

**Частное образовательное учреждение высшего образования
«ИНСТИТУТ БИЗНЕСА И ИННОВАЦИОННЫХ ТЕХНОЛОГИЙ»**

Одобрено
решением Ученого совета
от «29» июля 2023г.
протокол № 2



УТВЕРЖДАЮ
Ректор Института бизнеса
и инновационных технологий

А.И. Садыкова

«29» июля 2023 г.

**Рабочая программа дисциплины
«Системы обработки больших данных (Big Data)»**

Направление подготовки: 38.03.02 Менеджмент
Профиль подготовки: Управление проектами
Квалификация выпускника: Бакалавр
Форма обучения: очно-заочная

Вологда 2023

Содержание

1. АННОТАЦИЯ К ДИСЦИПЛИНЕ	3
2. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	5
3. ТЕМАТИЧЕСКИЙ ПЛАН	6
4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	7
5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ	9
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	12
7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	15

1. АННОТАЦИЯ К ДИСЦИПЛИНЕ

Рабочая программа дисциплины «Системы обработки больших данных (Big Data)» составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего образования по направлению подготовки 38.03.02 Менеджмент, утвержденным приказом Министерства науки и высшего образования Российской Федерации от 12.08.2020 N 970.

Основными потребителями технологий [Big Data](#) в являются [банки](#), [телеком](#)-операторы и крупные [ритейлеры](#). Главными проблемами развития направления больших данных в России являются нехватка квалифицированных кадров, отсутствие достаточного опыта российских внедрений, а также высокая стоимость решений.

Дисциплина «Большие данные в экономике» предоставляет необходимые знания для участия в проектах по анализу больших данных. Включает информацию о фазах жизненного цикла процессов аналитики больших данных при переходе бизнеса к использованию Big Data. Материал дисциплины обеспечивает слушателей знанием базовых и расширенных аналитических методов и техник, применяемых для поиска и извлечения знаний из больших массивов разнородных данных. Сравнение различных версий дистрибутивов платформы Hadoop, облачные сервисы для хранения и аналитики данных, "open source" и коммерческие инструменты, используемые для хранения, обработки, визуализации и аналитики больших данных.

Материал учебной дисциплины предназначен для использования в курсах, связанных с количественным анализом реальных экономических явлений, таких как, например, прикладная микро- и макроэкономика, маркетинг и других. Может быть использован в спецкурсах по теории случайных процессов, математическим моделям в экономике, оптимальному управлению, статистическому прогнозированию, применению методов теории вероятностей в финансовой математике, принятию решений в условиях неопределенности.

Место дисциплины в структуре образовательной программы

Настоящая дисциплина включена в учебные планы по программам подготовки бакалавров по направлению подготовки 38.03.02 Менеджмент и входит в обязательную часть Блока 1.

Дисциплина изучается на 3 курсе в 6 семестре.

Цель и задачи дисциплины.

Целью изучения дисциплины «Системы обработки больших данных (Big Data)» является теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов

структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу бакалавриата, для решения различных задач практической и научно-исследовательской деятельности.

Задачи дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- усвоение знаний о сущности, структуре и видах математических моделей принятия решений;
- приобретение практических навыков работы с программой SPSS, Stata.
- развитие навыков создания и решения моделей, необходимых в сфере экономики.

2. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Процесс изучения дисциплины направлен на формирование следующих компетенций, предусмотренных образовательной программой.

Результаты освоения ООП (содержание компетенций)	Код компетенции	Код и наименование индикатора достижения компетенций	Перечень планируемых результатов обучения по дисциплине			Формы образовательной деятельности
			выпускник должен знать	выпускник должен уметь	выпускник должен иметь практический опыт	
Способен использовать при решении профессиональных задач современные информационные технологии и программные средства, включая управление крупными массивами данных и их интеллектуальный анализ	ОПК-5	ОПК-5.2. Управляет крупными массивами данных и осуществляет их интеллектуальный анализ	принципы и методы работы с большими Данными на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности; принципы построения современных компьютерных кластеров для обработки Больших Данных	использовать параллельные и масштабируемые алгоритмы обработки информации различных видов при решении практических задач в области информационных систем и технологий	поиска и обработки информации различных видов с использованием современных компьютерных технологий в соответствии принятыми идеями и подходами к решению задачи	<u>Контактная работа:</u> Лекции Лабораторные практикумы <u>Самостоятельная работа</u>

3. ТЕМАТИЧЕСКИЙ ПЛАН

Наименование тем	Контактная работа обучающихся с преподавателем (по видам учебных занятий)								Самостоятельная работа обучающихся	Форма ТКУ Форма ПА, балл	
	Лекции	Семинары	Практикум по решению задач	Ситуационный практикум	Мастер-класс	Лабораторный практикум	Тренинг	Дидактическая игра			Из них в форме практической подготовки
<i>Очно-заочная форма</i>											
Тема 1. Введение в дисциплину <i>Big Data</i> в экономике. Обзор источников информации	2					2				18	Отчет по лабораторному практикуму /10
Тема 2. Техника и технология хранения и обработки <i>Big Data</i>	2					2				18	Отчет по лабораторному практикуму /10
Тема 3. Процесс анализа больших данных. Технологии анализа больших данных. Научные проблемы в области больших данных	1					1				18	Отчет по лабораторному практикуму /10
Тема 4. Статистические методы анализа данных	1					1				18	Отчет по лабораторному практикуму /10 Домашнее задание №1/5
Тема 5. Современные программные средства анализа больших объемов информации	2					2				20	Отчет по лабораторному практикуму /10 Домашнее задание №2/5
Всего:	8					8				92	100 (ТКУ+ПА)
Контроль, час	-									Зачет (40 из 100)	
Объем дисциплины (в академических часах)	108										
Объем дисциплины (в зачетных единицах)	3										

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Тема 1. Введение в дисциплину Big Data в экономике. Обзор источников информации.

Понятие, сущность, задачи и ключевые признаки больших данных. Роль и место больших данных в решении аналитических и исследовательских задач профессиональной деятельности. Вопросы безопасности. Понятие Data Mining. Когнитивный анализ данных. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации). Методики сбора данных.

Информационное общество и большие данные. М. МакЛюэн и роль медиа в больших данных. Современные концепции больших данных Информационно-технологическая структура Больших данных. Сферы применения Больших данных.

Тема 2. Техника и технология хранения Big Data

Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных. Техники больших данных. Консолидация данных. Визуализация. Классификация.

Технологии, используемые для сбора и обработки Больших Данных: программное обеспечение (ПО), технологическое оборудование и сервисные услуги. Наиболее распространенный инструмент обработки Больших Данных из группы ПО: SQL, NoSQL, MapReduce, Hadoop, SAP HANA. Блокчейн системы как приложение системы Больших данных. Структурированные и неструктурированные данные. Набор признаков Больших данных (VVV-подход).

Тема 3. Процесс анализа больших данных. Технологии анализа больших данных. Научные проблемы в области больших данных.

Методы и техники анализа, применимые к большим данным:

1. Методы класса Data Mining: обучение ассоциативным правилам (англ. association rule learning), классификация (методы категоризации новых данных на основе принципов, ранее применённых к уже наличествующим данным), кластерный анализ, регрессионный анализ;

2. Краудсорсинг - категоризация и обогащение данных силами широкого, неопределённого круга лиц, привлечённых на основании публичной оферты, без вступления в трудовые отношения;

3. Смешение и интеграция данных (англ. data fusion and integration) - набор техник, позволяющих интегрировать разнородные данные из разнообразных источников для возможности глубинного анализа, в качестве примеров таких техник, составляющих этот класс методов приводятся цифровая обработка сигналов и обработка естественного языка (включая

тональный анализ);

4. Машинное обучение, включая обучение с учителем и без учителя, а также Ensemble learning (англ.) - использование моделей, построенных на базе статистического анализа или машинного обучения для получения комплексных прогнозов на основе базовых моделей (англ. constituent models, ср. со статистическим ансамблем в статистической механике);

5. Искусственные нейронные сети, сетевой анализ, оптимизация, в том числе генетические алгоритмы; распознавание образов;

6. Прогнозная аналитика; имитационное моделирование; пространственный анализ (англ. Spatial analysis) - класс методов, использующих топологическую, геометрическую и географическую информацию в данных;

7. Статистический анализ, в качестве примеров методов приводятся A/B-тестирование и анализ временных рядов;

8. Визуализация аналитических данных - представление информации в виде рисунков, диаграмм, с использованием интерактивных возможностей и анимации как для получения результатов, так и для использования в качестве исходных данных для дальнейшего анализа

Тема 4. Статистические методы анализа данных

Основные понятия математической статистики. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).

