МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Санкт-Петербургский государственный экономический университет»

УТВЕРЖДАЮ
Проректор по учебной и методической работе
_____ В.Г. Шубаева
«26» мая 2021 г.

Большие данные

Рабочая программа дисциплины

Направление подготовки/ *Специальность*	*38.03.05 Бизнес-информатика*
Направленность (профиль) программы/ *Специализация*	*Деловая аналитика*
Уровень высшего образования	*Бакалавриат*
Форма обучения	*очная*

Составитель*(и)*:

к.т.н, Верзун Наталья Аркадьевна

Часов по учебному плану	108	**Виды контроля в семестрах:**
в том числе:		
контактная работа	48	Зачет: семестр 5
самостоятельная работа	60	
практическая подготовка	0	
часов на контроль	0	

Распределение часов дисциплины:

Семестр:	5
Вид занятий	Часы
Лекционные занятия	20
Практические занятия	28
Лабораторные работы	
Итого аудиторных часов	**48**
Самостоятельная работа	60
Часы на контроль	0
Итого академических часов	**108**
Общая трудоемкость в зачетных единицах	**3**

Санкт-Петербург
2021

СОДЕРЖАНИЕ

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ 3
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ 3
3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ 3
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ* 3
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ 5
5.1 Рекомендуемая литература 5
5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства 6
5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД) 6
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ 7
7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ 8
8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ 9
ФОНД ОЦЕНОЧНЫХ СРЕДСТВ 11
1.1 Контрольные вопросы и задания к промежуточной аттестации 11
1.2 Темы письменных работ 16
1.3 Контрольные точки 16
1.4 Другие объекты оценивания 17
1.5 Самостоятельная работа обучающегося 17
1.6 Шкала оценивания результата 17

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель:	Ознакомление с основами и принципами работы с большими данными. Получение знаний о сущности и признаках больших данных, о существующих и перспективных методах хранения, обработки, анализа разнородных данных большого объёма; формирование умений выбирать и применять математические методы и инструментальные средства для обработки и анализа данных большого объёма при решении бизнес-задач в различных сферах деятельности.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина Б1.В Большие данные относится к части, формируемой участниками образовательных отношений Блока 1.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Код и наименование компетенции выпускника	**Код и наименование индикатора достижения компетенций**	**Планируемые результаты обучения по дисциплине**
ПК-6 - Способен планировать, организовывать и проводить мероприятия по управлению ресурсами	*ПК-6.2 - Способен организовать мероприятия по управлению ресурсами в процессе экономической деятельности*	*Знать: Особенности и принципы работы с данными большого объема. Роль и место больших данных в решении аналитических и исследовательских задач в различных предметных областях. Существующие и перспективные математические методы обработки и анализа данных большого объема.* *Уметь: Организовать мероприятия по управлению ресурсами в процессе экономической деятельности. Применять математические методы обработки и анализа данных большого объема для решения бизнес-задач различных предметных областей. Выбирать и применять инструментальные средства для анализа, исследований и моделирования процессов и явлений в профессиональной деятельности..* *Владеть: Методами и инструментальными средствами обработки, анализа и визуализации данных и навыками их использования в профессиональной деятельности.*

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ*

Номер и наименование тем и/или	**Содержание дисциплины**	**Объем дисциплины (академические часы)**

разделов/тем		Контактная работа			СРО
		ЗЛТ	ПЗ	ЛР	
Раздел I. Большие данные как базовая технология цифровой экономики.					
Тема 1. Введение. Основные понятия.	Обзор технологий третьей платформы информатизации. Понятие Big Data. Роль и значение технологий Big Data в становлении цифровой экономики. Характеристики, особенности больших данных (объем, скорость обработки, разнородность, не структурированность). Задачи, возникающие при работе с большими данными. Источники больших данных. Применение Big Data в различных сферах: государственное управление, ритейл, ЖКХ, маркетинг, финансы, социальные сети и пр. Примеры.	2	2		6
Тема 2. Основные принципы работы с Big Data.	ИТ-инфраструктура для задач класса Big Data. Организация и управление Big Data. Аналитическая обработка Big Data и выявление закономерностей. Средства поддержки принятия решения. Проблемы технологий Big Data.	4	4		10
Раздел II. Задачи управления большими данными.					
Тема 3. Хранение и управление Big Data.	Платформа Apache Hadoop. Компоненты Hadoop. Принципы построения Hadoop систем. Распределённая файловая система HDFS. NoSQL базы данных: особенности, характеристики, преимущества использования. Типы хранилищ данных. Платформа Oracle для больших данных.	2	4		8
Тема 4. Неструктурированная информация.	Автоматическая обработка текстов. Лингвистическая обработка текстов. Основные задачи, подходы к их решению. Компьютерная лингвистика. Технологии анализа и поиска текстовой информации RCO (Russian Context Optimizer). Архитектура RCO. Возможности: поиск, анализ, синтез, нечёткий поиск, семантическая сеть, рубрицирование. ABBYY Compreno – семантический разбор текста и визуализация. Алгоритм извлечения информации в ABBYY Compreno. RDF (Resource Description Framework) – граф. Sentiment analysis – анализ тональности – автоматическое извлечение субъективных мнений из текста или речи.	2	4		12
Тема 5. Методы и техники анализа Big Data.	Методы и техники анализа: дескриптивные и предиктивные методы, предобработка данных; методы класса Data Mining; ассоциативные правила; секвенциальный анализ (анализ последовательностей), кластерный анализ; корреляционный и регрессионный анализ;	6	10		12

