

**МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНЖЕНЕРНЫХ
ТЕХНОЛОГИЙ»**

УТВЕРЖДАЮ
Проректор по учебной работе

(подпись) **Василенко В.Н.**
(Ф.И.О.)

«25» мая 2023 г.

**РАБОЧАЯ ПРОГРАММА
ДИСЦИПЛИНЫ**

Большие объемы данных

Направление подготовки

09.03.03 Прикладная информатика

Направленность (профиль) подготовки

Цифровизация бизнес-процессов

Квалификация выпускника

бакалавр

Воронеж

1. Целями освоения дисциплины

Целью освоения дисциплины “Большие объемы данных” является формирование компетенций обучающегося в области профессиональной деятельности и сфере профессиональной деятельности:

Об Связь, информационные и коммуникационные технологии в сфере исследования, разработки, внедрения и сопровождения информационных технологий и систем.

Дисциплина направлена на решение задач профессиональной деятельности следующих типов:

- производственно-технологический;
- организационно-управленческий;
- проектный.

Программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.03 «Прикладная информатика».

2. Перечень планируемых результатов обучения, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения дисциплины в соответствии с предусмотренными компетенциями обучающийся должен:

№п/п	Код компетенции	Наименование компетенции	Код и наименование индикатора достижения компетенции
1	2	3	4
1	УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИД2 _{УК-1} – Решает поставленные задачи, используя системный подход, на основе критического анализа и синтеза информации и оценивает последствия возможных решений
2	ПКв-1	Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе	ИД3 _{ПКв-1} - имеет навыки согласования требования к ИС с заказчиком
3	ПКв-9	Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач	ИД1 _{ПКв-9} - использование прикладного инструментария для разработки программного обеспечения в соответствии с назначением объекта проектирования

Код и наименование индикатора достижения компетенции	Результаты обучения (показатели оценивания)
ИД2 _{УК-1} – Решает поставленные задачи, используя системный подход, на основе критического анализа и синтеза информации и оценивает последствия возможных решений	Знает технологии хранения и обработки больших данных
	Умеет формулировать задачи в сфере анализа больших данных
	Владеет навыками использования современных информационных технологий для решения прикладных задач

ИДЗ _{ПКв-1} - имеет навыки согласования требования к ИС с заказчиком	Знает основные тенденции развития цифровых технологий в области анализа больших объемов данных
	Умеет использовать результаты анализа данных для принятия эффективных решений при согласовании требований к ИС с заказчиком
	Владеет навыками использования современных методов анализа больших данных
ИД1 _{ПКв-9} - использование прикладного инструментария для разработки программного обеспечения в соответствии с назначением объекта проектирования	Знает специфику анализа больших данных
	Умеет проводить сравнительный анализ и выбор цифровых технологий для решения прикладных задач
	Владеть инструментальными средствами обработки статистической информации, а также навыками применения вероятностно - статистического подхода к постановке и решению прикладных задач

3. Место дисциплины в структуре образовательной программы ВО

Дисциплина относится к части, формируемой участниками образовательных отношений - факультативы Блока ФТДООП. Дисциплина является не обязательной к изучению.

Изучение дисциплины основано на знаниях, умениях и навыках, полученных при изучении обучающимися следующих дисциплин: Компьютерные технологии, Базы данных, Web-технологии.

Дисциплина является предшествующей при подготовке, выполнении и защите выпускной квалификационной работы.

4. Объем дисциплины и виды учебных занятий

Общая трудоемкость дисциплины составляет 2 зачетные единицы.

Виды учебной работы	Всего ак. ч	Распределение трудоемкости по семестрам, ак. ч
		№ семестра 7
Общая трудоемкость дисциплины (модуля)	72	72
Контактная работа в т. ч. аудиторные занятия:	30,85	30,85
Лекции	15	15
<i>в том числе в форме практической подготовки</i>	-	-
Практические/лабораторные занятия	15	15
<i>в том числе в форме практической подготовки</i>	15	15
Консультации текущие	0,75	0,75
Консультации перед экзаменом	-	-
Вид аттестации (зачет/экзамен)	0,1	0,1
Самостоятельная работа:	41,15	41,15
Проработка материалов по лекциям, учебникам, учебным пособиям	10	10
Подготовка к практическим/лабораторным занятиям	21,15	21,15
Домашнее задание, реферат	10	10

