



Федеральное государственное бюджетное образовательное учреждение высшего образования  
«ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ ДВАЖДЫ ГЕРОЯ  
СОВЕТСКОГО СОЮЗА, ЛЕТЧИКА-КОСМОНАВТА А.А. ЛЕОНОВА»

«УТВЕРЖДАЮ»

И.о. проректора

\_\_\_\_\_ А.В. Троицкий

«\_\_\_» \_\_\_\_\_ 2023 г.

**ИНСТИТУТ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И  
ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И  
УПРАВЛЯЮЩИХ СИСТЕМ**

**РАБОЧАЯ ПРОГРАММА**

**ДИСЦИПЛИНЫ**

**«АНАЛИЗ БОЛЬШИХ ДАННЫХ»**

**Направление подготовки:** 27.03.04 – Управление в технических системах

**Профиль:** Информационные системы и средства управления  
технологическими процессами

**Уровень высшего образования:** бакалавриат

**Форма обучения:** очная

Королев  
2023

Рабочая программа является составной частью основной профессиональной образовательной программы и проходит рецензирование со стороны работодателей в составе основной профессиональной образовательной программы. Рабочая программа актуализируется и корректируется ежегодно.

**Автор: к.т.н., доцент Аббасова Т.С. Рабочая программа дисциплины: Анализ больших данных. – Королев МО: «Технологический университет», 2023 г.**

**Рецензент: д.т.н., профессор Артюшенко В.М.**

Рабочая программа составлена в соответствии с требованиями федерального государственного образовательного стандарта высшего образования (ФГОС ВО) по направлению подготовки бакалавров 27.03.04 «Управление в технических системах» и Учебного плана, утвержденного Ученым советом Технологического университета. Протокол № 9 от 11 апреля 2023 г.

**Рабочая программа рассмотрена и одобрена на заседании кафедры:**

Заведующий кафедрой (ФИО, ученая степень, звание, подпись)	Артюшенко В.М. д.т.н. профессор 			
Год утверждения (переутверждения)	2023			
Номер и дата протокола заседания кафедры	№ 12 от 05.04.2023			

**Рабочая программа согласована:**

**Руководитель ОПОП ВО**  **к.т.н., доц. Е.Н. Дмитренко**

**Рабочая программа рекомендована на заседании УМС:**

Год утверждения (переутверждения)	2023			
Номер и дата протокола заседания УМС	№ 5 от 11. 04.2023			

## **1. Перечень планируемых результатов обучения по модулю, соотнесенных с планируемыми результатами освоения ОПОП ВО**

**Целью изучения дисциплины «Анализ больших данных»** является – Формирование теоретических и методологических основ в области анализа неструктурированной информации, а также практических навыков, использования алгоритмов интеллектуального анализа данных, реализованных в специализированных программных продуктах.

В процессе обучения студент приобретает и совершенствует следующие компетенции.

### **Профессиональные компетенции:**

- ПК-1 - Способен проводить исследования электронных средств и электронных систем БКУ АКА.
- ПК-6 - Способен составлять проектно-сметной документации на проект или программу в РКП.

Основными **задачами** изучения дисциплины являются:

- освоение студентами теоретических методов дисциплины, применяемых при построении интерфейсов информационных систем;
- получение студентами умений и навыков, применяемых для решения практических задач в области построения и использования телекоммуникационных средств и их интерфейсов при организации хранения и поиска удаленных баз данных.

**После завершения освоения данной дисциплины студент должен:**

### **Необходимые знания:**

Языки программирования и языки поведенческого описания; аналоговую и цифровую схемотехнику, дисциплины естественнонаучного и математического цикла в рамках основной профессиональной образовательной программы

Программные продукты: специализированное программное обеспечение по управлению проектами и программам, офисный пакет приложений для операционных систем. Технические аспекты аналогичных программ организации.

### **Необходимые умения:**

Осуществлять поиск, хранение, обработку и анализ информации из различных источников и баз данных, применяя современные, компьютерные и сетевые технологии.. Работать с современными системами автоматизированного проектирования и системами электронного документооборота.

Осваивать новые образцы программных, технических и информационных технологий.

Работать с информационным пространством на сервере организации для хранения, обмена и совместного использования информации по проекту или программе в РКП.

### **Трудовые действия:**

Владеет методами теоретических исследований электронных систем БКУ АКА.

Навыками разработки рекомендаций и заключений по использованию результатов научно-исследовательских и опытно-конструкторских работ электронных средств и электронных систем БКУ АКА

Владеет методами анализа чувствительности проекта или программы к изменению факторов, влияющих на параметры проекта или программы в РКП.

## **2. Место дисциплины в структуре ОПОП ВО**

Дисциплина «Анализ больших данных» относится к факультативу основной образовательной программы подготовки бакалавров по направлению подготовки 27.03.04 «Управление в технических системах».

Дисциплина базируется на дисциплинах: «Информатика», «Теория вероятностей и математическая статистика», «Математическая логика и теория алгоритмов», а также компетенциях: УК-1, ОПК-1,4,6,11.