	краудсорсинг, смешение и интеграция данных; машинное обучение (с учителем и без учителя); искусственные нейронные сети; сетевой анализ; оптимизация; генетические алгоритмы; распознавание образов; прогнозная аналитика; имитационное моделирование; пространственный анализ; статистический анализ; визуализация аналитических данных.				
Тема 6. Технологии и инструментальные средства.	Система MapReduce (модель распределённых вычислений): алгоритмы, графы. Statistica Big Data Analytics инструмент статистического и графического анализа, прогнозирования, data mining и др. Возможности. Примеры применения. Аналитическая платформа Deductor. Состав. Назначение. Возможности. Примеры применения. Язык программирования R для статистической обработки данных и работы с графикой. Возможности R для статистической обработки данных. Методы обработки информационных блоков.	4	4		12
Контроль:					**0**
Всего по дисциплине:		**20**	**28**	**0**	**60**

*ЗЛТ – занятия лекционного типа, ПЗ – все виды занятий семинарского типа, кроме лабораторных работ, ЛР – лабораторные работы, СРО – самостоятельная работа обучающегося

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Рекомендуемая литература

Библиографическое описание издания (автор, заглавие, вид, место и год издания, кол. стр.)	**Электронные ресурсы**
1. Парфенов Ю.П. Постреляционные хранилища данных: учебное пособие для вузов / Ю. П. Парфенов; под науч. ред. Н. В. Папуловской. — М.: Издательство Юрайт, 2019 — 121 с.	https://www.urait.ru/bcode/438577
2. Гасанов Э.Э. Интеллектуальные системы. Теория хранения и поиска информации : учебник для бакалавриата и магистратуры / Э. Э. Гасанов, В. Б. Кудрявцев. — 2-е изд., испр. и доп. — Электрон. дан. — Москва : Издательство Юрайт, 2019. — 271 с.	https://www.urait.ru/bcode/437023
3. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Электрон. дан. — М.: Издательство Юрайт, 2019. — 174 с.	https://www.urait.ru/bcode/432851

4. Верзун Н.А. Введение в инфокоммуникационные технологии и сети Future Networks : учебное пособие / Н.А.Верзун , М.О.Колбанев, А.В.Омельян .— Санкт-Петербург : Изд-во СПбГЭУ, 2016 .— 51 с.	http://opac.unecon.ru/elibrary ... BE%D0%BD%D0%BD%D1%8B%D0%B5.pdf
5. Верзун Н.А. Сетевая архитектура цифровой экономики : [монография] / Н.А.Верзун, М.О.Колбанев , А.В.Омельян .— Санкт-Петербург : Изд-во СПбГЭУ, 2018 .— 156 с.	http://opac.unecon.ru/elibrary ... 80%D0%BE%D0%B2%D0%BE%D0%B9.pdf

5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства

- 7-Zip
- Microsoft Office Professional
- Microsoft Windows Professional
- Deductor
- Loginom
- R

5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД)

№	Наименование СПБД/ ИСС
1.	Электронная библиотека Grebennikon.ru – www.grebennikon.ru
2.	Научная электронная библиотека eLIBRARRY – www.elibrary.ru
3.	Научная электронная библиотека КиберЛеника – www.cyberleninka.ru
4.	База данных ПОЛПРЕД Справочники – www.polpred.com
5.	База данных OECD Books, Papers & Statistics на платформе OECD iLibrary www.oecd-ilibrary.org
6.	Справочная правовая система КонсультантПлюс (инсталлированный ресурс СПбГЭУ или www.consultant.ru)
7.	Справочная правовая система «ГАРАНТ» (инсталлированный ресурс СПбГЭУ или www.garant.ru)
8.	Информационно-справочная система «Кодекс» (инсталлированный ресурс СПбГЭУ или www.kodeks.ru)
9.	Электронная библиотечная система BOOK.ru - www.book.ru
10.	Электронная библиотечная система ЭБС ЮРАЙТ – www.urait.ru
11.	Электронно-библиотечная система ЗНАНИУМ (ZNANIUM) – www.znanium.com
12.	Электронная библиотека СПбГЭУ– opac.unecon.ru

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для реализации данной дисциплины имеются специальные помещения для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ) групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.