Тема 5. Современные программные средства анализа больших объемов информации

Сущность и принцип работы аналитических платформ SPSS/Stata. Основные функции и инструменты аналитических платформ SPSS/Stata для целей анализа и исследования финансово-экономических процессов и явлений в деятельности предприятий и банков. Моделирование финансово-экономических процессов и явлений в деятельности предприятий и банков с помощью платформ SPSS/Stata. Инструментарий прикладного компьютерного анализа и моделирования в SPSS/Stata.

Статистическая обработка данных в программах SPSS и Stata: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы).

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

В процессе изучения дисциплины используются такие виды учебной работы, как лекции, лабораторные практикумы, а также различные виды самостоятельной работы обучающихся по заданиям преподавателя.

Методические указания для обучающихся при работе над конспектом лекций во время проведения лекции.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект, что позволит впоследствии вспомнить изученный учебный материал, дополнить содержание при самостоятельной работе с литературой, подготовиться к экзамену.

Следует также обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации, положительный опыт в ораторском искусстве. Желательно оставить в рабочих конспектах поля, на которых делать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений.

Любая лекция должна иметь логическое завершение, роль которого выполняет заключение. Выводы по лекции подытоживают размышления преподавателя по учебным вопросам. Формулируются они кратко и лаконично, их целесообразно записывать. В конце лекции обучающиеся имеют возможность задать вопросы преподавателю по теме лекции.

Методические указания для обучающихся по выполнению лабораторных практикумов.

Лабораторные практикумы выполняются в соответствии с рабочим учебным планом при последовательном изучении тем дисциплины.

Порядок проведения практикума.

Получение задания и рекомендаций к выполнению практикума.

Настройка инструментальных средств, необходимых для выполнения практикума.

Выполнение заданий практикума.

Подготовка отчета в соответствии с требованиями.

Сдача отчета преподавателю.

В ходе выполнения практикума необходимо следовать технологическим инструкциям, использовать материал лекций, рекомендованных учебников, источников интернета, активно использовать помощь преподавателя на занятии.

Требования к оформлению результатов практикумов (отчет)

При подготовке отчета: изложение материала должно идти в логической последовательности, отсутствие грамматических и синтаксических ошибок, шрифт Times New Roman, размер – 14, выравнивание по ширине, отступ первой строки – 1,25, междустрочный интервал – 1,5, правильное оформление рисунков (подпись, ссылка на рисунок в тексте).

При подготовке презентации: строгий дизайн, минимум текстовых элементов, четкость формулировок, отсутствие грамматических и синтаксических ошибок, воспринимаемая графика, умеренная анимация.

Методические указания для обучающихся по выполнению домашних заданий

Домашнее задание выполняется в соответствии с рабочим учебным планом при последовательном изучении тем дисциплины.

Прежде чем приступить к выполнению домашнего задания, обучающемуся необходимо:

- ознакомиться с соответствующими разделами программы дисциплины по учебной литературе, рекомендованной программой курса;
- получить от преподавателя рекомендации о порядке выполнения заданий;
- получить от преподавателя конкретное задание и информацию о сроках выполнения, требованиях к оформлению, форме представления и критериях оценки результатов работы.

В ходе выполнения домашнего задания необходимо следовать технологическим инструкциям, использовать материал лекций и практических занятий, рекомендованных учебников, источников интернета.

Требования к оформлению результатов домашних заданий

В зависимости от темы результаты домашнего задания представляются в виде:

- результатов расчетов и модели, полученных при помощи пакета прикладных статистических программ;
- аналитической записки, подкрепленной результатами и протоколом расчетов.

Методические указания для обучающихся по организации самостоятельной работы.

Самостоятельная работа обучающихся направлена на самостоятельное изучение отдельных тем/вопросов учебной дисциплины.

Самостоятельная работа является обязательной для каждого обучающегося, ее объем по дисциплине определяется учебным планом.

При самостоятельной работе обучающийся взаимодействует с рекомендованными материалами при минимальном участии преподавателя.

Работа с литературой (конспектирование)

Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий

и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме.

Изучая материал по учебной книге (учебнику, учебному пособию, монографии и др.), следует переходить к следующему вопросу только после полного уяснения предыдущего, фиксируя выводы и вычисления, в том числе те, которые в учебнике опущены или на лекции даны для самостоятельного вывода.

Особое внимание обучающийся должен обратить на определение основных понятий курса. Надо подробно разбирать примеры, которые поясняют определения. Полезно составлять опорные конспекты.

Выводы, полученные в результате изучения учебной литературы, рекомендуется в конспекте выделять, чтобы при перечитывании материала они лучше запоминались.

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса.

Вопросы, которые вызывают у обучающегося затруднение при подготовке, должны быть заранее сформулированы и озвучены во время занятий в аудитории для дополнительного разъяснения преподавателем.

Навигация для обучающихся по самостоятельной работе в рамках изучения дисциплины

Наименование темы	Вопросы, вынесенные на самостоятельное изучение	Формы самостоят. работы	Форма текущего контроля
<i>Тема 1. Введение в дисциплину Big Data в экономике. Обзор источников информации</i>	Когнитивный анализ данных	Работа с литературой, источниками в сети Internet. Подготовка к тесту. Выполнение теста.	Отчёт по лабораторному практикуму.
<i>Тема 2. Техника и технология хранения и обработки Big Data</i>	Инструменты обработки Больших Данных из группы ПО: MapReduce, SAP HANA	Работа с литературой, источниками в сети Internet. Подготовка к тесту. Выполнение теста.	Отчёт по лабораторному практикуму.
<i>Тема 3. Процесс анализа больших данных. Технологии анализа больших данных. Научные проблемы в области больших данных</i>	Методы и техники анализа, применимые к большим данным: смешение и интеграция данных; искусственные нейронные сети, сетевой анализ	Работа с литературой, источниками в сети Internet. Подготовка к кейсу. Выполнение кейса.	Отчёт по лабораторному практикуму.

Наименование темы	Вопросы, вынесенные на самостоятельное изучение	Формы самостоят. работы	Форма текущего контроля
<i>Тема 4. Статистические методы анализа данных</i>	Методы анализа данных: непараметрические и номинальные методы	Работа с литературой, источниками в сети Internet. Подготовка к практикуму. Подготовка отчета по практикуму. Подготовка к домашнему заданию. Выполнение домашнего задания.	Отчёт по лабораторному практикуму. Выполненное домашнее задание.
<i>Тема 5. Современные программные средства анализа больших объемов информации</i>	Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий: дискриминантный и факторный анализы.	Работа с литературой, источниками в сети Internet. Подготовка к практикуму. Подготовка отчета по практикуму. Подготовка к тесту. Подготовка к домашнему заданию. Выполнение домашнего задания.	Отчёт по лабораторному практикуму. Выполненное домашнее задание.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Перечень основной и дополнительной литературы

Основная литература:

1. Информационные системы и технологии управления : учебник / ред. Г. А. Титоренко. – 3-е изд., перераб. и доп. – Москва : Юнити, 2015. – 591 с. : ил., табл., схемы – (Золотой фонд российских учебников). – Режим доступа: по подписке. – URL: <https://biblioclub.ru/>

2. Ясенев, В. Н. Информационные системы и технологии в экономике : учебное пособие / В. Н. Ясенев. – 3-е изд., перераб. и доп. – Москва : Юнити, 2015. – 560 с. : табл., граф., ил., схемы – Режим доступа: по подписке. – URL: <https://biblioclub.ru/>

Зехин, В. А. Практикум по многомерным статистическим методам : учебное пособие / В. А. Зехин, В. С. Мхитарян, С. А. Айвазян. – 1-е изд. – Москва : Московский государственный университет экономики, статистики и информатики, 2003. – 76 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/>

Дополнительная литература:

1. Гринберг, А. С. Информационные технологии управления : учебное пособие / А. С. Гринберг, А. С. Бондаренко, Н. Н. Горбачёв. – Москва : Юнити, 2015. – 479 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/>

2. Белов, В. С. Информационно-аналитические системы: основы проектирования и применения : учебное пособие : [16+] / В. С. Белов. – 2-е изд., перераб. и доп. – Москва : Евразийский открытый институт, 2010. – 111 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/>

3. Мхитарян, В. С. Эконометрика: учебно-практическое пособие : учебное пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин. – Москва : Евразийский открытый институт, 2012. – 221 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/>

6.2. Перечень ресурсов информационно-коммуникационной сети «Интернет»