Основные положения и знания, полученные при освоении дисциплины должны быть использованы при выполнении выпускной квалификационной работы бакалавра.

## **3. Объем дисциплины и виды учебной работы**

Общая трудоемкость дисциплины составляет 2 зачетные единицы, 72 часа.

Таблица 1

<b>Виды занятий</b>	<b>Всего часов</b>	<b>Семестр пятый</b>
<b>Общая трудоемкость</b>	<b>72</b>	<b>72</b>
<b>Аудиторные занятия</b>	<b>32</b>	<b>32</b>
Лекции (Л)	<b>16</b>	<b>16</b>
Практические занятия (ПЗ)	<b>16</b>	<b>16</b>
Семинары (С)	-	-
Лабораторные работы (ЛР) и (или) другие виды аудиторных занятий	-	-
<b>Самостоятельная работа</b>	<b>40</b>	<b>40</b>
<b>Курсовые работы, проекты</b>	-	-
<b>Расчетно-графические работы</b>	-	-
<b>Контрольная работа, домашнее задание</b>	+ -	+ -
<b>Текущий контроль знаний (7 - 8, 15 - 16 недели)</b>	<b>Тест</b>	<b>Тест</b>
<b>Вид итогового контроля</b>	<b>Зачет</b>	<b>Зачет</b>

## 4. Содержание дисциплины

### 4.1. Темы дисциплины и виды занятий

Таблица 2

Наименование тем	Лекции, час.	Практические занятия, час	Занятия в интерактивной форме	Код компетенций
Тема 1. . Концепция «Больших Данных»	2	2	1	ПК-1 ПК-6
Тема 2. Неструктурированная информация	2	-	1	
Тема 3. . Аппаратное и программное обеспечение «Больших Данных»	1	2	1	
Тема 4. . Масштабирование и многоуровневое хранение «Больших Данных»	1	2	1	
Тема 5. Практическое применение «Больших Данных»	4	2	1	
<b>Итого:</b>	<b>16</b>	<b>16</b>	<b>8</b>	

## 4.2. Содержание тем дисциплины

### Тема 1. Концепция «Больших Данных»

Что такое «Большие данные», и что они нам сулят. Разница между бизнес-аналитикой и «Большими данными». Устаревание информации. Рост объемов данных на фоне вытеснения аналоговых средств хранения. Корректная интерпретация информационных потоков. Обработка информационных потоков. Предпосылки применения контент-анализа в различных исследованиях. Необходимость в аналитической работе с большими данными. Явная (выраженная) и скрытая (структурная) информация. Количественная и качественная стратегия анализа текстов. Возможности и ограничения каждого из подходов. Процедура контент-анализа. Определение круга проблем для контент-анализа. Начальный этап исследования: формулирование целей и задач исследования, выбор эмпирического материала, выдвижение рабочих гипотез. Операциональный этап исследования: определение категорий и подкатегорий, выбор единиц анализа, установление правил кодирования. Этап счета. Этап интерпретации результатов. Презентация результатов. Типичные ошибки при проведении контент-анализа. Технические признаки, характеризующие «Большие данные». Принцип V3 – Volume (объем данных), Variety (разнообразие данных) и Velocity (скорость генерации и работы с данными). Интеграция, миграция и построение хранилищ данных. Высокопроизводительные вычисления (High Performance Computing, HPC) при выполнении аналитических исследований. Grid computing (распределенные вычисления на нескольких серверах), in-database analytics (частичный перевод нагрузки при аналитических вычислениях в СУБД, а также регламентное применение готовых аналитических моделей к новым данным полностью на стороне СУБД) и in-memory analytics (применение аналитики прямо в оперативной памяти сервера СУБД)

### Тема 2. Неструктурированная информация

Эвристические алгоритмы поиска, эволюционное вычисление, этапы генетического алгоритма: задание целевой функции (приспособленности) для особей популяции, создание начальной популяции, размножение (скрещивание), мутирование, вычисление значения целевой функции для всех особей, формирование нового поколения (селекция). Задача кластеризации, методы кластеризации, иерархическая кластеризация, алгоритм k-средних, зонтичная кластеризация, методы ненаправленного обучения (Unsupervised Learning). Постановка задачи классификации, подходы и применения, построение и обучение классификатора, оценка качества классификации, рубрикации тренировочных данных (Training Data Set), методы управляемого (направляемого) обучения (Supervised Learning). Методы распознавания образов, дискриминантный анализ, нелинейная оптимизация, этапы формирования нейронных сетей: сбор данных для обучения, подготовка и нормализация данных, выбор топологии сети,

экспериментальный подбор характеристик сети, экспериментальный подбор параметров обучения, собственно обучение, проверка адекватности обучения, корректировка параметров, окончательное обучение, вербализация сети с целью дальнейшего использования. Совместное использование компьютерных технологий и лингвистики для создания алгоритмов, позволяющих анализировать естественные (человеческие) языки. Применение методов обработки естественных языков и других аналитических методов для выявления и извлечения из анализируемого текста субъективной информации, характеризующей настроения, мнения, отношение людей к проблеме. Рассмотрение следующих основных задач: синтез речи, распознавание речи, анализ текста, синтез текста, машинный перевод, вопросно-ответные системы, информационный поиск, извлечение информации, анализ тональности текста, анализ высказываний, упрощение текста.