Помещения оснащены оборудованием и техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование учебных аудиторий, перечень	**Адрес (местоположение) учебных аудиторий**
Ауд. 2018 Компьютерный класс (для проведения практических занятий, курсового проектирования (выполнения курсовых работ) с применением вычислительной техники). Оборудован мультимедийным комплексом. Специализированная мебель и оборудование: Учебная мебель на 16 посадочных мест (16 компьютерных столов, 16 черных кресел), рабочее место преподавателя 2стола+1кресло, доска меловая 1 шт., доска маркерная на колесиках 1шт., вешалка стойка 1шт., стул изо10шт.Компьютер Intel I5-7400/16Gb/1Tb/ видеокарта NVIDIA GeForce GT 710/Монитор. DELL S2218H - 17 шт., Точка беспроводного доступа Wi-Fi Тип1 UBIQUITI UAP-AC-PRO - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	191023, г. Санкт-Петербург, ул. Канал Грибоедова, 30/32, литер «А», «Б», «Р»
Ауд. 2022 Лаборатория "Лабораторный комплекс "Специализированная мебель и оборудование: Учебная мебель на 19 посадочных мест (19 компьютерных стола, 19 черных кресела) рабочее место преподавателя (компьютерный стол 1шт., кресло 1шт.), доска меловая односекционная 1шт., доска маркерная на колесиках 1 шт., стол 1шт., стул 1шт., жалюзи 1шт., вешалка стойка 1шт.Компьютер Intel i5 4460/1Тб/8Гб/монитор Samsung 23" - 1 шт., Компьютер Intel i5 4460/1Тб/8Гб/ монитор Samsung 23" - 18 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	191023, г. Санкт-Петербург, ул. Канал Грибоедова, 30/32, литер «А», «Б», «Р»

Ауд. 0007 Компьютерный класс (для проведения практических занятий, курсового проектирования (выполнения курсовых работ) с применением вычислительной техники). Оборудован мультимедийным комплексом. Специализированная мебель и оборудование: Учебная мебель на 33 посадочных места, рабочее место преподавателя, доска меловая - 1 шт., доска маркерная на колесиках - 1 шт., вешалка стойка - 3 шт., жалюзи - 3 шт., Компьютер Intel Core i3 6100/ MSI H110M PRO-D/ ОЗУ DDR4 8GB 2400MHz/SSD SATA III 240Gb/Aerocool Qs-180 400W/Клавиатура + мышь Microsoft400 for Business/монитор Asus VS228DE - 24 шт., Мультимедийный проектор Тип 1 Optoma x 400 - 1 шт., Ноутбук HP 250 G6 1WY58EA - 2 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	191023, г. Санкт-Петербург, ул. Канал Грибоедова, 30/32, литер «А», «Б», «Р»

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Приступая к изучению дисциплины, обучающемуся необходимо ознакомиться со следующими документами:

– учебно-методической документацией;

– локальными нормативными актами, регламентирующими основные вопросы организации и осуществления образовательной деятельности, в том числе регламентирующие порядок проведения текущего контроля успеваемости и промежуточной аттестации обучающихся;

– графиком консультаций сотрудников профессорско-преподавательского состава.

Уровень и глубина освоения дисциплины определяются активной и систематической работой обучающихся на лекционных занятиях, занятиях семинарского типа, выполнением самостоятельной работы, в том числе в части выделения наиболее значимых и актуальных проблем для дальнейшего изучения. Особым условием качественного освоения дисциплины является эффективная организация труда, позволяющая распределить учебную нагрузку равномерно в соответствии с графиком учебного процесса.

При подготовке к учебным занятиям обучающимся предоставляется возможность посещения консультаций сотрудников профессорско-преподавательского состава СПбГЭУ согласно расписанию, установленному в графике консультаций.

Аудиторная и внеаудиторная работа обучающихся должна быть направлена на формирование:

- фундаментальных основ мировоззрения обучающихся и естественнонаучного познания;
- базисных знаний, соответствующих направлению подготовки и заявленной профессиональной области, формирующих целевую и профессиональную основу для подготовки кадров;
- профессиональных компетенций ориентированных на удовлетворение потребностей рынка труда;
- индивидуальной траектории посредством освоения уникального набора профессиональных компетенций дополняющих компетентностную модель обучающегося, за счет ориентации на конкретные профессиональные специализированные области знаний, определяемые представителями рынка труда;
- метанавыков обучающихся, таких как: командная работа и лидерство, анализ данных, цифровые навыки, разработка и реализация проектов, межкультурное взаимодействие.