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Центральный банк Российской Федерации	http://www.cbr.ru/
2.	Министерство финансов РФ	https://www.minfin.ru/ru/
3	Центральный экономико-математический институт (ЦЭМИ) РАН	https://www.cemi.rssi.ru/
4	Большие данные	https://habrahabr.ru/hub/bigdata/
5	Центр управления финансами - Методы прогнозирования	http://center-yf.ru/data/Marketologu/Metody-prognozirovaniya.php
6	11 текстов, которые помогут разобраться в больших данных -	https://blog.relap.ru/2016/09/big-data-articles/

6.3. Описание материально-технической базы

<p><i>Специализированная многофункциональная учебная аудитория № 2 для проведения учебных занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной/итоговой аттестации (с возможностью обучения лиц с ОВЗ), в том числе для организации практической подготовки обучающихся, с перечнем основного оборудования:</i></p> <p>Письменные столы обучающихся;</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 2 (25,2 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года (Ссылка на файл договора)</i></p>
--	--	---------------	--	--

<p>Стулья обучающихся; Столы для обучающихся с ограниченными возможностями здоровья; Стулья для обучающихся с ограниченными возможностями здоровья; Письменный стол педагогического работника; Стул педагогического работника; Кафедра; Магнитно-маркерная доска; Мультимедийный проектор; Экран; Ноутбук с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>				
<p><i>Специализированная многофункциональная учебная аудитория № 4 для проведения учебных занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной/ итоговой аттестации (с возможностью обучения лиц с ОВЗ), в том числе для организации практической подготовки обучающихся, с перечнем основного оборудования:</i> Письменные столы обучающихся; Стулья обучающихся; Столы для обучающихся с ограниченными возможностями здоровья; Стулья для обучающихся с ограниченными возможностями здоровья; Письменный стол педагогического работника; Стул педагогического работника; Кафедра; Магнитно-маркерная доска; Мультимедийный проектор; Экран; Ноутбук с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 4 (21,6 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года (Ссылка на файл договора)</i></p>
<p><i>Специализированная многофункциональная учебная аудитория № 5 для проведения учебных занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и</i></p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж,</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i></p>

<p><i>промежуточной/ итоговой аттестации (с возможностью обучения лиц с ОВЗ), в том числе для организации практической подготовки обучающихся, с перечнем основного оборудования:</i></p> <p>Письменные столы обучающихся; Стулья обучающихся; Столы для обучающихся с ограниченными возможностями здоровья; Стулья для обучающихся с ограниченными возможностями здоровья; Письменный стол педагогического работника; Стул педагогического работника; Кафедра; Магнитно-маркерная доска; Мультимедийный проектор; Экран; Ноутбук с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>	<p>помещение № 5 (19,6 кв.м)</p>			<p>(Ссылка на файл договора)</p>
<p><i>Специализированная многофункциональная учебная аудитория № 6 для проведения учебных занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной/ итоговой аттестации (с возможностью обучения лиц с ОВЗ), в том числе для организации практической подготовки обучающихся, с перечнем основного оборудования:</i></p> <p>Компьютерные столы обучающихся; Стулья обучающихся; Компьютерные столы для обучающихся с ограниченными возможностями здоровья; Стулья для обучающихся с ограниченными возможностями здоровья; Письменный стол педагогического работника; Стул педагогического работника; Стеллаж для учебно-методических материалов, в том числе учебно-наглядных пособий; Многофункциональное устройство (принтер, сканер, ксерокс); Интерактивная доска;</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 6 (18 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i></p> <p>(Ссылка на файл договора)</p>

<p>Мультимедийный проектор; Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>				
<p><i>Специализированная многофункциональная учебная аудитория № 7 для проведения учебных занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной/ итоговой аттестации (с возможностью обучения лиц с ОВЗ), в том числе для организации практической подготовки обучающихся, с перечнем основного оборудования:</i></p> <p>Компьютерные столы обучающихся; Стулья обучающихся; Компьютерные столы для обучающихся с ограниченными возможностями здоровья; Стулья для обучающихся с ограниченными возможностями здоровья; Письменный стол педагогического работника; Стул педагогического работника; Стеллаж для учебно-методических материалов, в том числе учебно-наглядных пособий; Многофункциональное устройство (принтер, сканер, ксерокс); Интерактивная доска; Мультимедийный проектор; Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 7 (18,5 кв..м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i> (Ссылка на файл договора)</p>
<p><i>Специализированная многофункциональная учебная аудитория № 13 для проведения учебных занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной/ итоговой аттестации (с возможностью обучения лиц с ОВЗ), в том числе для организации практической подготовки обучающихся, с перечнем основного оборудования:</i></p> <p>Компьютерные столы обучающихся; Стулья обучающихся;</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 13 (19,7 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i> (Ссылка на файл договора)</p>

<p>Компьютерные столы для обучающихся с ограниченными возможностями здоровья;</p> <p>Стулья для обучающихся с ограниченными возможностями здоровья;</p> <p>Письменный стол педагогического работника;</p> <p>Стул педагогического работника;</p> <p>Стеллаж для учебно-методических материалов, в том числе учебно-наглядных пособий;</p> <p>Многофункциональное устройство (принтер, сканер, ксерокс);</p> <p>Интерактивная доска;</p> <p>Мультимедийный проектор;</p> <p>Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>				
<p>Помещение № 1 для самостоятельной работы обучающихся (с возможностью обучения лиц с ОВЗ) с перечнем основного оборудования:</p> <p>Письменный стол обучающегося;</p> <p>Стул обучающегося;</p> <p>Письменный стол обучающегося с ограниченными возможностями здоровья;</p> <p>Стул обучающегося с ограниченными возможностями здоровья;</p> <p>Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата;</p> <p>Моноблок (в том числе, клавиатуры, мыши, наушники) с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 1 (12,2 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i> (Ссылка на файл договора)</p>
<p>Помещение № 3 для самостоятельной работы обучающихся (с возможностью обучения лиц с ОВЗ) с перечнем основного оборудования:</p> <p>Письменный стол обучающегося;</p> <p>Стул обучающегося;</p> <p>Письменный стол обучающегося с ограниченными возможностями здоровья;</p> <p>Стул обучающегося с ограниченными возможностями</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 3 (16,2 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i> (Ссылка на файл договора)</p>

<p>здоровья; Стеллаж для учебно-методических материалов; Многофункциональное устройство (принтер, сканер, ксерокс); Моноблоки (в том числе, клавиатуры, мыши, наушники) с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>				
<p>Помещение № 12 для самостоятельной работы обучающихся (с возможностью обучения лиц с ОВЗ) с перечнем основного оборудования: Письменные столы; Стулья; Письменный стол обучающегося с ограниченными возможностями здоровья; Стул обучающегося с ограниченными возможностями здоровья; Стеллаж для учебно-методических материалов; Многофункциональное устройство (принтер, сканер, ксерокс); Ноутбуки с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата; Моноблок (в том числе, клавиатуры, мыши, наушники) с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде лицензиата</p>	<p>160019, Вологодская область, г.о. город Вологда, г Вологда, ул Добролюбова, д. 68а. БТИ: 1 этаж, помещение № 12 (18,1 кв.м.)</p>	<p>Аренда</p>	<p>Общество с ограниченной ответственностью «Балтэстейт»</p>	<p><i>Договор аренды нежилого помещения от 1 августа 2024 года, срок действия с 1 августа 2024 года по 30 июня 2025 года</i> (Ссылка на файл договора)</p>

6.4. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, в том числе комплект лицензионного программного обеспечения, электронно-библиотечные системы, современные профессиональные базы данных и информационные справочные системы

Обучающиеся обеспечены доступом к электронной информационно-образовательной среде института из любой точки, в которой имеется доступ к сети «Интернет», как на территории организации, так и вне ее.

лицензионное программное обеспечение:

- Программное обеспечение Microsoft Office
- Программное обеспечение Microsoft Office
- Комплексная Система Антивирусной Защиты Kaspersky Total Security

для бизнеса Russian Edition

лицензионное программное обеспечение отечественного производства:

- Антивирусная программа Dr. Web;

свободно-распространяемое программное обеспечение:

- 7-ZIP – архиватор <https://7-zip.org.ua/ru/>
- Inkscape – векторный графический редактор <https://inkscape.org/ru/o-programme/>

- Gimp – растровый графический редактор <http://www.progimp.ru/>

электронно-библиотечная система:

- Электронная библиотечная система (ЭБС) «Университетская библиотека ONLINE» <http://biblioclub.ru/>.