5 Содержание дисциплины

5.1 Содержание разделов дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела	Часы
1	Большие данные	Понятие о BigData. Причины появления больших данных. Примеры возможностей для промышленности. Понятие жизненного цикла аналитики данных. Когнитивный анализ данных. Парадигма MapReduce. Проекты ApacheHadoop, ApacheSpark.	45,15
2	Машинное обучение	Классификация задач обработки данных. Информативность и выбор признаков. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация. Основы языка R: основные функции, синтаксис, типы данных. DataMininng: инструменты, основные библиотеки. Технологии хранения больших данных. Парадигма NoSQL.	26
		<i>Консультации текущие</i>	0,75
		<i>Вид аттестации (зачет/экзамен)</i>	0,1

5.2 Разделы дисциплины и виды занятий

№ п/п	Наименование раздела дисциплины	Лекции, час	ПЗ, час	СРО, час
1	Большие данные	7	7	20
2	Машинное обучение	8	8	21,15
	<i>Консультации текущие</i>	0,75		
	<i>Вид аттестации (зачет/экзамен)</i>	0,1		

5.2.1 Лекции

№ п/п	Наименование раздела дисциплины	Тематика лекционных занятий	Трудоемкость, час
1	Большие данные	Понятие о BigData. Причины появления больших данных. Примеры возможностей для промышленности. Понятие жизненного цикла аналитики данных. Когнитивный анализ данных. Парадигма MapReduce. Проекты ApacheHadoop, ApacheSpark.	7
2	Машинное обучение	Классификация задач обработки данных. Информативность и выбор признаков. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация. Основы языка R: основные функции, синтаксис, типы данных. DataMininng: инструменты, основные библиотеки. Технологии хранения больших данных. Парадигма NoSQL.	8

5.2.2 Практические занятия

№ п/п	Наименование раздела дисциплины	Тематика практических занятий	Трудоемкость, час
1	Большие данные	Жизненный цикл аналитики данных. Когнитивный анализ данных. MapReduce. ApacheHadoop, ApacheSpark.	7

2	Машинное обучение	Задачи классификации и кластеризации. K-means и C-means кластеризация, классификация. Основы языка R: основные функции, синтаксис, типы данных языка R. Инструменты, основные библиотеки для работы с данными.	8
---	-------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---

5.3.3 Лабораторный практикум не предусмотрен.

5.2.4 Самостоятельная работа обучающихся (СРО)

№ п/п	Наименование раздела дисциплины	Вид СРО	Трудоемкость, час
Четвертый семестр			
1.	Большие данные	Проработка материалов по конспекту лекций, учебникам, учебным пособиям	10
		Домашняя работа	10
		Подготовка к защите работ, выполняемых на практических работах	5
2.	Машинное обучение	Проработка материала по конспекту лекций, учебникам, учебным пособиям	11,5
		Домашняя работа	10
		Подготовка к защите работ, выполняемых на практических работах	5

6 Учебно-методическое обеспечение и информационное обеспечение дисциплины

Для освоения дисциплины обучающийся может использовать:

6.1 Основная литература

1. Колемаев, В.А. Теория вероятностей и математическая статистика : учебник / В.А. Колемаев, В.Н. Калинина. - Москва :Юнити-Дана, 2015. - 352 с. : табл. - ISBN 5-238-00560-1 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=436721>

2. Гусева, Е.Н. Теория вероятностей и математическая статистика : учебное пособие / Е.Н. Гусева. - Москва : Флинта, 2011. - 220 с. - ISBN 978-5-9765-1192-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=83543>

6.2 Дополнительная литература

1. Ю. Лесковец, А. Раджараман, Д. Ульман. Анализ больших наборов данных. / Пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 498 с.

6.3 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся.

Никитин Б. Е., Методические указания к самостоятельной работе обучающихся по дисциплине [Электронный ресурс] : для студентов, обучающихся по направлению 09.03.02 очной и заочной формы обучения / Б. Е. Никитин; ВГУИТ, Кафедра информационных технологий, моделирования и управления. - Воронеж : ВГУИТ, 2015. - Режим доступа <https://education.vsuet.ru/mod/glossary/view.php?id=48526>