8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

В целях освоения учебной программы дисциплины инвалидами и лицами с ограниченными возможностями здоровья Университет обеспечивает:

– для инвалидов и лиц с ограниченными возможностями здоровья по зрению: размещение в доступных для обучающихся, являющихся слепыми или слабовидящими, местах и в адаптированной форме справочной информации о расписании учебных занятий; присутствие ассистента, оказывающего обучающемуся необходимую помощь; выпуск альтернативных форматов методических материалов (крупный шрифт или аудиофайлы);

– для инвалидов и лиц с ограниченными возможностями здоровья по слуху: надлежащими звуковыми средствами воспроизведение информации;

– для инвалидов и лиц с ограниченными возможностями здоровья, имеющих нарушения опорно-двигательного аппарата: возможность беспрепятственного доступа обучающихся в учебные помещения, туалетные

комнаты и другие помещения кафедры, а также пребывание в указанных помещениях.

Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечены печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья. Образование обучающихся с ограниченными возможностями здоровья может быть организовано как совместно с другими обучающимися, так и в отдельных группах или в отдельных организациях.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1.1 Контрольные вопросы и задания к промежуточной аттестации

Промежуточная аттестация проводится в форме зачета. Перечень всех тестовых вопросов по дисциплине «Большие данные» разбит на 25 категорий в каждой из которых от 3 до 10 тестовых заданий. Выполняется тест в ЭИОС, где для каждого студента автоматически формируется тест из Банка вопросов – всего каждому студенту задается 25 вопросов. Время выполнения теста – 1 час 30 минут.

Ниже представлен типовой (автоматически сформированный в Moodle) вариант теста.

1. Что из нижеперечисленного является Мастером аналитической платформы Deductor
Выберите один или несколько ответов:
1 - Мастер обработки
2 - Мастер тиражирования
3 - Мастер агрегации
4 - Мастер трансформации
5 - Мастер импорта
6 - Мастер очистки данных
2. - Какие из нижеперечисленных процедур в Deductor не относятся к процедурам трансформации данных.
Выберите один или несколько ответов:
1 - квантование
2 - сортировка
3 - замена данных
4 - редактирование выбросов
5 - выявление дубликатов
3. Найдите в таблице строки противоречия и строки дубликаты

	Входные поля			Выходное поле	Информационное поле
	Фамилия	Имя	Отчество	Оклад, руб	Должность
1	Иванов	Иван	Сергеевич	30 000	Инженер
2	Петров	Сергей	Андреевич	25 000	Диспетчер
3	Петров	Сергей	Андреевич	43 000	Механик
4	Иванов	Иван	Сергеевич	30 000	Механик
5	Петров	Сергей	Андреевич	28 000	Механик

Ответ дайте в таком виде: П-1,2,3,Д-3,4,5 (без пробелов, через запятую)
4. Дана транзакционная база данных. Чему равна поддержка набора {огурцы, соль}, выраженная в процентах. Ответ округлите до целого числа

TID	**Покупки**
1	Спагетти соус соль

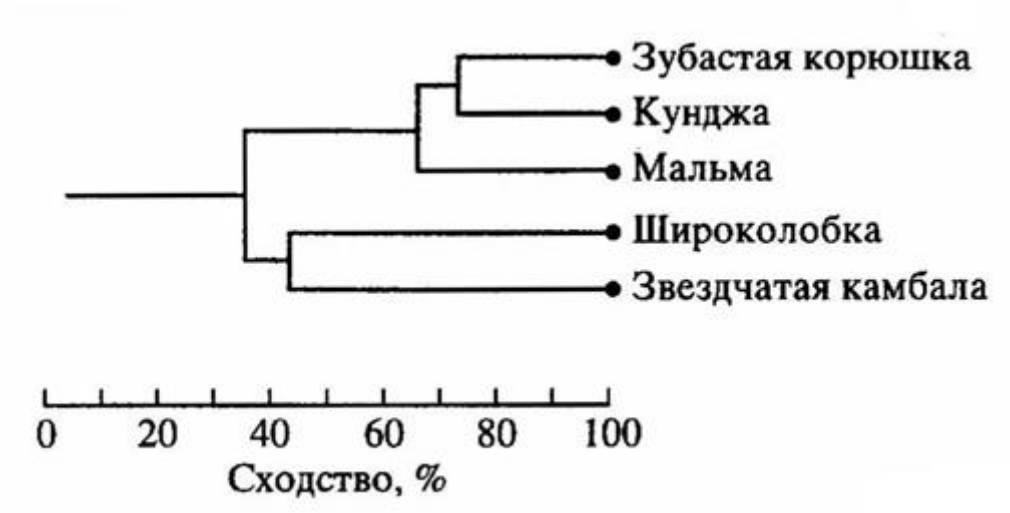
2	Хлеб соус огурцы
3	Огурцы помидоры соль
4	Спагетти помидоры
5	Хлеб огурцы соль
6	Хлеб спагетти соус
7	Помидоры огурцы соль