- Образовательная платформа «Юрайт». Для вузов и ссузов. Электронная библиотечная система (ЭБС) <https://urait.ru/>

современные профессиональные баз данных:

- Официальный интернет-портал базы данных правовой информации <http://pravo.gov.ru>.

- Портал Единое окно доступа к образовательным ресурсам <http://window.edu.ru/>

информационные справочные системы:

- Портал Федеральных государственных образовательных стандартов высшего образования <http://fgosvo.ru>.

- Компьютерная справочная правовая система «КонсультантПлюс» (<http://www.consultant.ru/>).

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание оценочных средств для проведения текущего

контроля успеваемости в процессе освоения дисциплины

№ п/п	Наименование оценочного средства	Шкала и критерии оценки, балл
1.	Лабораторный практикум	<p>10-8 – работа выполнена в срок, самостоятельно, правильно поняты и использованы соответствующие формулы, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, правильно выбраны совместимые комплектующие, сделаны необходимые выводы, хорошо аргументированы, даны исчерпывающие ответы на все поставленные вопросы;</p> <p>7-5 – работа выполнена в срок, самостоятельно, правильно поняты и использованы соответствующие формулы, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, правильно выбраны совместимые комплектующие, необходимые выводы сделаны частично, хорошо аргументированы, даны ответы на все поставленные вопросы;</p> <p>4-1 – работа выполнена в срок, в основном самостоятельно, использованы соответствующие формулы; определены соответствующие спецификации, имеются ошибки в расчетах; выбраны совместимые комплектующие необходимые, выводы сделаны частично, слабо аргументированы, даны ответы не на все вопросы;</p> <p>0 – обучающийся подготовил работу несамостоятельно или не завершил в срок, описание спецификации содержит незначительные ошибки, выводы и ответы на вопросы отсутствуют.</p>
3.	Домашнее задание	<p>5 – результат предоставлен в установленные сроки, задание выполнено полностью, самостоятельно</p> <p>4 – результат предоставлен с нарушением сроков, задание выполнено полностью, самостоятельно</p> <p>3 – результат предоставлен с нарушением сроков, выполнено только 50%-90% задания, задание выполнено самостоятельно</p> <p>2-1 – результат предоставлен с нарушением сроков, задание выполнено менее чем на 50%, результаты полностью или частично заимствованы.</p> <p>0 – домашнее задание не выполнено.</p>

Типовые контрольные задания или иные материалы в рамках текущего контроля успеваемости

Типовые задания к лабораторному практикуму.

Задание №1.

Поиск источников информации в сети Интернет: открытые и закрытые источники данных. Портал открытых данных РФ. Сохранение данных в программе MS Excel. Преобразование и первичная обработка данных.

Задание №2.

Возможности графического представления информации в программе MS Excel: графические функции отображения одномерных и многомерных

данных, графический вывод с использованием графических параметров.

Задание №3.

Представление исходных данных в программах MS Excel (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах MS Excel: подсчет описательных статистик, графическое представление данных. Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа.

Задание №4.

Представление исходных данных в программах MS Excel (массивы, таблицы). Применение различных методов выделения, извлечения и группировки данных в программах MS Excel, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (регрессионный, дисперсионный, кластерный, дискриминантный, факторный анализы).

Задание №5.

Ознакомьтесь с нижеприведенной информацией. Как государственные органы власти в России могут использовать базовый инструментарий технологии Big Data в интересах граждан?

Приведите примеры. Как Вы считаете, стоит ли тратить бюджетные средства на разработки систем Big Data?

«Анализируя большие данные интернет-запросов, исследователи обнаружили странный феномен. Уже несколько лет всплеск поисковых запросов Google по таким терминам, как лечение гриппа, симптомы гриппа и т.п. на несколько недель предвещает начало стремительного нарастания эпидемии гриппа. Эта закономерность уже сегодня используется для проведения превентивных мер по предотвращению во многих штатах эпидемии гриппа, подготовке врачей, освобождению лечебных коек и т.п. Следует отметить, что используемая до этого информация, поступающая от участковых врачей и пунктов неотложной помощи, как правило, отставала от реальной картины. Специалисты Федеральной резервной системы выяснили, что статистика поисковых запросов Google относительно покупки домов является более надежным источником для определения тенденций в увеличении или уменьшении объемов продаж недвижимости и динамики жилищного строительства, чем прогнозы наиболее известных экономистов. По мнению участников Всемирного экономического форума 2012 года в Давосе, те, кто оседлает тему 14 интеллектуального анализа больших данных, станут хозяевами информационного пространства. Этой теме был посвящен специальный доклад на Форуме «Большие данные – большое влияние». Ключевой вывод доклада – цифровые активы становятся не менее значимым экономическим активом, чем золото или валюта. Исследования, проведенные профессором Бринйолфсоном (E.Brunjolfsson) и двумя его коллегами, показали, что анализ и прогнозирование на основе больших данных берется на вооружение корпоративной Америкой. Они изучили 179

крупных компаний и обнаружили, что те из них, кто взял в последние год-полтора на вооружение интеллектуальный анализ больших данных получил немедленное улучшение экономических показателей на 5-6%. С учетом оборота и размеров этих компаний это очень и очень много и показывает сумасшедшую рентабельность вложений в интеллектуальный анализ больших данных. И если сегодня, пока, как показывают исследования агентства Wikibon research, компании не получают должной отдачи от инвестиций в технологии Big Data и от каждого вложенного доллара пока возвращается половина, то в следующие три – пять лет ситуация кардинально изменится, и ROI составит не менее 3,5 долларов на один доллар инвестиций».

Типовые задания к домашним заданиям.

Домашнее задание № 1.

«Статистические методы анализа данных»

Задание №1.

При исследовании взаимосвязи цен на следующие виды продовольственных товаров: говядина (x_1), растительное масло (x_2), сахар-песок (x_3) и хлеб белый в/с (x_4) в $n=22$ городах Центрального района России получена матрица парных коэффициентов корреляции:

$$R = \begin{pmatrix} 1,00 & -0,18 & 0,01 & 0,61 \\ -0,18 & 1,00 & 0,82 & 0,24 \\ 0,01 & 0,82 & 1,00 & -0,05 \\ 0,61 & 0,24 & -0,05 & 1,00 \end{pmatrix}$$

Для трехмерной совокупности x_1, x_2, x_4 требуется:

- построить матрицу парных коэффициентов корреляции;
- при $\alpha=0,1$ проверить значимость частного коэффициента корреляции и найти его интервальную оценку при $\gamma=0,954$. Сравнить полученные результаты. Как влияет x_4 на тесноту связи между x_1 и x_2 ?
- при $\alpha=0,05$ проверить значимость множественного коэффициента корреляции $R_{4(1,2)}$.

Задание №2.

По данным задания 1 для трехмерной совокупности x_2, x_3, x_4 требуется:

- построить матрицу парных коэффициентов корреляции R ;
- при $\alpha=0,1$ проверить значимость частного коэффициента корреляции $r_{2,3(4)}$ и найти его интервальную оценку при $\gamma=0,9$. Сравнить полученные результаты. Как влияет x_4 на тесноту связи между x_3 и x_2 ?
- при $\alpha=0,05$ проверить значимость множественного коэффициента корреляции $R_{2(3,4)}$. Дайте интерпретацию $R^{\wedge}_{2(3,4)}$.

Задание №3.

На основании данных о динамике темпов прироста курса акций за 5 месяцев, приведенных в таблице и предположения, что генеральное уравнение регрессии имеет вид $\hat{y}_x = a_0 + a_1x$, требуется:

Месяц (x)	0	1	2	3	4
y (%)	10	8	5	3	4

- а) определить оценки параметров уравнения регрессии a_0 и a_1 и остаточной дисперсии S^2 ;
- б) проверить при $\alpha=0,01$ значимость коэффициента регрессии, т.е. гипотезы $H_0=a_1=0$;
- в) с надежностью $\gamma=0,95$ найти интервальные оценки параметров a_0 и a_1 ;
- г) с надежностью $\gamma=0,9$ установить интервальную оценку условного математического ожидания y при $x_0=4$;
- д) определить при $\gamma=0,9$ доверительный интервал предсказания \tilde{y}_{n+1} в точке $x=5$.

Домашнее задание № 2.

«Современные программные средства анализа больших объемов информации»

Задание №1.