### **Тема 3. Аппаратное и программное обеспечение «Больших Данных»**

Вычисления некоторых наборов распределенных задач с использованием большого количества компьютеров, образующих кластер. Шаги Map и Reduce. Предварительная обработка входных данных и свёртка данных. Концепция параллелизма. Шаблоны доступа к данным, хеш-таблица, деревья, таксономия NoSQL, колоночные СУБД, bigtable. Разработка и выполнение распределённых программ, расширение вычислительных мощностей посредством добавления в кластер дополнительных узлов, технология Hadoop, распределённая файловая система HDFS (Hadoop Distributed File System), интеграция с NoSQL и MapReduce.

### **Тема 4. Масштабирование и многоуровневое хранение «Больших Данных»**

Модели развёртывания: частное облако, публичное облако, гибридное облако, общественное облако. Модели обслуживания: программное обеспечение, платформа, инфраструктура. Экономические аспекты центров обработки данных. Безопасность при хранении и пересылке данных. Проблема «последней мили». Обработка Fast Data, подтверждение и корректировка априорных знаний и гипотез, синхронизация скорости работы с ростом объема данных. Получение знаний посредством Big Analytics, преобразования зафиксированной в данных информации в новое знание, принцип «обучения с учителем». Высший уровень работы с данными Deep Insight, обучение без учителя (unsupervised learning), использование современных методов аналитики, а также различные способы визуализации, обнаружение знаний и закономерностей, априорно неизвестных.

### **Тема 5. Практическое применение «Больших Данных»**

Практическое применение решений IBM Cognos Analytics и ресурсов платформы IBM Bluemix. Понятие шаблона, создание правил и категорий.

Персональная база данных, фразовый поиск, нечеткий поиск. Возможности уточнения результатов запросов с учетом структуры текста. Анализ совместной встречаемости (collocate analysis) и коэффициент связи категорий (Z-score). Практическое применение решений IBM Cognos Analytics и ресурсов платформы IBM Bluemix. Контент-анализ массовой корреспонденции и социологических опросов. Прямые пропорциональные закономерности, аддитивные закономерности, мультипликативные закономерности.

## **5. Перечень учебно-методического обеспечения для самостоятельной работы по дисциплине**

«Методические указания для обучающихся по освоению дисциплины «Анализ больших данных» приведены в Приложении 2.

## **6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Структура фонда оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведена в Приложении 1.

## **7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины**

### **Основная литература:**

1. Каган, Е. С. Прикладной статистический анализ данных : учебное пособие / Е. С. Каган. — Кемерово : КемГУ, 2021. — 235 с. — ISBN 978-5-8353-2413-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/13431>. — Режим доступа: для авториз. пользователей.

2. Храмов, А. Г. Методы и алгоритмы интеллектуального анализа данных : учебное пособие / А. Г. Храмов. — Самара : СамГУ, 2019. — 176 с. — ISBN 978-5-7883-1414-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/148603>. — Режим доступа: для авториз. пользователей.

### **Дополнительная литература:**

1 Долматов, А. В. Статистический анализ данных в среде MATLAB : учебно-методическое пособие / А. В. Долматов. — Ханты-Мансийск : ЮГУ, 2016. — 34 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/149002>. — Режим доступа: для авториз. пользователей.

## **8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

1. <http://www.rusedu.info/> Сайт «Информационные технологии в образовании»
2. <http://nit.miem.edu.ru/> Сайт «Международная студенческая школа-семинар «Новые информационные технологии»».
3. 1. Электронно-библиотечная система ЭБС Университетская библиотека онлайн <http://www.biblioclub.ru>
4. 2. Электронно-библиотечная система ЭБС ZNANIUM.COM
5. <http://www.znanium.com>

## **9. Методические указания для обучающихся по освоению дисциплины**

Методические указания для обучающихся по освоению дисциплины (модуля) приведены в Приложении 2 к настоящей рабочей программе.

## **10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине**

**Перечень программного обеспечения:** LibreOffice, Microsoft Power BI - Свободно распространяемое ПО.

**Информационные справочные системы:** не предусмотрено курсом данной дисциплины

Ресурсы информационно-образовательной среды ТУ:

Рабочая программа и методическое обеспечение по дисциплине «Анализ больших данных».

## **11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

**Лекционные занятия:**

- аудитория, оснащенная презентационной техникой (проектор, экран).

**Практические занятия:**

- учебный класс, оснащенный вычислительной техникой (ПК) и доступом к Интернет-ресурсам, интерактивной доской Smart Board.

- рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;

- рабочие места студентов, оснащенные компьютерами с доступом в Интернет.

**Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине (модулю)**

**ИНСТИТУТ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И  
ТЕХНОЛОГИЙ**

**КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И  
УПРАВЛЯЮЩИХ СИСТЕМ**

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ  
ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО  
ДИСЦИПЛИНЕ  
«АНАЛИЗ БОЛЬШИХ ДАННЫХ»**

**Направление подготовки:** 27.03.04 – Управление в технических системах

**Профиль:** Информационные системы и средства управления технологическими процессами

**Уровень высшего образования:** бакалавриат

**Форма обучения:** очная

Королев  
2023

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

№ п/п	Индекс компетенции	Содержание компетенции (или ее части)*	Раздел дисциплины, обеспечивающий формирование компетенции (или ее части)	В результате изучения раздела дисциплины, обеспечивающего формирование компетенции (или ее части), обучающийся должен:		
				Необходимые знания	Необходимые умения	Трудовые действия
1	ПК-1	Способен проводить исследования электронных средств и электронных систем БКУ АКА.	Тема 1. Тема 2. Тема 3. Тема 4. Тема 5. Тема 6. Тема 7. Тема 8.	Языки программирования и языки поведенческого описания; аналоговую и цифровую схемотехнику, дисциплины естественнонаучного и математического цикла в рамках основной профессиональной образовательной программы	Осуществлять поиск, хранение, обработку и анализ информации и из различных источников и баз данных, применяя современные, компьютерные и сетевые технологии .. Работать с современными системами автоматизированного проектирования и системами электронного документооборота.	Владеет методами теоретических исследований электронных систем БКУ АКА. Навыками разработки рекомендаций и заключений по использованию результатов научно-исследовательских и опытно-конструкторских работ электронных средств и электронных систем БКУ АКА

2.	ПК-6	способность составлять проектно-сметную документацию на проект или программу в РКП	Тема 1. Тема 2. Тема 3. Тема 4. Тема 5. Тема 6. Тема 7. Тема 8.	Программные продукты: специализированное программное обеспечение по управлению проектами и программам, офисный пакет приложений для операционных систем. Технические аспекты аналогичных программ организации.	Осваивать новые образцы программных, технических и информационных технологий. Работать с информационным пространством на сервере организации и для хранения, обмена и совместного использования информации по проекту или программе в РКП.	Владеет методами анализа чувствительности проекта или программы к изменению факторов, влияющих на параметры проекта или программы в РКП.
----	------	--	--	--	--	--

## 2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Код компетенции	Инструменты, оценивающие сформированность компетенции	Показатель оценивания компетенции	Критерии оценки
ПК-1 ПК-6	Доклад в форме презентации	<p><i>А) полностью сформирована (компетенция освоена на <u>высоком</u> уровне) – 5 баллов</i></p> <p><i>Б) частично сформирована:</i></p> <ul style="list-style-type: none"> <li>• <i>компетенция освоена на <u>продвинутом</u> уровне – 4 балла;</i></li> <li>• <i>компетенция освоена на <u>базовом</u> уровне – 3 балла;</i></li> </ul> <p><i>В) не сформирована (компетенция <u>не сформирована</u>) – 2 и менее баллов</i></p>	<p>Проводится устно с использованием мультимедийных систем, а также с использованием технических средств</p> <p>Время, отведенное на процедуру – 10 - 15 мин.</p> <p>Неявка – 0.</p> <p>Критерии оценки:</p> <ol style="list-style-type: none"> <li>1.Соответствие представленной презентации заявленной тематике (1 балл).</li> <li>2.Качество источников и их количество при подготовке доклада и разработке презентации (1 балл).</li> <li>3.Владение информацией и способность отвечать на вопросы аудитории (1 балл).</li> <li>4.Качество самой представленной презентации (1 балл).</li> <li>5.Оригинальность подхода и всестороннее раскрытие выбранной тематики (1 балл).</li> </ol> <p>Максимальная сумма баллов - 5 баллов.</p> <p>Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.</p>
ПК-1 ПК-6	Контрольная работа	<p><i>А) полностью сформирована (компетенция освоена на <u>высоком</u> уровне) – 5 баллов</i></p> <p><i>Б) частично сформирована:</i></p>	<ol style="list-style-type: none"> <li>1. Проводится в форме письменной работы</li> <li>2.Время, отведенное на процедуру – семестр.</li> </ol> <p>Неявка на защиту контрольной работы – 0.</p> <p>Критерии оценки:</p> <ol style="list-style-type: none"> <li>1.Соответствие содержания контрольной работы заявленной тематике (1 балл).</li> </ol>

		<ul style="list-style-type: none"> <li>• компетенция освоена на продвинутом уровне – 4 балла;</li> <li>• компетенция освоена на базовом уровне – 3 балла;</li> </ul> <p>В) не сформирована (компетенция не сформирована) – 2 и менее баллов</p>	<p>2. Качество источников и их количество при подготовке работы (1 балл).</p> <p>3. Владение информацией и способность отвечать на вопросы аудитории (1 балл).</p> <p>4. Использование специализированного программного обеспечения (1 балл).</p> <p>5. Оригинальность подхода и всестороннее раскрытие выбранной тематики (1 балл).</p> <p>Максимальная сумма баллов - 5 баллов.</p> <p>Результаты оценочной процедуры представляются обучающимся в срок не позднее 1 недели после проведения процедуры – для текущего контроля. Оценка проставляется в электронный журнал.</p>
--	--	---	--

