным электронным ресурсам:

Дополнительная литература и интернет-ресурсы

1. Курс “Introduction to Python for Data Science”

<https://www.datacamp.com/courses/intro-to-python-for-data-science>

2. Специализация “Python for Everybody”

<https://www.coursera.org/specializations/python>

3. Wes McKinney (2011). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.

4. Charles Severance (2016). Python for Everybody: Exploring Data in Python 3.

5. Anastasopoulos et al. (2017). Political image analysis with deep neural networks.

6. Sebastian Raschka (2015). Python Machine Learning.

7. Эдвард Тафти (1983). Визуальное представление больших объемов информации. 8. Guillermo Moncecchi, Raul Garreta (2013). Learning scikit-learn: Machine Learning in Python.

9. Aurélien Géron (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow.

10. Марк Лутц (2011). Изучаем Python.

11. Дж. Вандер Плас (2017). Python для сложных задач. Наука о данных и машинное обучение.

12. Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda (2018). Applied Text Analysis with Python.

Русскоязычные ресурсы

- Электронные учебники электронно - библиотечной системы (ЭБС) «Айбукс»
- Электронные учебники электронно – библиотечной системы (ЭБС) «Лань»
- Рекомендуется использовать следующий интернет-ресурс
- <http://serg.fedosin.ru/ts.htm>
- <http://window.edu.ru/resource/188/64188/files/chernyshov.pdf>

6.6. Иные источники.

Не используются.

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

Курс включает использование программного обеспечения Microsoft Excel, Microsoft Word, Microsoft Power Point для подготовки текстового и табличного материала, графических иллюстраций. При проведении занятий используются средства бизнес-аналитики.

Методы обучения с использованием информационных технологий (компьютерное тестирование, демонстрация мультимедийных материалов).

Интернет-сервисы и электронные ресурсы (поисковые системы, электронная почта, профессиональные тематические чаты и форумы, системы аудио и видео конференций, онлайн энциклопедии, справочники, библиотеки, электронные учебные и учебно-методические материалы).

Для организации дистанционного обучения используется система Moodle.

№ п/п	Наименование
1.	Компьютерные классы с персональными ЭВМ, объединенными в локальные сети с выходом в Интернет
2.	Пакет Excel -2013, 2016, professional plus
3.	Аналитическая платформа Qlik View, MS BI
4.	Система бизнес-аналитики Deductor Academic
5.	SPSS
6.	Язык Python, платформа Anaconda
7.	Язык R, IDE Rstudio
8.	Мультимедийные средства в каждом компьютерном классе и в лекционной аудитории
9.	Браузер, сетевые коммуникационные средства для выхода в Интернет
10.	Система дистанционного обучения Moodle

Компьютерные классы из расчета 1 ПЭВМ для одного обучаемого. Каждому обучающемуся должна быть предоставлена возможность доступа к сетям типа Интернет в течение не менее 20% времени, отведенного на самостоятельную подготовку.