Условия жизни населения $n=10$ стран характеризуется тремя показателями: оценка ВВП по паритету покупательской способности в 2014г. на душу населения (в % к США) (x_1); расходы на здравоохранение (в % от ВВП) (x_2); численность врачей на 10.000 тыс. населения (x_3), значения которых приводятся в таблице:

№ п/п	Страна	x ₁	x ₂	x ₃
1	Россия	20,4	3,2	44,5
2	Австралия	71,4	8,5	32,5
3	Австрия	78,7	9,2	33,9
4	Азербайджан	12,1	3,3	38,8
5	Армения	10,9	3,2	34,4
6	Белоруссия	20,4	5,4	43,6
7	Бельгия	79,7	8,9	41,0
8	Болгария	17,3	5,4	36,4
9	Великобритания	69,7	7,1	17,9
10	Венгрия	24,5	6,0	32,1

Требуется по показателям x₁ и x₂ определить:

- собственные значения главных компонент и их вклад в суммарную дисперсию исходных показателей;
- определить ортогональную матрицу собственных векторов корреляционной матрицы R;
- дать экономическую интерпретацию главных компонент.

Задание №2.

При расчете матрицы факторных нагрузок A по матрицы нормированных значений X* исходных признаков, оказались пропущенными диагональные элементы:

$$A = \begin{pmatrix} a_{11} & 0,68 & -0,14 \\ 0,88 & a_{22} & 0,18 \\ 0,95 & 0,14 & a_{33} \end{pmatrix}$$

Требуется:

- восстановить пропущенные значения a₁₁, a₂₂, a₃₃, с точностью до знака;
- определить дисперсии главных компонент и относительных вклад двух первых главных компонент в суммарную дисперсию;
- рассчитать матрицу парных коэффициентов корреляции R[^] исходных показателей.

Задание №3.

По иерархическому агломеративному алгоритму провести классификацию n=6 предприятий машиностроения, данные о деятельности которых характеризуются показателями: x₁ – рентабельность (%), x₂ – производительность труда (млн.руб./чел.) и представлены в таблице:

Номер предприятия	1	2	3	4	5	6
x ₁	23,4	17,5	9,7	18,2	6,6	8
x ₂	9,1	5,2	5,5	9,4	7,5	5,7

В качестве расстояния между объектами принять:

- обычное евклидово расстояние;
- взвешенное евклидово расстояние с весами ω₁=0,1 и ω₂=0,9;
- сравнить разбиение на два кластера по критерию минимума суммы внутрикласовых дисперсий.

Расстояние между кластерами определить по принципу «ближнего соседа»

не нормализуя данные.

7.2. Описание оценочных средств для проведения промежуточной аттестации

Промежуточная аттестация по дисциплине проводится в форме зачета.

Процедура оценивания	Шкала и критерии оценки, балл
<p>Зачет представляет собой выполнение обучающимся заданий билета, включающего в себя.</p> <p>Задание №1 – теоретический вопрос на знание базовых понятий предметной области дисциплины, а также позволяющий оценить степень владения обучающегося принципами предметной области дисциплины, понимание их особенностей и взаимосвязи между ними;</p> <p>Задание №2 – задание на анализ ситуации из предметной области дисциплины и выявление способности обучающегося выбирать и применять соответствующие принципы и методы решения практических проблем, близких к профессиональной деятельности;</p> <p>Задания №3 – задания на проверку умений и навыков, полученных в результате освоения дисциплины</p>	<p>Выполнение обучающимся заданий билета оценивается по следующей балльной шкале:</p> <p>Задание 1: 0-10 баллов Задание 2: 0-10 баллов Задание 3: 0-20 баллов</p> <p>Суммарное количество баллов ТКУ и ПА «Зачтено»</p> <p>– 90-100 – ответ правильный, логически выстроен, использована профессиональная терминология. Задания решены правильно. Обучающийся правильно интерпретирует полученный результат.</p> <p>– 70 -89 – ответ в целом правильный, логически выстроен, использована профессиональная терминология. Ход решения заданий правильный, ответ неверный. Обучающийся в целом правильно интерпретирует полученный результат.</p> <p>– 50 - 69 – ответ в основном правильный, логически выстроен, использована профессиональная терминология. Задание решено частично.</p> <p>«Не зачтено»</p> <p>– менее 50 – ответы на теоретическую часть неправильные или неполные. Задания не решены.</p>

Типовые задания для проведения промежуточной аттестации обучающихся

Задания 1 типа.

1. Определите сущность понятия «большие данные».
1. Опишите методики анализа больших данных.
2. Процесс аналитики анализа больших данных.
3. Дайте характеристику Big Data на мировом рынке.
4. Охарактеризуйте Big Data в России.
5. Определите понятие Data Mining.
6. Вопросы безопасности больших данных.
7. Информационное общество и большие данные.
8. Современные способы хранения и обработки больших данных.
9. Современные способы анализа массивов данных.
10. Современные концепции больших данных.
11. В чем состоит когнитивный анализ данных.
12. Какие модели данных вы знаете?
13. Основные описательные статистики.
14. Определите различия между параметрическими,

непараметрическими и номинальными методами.

15. Опишите основную идею корреляционного анализа.
16. Опишите основную идею Регрессионный анализ.
17. Основная идея дисперсионного анализа.
18. Сущность кластерного анализа.
19. Дискриминантный анализ: модель и общая процедура выполнения.
20. Цели факторного анализа.
21. Программные средства анализа данных: Stata, SPSS, Excel; их преимущества и недостатки.
22. Преимущества работа с данными в программах Stata и SPSS.
23. Представление исходных данных в программах Stata и SPSS.
24. Выполнение анализа данных в программах Stata и SPSS.

Задания 2 типа.

1. В каком виде к 2020г. содержится большинство данных в мире? В каком веке произошёл перевес объёмов накопленных человечеством данных в сторону цифровых? В чем измеряется объём накопленных человечеством цифровых данных на 2020 год? Сколько Петабайт в Зеттабайте?

1. Какой из факторов, по вашему мнению, способствовал появлению тренда больших данных: маркетинговые кампании крупных корпораций; снижение издержек на хранение данных; появление новых технологий обработки потоковых данных или выпуск баз данных с обработкой данных в памяти? Ответ аргументируйте.

2. Какие на сегодняшний день можно выделить вероятные разочарования тренда больших данных? Ответ аргументируйте.

3. Какие из данных событий можно отнести к наиболее значимым, повлиявшие на формирование тренда больших данных: разработка Hadoop; изобретение принципа MapReduce; разработка языка Python или победа Deepblue в матче с Г.Каспаровым? Ответ аргументируйте.

4. Приведены четыре утверждения:

а) Большие данные это обработка или хранение более 1 Тб информации.
б) Проблема больших данных это такая проблема, когда при существующих технологиях хранения и обработки сущностная обработка данных затруднена или невозможна.

в) Большие данные это огромная PR - акция крупных вендоров и не более того.

г) Большие данные это явление, когда цифровые данные наиболее полно представляют изучаемый объект.

С какими из данных утверждений вы можете согласиться, а с какими нет? Ответ объясните.

5. Приведены четыре утверждения:

а) Большие данные это данные объёма свыше 1 Тб.
б) Проблема больших данных это проблема, когда при существующих технологиях хранения и обработки сущностная обработка данных затруднена или невозможна.

в) Большие данные - это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров.

г) Большие данные как правило не структурированы.

С какими из данных утверждений вы можете согласиться, а с какими нет? Ответ объясните.

6. В каких из приведенных ниже случаев речь идет структурированных данных: данные о продажах компании, представленные в виде помесечных отчетов в формате MS Word; таблица с ежедневными показаниями температуры помещения за год в файле формата CSV; текст педагогической поэмы А.С. Макаренки, представленный в формате PDF; библиотека фильмов, представленных в формате mp4 на одном жестком диске? Ответ обоснуйте.

7. Сфера Big Data характеризуется пятью основными признаками. Именно при соблюдении этих пяти условий объем данных можно отнести к числу больших. Соотнесите соответствующий признак с его описанием:

1. Volume	А. Данный признак указывает как на увеличивающуюся скорость накопления данных (90% информации было собрано за последние 2 года), так и на скорость обработки данных, в последнее время стали более востребованы технологии обработки данных в реальном времени.
2. Velocity	Б. Все большее значение пользователи стали придавать значимости достоверности имеющихся данных. Так, у интернет-компаний есть проблема по разделению действий, проводимых роботом и человеком на сайте компании, что приводит в конечном счете к затруднению анализа данных.
3. Variety	В. Накопленная база данных представляет собой большой объем информации, который трудоемко обрабатывать и хранить традиционными способами, для них требуются новый подход и усовершенствованные инструменты.
4. Veracity	Г. Возможность одновременной обработки структурированной и неструктурированной разноформатной информации. Главное отличие структурированной информации – это то, что она может быть классифицирована. Примером такой информации может служить информация о клиентских транзакциях. Неструктурированная информация включает в себя видео, аудио файлы, свободный текст, информацию, поступающую из социальных сетей. На сегодняшний день 80% информации входит в группу неструктурированной. Данная информация нуждается в комплексном анализе, чтобы сделать ее полезной для дальнейшей обработки.
5 Value	Д. Большие Данные должны быть полезны компании и приносить определенную ценность для нее. К примеру, помогать в усовершенствовании бизнес-процессов, составлении отчетности или оптимизации расходов.