**3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы**

**Тематика докладов в презентационной форме**

1. Проектирование идеального интерфейса для сотового телефона.
2. Анализ свойств человеческого тела как системы интерфейса между психическим и физическим мирами.
3. Мир как интерфейс.
4. Этапы эволюции интерфейсов операционной системы Windows.
5. «Пользовательские качества» интерфейса и их зависимость от опыта пользователя.
6. Организация эффективной процедуры юзабилити-тестирования сотового телефона?
7. Оснащение и назначение лабораторий тестирования юзабилити.
8. Правила поведения персонала юзабилити-лаборатории в процессе юзабилити-тестирования.
9. Полное и промежуточное юзабилити-тестирование.
10. Методы статистической обработки при анализе результатов юзабилити-тестирования.
11. Оценка юзабилити путем анализа процесса взаимодействия между пользователем и продуктом.
12. Этапы проектирования программных интерфейсов и методы юзабилити, используемые на этих этапах.
13. Методы юзабилити для составления требований к проекту.
14. Методы юзабилити для тестирования и оценки программного продукта.
15. Применение метода «мозгового штурма» для оценки юзабилити.
16. Применение метода сценариев для оценки юзабилити.
17. Методы эвристической и экспертной оценки юзабилити.
18. Метод параллельного дизайна для оценки юзабилити.
19. Методы формирования репрезентативной выборки для проведения юзабилити-экспериментов.
21. Методы «исследования контекста» для оценки юзабилити.
22. Перспективы автоматизации классических методов юзабилити-тестирования.
23. Методы Ай-трекинга в юзабилити. Области применения.
24. Разработка плана и состава программно-аппаратных средств для юзабилити-лаборатории по исследованию и проектированию веб-сайтов.
25. Тенденции и направления развития технологий юзабилити-тестирования.

#### 4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Формой контроля знаний по дисциплине «Анализ больших данных» являются две текущие аттестации в виде тестов и одна промежуточная аттестация в виде зачета в письменной форме.

Неделя текущего контроля	Вид оценочного средства	Код компетенций, оценивающих знания, умения, навыки	Содержание оценочного средства	Требования к выполнению	Срок сдачи (неделя семестра)	Критерии оценки по содержанию и качеству с указанием баллов
Согласно графику учебного процесса	тестирование	ПК-1 ПК-6	20 вопросов	Компьютерное тестирование ; время отведенное на процедуру -30 минут	Результаты тестирования предоставляются в день проведения процедуры	Критерии оценки определяются процентным соотношением. Не явка -0 Удовлетворительно - от 51% правильных ответов. Хорошо - от 70%. Отлично – от 90%.
Согласно графику учебного процесса	тестирование	ПК-1 ПК-6	20 вопросов	Компьютерное тестирование; время отведенное на процедуру – 30 минут	Результаты тестирования предоставляются в день проведения процедуры	Критерии оценки определяются процентным соотношением. Не явка -0 Удовлетворительно - от 51% правильных ответов. Хорошо - от 70%. Отлично – от 90%. Максимальная оценка – 5 баллов.
Согласно графику учебного процесса	Зачет	ПК-1 ПК-6	практическое задание	Зачет проводится в письменной форме, путем ответа на вопросы. Время, отведенное на процедуру – 30 минут.	Результаты предоставляются в день проведения зачета	Критерии оценки: <b>«Зачтено»:</b> <ul style="list-style-type: none"> <li>• знание основных понятий предмета;</li> <li>• умение использовать и применять полученные знания на практике;</li> <li>• работа на семинарских занятиях;</li> <li>• знание</li> </ul>

						<p>основных научных теорий, изучаемых предметов;</p> <ul style="list-style-type: none"> <li>• ответ на вопросы билета.</li> </ul> <p>«Не зачтено»:</p> <ul style="list-style-type: none"> <li>• демонстрирует частичные знания по темам дисциплин;</li> <li>• незнание основных понятий предмета;</li> <li>• неумение использовать и применять полученные знания на практике;</li> <li>• не работал на семинарских занятиях;</li> <li>• не отвечает на вопросы.</li> </ul>
--	--	--	--	--	--	--

Итоговое начисление баллов по дисциплине осуществляется в соответствии с разработанной и внедренной балльно-рейтинговой системой контроля и оценивания уровня знаний и внеучебной созидательной активности обучающихся.

### **Примеры тестовых заданий:**

Тесты используются в режиме промежуточного контроля. По форме заданий выбраны закрытые тесты (с выборочным ответом). Каждому вопросу соответствует один вариант ответа.

1. Аналитик это ...

- а) специалист в области анализа и моделирование
- б) специалист в предметной области;
- в) человек, решающий определенные задачи;
- г) человек, который имеет опыт в программировании.