6.4 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Сайт научной библиотеки ВГУИТ <<http://biblos.vsu.ru>>.
2. Базовые федеральные образовательные порталы. <http://www.edu.ru/db/portal/sites/portal_page.htm>.
3. Государственная публичная научно-техническая библиотека. <www.gpntb.ru>.
4. Информационно-коммуникационные технологии в образовании. Система федеральных образовательных порталов. <<http://www.ict.edu.ru>>.
5. Национальная электронная библиотека. <www.nns.ru>..
6. Российская государственная библиотека. <www.rsl.ru>.
7. Российская национальная библиотека. <www.nlr.ru>.
8. Информационно-поисковая система ФИПС. <<http://www1.fips.ru>>
9. Европейская патентная поисковая система ЕРО — European Patent Office <<http://ep.espacenet.com>>
10. Ведомство патентов и торговых марок США US Patent and Trademark Office (USPTO) <<http://www.uspto.gov>>
11. Список поисковых систем патентов <http://www.borovic.ru/index_p_14_p_2.html>
12. Поисковая система «Google». <<https://www.google.ru/>>.
13. Поисковая система «Рамблер». <www.rambler.ru/>.
14. Поисковая система «Яндекс». <www.yandex.ru/>.

6.5 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения, современных профессиональных баз данных и информационных справочных систем

При изучении дисциплины используется программное обеспечение, современные профессиональные базы данных и информационные справочные системы: ЭИОС университета, в том числе на базе программной платформы «Среда электронного обучения ЗКЛ», автоматизированная информационная база «Интернет-тренажеры», «Интернет-экзамен» и пр. (указать средства, необходимы для реализации дисциплины).

При освоении дисциплины используется лицензионное и открытое программное обеспечение – н-р, ОС Windows, ОС ALT Linux.

7 Материально-техническое обеспечение дисциплины

Аудитории для проведения лекционных занятий (а.334)	Комплекты мебели для учебного процесса- 30 шт.	
Читальные залы библиотеки	Компьютеры со свободным доступом в сеть Интернет и Электронными библиотечными и информационно справочными системами.	
Аудитории для проведения практических занятий (а.339 - компьютерный класс каф.ИТМУ)	Количество ПЭВМ -12 (Coreis 540), проектор – 1 (ViewSonikPJD5255)	Microsoft Office 2007, Microsoft Office Professional Plus 2007 (Access, Visio, Project), Microsoft Share Point Designer 2013 Russian Academic OPEN No Level #44822753 от 17.11.2008 http://eopen.microsoft.com
Помещение для самостоятельной работы	Количество ПЭВМ -12 (Coreis 540), проектор – 1	Microsoft Office 2007, Microsoft Office Professional Plus 2007 (Access, Visio,

(a.339)	(ViewSonikPJD5255	Project), Microsoft Share Point Designer 2013 Russian Academic OPEN No Level #44822753 от 17.11.2008 http://eopen.microsoft.com
---------	-------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------

8. Оценочные материалы для промежуточной аттестации обучающихся по дисциплине

Оценочные материалы (ОМ) для дисциплины (модуля) включают в себя:

- перечень компетенций с указанием индикаторов достижения компетенций, этапов их формирования в процессе освоения образовательной программы;
- описание шкал оценивания;
- типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков;
- методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности.

ОМ представляются отдельным комплектом и **входят в состав рабочей программы дисциплины** в виде приложения.

Оценочные материалы формируются в соответствии с П ВГУИТ «Положение об оценочных материалах».

ПРИЛОЖЕНИЕ
к рабочей программе

1. Организационно-методические данные дисциплины для заочной формы обучения

1.1 Объемы различных форм учебной работы и виды контроля в соответствии с учебным планом (заочная форма)

Общая трудоемкость дисциплины составляет 2 зачетных единицы

Виды учебной работы	Всего ак. ч	Распределение трудоемкости по семестрам, ак. ч
		№ семестра 7
Общая трудоемкость дисциплины (модуля)	72	72
Контактная работа в т. ч. аудиторные занятия:	13	13
Лекции	6	6
<i>в том числе в форме практической подготовки</i>	-	-
Практические/лабораторные занятия	6	6
<i>в том числе в форме практической подготовки</i>	6	6
Консультации текущие	0,1	0,1
Консультации перед экзаменом	-	-
Контрольная работа	0,8	0,8
Вид аттестации (зачет/экзамен)	0,1	0,1
Самостоятельная работа:	55,1	55,1
Проработка материалов по лекциям, учебникам, учебным пособиям	15	15
Подготовка к практическим/лабораторным занятиям	15,1	15,1
Домашнее задание, реферат	25	25
Контроль зачет	3,9	3,9