5. Дана транзакционная база данных. Чему равна достоверность следующего ассоциативного правила:
"Из покупки ручки следует покупка блокнота", выраженная в процентах?
В ответе написать только число (округлить до целого), БЕЗ символа %.

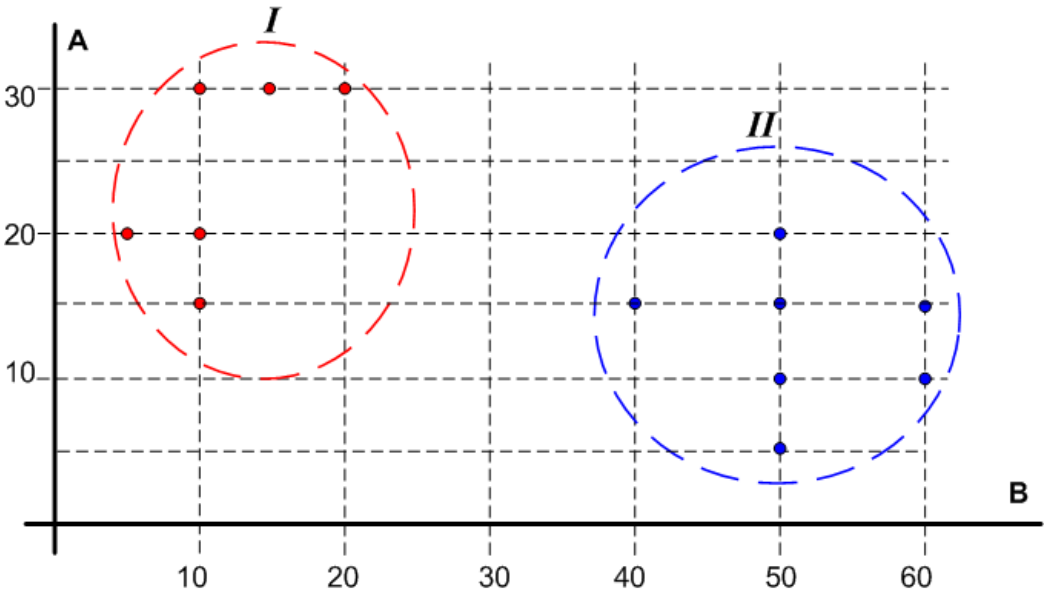
TID	**Покупки**
1	Резинка блокнот пенал ручка
2	Тетрадь ручка карандаш резинка
3	Карандаш пенал линейка ручка
4	Линейка блокнот тетрадь пенал
5	Ручка карандаш пенал
6	Резинка линейка карандаш
7	Карандаш резинка пенал тетрадь

6. Фермер может выращивать или морковь, или свеклу.
Вероятности того, что цены на будущий урожай этих культур повысятся, останутся на том же уровне или понизятся, равны соответственно 0,3, 0,4 и 0,3.
Если цены возрастут, урожай моркови даст 200 000 руб. чистого дохода, а урожай свеклы — 120 000 руб. Если цены останутся неизменными, фермер лишь покроет расходы. Но если цены станут ниже, урожай моркови и свеклы приведет к потерям в 23 000 и 19 000 руб. соответственно.
Дается уже частично построено дерево решения для этой задачи.
Какую культуру следует выращивать фермеру? Каково ожидаемое значение дохода от реализации культур?
В ответе достаточно написать название культуры и сумму. Расчет приводить не надо.
7. Какие из нижеприведенных алгоритмов относятся к методам Data Mining (например в Deductor)
Выберите один или несколько ответов:
1 - кластеризация
2 - редактирование выбросов
3 - ассоциация
4 - заполнение пропущенных данных
5 – квантование
8. Целями кластеризации могут быть…. Выберите из нижепредложенного
1 - Поиск ассоциативных правил.
2 - Отнесение каких-либо объектов к одной из заранее известных групп.
3 - Обнаружение новизны – выделение нетипичных объектов.
4 - Регрессионный анализ.
5 - Разбиение выборки на группы схожих объектов, что позволяет упростить дальнейшую обработку данных.