8. На основании выжимки из аналитического обзора рынка Big Data оцените его возможное развитие в 2021-2022гг. Можно ли утверждать, что несмотря на относительно молодой ее возраст, Большие данные это устоявшаяся сфера технологий, получившая распространение во многих сферах бизнеса и играющая немаловажную роль в развитии компаний? Ответ

аргументируйте.

«Большинство компаний используют Большие Данные в сфере клиентского сервиса (53%), второе по популярности направление – операционная эффективность (40%), в сфере управления рисками Большие Данные менее распространены на текущий момент (7%). Следует также отметить, что Big Data являются одной из самых быстрорастущих сфер информационных технологий, согласно статистике, общий объем получаемых и хранимых данных удваивается каждые 1,2 года. За период с 2012 по 2014 год количество данных, ежемесячно передаваемых мобильными сетями, выросло на 81%. По оценкам Cisco, в 2014 году объем мобильного трафика составил 2,5 эксабайта (единица измерения количества информации, равная 10^{18} стандартным байтам) в месяц, а уже в 2019 году он будет равен 24,3 эксабайтам».

9. Соотнесите соответствующий термин с соответствующим подходом обработки данных (ПО):

1. SQL	А. Модель распределения вычислений. Используется для параллельных вычислений над очень большими наборами данных (петабайты и более). В программном интерфейсе не данные передаются на обработку программе, а программа – данным. Таким образом, запрос представляет собой отдельную программу.
2. NoSQL	Б. Высокопроизводительная NewSQL платформа для хранения и обработки данных. Обеспечивает высокую скорость обработки запросов. Еще одним отличительным признаком является то, что она упрощает системный ландшафт, уменьшая затраты на поддержку аналитических систем.
3. MapReduce	В. Включает в себя ряд подходов, направленных на реализацию базы данных, имеющих отличия от моделей, используемых в традиционных, реляционных СУБД. Их удобно использовать при постоянно меняющейся структуре данных. Например, для сбора и хранения информации в социальных сетях.
4. Hadoop	Г. Язык структурированных запросов, позволяющий работать с базами данных. С помощью него можно создавать и модифицировать данные, а управлением массива данных занимается соответствующая система управления базами данных.
5. SAP HANA	Д. Используется для реализации поисковых и контекстных механизмов высоконагруженных сайтов – Facebook, eBay, Amazon и др. Отличительной особенностью является то, что система защищена от выхода из строя любого из узлов кластера, так как каждый блок имеет, как минимум, одну копию данных на другом узле.

10. Большие Данные получили широкое распространение во многих отраслях бизнеса. Приведите несколько примеров использования Больших Данных в таких отраслях как розничная торговля, финансы и связь.

11. Главной целью анализа данных является удержание существующих клиентов и привлечение новых. Какие мероприятия для этого реализуют телекоммуникационные компании? Как еще могут быть ими использованы технологии Big Data?

12. Большие Данные получили широкое распространение во многих

отраслях бизнеса. Приведите несколько примеров использования Больших Данных в горнодобывающей и нефтяной промышленности.

13. В 2015 году мировой рынок продуктов и услуг для работы с Big Data составлял \$33,3 млрд. Такая цифра приводится в мартовском исследовании американского агентства Wikibon. По их прогнозу, к 2020 году объем индустрии больших данных вырастет до \$61 млрд. На сколько процентов ежегодно рынок должен прибавлять для получения такой оценки? Каков будет прогноз мирового рынка Big Data в 2026 году если тенденция сохранится?

14. В мировом масштабе российский рынок услуг и технологий Big Data исчезающе мал. В 2018 году американская компания IDC оценивала его в \$340 млн. По некоторым данным, по итогам 2019 года он увеличится до \$500 млн. Каков уровень среднего абсолютного прироста и среднего темпа роста в этом случае? Если тенденция не изменится, какой объем российского рынка услуг и технологий Big Data можно ожидать в 2022г.?

15. Распределите нижеприведенные результаты внедрения Big Data по степени эффективности (наиболее эффективные – 1 место, наименее – 7). Ответ объясните.

1. Улучшение клиентского сервиса	
2. Улучшение реагирования на запросы клиентов	
3. Рост эффективности обработки клиентских запросов	
4. Улучшение интеграции в цепи поставок	
5. Оптимизация запасов и продуктивности основных активов	
6. Улучшение процессов планирования компании	
7. Оптимизация затрат на сервис	

16. Установите соответствие между алгоритмами объединения двух кластеров и их характеристиками. Ответ объясните.

1. Метод дальнего соседа	а. Степень близости оценивается как средняя величина степеней близости между объектами кластеров.
2. Метод средней связи b.	б. Расстояние между любым кластером S и новым кластером, который получился в результате объединения кластеров P и Q, определяется как расстояние от центра кластера S до середины отрезка, соединяющего центры кластеров P и Q.
3. Метод медианной связи	с. Степень близости оценивается по степени близости между наиболее отдаленными объектами кластеров.

17. Соотнесите задачи инфраструктуры Больших Данных с соответствующим этапом аналитики Big Data. Ответ объясните.

1. Сбор данных	А. Построение специализированных витрин данных.
	Б. Обеспечение защиты и ограничения доступа к данным.
2. Обработка данных	В. Выявление закономерностей.
	Г. Разработка решения по доставке (загрузке) данных в инфраструктуру Больших данных
3. Контроль качества данных	Д. Преобразование данных в удобную для анализа форму.
4. Аналитика и машинное обучение	Е. Обеспечить контроль и неизменность качества данных от Источника.
	Ж. Формирование отчетов и предоставление рекомендаций.

18. Соотнесите проблемы инфраструктуры Больших Данных с соответствующим этапом аналитики Big Data. Ответ объясните.

1. Сбор данных	А. Сложные процедуры согласования доступа к источнику.
	Б. Отсутствие единого универсального инструмента для контроля качества данных на всех этапах преобразования.
2. Обработка данных	В. Недостаточное качество данных по сравнению с Источником.
	Г. Обеспечение работоспособности ETL инструментов Больших Данных в текущей инфраструктуре.
3. Контроль качества данных	Д. Необходимость исключения ряда атрибутов загружаемых таблиц в зависимости от технологии загрузки.
4. Аналитика и машинное обучение	Е. Обеспечение неразрывной историчности данных в инфраструктуре (без дыр).
	Ж. Значительные или недостаточные объемы данных, скорость их поступления.

19. У предприятия, осуществляющего свою деятельность на рынке массовых розничных услуг, стоит задача формирования и анализа клиентской базы с целью определения кластеров покупателей и подготовки отдельных предложений для каждого кластера. По каким критериям возможно формирование кластеров? Какие данные можно использовать в этих целях? Предложите свой вариант кластерной сегментации на любом розничном рынке товаров или услуг.

20. На рынке кредитования физических лиц в России осуществляют свою деятельность банки и микрофинансовые организации. Существует ли разница в скоринговых моделях этих организаций? По каким ключевым критериям формируются данные скоринговые модели? Предложите свой вариант для основы скоринговой модели для микрофинансовой организации.

21. Результаты опроса Tech Pro Research по внедрению Больших Данных в различные отрасли представлены в таблице:

Отрасли	Доля компаний, в том числе		
	Внедривших БД	Не внедривших БД	Затруднившихся ответить
1. Телекоммуникационные предприятия	58	21	21
2. Инжиниринг и конструкторское бюро	45	55	-
3. Государственные предприятия	38	62	-
4. IT	36	59	5
5. Финансы и страхование	33	50	17
6. Логистика и транспорт	33	67	-
7. FMCG	21	57	21
8. Бизнес-сервис/Консалтинг	18	64	18
9. Образование	15	67	18
10. Здравоохранение	13	75	13

А) Постройте секторную или столбиковую диаграмму структуры с накоплением по полученным данным.

Б) Сделайте выводы о том, в каких отраслях Большие Данные получили

наибольшее распространение, а в каких наименьшее.