2 Эксперт это ...

- а) специалист в области анализа и моделирование;
- б) специалист в предметной области;
- в) человек, решать определенные задачи;
- г) человек, который имеет опыт в программировании.

3 Задача классификации сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

4 Задача регрессии сводится к ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

5. Задача кластеризации заключается в ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

6 Целью поиска ассоциативных правил является ...

- а) нахождения частых зависимостей между объектами или событиями;
- б) определения класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиска независимых групп и их характеристик в всем множестве анализируемых данных.

7. До предполагаемых моделей относятся такие модели данных:

- а) модели классификации и последовательностей;

- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
- в) классификации, кластеризации, исключений, итоговые и ассоциации;
- г) модели классификации, последовательностей и исключений.

8 В описательных моделях относятся следующие модели данных:

- а) модели классификации и последовательностей;
- б) регрессивные, кластеризации, исключений, итоговые и ассоциации;
- в) классификации, кластеризации, исключений, итоговые и ассоциации;
- г) модели классификации, последовательностей и исключений.

9 Модели классификации описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

10. Модели последовательностей описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
- в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
- г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

11 Регрессивные модели описывают ...

- а) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
- б) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;

в) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

## 12. Виды лингвистической неопределенности:

а) неточность измерений значений определенной величины, выполняемых физическими приборами;

б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);

в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость)

г) неоднозначность смысла фраз (Синтаксическая и семантическая).

## 13. Модели исключений описывают ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

## 14. Итоговые модели обнаружат ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

## 15. Модели ассоциации проявляют ...

а) исключительные ситуации в записях, которые резко отличаются произвольной признаку от основной множества записей;

б) ограничения на данные анализируемого массива;

в) закономерности между связанными событиями;

г) группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

## 16. Виды физической неопределенности данных:

а) неточность измерений значений определенной величины, выполняемых физическими приборами; случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью)

б) неопределенность значений слов (Многозначность, размытость, непонятность, нечеткость); неоднозначность смысла фраз (Синтаксическая и семантическая);

в) случайность (или наличие в внешней среде нескольких возможностей, каждая из которых случайным образом может стать действительностью); неопределенность значений слов (многозначность, размытость, неясность, нечеткость);

г) неоднозначность смысла фраз (Синтаксическая и семантическая).

## 17 Очистка данных — ...

а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач

в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи

г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

## 18 Обогащение — ...

а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач

в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи

г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему.

## 19 Консолидация — ...

а) комплекс методов и процедур, направленных на устранение причин, мешающих корректной обработке: аномалий, пропусков, дубликатов, противоречий, шумов и т.д.

б) процесс дополнения данных некоторой информацией, позволяющей повысить эффективность развязку аналитических задач

в) объект, содержащий структурированные данные, которые могут оказаться полезными для развязку аналитического задачи

г) комплекс методов и процедур, направленных на извлечение данных из различных источников, обеспечение необходимого уровня их информативности и качества, преобразования в единый формат, в котором они могут быть загружены в хранилище данных или аналитическую систему

## 20 Транзакция — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов.

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных

## 21 Метаданные — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных

## 22 Классификация — ...

а) некоторый набор операций над базой данных, который рассматривается как единственное завершено, с точки зрения пользователя, действие над некоторой информацией, обычно связано с обращением к базе данных

б) разновидность систем хранения, ориентирована на поддержку процесса анализа данных целостность, обеспечивает, непротиворечивость и хронологию данных, а также высокую скорость выполнения аналитических запросов

в) высокоуровневые средства отражения информационной модели и описания структуры данных

г) это установление зависимости дискретной выходной переменной от входных переменных.

### 23 Регрессия — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

### 24 Кластеризация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных.

### 25 Ассоциация — ...

а) это установление зависимости непрерывной выходной переменной от входных переменных

б) эта группировка объектов (наблюдений, событий) на основе данных, описывающих свойства объектов

в) выявление закономерностей между связанными событиями

г) это установление зависимости дискретной выходной переменной от входных переменных

### 26 Машинное обучение — ...

а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.

г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

## 27 Аналитическая платформа — ...

а) специализированный программный решение (или набор решений), который включает в себя все инструменты для извлечения закономерностей из сырых данных

б) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.

г) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

## 28 Обучающая выборка — ...

а) эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов

б) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

в) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, что и отвечает ему правильный выходной результат.

г) выявление в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

## 29 Ошибка обучения — ...

а) это ошибка, допущенная моделью на учебной множества.

б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества

- в) имена, типы, метки и назначения полей исходной выборки данных
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

30 Ошибка обобщения — ...