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

по дисциплине

Большие данные

1 Перечень компетенций с указанием этапов их формирования

№п/п	Код компетенции	Наименование компетенции	Код и наименование индикатора достижения компетенции
1	2	3	4
1	УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИД2 _{УК-1} – Решает поставленные задачи, используя системный подход, на основе критического анализа и синтеза информации и оценивает последствия возможных решений
2	ПКв-1	Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе	ИД3 _{ПКв-1} - имеет навыки согласования требования к ИС с заказчиком
3	ПКв-9	Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач	ИД1 _{ПКв-9} - использование прикладного инструментария для разработки программного обеспечения в соответствии с назначением объекта проектирования

Код и наименование индикатора достижения компетенции	Результаты обучения (показатели оценивания)
ИД2 _{УК-1} – Решает поставленные задачи, используя системный подход, на основе критического анализа и синтеза информации и оценивает последствия возможных решений	Знает технологии хранения и обработки больших данных
	Умеет формулировать задачи в сфере анализа больших данных
	Владеет навыками использования современных информационных технологий для решения прикладных задач
ИД3 _{ПКв-1} - имеет навыки согласования требования к ИС с заказчиком	Знает основные тенденции развития цифровых технологий в области анализа больших объемов данных
	Умеет использовать результаты анализа данных для принятия эффективных решений при согласовании требований к ИС с заказчиком
	Владеет навыками использования современных методов анализа больших данных
ИД1 _{ПКв-9} - использование прикладного инструментария для разработки программного обеспечения в соответствии с назначением объекта проектирования	Знает специфику анализа больших данных
	Умеет проводить сравнительный анализ и выбор цифровых технологий для решения прикладных задач
	Владеть инструментальными средствами обработки статистической информации, а также навыками применения вероятностно - статистического подхода к постановке и решению прикладных задач

2 Паспорт оценочных материалов по дисциплине

№ п/п	Разделы дисциплины	Индекс контролируемой компетенции (или ее части)	Оценочные средства		Технология/процедура оценивания (способ контроля)
			наименование	№№ заданий	
1	Большие	УК1	Тест	9-19	Компьютерное тестирование

	данные				Процентная шкала. 0-100 %; 0-59,99% - неудовлетворительно; 60-74,99% - удовлетворительно; 75- 84,99% -хорошо; 85-100% - отлично.
			Собеседование (вопросы для зачета)	47-50	Проверка преподавателем Отметка в системе «зачтено – не зачтено»
2	Машинное обучение	ПКв-9, ПКв-1	Тест	1-8, 20-40	Компьютерное тестирование Процентная шкала. 0-100 %; 0-59,99% - неудовлетворительно; 60-74,99% - удовлетворительно; 75- 84,99% -хорошо; 85-100% - отлично.
			Собеседование (вопросы для зачета)	51-58	Проверка преподавателем Отметка в системе «зачтено – не зачтено»
			Домашнее задание	44-45	Проверка преподавателем Отметка в системе «зачтено – не зачтено»

3 Оценочные материалы для промежуточной аттестации

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

Для оценки знаний, умений, навыков студентов по дисциплине применяется бально-рейтинговая система оценки сформированности компетенций студента.

Бально-рейтинговая система оценки осуществляется в течение всего семестра при проведении аудиторных занятий и контроля самостоятельной работы.

Показателями ОМ являются: текущий опрос в виде собеседования на лабораторных работах, тестовые задания и самостоятельно (домашнее задание). Оценки выставляются в соответствии с графиком контроля текущей успеваемости студентов в автоматизированную систему баз данных (АСУБД) «Рейтинг студентов».

Обучающийся, набравший в семестре более 60 % от максимально возможной бально-рейтинговой оценки работы в семестре получает зачет автоматически.

Студент, набравший за текущую работу в семестре менее 60 %, т.к. не выполнил всю работу в семестре по объективным причинам (болезнь, официальное освобождение и т.п.) допускается до зачета, однако ему дополнительно задаются вопросы на собеседовании по разделам, выносимым на зачет.

Аттестация обучающегося по дисциплине проводится в форме тестирования и предусматривает возможность последующего собеседования (зачета). Зачет проводится в виде тестового задания.

Каждый вариант теста включает 30 контрольных заданий, из них:

- 10 контрольных заданий на проверку знаний;
- 10 контрольных заданий на проверку умений;
- 10 контрольных заданий на проверку навыков;

В случае неудовлетворительной сдачи зачета студенту предоставляется право повторной сдачи в срок, установленный для ликвидации академической задолженности по итогам соответствующей сессии. При повторной сдаче зачета количество набранных студентом баллов на предыдущем зачете не учитывается.