6 - Сжатие данных.
7 - Выявление “цепочек”, связанных во времени событий.
9. Ниже представлен процесс кластеризации объектов (рыб).
При разбиении рыб на два кластера, какие рыбы попадут в более мощный кластер?
В ответе ОБЯЗАТЕЛЬНО (для корректной оценки) названия рыб написать по алфавиту, через запятую.

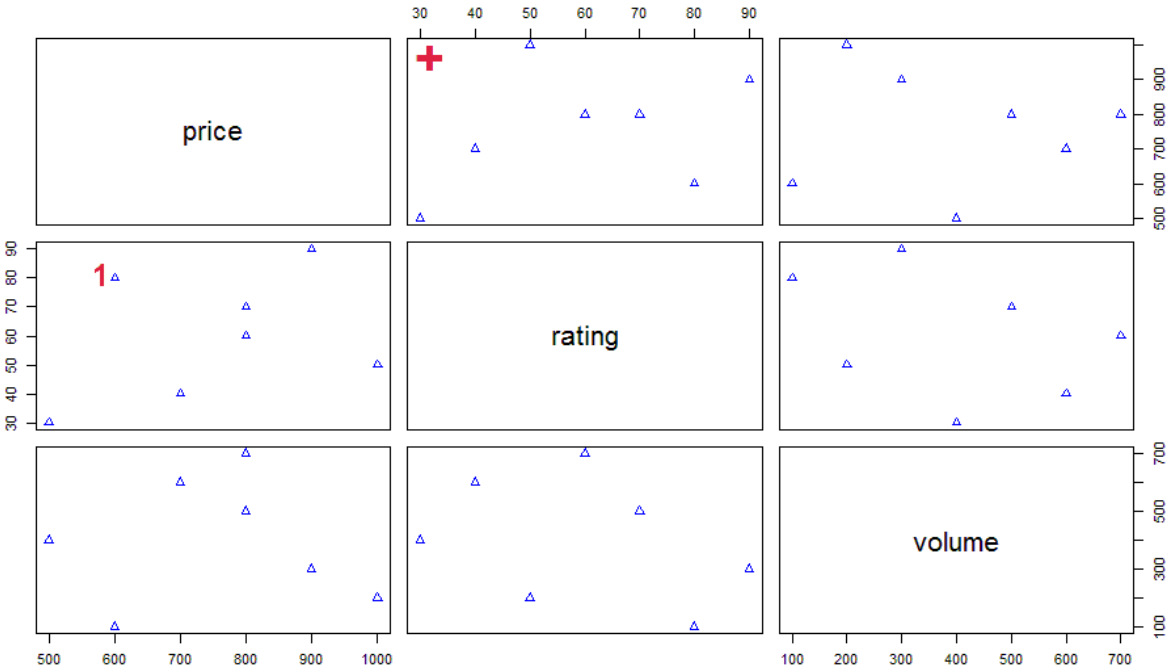


10. Выше показана диаграмма.
Результат какого вида/типа/класса анализа(алгоритма) она позволяет визуализировать?
Выберите из предложенных вариантов один ответ
1 - неиерархические алгоритмы кластеризации
2 - алгоритмы построения дерева правил
3 - нечеткие алгоритмы кластеризации
4 - алгоритмы построения дерева решений
5 - иерархические алгоритмы кластеризации
6 - итеративные алгоритмы кластеризации
11. На рисунке ниже показан результат кластеризации объектов по двум параметрам A и B.
В результате объекты были разбиты на два кластера I и II.
Найдите расстояние между двумя этими кластерами рассчитанное по методу ближнего соседа. В ответе напишите только число.

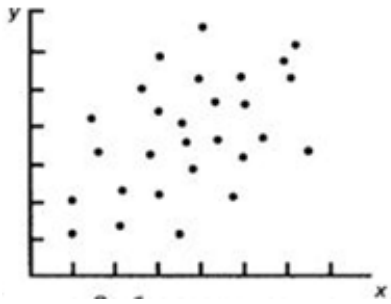


12. На рисунке ниже показаны диаграммы рассеяния для трех параметров: цена, рейтинг, объем.

Зависимость каких двух параметров показана на диаграмме отмеченном красным крестиком (что по оси y? и что по оси x?).
Напишите значение всех трех координат(цена, рейтинг, объем) для 1-го наблюдения - отмечен красной цифрой 1.