22. Перед вами стоит задача исследовать имидж компании. Клиенту предлагается оценить данную компанию по целому ряду критериев, общее число которых может превышать несколько десятков. Применение какого из методов многомерного анализа данных в данном случае позволяет снизить общее количество переменных путем распределения их в обобщенные пучки факторов, например, «материальные условия компании», «взаимодействие с персоналом», «удобство обслуживания»? Ответ объясните.

23. Перед вами стоит задача составления социально-психологических портретов потребителей. Респонденту необходимо выразить степень своего согласия/несогласия с перечнем высказываний о стиле жизни. Применение какого из методов многомерного анализа данных в данном случае позволит выделить, например, целевые группы потребителей: «новаторы», «прогрессисты» и «консерваторы»? Ответ обоснуйте.

24. Допустим, вы планируете провести опрос потребителей, (а все потребители разные), и вам, соответственно, необходимы различные стратегии для их привлечения. Для решения данной задачи мы предлагаем сегментировать клиентов, прибегнув к методу кластеризации. В результате применения кластерного анализа получены следующие данные:

	Уровень дохода, тыс. руб.			Возраст, лет			Ежемесячные затраты на покупку одежды		
	15-25	25-35	35-45	25-30	30-35	35-40	до 10	до 20	до 40
Кластер 1	72%	18%	10%	73%	20%	7%	75%	15%	10%
Кластер 2	20%	64%	16%	17%	67%	16%	15%	73%	12%
Кластер 3	9%	14%	77%	9%	11%	80%	8%	10%	82%

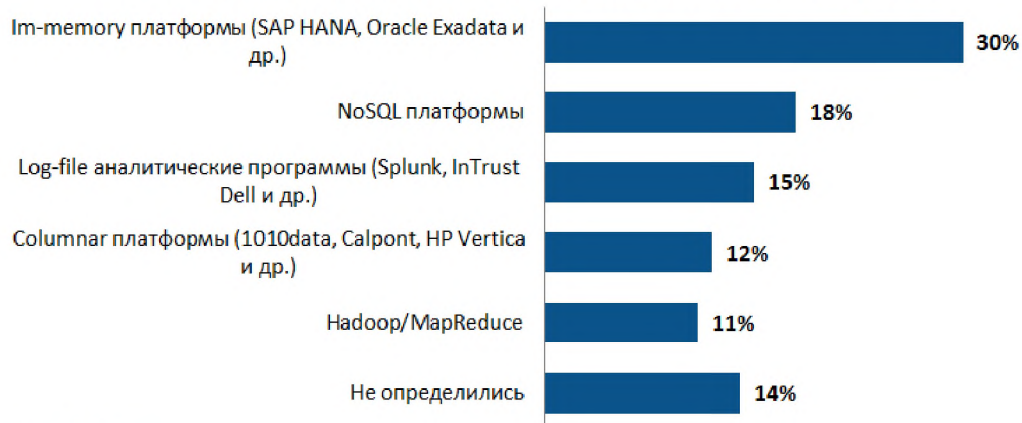
Сделайте выводы по результатам кластеризации. Опишите портрет клиента каждого из кластеров, которые впоследствии необходимо учитывать при составлении стратегии успешного продвижения продукта на рынке.

Задания 3 типа.

Задание №1.

Результаты опроса Tech Pro Research по видам технологий, наиболее востребованным при использовании Больших Данных представлены на рисунке.

Какие технологии востребованы при использовании Big Data



Источник: T-Systems

Рассчитайте показатели уровня концентрации использования технологий Больших Данных по видам. Расчет можно проводить по индексу Херфиндаля-Хиршмана или Розенблота. Проранжируйте технологии по степени популярности использования при работе с Большими Данными и их концентрацию. Сделайте выводы.

Задание №2.

По данным EMC, объем накопленных данных в Германии по итогам 2014 года составляет 230 эксабайт, что составляет 4% от общего объема информации в мире. К 2020 году объем информации вырастет до 1100 эксабайт и составит 2%. На рынке Германии большую долю выручки, по прогнозам Experton Group, будет генерировать сегмент сервисных услуг, доля которых в 2015 году составит 54%, а в 2019 году увеличится до 59%, доли программного обеспечения и оборудования, наоборот, уменьшатся.

Объем рынка Big Data Германии (млн евро)



Источник: Experton Group

Рассчитайте аналитические показатели динамики (абсолютные, относительные и средние темпы роста и прироста) прогноза объема рынка Больших Данных Германии. Если тенденция сохранится каков будет размер выручки рынка Больших Данных в Германии в 2020г. по видам. Сделайте выводы.

Задание №3.

По данным EMC, объем накопленных данных в Германии по итогам 2014 года составляет 230 эксабайт, что составляет 4% от общего объема информации в мире. К 2020 году объем информации вырастет до 1100 эксабайт и составит 2%. На рынке Германии большую долю выручки, по прогнозам Experton Group, будет генерировать сегмент сервисных услуг, доля которых в 2015 году составит 54%, а в 2019 году увеличится до 59%, доли программного обеспечения и оборудования, наоборот, уменьшатся.

Объем рынка Big Data Германии (млн евро)



Источник: Experton Group

А). Рассчитайте структуру объема рынка Больших данных в 2015-2019гг. Данные представьте в таблице и сделайте выводы.

Б). Проведите анализ структуры и структурных сдвигов объема рынка с помощью индивидуальных и обобщающих показателей структурных различий:

1) абсолютных и относительных индивидуальных показателей структуры;

2) линейного и квадратического коэффициента структурных сдвигов;

3) интегрального коэффициента структурных сдвигов К. Гатева или Салаи.

В). Сделайте выводы.

Задание №4.

В таблице представлены темпы прироста следующих макроэкономических показателей десяти различных стран мира: ВВП (x_1), промышленное производство (x_2), индекс цен (x_3) и доли безработных (x_4).

Страна	ВВП (x_1)	Промышленное производство (x_2)	Индекс цен (x_3)	Доля безработных (x_4)
Япония	3,5	4,3	2,1	2,3
США	3,1	4,6	3,9	6,3
Германия	2,2	2	3,4	5,1
Франция	2,7	3,1	2,9	9,7
Италия	2,7	3,0	5,6	11,1

Великобритания	1,6	1,4	4,0	9,5
Канада	3,1	3,4	3,0	10
Австралия	1,8	2,6	4,0	2,6
Бельгия	2,3	2,6	3,4	8,9
Нидерланды	2,3	2,4	3,5	6,4

А) Найдите оценку коэффициента корреляции между темпами прироста ВВП (x_1) и промышленного производства (x_2); при $\alpha=0,05$ проверить его значимость; при $\gamma=0,923$ найти его интервальную оценку.

Б) Определите долю дисперсии x_1 , обусловленную влиянием x_2 .

Задание №5.

В таблице представлены темпы прироста следующих макроэкономических показателей десяти различных стран мира: ВВП (x_1), промышленное производство (x_2), индекс цен (x_3) и доли безработных (x_4).

Страна	ВВП (x_1)	Промышленное производство (x_2)	Индекс цен (x_3)	Доля безработных (x_4)
Япония	3,5	4,3	2,1	2,3
США	3,1	4,6	3,9	6,3
Германия	2,2	2	3,4	5,1
Франция	2,7	3,1	2,9	9,7
Италия	2,7	3,0	5,6	11,1
Великобритания	1,6	1,4	4,0	9,5
Канада	3,1	3,4	3,0	10
Австралия	1,8	2,6	4,0	2,6
Бельгия	2,3	2,6	3,4	8,9
Нидерланды	2,3	2,4	3,5	6,4

А) Оценить тесноту связи между ВВП (x_1) и индексом цен (x_3); при $\alpha=0,05$ проверить его значимость; при $\gamma=0,857$ найти его интервальную оценку для r_{13} .

Б) Определите долю дисперсии x_1 , обусловленную влиянием x_3 .

Задание №6.

В таблице представлены темпы прироста следующих макроэкономических показателей десяти различных стран мира: ВВП (x_1), промышленное производство (x_2), индекс цен (x_3) и доли безработных (x_4).

Страна	ВВП (x_1)	Промышленное производство (x_2)	Индекс цен (x_3)	Доля безработных (x_4)
Япония	3,5	4,3	2,1	2,3
США	3,1	4,6	3,9	6,3
Германия	2,2	2	3,4	5,1
Франция	2,7	3,1	2,9	9,7
Италия	2,7	3,0	5,6	11,1
Великобритания	1,6	1,4	4,0	9,5
Канада	3,1	3,4	3,0	10
Австралия	1,8	2,6	4,0	2,6
Бельгия	2,3	2,6	3,4	8,9
Нидерланды	2,3	2,4	3,5	6,4

А) Оценить тесноту связи между промышленным производством (x_2) и долей безработных (x_4); при $\alpha=0,05$ проверить его значимость; при $\gamma=0,95$ найти его интервальную оценку для r_{24} .