3.1 Тесты (тестовые задания и кейс-задания)

3.1.1 Шифр и наименование компетенции

ПКв-9 Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач

№ задания	Тестовое задание
1.	Какие оконные функции можно использовать для того, чтобы сравнивать текущее значение строки с предыдущим или следующим? Выберите один вариант: 1. LAG и LEAD 2. RANK и DENSE_RANK 3. NTAILE и NTILE
2.	Какой оператор используется для возврата в первом операторе SELECT всех строк, которые не возвращаются вторым оператором SELECT? Выберите один вариант: 1. UNION ALL 2. MINUS 3. INTERSECT
3.	При каком условии выполнится оператор соединения UNION? Выберите один вариант: 1. Запросы имеют одно общее поле 2. Равное количество столбцов в обоих запросах 3. Запросы имеют одинаковое количество строк
4.	Какая команда используется для создания виртуальной таблицы? Выберите один вариант: 1. CREATE VIEW 2. CREAT VIRTUAL TABLE 3. ALTER VIEW
5.	Какой оператор объединяет таблицы без удаления дубликатов? Выберите один вариант: 1. UNION ALL 2. UNION 3. FULL JOIN
6.	Зачем нужен оператор INTERSECT? Выберите один вариант: 1. Для удаления дублированных строк 2. Для нахождения пересечения двух множеств 3. Для полного соединения двух таблиц
7.	В чем разница между операторами UNION и JOIN? Выберите один вариант: 1. UNION можно использовать только для объединения подзапросов 2. UNION возвращает дубликаты, а JOIN - нет 3. JOIN объединяет столбцы, а UNION – строки таблиц
8.	Структурированные данные - это Выберите один вариант: 1. Данные, которые соответствуют общеизвестным форматам (doc, docx, pdf, jpeg и тд) 2. Данные с четко заданным форматом и однозначным обращением к ним 3. Данные с гибким форматом, в котором информация не представлена в однозначном виде
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	
9.	Неструктурированные данные - это 1. Данные, которые соответствуют общеизвестным форматам (doc, docx, pdf, jpeg и тд) 2. Данные с четко заданным форматом и однозначным обращением к ним 3. Данные с гибким форматом, в котором информация не представлена в однозначном виде

10.	Какие наборы признаков заложены в 3V? 1. vocal, velocity, variety 2. vacation, velocity, variety 3. volume, velocity, variety 4. violence, velocity, variety
11.	В какой компании создали MapReduce? 1. Google 2. Yahoo 3. Amazon 4. Facebook
12.	Из чего состоит вычислительная модель MapReduce? 1. Map, reduce 2. Map, resort, reduce 3. Map, shuffle, reduce 4. Map, calculation, reduce
13.	Из каких узлов состоит MapReduce? 1. Master, work 2. Master, slave 3. Master, pork
14.	Где используется файловая система HDFS? 1. Apache Hadoop 2. Oracle DB 3. Mongo DB
15.	Из чего состоит архитектура master/slave HDFS? 1. MainNode, StorageNodes, Client 2. MainNode, DataNodes, Client 3. NameNode, DataNodes, Client 4. NameNode, DataNodes, Patient
16.	Power BI – это 1. Семейство продуктов от Microsoft для бизнес-аналитики 2. Desktopное приложение от Microsoft для бизнес-аналитики 3. Надстройка для MS Excel 4. Веб-приложение от Microsoft для бизнес-аналитики
17.	Power BI предназначен для: 1. Создания приложений 2. Обработки данных 3. Визуализации 4. Извлечения данных из источников
18.	Верно ли следующее утверждение? Бизнес-аналитические задачи можно решать только при помощи Power BI. 1. Да 2. Нет
19.	Power Query предназначен для: 1. Создания модели данных и связей между ними 2. Извлечения, преобразования и очистки данных
ПКв-1 Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе	
20.	Чем отличаются меры от вычисляемых столбцов? 1. В случае мер хранится только формула 2. Меры хранятся со всей моделью 3. Меры рассчитывается для каждой строки таблицы 4. Меры вычисляются только при использовании 5. Меры вычисляются моментально после создания
21.	Дах – это 1. Язык запросов для Power Pivot 2. Язык запросов для Power Query
22.	M language – это 1. Язык запросов для Power Pivot 2. Язык запросов для Power Query