- а) это ошибка, допущенная моделью на учебной множества.
- б) это ошибка, полученная на тестовых примерах, то есть, что вычисляется по тем же формулам, но для тестовой множества
- в) имена, типы, метки и назначения полей исходной выборки данных
- г) набор данных, каждая запись которого представляет собой учебный пример, содержащего заданный входной влияние, и соответствующий ему правильный выходной результат

## 4.2. Типовые вопросы, выносимые на зачет

### Вопросы к Теме 1. Концепция «Больших Данных»

1. В чем принципиальное отличие концепции Big Data от традиционного подхода BI?
2. Понятие явной (выраженной) и скрытой (структурной) информации.
3. Определение контент-анализа.
4. Каковы основные понятия контент-анализа?
5. Какие существуют виды контент-анализа?
6. Какие существуют этапы контент-анализа?
7. Каковы основные признаки, характеризующие «Большие данные»?

### Вопросы к Теме 2. Неструктурированная информация.

1. Сущность и задачи кластеризации.
2. Основные понятия, принципы и предпосылки генетических алгоритмов.
3. Достоинства и недостатки генетических алгоритмов.
4. Классификация нейронных сетей и принципы построения.
5. Искусственная нейронная сеть прямого прохода.
6. Использование генетических алгоритмов для обучения искусственных нейронных сетей
7. Кластеризация как инструмент предварительной обработки данных для искусственной нейронной сети
8. Какова цель синтаксического анализа?
9. Общая схема алгоритма синтаксического анализа «сверху-вниз» и «снизу-вверх».

### Вопросы к Теме 3. Аппаратное и программное обеспечение «Больших Данных»

1. Схема работы фаз map( $f$ ,  $c$ ) и reduce( $f$ ,  $c$ ).

2. Преимущества, ограничения и недостатки парадигмы MapReduce.
3. Какие бывают модели данных и запросов в NoSQL?
4. Какие бывают системы хранения данных в NoSQL?
5. Основные принципы работы фреймворка Hadoop.
6. Репликация данных в распределенной файловой системе HDFS.

#### **Вопросы к Теме 4. Масштабирование и многоуровневое хранение «Больших Данных»**

1. Модели развертывания облачных хранилищ.
2. Модели обслуживания облачных хранилищ.
3. Постановка и описание проблемы «последней мили».
4. Безопасность, производительность и надежность при работе с облачными данными.
5. Экономическая составляющая облачных подходов.
6. Способы машинного обучения.
7. Основные фазы обработки «больших данных».

#### **Вопросы к Теме 5. Практическое применение «Больших Данных»**

1. Чем отличаются текстовая и персональная базы данных?
2. Метод анализа комбинации слов (collocate analysis).
3. Понятие «сила связи».
4. Статистическая мера совместной встречаемости слов и категорий (Z-score).
5. Реализация закономерностей в системе IBM Cognos Analytics.

**Методические указания для обучающихся по освоению дисциплины**

**ИНСТИТУТ ИНФОКОММУНИКАЦИОННЫХ СИСТЕМ И  
ТЕХНОЛОГИЙ  
КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И  
УПРАВЛЯЮЩИХ СИСТЕМ**

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО  
ОСВОЕНИЮ ДИСЦИПЛИНЫ  
«АНАЛИЗ БОЛЬШИХ ДАННЫХ»  
(Приложение 2 к рабочей программе)**

**Направление подготовки:** 27.03.04 «Управление в технических системах»

**Квалификация (степень) выпускника:** бакалавр

**Профиль:** Информационные системы и средства управления технологическими процессами

**Уровень высшего образования:** бакалавриат

**Форма обучения:** очная

Королев  
2023

## 1. Общие положения

### **Цель дисциплины:**

- приобретение студентами знаний и представлений об основных принципах, закономерностях, методах организации беспроводных сетей;
- приобретение студентами теоретических сведений и практических навыков, позволяющих проводить как расчет, так и измерения основных параметров и характеристик беспроводных сетей.

### **Задачи дисциплины:**

- освоение студентами теоретических методов дисциплины, применяемых при проектировании беспроводных сетей;
- получение студентами умений и навыков, применяемых для решения практических задач в области беспроводных сетей.

## 2. Указания по проведению практических занятий

### **Практическое занятие 1 - 4.**

#### **Тема: Параметрические и непараметрические статистические критерии**

**Вид практического занятия:** смешанная форма практического занятия.

Образовательные технологии: практическая работа в группах.

Понятие статистических критериев. Виды критериев: параметрические и непараметрические. Критерии для зависимых и независимых выборок. Многофункциональные статистические критерии.

Продолжительность занятия 8 часов.

### **Практическое занятие 5-8-**

#### **Тема практического занятия: Корреляционный анализ**

**Вид практического занятия:** смешанная форма практического занятия.

Образовательные технологии: практическая работа в группах.

Понятие и виды корреляции. Линейная и нелинейная связи. Типы корреляционных связей и их графическое представление. Выбросы. Основные коэффициенты корреляции и их характеристика. Частный коэффициент корреляции.

Продолжительность занятия: 8 часов.

## 3. Указания по проведению лабораторного практикума

Не предусмотрен учебным планом.