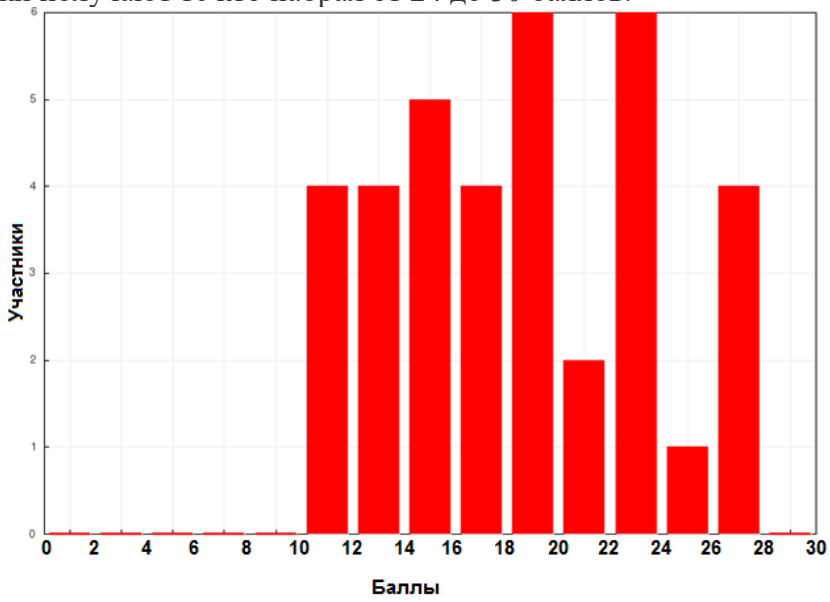
13. Два документа представлены в виде бинарных векторов:
D1 ={0,0,0,1,0,1,0,1,1,1,0,0}
D2 ={0,1,0,1,0,0,0,0,0,1,0,1}
Чему равно расстояние Хэмминга между этими документами
14. По виду диаграммы рассеяния **на рисунке** выберите какая зависимость между величинами X и Y.



15. Какой оператор R позволяет найти коэффициент корреляции Пирсона для двух величин?
Какой оператор позволяет вывести на экран рабочую директорию?
16. Ниже приведен фрагмент кода. Напишите выражение для уравнения линейной регрессии: y(x)

```
…
Call:
lm(formula =y~x)
Residuals:
    Min       1Q     Median      3Q     Max
-7.2002    -1.3339   0.0513    1.8944  4.0225
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -40.11165    6.03121  -3.448  0.00149 **
x             1.50001    0.04131  10.005  1.23e-05 ***
-------
```

17. Для визуального представления результатов сдачи 2 контрольной точки по ПОИС de.unecon.ru сформировал гистограмму представленную ниже.
Сколько сдающих получили 5 за эту контрольную точку если было установлено, что отличные оценки получают те кто набрал от 24 до 30 баллов.