23.	<p>Как выбирать визуализацию в Power BI?</p> <ol style="list-style-type: none"> 1. Используя готовые шаблоны выбора визуализаций 2. По принципу "что больше нравится" 3. Поняв предварительно, какие данные визуализируем (сравнение, распределение, связь переменных, структуру и т.д.), и кто будет «потребителем» результатов анализа
24.	<p>Почему Apache Spark называют "ленивым"?</p> <ol style="list-style-type: none"> 1. Трансформации данных накапливаются, а исполняются лишь при вызове действия 2. Самые затратные трансформации не исполняются 3. Spark работает медленнее Pandas 4. Трансформацию данных всегда нужно ждать большое количество времени
25.	<p>Что такое "драйвер" в Apache Spark?</p> <ol style="list-style-type: none"> 1. Связующее звено между распределителем ресурсов и контроллером отказоустойчивости приложения 2. Процесс, выполняющий задачи и конфигурирующий работу приложения 3. Планировщик исполнения запросов к базам данных 4. Программа, которая создает SparkContext и распределяет задачи между исполнителями
26.	<p>Что делают исполнители при отказе работы одного из них?</p> <ol style="list-style-type: none"> 1. Перехватывают работу на рабочие узлы 2. Завершают работу основной программы 3. Останавливаются и ждут перезагрузку вышедшего из строя исполнителя 4. Перезагружают программу и начинают всю работу заново
27.	<p>Формат хранения данных, разработанный для Apache Spark – это</p> <ol style="list-style-type: none"> 1. Xlsx 2. Csv 3. Json 4. Parquet
28.	<p>Что такое A/B тестирование?</p> <ol style="list-style-type: none"> 1. Метод неформальной оценки 2. Метод проверки идей или гипотез 3. Алгоритмическое тестирование 4. Вербальное тестирование
29.	<p>Какие критерии позволяют сказать, есть ли статистически значимые различия между двумя группами?</p> <ol style="list-style-type: none"> 1. Критерий Стьюдента 2. Критерий Манна-Уитни 3. Критерий Менделеева-Клайперона 4. Критерий Шапиро-Уилка
30.	<p>Какой критерий является легко интерпретируемым?</p> <ol style="list-style-type: none"> 1. Критерий Уилкоксона 2. Критерий Коши 3. t- критерий Стьюдента 4. Критерий Манна-Уитни
31.	<p>Какой метод лучше подходит для поиска различий между 10 группами?</p> <ol style="list-style-type: none"> 1. Критерий Уилкоксона 2. Критерий Манна-Уитни 3. Дисперсионный анализ 4. t- критерий Стьюдента
32.	<p>Какой из критериев чувствителен к распределению в данных?</p> <ol style="list-style-type: none"> 1. t- критерий Стьюдента 2. Критерий Уилкоксона 3. Дисперсионный анализ 4. Критерий Манна-Уитни
33.	<p>Какую архитектуру модели стоит использовать для определения вероятности оттока клиента?</p> <ol style="list-style-type: none"> 1. Градиентный бустинг 2. Модель арифметической прогрессии 3. Модель геометрической прогрессии 4. Нейросеть

34.	Какую задачу должна решать ML-модель при определении точной стоимости объекта недвижимости? 1. Персонализация 2. Регрессия 3. Классификация 4. Кластеризация
35.	Как называется математический подход, при котором определяются "плотные острова" в неразмеченных данных? 1. Машина опорных векторов 2. Кластеризация 3. Логистическая регрессия 4. Персонализация
36.	Какой тип архитектуры используется для машинного перевода? 1. Трансформер 2. Байесовский классификатор 3. Метод ближайших соседей 4. Рекуррентная нейросеть 5. Логистическая регрессия
37.	Вы построили нейронную сеть, в которой к входным данным добавляется выход от предыдущего слоя. Как называется данная архитектура? 1. Сверточная нейронная сеть 2. Рекуррентная нейронная сеть
38.	При решении какой задачи обычно используют сверточные нейронные сети? 1. При решении задачи детекции мошеннических операций на данных по финансовым транзакциям 2. При решении задачи детекции автомобилей на изображениях
39.	Может ли помочь автоэнкодер служить инструментом для понижения размерности? 1. Может 2. Не может
40.	Какую проблему помогает решить padding в операции свертки? 1. Позволяет уменьшить входное изображение путем выделения максимального элемента из блока 2. Позволяет чаще задействовать данные на границах изображения в операциях свертки