#### 4. Указания по проведению самостоятельной работы студентов

№ п/п	Наименование блока (раздела) дисциплины	Виды СРС
1.	Тема 1.	1. Источник данных студент выбирает самостоятельно также, как и предметную область. Среди обязательных требований наличие не менее 10000 записей, а также геопривязок данных любого уровня грануляции.
2.	Тема 2.	2. Модификация данных и первичный анализ Модификация исходных данных, а также предварительный анализ в форме отчета верхнего уровня формируется либо в системе Microsoft Power BI, либо IBM Cognos Analytics на выбор студента
3.	Тема 3.	3. Внедрение вычисляемых полей Вычисляемые поля могут представлять из себя как промежуточные вычисления в рамках статистического анализа (п.4), так и введенные KPI или иные меры на исходных данных.
4.	Тема 4.	4. Проведение статистического анализа Статистический анализ проводится с целью выявления скрытых зависимостей или для подтверждения гипотез, сформированных по итогам первичного анализа (п.2), в частности, допустимо проведение регрессионного анализа
5.	Тема 5.	5.Создание итогового видео ролика Видео ролик представляет из себя вынесенные на карту статистические данные, а также итоги проведенного анализа

## **5. Указания по проведению контрольных работ**

### **5.1. Требования к структуре**

Структура контрольной работы должна способствовать раскрытию темы: иметь титульный лист, содержание, введение, основную часть, заключение, список литературы.

### **5.2. Требования к содержанию**

5.2.1. Во введении обосновывается актуальность темы, определяется цель работы, задачи и методы исследования.

5.2.2. При определении целей и задач исследования необходимо правильно их формулировать. Так, в качестве цели не следует употреблять глагол «сделать». Правильно будет использовать глаголы: «раскрыть», «определить», «установить», «показать», «выявить» и т.д.

5.2.3. Основная часть работы включает вопросы, каждый из которых посвящается решению задач, сформулированных во введении, и заканчивается констатацией итогов.

5.2.4. Приветствуется иллюстрация содержания работы таблицами, графическим материалом (рисунками, схемами и т.п.).

5.2.5. Необходимо давать ссылки на используемую Вами литературу.

5.2.6. Заключение должно содержать сделанные автором работы выводы, итоги исследования.

5.2.7. Вслед за заключением идет список литературы, который должен быть составлен в соответствии с установленными требованиями. Если в работе имеются приложения, они оформляются на отдельных листах, и должны быть соответственно пронумерованы.

### **5.3. Требования к оформлению**

Объём контрольной работы – 5...10 страниц формата А4, напечатанного с одной стороны текста (1,5 интервал, шрифт Times New Roman). Контрольная работа должна быть также представлена в электронном виде.

### **Тематика контрольных работ**

1. Распознавание лица с помощью нейронных сетей.
2. Влияние GDPR на сбор и обработку больших данных.
3. Разработка лекарственных препаратов с помощью искусственного интеллекта.
4. Влияние технологии смарт-контрактов на межорганизационные процессы.
5. Встроенный инструмент удаленного web-управления CMS.
6. Встроенный инструмент удаленного web-управления CMS.
7. Решение задач с использованием параметрических и непараметрических статистических критериев.

8. Решение задач с использованием корреляционного анализа.
9. Анализ результатов описательной статистики.
10. Решение задач с применением индексного метода.
11. Решение задач на ранжирование социологических данных.
12. Решение задач с применением корреляционного анализа.
13. Решение задач с применением факторного анализа.
14. Решение задач с применением регрессионного
15. Организации очередей и управления перегрузками в сети.

## **6. Перечень основной и дополнительной учебной литературы**

### **Основная литература:**

1. Каган, Е. С. Прикладной статистический анализ данных : учебное пособие / Е. С. Каган. — Кемерово : КемГУ, 2018. — 235 с. — ISBN 978-5-8353-2413-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/13431>). — Режим доступа: для авториз. пользователей.

2. Храмов, А. Г. Методы и алгоритмы интеллектуального анализа данных : учебное пособие / А. Г. Храмов. — Самара : СамГУ, 2019. — 176 с. — ISBN 978-5-7883-1414-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/148603>. — Режим доступа: для авториз. пользователей.

### **Дополнительная литература:**

1 Долматов, А. В. Статистический анализ данных в среде MATLAB : учебно-методическое пособие / А. В. Долматов. — Ханты-Мансийск : ЮГУ, 2016. — 34 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/149002>. — Режим доступа: для авториз. пользователей.

## **7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

2. <http://www.rusedu.info/> Сайт «Информационные технологии в образовании»
3. <http://nit.miem.edu.ru/> Сайт «Международная студенческая школа-семинар «Новые информационные технологии»».
4. 1. Электронно-библиотечная система ЭБС Университетская библиотека онлайн <http://www.biblioclub.ru>
5. 2. Электронно-библиотечная система ЭБС ZNANIUM.COM
6. <http://www.znanium.com>

## **8. Перечень информационных технологий**

**Перечень программного обеспечения:** LibreOffice, ВІ - Свободно распространяемое ПО.

**Информационные справочные системы:** не предусмотрено курсом данной дисциплины

Ресурсы информационно-образовательной среды Университета.

Рабочая программа и методическое обеспечение по дисциплине «Анализ больших данных».