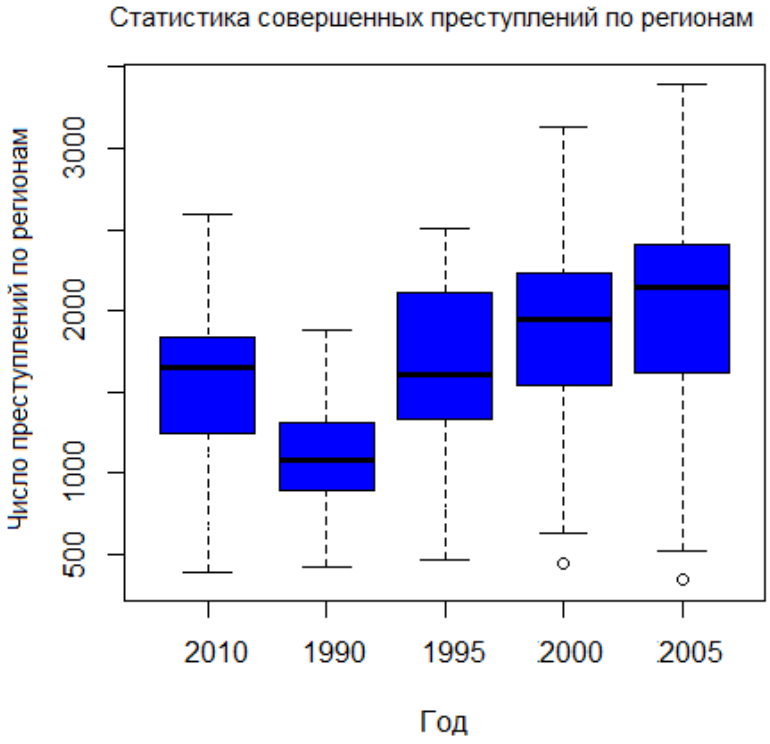


18. Для поиска кратчайшего пути на графе используют. Найдите лишние.
Выберите один или несколько ответов:
1 - Алгоритм почтальона
2 - Алгоритм ранца
3 - Метод ветвей и границ
4 - Алгоритм Дейкстры
5 - Алгоритм Ли
19. “Встретить” необходимость решения задачи о рюкзаке можно...
Выберите один или несколько ответов:
1 - В работе GPS навигатора
2 - При выборе лучшего товара из нескольких предложенных
3 - При раскрое материала (н-р, ткани)
4 - При оптимальной загрузке склада
5 - При трассировке печатных плат
6 - В протоколе OSPF стека TCP/IP
20. Есть следующие предметы:
1 предмет с весом 5 кг и ценой 50
2 предмет с весом 10 кг и ценой 30
3 предмет с весом 10 кг и ценой 80
4 предмет с весом 15 кг и ценой 60
Известна вместимость рюкзака: 35 кг
Какие предметы и в каком порядке попадут в рюкзак если его загружать, используя жадный алгоритм?
Чему будет равен вес и стоимость собранного в рюкзак?
В ответе надо через запятую без пробелов написать номера предметов, вес, стоимость.
Например так
1,2,3,40,90
21.Это система оценки рисков. Например: оценка риска невозврата кредита.
22. Предприятие оптовой торговли разработало 4 варианта плана продажи товаров (1, 2, 3, 4) в предстоящем сезоне с учетом меняющейся конъюнктуры рынка и спроса покупателей (A, B, C, D, E). Получающиеся от их возможных сочетаний величины прибыли представлены в виде матрицы выигрышей. Определить оптимальный план продажи товаров по критерию максимакса.
В ответе написать номер плана и ожидаемое значение прибыли (оптимистический сценарий) через пробел.

Например так: 3 18

	A	B	C	D	E
1	15	23	22	17	10
2	21	14	20	19	24
3	11	17	19	22	21
4	23	19	21	23	15

23. Для более глубокого (по сравнению с ассоциативными правилами) понимания (и сжатого описания) логики действий покупателя могут применяться ...
Это ассоциативные правила, в которых важен порядок следования событий (покупок).
Вставьте пропущенные слова.
24. В результате работы некоторой информационной поисковой системы в наборе из 1 500 000 документов было найдено 15 документов, содержащих искомое слово. Документ А содержит 3000 слов и искомое слово встречается в нем 30 раз. Определите релевантность (искомому слову) Документа А методом TF-IDF. Ответ округлите до двух знаков после запятой.
25. Ниже приведена статистика числа совершенных преступлений по регионам России.
Сравните три года: 1995, 2000, 2010. В каком году (из этих трех) в среднем по регионам было совершено меньше преступлений?
В ответе напишите просто год без комментариев (Например: 1999).



1.2 Темы письменных работ

Рабочей программой дисциплины не предусмотрено.

1.3 Контрольные точки

Номер контрольной точки	Тип контрольной точки	Способ проведения	Номера тем
1	*Доклад*	*устно*	*1,2,3,5,6*
2	*Защита практикума*	*устно*	*5,6*

1.4 Другие объекты оценивания

Рабочей программой дисциплины не предусмотрено.

1.5 Самостоятельная работа обучающегося

Наименования самостоятельной работы	Номера тем
Написание реферата	*1,2,3,5,6*
Выполнение расчетных, аналитических, расчетно-графических и др. заданий	*4,5*
Подготовка к лекционным и практическим занятиям	*2-6*

1.6 Шкала оценивания результата

Шкалы оценивания и процедуры оценивания результатов обучения **по дисциплине** регламентируются Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся по программам высшего образования и Положением о балльно-рейтинговой системе.

Для оценки сформированности результатов обучения по дисциплине используется **балльно-рейтинговая система успеваемости обучающихся**:

Формой итогового контроля по дисциплине является зачет, итоговый результат формируется в соответствии со шкалой, приведенной ниже в таблице:

Баллы	Оценка
<55	Незачет
>=55	Зачет

Шкала оценивания результата

2 (балл до 54)	Демонстрирует непонимание проблемы. Многие требования, предъявляемые к заданию не выполнены. Демонстрируется первичное восприятие материала. Работа незакончена и /или это плагиат.
3 (балл 55-69)	Демонстрирует частичное понимание проблемы. Большинство требований, предъявляемых, к заданию выполнены. Владение элементами заданного материала. В основном выполненный материал понятен и носит целостный характер.
4 (балл 70-84)	Демонстрирует значительное понимание проблемы обозначенной дисциплиной. Все требования, предъявляемые к заданию выполнены. Содержание выполненных заданий раскрыто и рассмотрено с разных точек зрения.
5 (балл 85-100)	Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию выполнены. Продемонстрировано уверенное владение материалом дисциплины. Выполненные задания носят целостных характер, выполнены в полном объеме, структурированы, представлены различные точки зрения, продемонстрирован творческий подход.