3.2. Кейс- задания (примеры)

ПКв-1 Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе

№ задания	Формулировка задачи
41.	<p>Дана выгрузка по продажам компании-поставщика тяжелой техники в одном из регионов России. Требуется посчитать какой доход принесла каждая продаваемая модель техники.</p> <p>В среде MSExcel (версия не ниже 10) решение нужно выполнить следующие действия:</p> <ol style="list-style-type: none"> 1. Отформатировать выгрузку как таблицу. 2. Функцией ВПР «вытянуть» в выгрузку цены из листа с ценами моделей 3. Добавить в выгрузку столбец с суммой сделки 4. На листе с ценами моделей добавить столбец с суммарной выручкой по каждой модели (используя функцию СУММЕСЛИ). 5. Добавить строку Итого.
42.	<p>Дана выгрузка по продажам компании-поставщика тяжелой техники в одном из регионов России. Требуется узнать сколько денег заработала фирма в каждом городе на каждой модели в каждом месяце каждым продавцом.</p> <p>В среде MSExcel (версия не ниже 10) нужно выполнить следующие действия:</p> <ol style="list-style-type: none"> 1. Подключится к исходному диапазону с помощью Power Query 2. Подключится к диапазону с ценами с помощью Power Query 3. Объединить диапазоны из п.п. 1 и 2 4. Добавить столбец с произведением цены на количество

	5. Построить сводную таблицу с отображением выручки в разрезе городов, месяцев, продавцов и товаров
43.	<p>Дана выгрузка по продажам компании-поставщика тяжелой техники в одном из регионов России. Требуется узнать сколько денег заработала фирма в каждом городе на каждой модели в каждом месяце каждым продавцом.</p> <p>В среде MSExcel (версия не ниже 10) нужно выполнить следующие действия:</p> <ol style="list-style-type: none"> 1. Загрузить исходные данные в Power Query 2. В редакторе Power Query заполнить пропущенные данные в столбце “Модель” (<i>Преобразование/Заполнить/Вниз</i>) 3. Кома первых двух столбцов (“Модель” и “Дата продажи”) к остальным применить <i>Отменить свертывание столбцов/ Отменить свертывание других</i> 4. Изменить наименование получившихся двух последних столбцов На “Город” и “Количество” соответственно. 5. Загрузить данные в Power Pivot (в усиленную сводную таблицу) <ol style="list-style-type: none"> 6. Связать таблицы <i>Выгрузка</i> и <i>Цены моделей</i> (<i>Представление диаграммы</i> в группе <i>Просмотр</i>). 7. Добавить и связать третью таблицу <i>Calendar</i> (<i>Конструктор/Таблица дат/Создать</i>) 8. На языке DAX ввести формулу для расчета выручки: 9. Создать сводную таблицу и сохранить. 10. Импортировать в Power BI сводную таблицу. 11. Создать интерактивный отчет (интерактивные элементы <i>Карта, Гистограмма, Срез</i>)

3.3 Домашнее задание (примеры)

ПКВ-9 Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач

№ задания	Формулировка задачи
44.	<ol style="list-style-type: none"> 1. Найти набор данных в репозитории Kaggle [2] или [3] и скачать его на свой компьютер 2. Выполнить загрузку в разных вариантах: полной, отдельных столбцов, части строк 3. Выполнить обработку данных и сохранить результаты во внешнем файле
45.	Изучить основные функции библиотеки numpy и научиться их использовать при выполнении операций с многомерными массивами
46.	Изучить функции библиотеки numpy по агрегированию данных. Научиться выполнять агрегирование данных, содержащихся в многомерных массивах.

3.4 Вопросы к зачёту

УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач

№	Текст вопроса
47.	Понятие Больших данных
48.	Особенности сбора, хранения, обработки и анализа Больших данных.
49.	Требования к распределенным информационным системам
50.	Средства построения распределенных информационных систем
ПКВ-1 Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе	
51.	Технология Map-Reduce

52.	Система Apache Hadoop
ПКв-9 Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач	
53.	Базы данных NoSQL. Особенности, классификация.
54.	Возможности NoSQL-баз данных по обеспечению целостности, доступности.
55.	Способы репликации и кластеризации баз данных.
56.	Документно-ориентированные базы данных
57.	Возможности СУБД MongoDB
58.	Работа с документно-ориентированными БД.