Б) Определите долю дисперсии x_2 , обусловленную влиянием x_4 .

Задание №7.

В таблице представлены темпы прироста следующих макроэкономических показателей десяти различных стран мира: ВВП (x_1), промышленное производство (x_2), индекс цен (x_3) и доли безработных (x_4).

Страна	ВВП (x_1)	Промышленное производство (x_2)	Индекс цен (x_3)	Доля безработных (x_4)
Япония	3,5	4,3	2,1	2,3
США	3,1	4,6	3,9	6,3
Германия	2,2	2	3,4	5,1
Франция	2,7	3,1	2,9	9,7
Италия	2,7	3,0	5,6	11,1
Великобритания	1,6	1,4	4,0	9,5
Канада	3,1	3,4	3,0	10
Австралия	1,8	2,6	4,0	2,6
Бельгия	2,3	2,6	3,4	8,9
Нидерланды	2,3	2,4	3,5	6,4

А) Оценить тесноту связи между индексом (x_3) и долей безработных (x_4); при $\alpha=0,05$ проверить его значимость; при $\gamma=0,888$ найти его интервальную оценку для r_{34} .

Б) Определите долю дисперсии x_4 , обусловленную влиянием x_3 .

Задание №8.

На основании данных о загрузке компьютера (x) и времени его реакции (y) найти точечную оценку условного математического ожидания в предположении, что генеральное уравнение регрессии линейное:

Загрузка компьютера, в % (x)	10	23	36	48	50
Время реакции, в сек. (y)	1,6	2,4	2,8	3,2	3,3

а) определить оценки параметров уравнения регрессии a_0 и a_1 и остаточной дисперсии S^2 , в предположении, что генеральное уравнение регрессии имеет вид $\hat{y}_x = a_0 + a_1x$.

б) проверить при $\alpha=0,01$ значимость коэффициента регрессии, т.е. гипотезы $H_0=a_1=0$;

в) с надежностью $\gamma=0,95$ найти интервальные оценки параметров a_0 и a_1 ;

г) с надежностью $\gamma=0,9$ установить интервальную оценку условного математического ожидания y при $x_0=36$;

д) определить при $\gamma=0,9$ доверительный интервал предсказания \tilde{y}_{n+1} в точке $x=65$.

Задание №9.

Себестоимость (y) одного экземпляра книги в зависимости от тиража (x)

(тыс. экз.) характеризуется данными, собранные издательством. Определите МНК оценки параметров уравнения регрессии a_0 и a_1 гиперболического вида. С надежностью $\gamma=0,9$ построить доверительные интервалы для параметров a_0 и a_1 , а также условного математического ожидания \tilde{y}_{n+1} в точке $x=10$.

Тираж, тыс. экз. (x)	1	2	3	5	10	20	30	50
Себестоимость, руб. (y)	9,1	5,3	4,11	2,83	2,11	1,62	1,41	1,3

Задание №10.

По данным $n=15$ предприятий отрасли рассчитан коэффициент корреляции между себестоимостью продукции (x_1) и производительности труда (x_2), который оказался равным $\hat{r}(x_1, x_2) = -0,7$. Требуется:

А) построить матрицу собственных значений Λ и матрицу факторных нагрузок A ;

Б) Определить дисперсию и относительный вклад первой главной компоненты в обобщенную дисперсию исходных признаков.

Задание №11.

Хозяйственная деятельность $n=10$ предприятий приборостроения характеризуется $p=2$ показателями:

№ п/п	Общие затраты на рубль товарной продукции (x_1)	Фондоотдача (x_2)
1	0,92	0,51
2	0,72	0,59
3	0,83	1,03
4	0,81	1,21
5	0,82	0,63
6	0,93	0,68
7	0,84	0,57
8	0,89	1,52
9	0,89	1,04
10	0,95	0,99

Проведите ранжирование предприятий по первой главной компоненте. Графически изобразите предприятия в двух главных компонент. Дайте интерпретацию полученным результатам.

Задание №12.

Найдены значения главных компонент для i -го предприятия $z_{i1}=0,661$, $z_{i2}=2,151$ и матрица факторных нагрузок $A = \begin{pmatrix} -0,756 & 0,654 \\ 0,756 & 0,654 \end{pmatrix}$.

Требуется найти значения исходных показателей для этого предприятия (x_{i1} и x_{i2}), если выборка характеризуется векторами средних $\bar{x} = \begin{pmatrix} 10 \\ 20 \end{pmatrix}$ и среднеквадратических отклонений $S = \begin{pmatrix} 0,072 \\ 0,333 \end{pmatrix}$.

Задание №13.

По данным $n=20$ стран был рассчитан коэффициент корреляции между среднедушевым потреблением мяса (x_1) и среднедушевым потреблением

фруктов (x_2), который оказался равным $\hat{r}(x_1, x_2) = -0,63$. Требуется:

А) построить матрицу собственных значений Λ и матрицу факторных нагрузок A ;

Б) Определить дисперсию и относительный вклад первой главной компоненты в обобщенную дисперсию исходных признаков.

Задание №14.

По иерархическому агломеративному алгоритму провести классификацию $n=7$ семей по двум показателям: x_1 – уровень расходов за летние месяцы на спорт, отдых и культурные нужды (тыс. руб.), x_2 – уровень расходов на продукты питания (тыс. руб.) представлены в таблице:

Номер семьи (i)	1	2	3	4	5	6	7
x_1	2	4	8	12	13	15	16
x_2	10	7	6	11	9	11	18

В качестве расстояния между объектами принять обычное евклидово расстояние принцип ближнего соседа. Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий.

Задание №15.

По иерархическому агломеративному алгоритму провести классификацию $n=7$ семей по двум показателям: x_1 – уровень расходов за летние месяцы на спорт, отдых и культурные нужды (тыс. руб.), x_2 – уровень расходов на продукты питания (тыс. руб.) представлены в таблице:

Номер семьи (i)	1	2	3	4	5	6	7
x_1	2	4	8	12	13	15	16
x_2	10	7	6	11	9	11	18

В качестве расстояния между объектами принять взвешенное евклидово расстояние с весами $\omega_1=0,1$ и $\omega_2=0,9$ и принцип ближнего соседа.

Задание №16.

По иерархическому агломеративному алгоритму провести классификацию $n=7$ семей по двум показателям: x_1 – уровень расходов за летние месяцы на спорт, отдых и культурные нужды (тыс. руб.), x_2 – уровень расходов на продукты питания (тыс. руб.) представлены в таблице:

Номер семьи (i)	1	2	3	4	5	6	7
x_1	2	4	8	12	13	15	16
x_2	10	7	6	11	9	11	18

В качестве расстояния между объектами принять обычное евклидово расстояние принцип центра тяжести. Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий.

Задание №17.

По иерархическому агломеративному алгоритму провести классификацию $n=7$ семей по двум показателям: x_1 – уровень расходов за летние месяцы на спорт, отдых и культурные нужды (тыс. руб.), x_2 – уровень расходов на продукты питания (тыс. руб.) представлены в таблице:

Номер семьи (i)	1	2	3	4	5	6	7
x_1	2	4	8	12	13	15	16
x_2	10	7	6	11	9	11	18

Расстояние между кластерами определить по принципу «дальнего соседа» не нормализуя данные. Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий.

Задание №18.

По иерархическому агломеративному алгоритму провести классификацию $n=7$ семей по двум показателям: x_1 – уровень расходов за летние месяцы на спорт, отдых и культурные нужды (тыс. руб.), x_2 – уровень расходов на продукты питания (тыс. руб.) представлены в таблице:

Номер семьи (i)	1	2	3	4	5	6	7
x_1	2	4	8	12	13	15	16
x_2	10	7	6	11	9	11	18

В качестве расстояния между объектами принять обычное евклидово расстояние принцип «средней связи». Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий.

Задание №19.

По иерархическому агломеративному алгоритму провести классификацию $n=7$ семей по двум показателям: x_1 – уровень расходов за летние месяцы на спорт, отдых и культурные нужды (тыс. руб.), x_2 – уровень расходов на продукты питания (тыс. руб.) представлены в таблице:

Номер семьи (i)	1	2	3	4	5	6	7
x_1	2	4	8	12	13	15	16
x_2	10	7	6	11	9	11	18

В качестве расстояния между объектами принять взвешенное евклидово расстояние с весами $\omega_1=0,1$ и $\omega_2=0,9$ и принцип дальнего соседа.