Критерии и шкалы оценки:

- **оценка «зачтено»** выставляется студенту, если домашнее задание является самостоятельным, оригинальным текстом, в котором прослеживается авторская позиция, продуманная система аргументов, а также наличествует обоснованные выводы; используются термины, понятия по дисциплине, в рамках которой выполняется работа; полностью соответствует выбранной теме, цели и задачам; текст домашнего задания логически выстроен, имеет четкую структуру; работа соответствует всем техническим требованиям; домашнее задание выполнено в установленный срок.

- **оценка «не зачтено»**, выставляется студенту, если домашнее задание не является самостоятельным, оригинальным текстом, в котором не прослеживается авторская позиция, не продумана система аргументов, а также отсутствуют обоснованные выводы; не используются термины, понятия по дисциплине, в рамках которой выполняется работа; не соответствует выбранной теме, цели и задачам; текст домашнего задания композиционно не выстроен; работа не соответствует техническим требованиям; домашнее задание не выполнено в установленный срок.

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

Процедуры оценивания в ходе изучения дисциплины знаний, умений и навыков, характеризующих этапы формирования компетенций, регламентируются положениями:

- П ВГУИТ 2.4.03 Положение о курсовых экзаменах и зачетах;
- П ВГУИТ 4.1.02 Положение о рейтинговой оценке текущей успеваемости.

Для оценки знаний, умений, навыков обучающихся по дисциплине применяется рейтинговая система. Итоговая оценка по дисциплине определяется на основании определения среднеарифметического значения баллов по каждому заданию.

Зачет по дисциплине выставляется в зачетную ведомость по результатам работы в семестре после выполнения всех видов учебной работы, предусмотренных рабочей программой дисциплины (с отметкой «зачтено») и получении по результатам тестирования по всем разделам дисциплины не менее 60 %.

5. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания для каждого результата обучения по дисциплине

Результаты обучения по этапам формирования компетенций	Предмет оценки (продукт или процесс)	Показатель оценивания	Критерии оценивания сформированности компетенций	Шкала оценивания	
				Академическая оценка или баллы	Уровень освоения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач					
Знать технологии хранения и обработки больших данных	Собеседование (зачет)	Корректно работает с разнородными данными	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (базовый)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Уметь формулировать задачи в сфере анализа больших данных	Выполнение заданий на практических занятиях	Применение знаний по обработке и хранению больших данных к решению заданий на практических занятиях	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Владеть навыками использования современных информационных технологий для решения прикладных задач	Домашнее задание	Демонстрация навыков владения технологиями хранения и обработки больших данных	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
ПКв-1 Способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе					
Знать основные тенденции развития	Собеседование (зачет)	Корректно формирует требования к информационным	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (базовый)

цифровых технологий в области анализа больших объемов данных		системам	Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Уметь использовать результаты анализа данных для принятия эффективных решений при согласовании требований к ИС с заказчиком	Выполнение заданий на практических занятиях	Применение знаний об основных тенденциях развития цифровых технологиях к решению заданий на практических занятиях	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Владеть навыками использования современных методов анализа больших данных	Домашнее задание	Демонстрация навыков использования современных методов анализа больших данных	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
ПКв-9 Способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач					
Знать специфику анализа больших данных	Собеседование (зачет)	Корректно применяет знания о специфике анализа больших данных	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)
Уметь формулировать задачи в сфере анализа больших данных	Выполнение заданий на практических занятиях	Корректно формулирует задачи анализа больших данных при выполнении заданий на практических занятиях	Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности	Зачтено/ 60-100	Освоена (повышенный)
			Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины	Не зачтено/ 0-59,99	Не освоена (недостаточный)

<p>Владеть инструментальными средствами обработки статистической информации, а также навыками применения вероятностно - статистического подхода к постановке и решению прикладных задач</p>	<p>Домашнее задание</p>	<p>Демонстрация навыков владения инструментальными средствами обработки статистической информации при решении прикладных задач</p>	<p>Обучающийся твердо знает материал, грамотно и по существу излагает его, но допускает в ответе некоторые неточности</p>	<p>Зачтено/ 60-100</p>	<p>Освоена (повышенный)</p>
			<p>Обучающийся не раскрыл содержание материала, допускает грубые ошибки в формулировках основных понятий дисциплины</p>	<p>Не зачтено/ 0-59,99</p>	<p>Не освоена (недостаточный)</p>