

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

**УТВЕРЖДАЮ**  
Проректор по учебной работе

\_\_\_\_\_ Василенко В.Н.

«25» мая 2023 г.

**РАБОЧАЯ ПРОГРАММА**  
**ДИСЦИПЛИНЫ**

**БОЛЬШИЕ ДАННЫЕ**

Направление подготовки

**09.03.02 Информационные системы и технологии**

---

Направленность (профиль) подготовки

**Разработка информационных систем и технологий**

---

Квалификация выпускника

**Бакалавр**

---

Воронеж

## 1. Цели и задачи дисциплины

Целью дисциплины «Большие данные» является формирование компетенций обучающегося в области профессиональной деятельности и сфере профессиональной деятельности:

*06 Связь, информационные и коммуникационные технологии (в сфере проектирования, разработки, внедрения и эксплуатации средств вычислительной техники и информационных систем, управления их жизненным циклом);*

*40 Сквозные виды профессиональные деятельности в промышленности.*

*(в сфере организации и проведения научно-исследовательских и опытно-конструкторских работ в области информатики и вычислительной техники).*

Дисциплина направлена на решение задач профессиональной деятельности следующих типов:

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

Программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.02 Информационные системы и технологии, утвержденного приказом Министерства образования и науки Российской Федерации от 19.09.2017г. № 926.

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

№ п/п	Компетенции	Формулировка компетенции	Код и наименование индикатора достижения компетенции
1	ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ИД1 <sub>ОПК-2</sub> – Демонстрирует знания принципов работы современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности ИД2 <sub>ОПК-2</sub> – Применяет знания принципов работы современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности ИД3 <sub>ОПК-2</sub> – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности

Код и наименование индикатора достижения компетенции	Результаты обучения (показатели оценивания)
ИД1 <sub>ОПК-2</sub> – Демонстрирует знания принципов работы современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	Знает: современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	Умеет: выбирать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	Владеет: навыками применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности

ИД <sub>20пк-2</sub> – Применяет знания принципов работы современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	Знает: современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности
	Умеет: выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности.
	Владеет: навыками применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности
ИД <sub>30пк-2</sub> – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	Знает: возможности поиска современных информационных средств и программных средств для решения поставленной задачи
	Умеет: определять необходимые программные средства для решения поставленной задачи
	Владеет: навыками использования современных информационных технологии и программные средства при решении поставленной задачи

### 3. Место дисциплины в структуре ООП ВО

Дисциплина «Большие данные» относится к *Обязательной части* Блока 1 ООП, модуль «Введение в информационные системы». Дисциплина является обязательной к изучению.

Изучение дисциплины основано на знаниях, умениях и навыках, полученных обучающимися при изучении дисциплин предметной области по направлению подготовки бакалавров.

Дисциплина является основана на данных дисциплин: Алгоритмы и структуры данных, Планирование и организация производственных процессов

### 4. Объем дисциплины и виды учебных занятий

Общая трудоемкость дисциплины составляет 5 зачетных единицы.

Виды учебной работы	Всего ак. ч.	Распределение трудоемкости по семестрам, ак. ч.
		6 семестр
Общая трудоемкость дисциплины	<b>180</b>	<b>180</b>
<b>Контактная работа в т.ч. аудиторные занятия:</b>	<b>76</b>	<b>76</b>
Лекции	36	36
<i>в том числе в форме практической подготовки</i>	-	-
Практические занятия	36	36
<i>в том числе в форме практической подготовки</i>	-	-
Консультации перед экзаменом	2	2
Консультации текущие	1,8	1,8
Вид аттестации - экзамен	0,2	0,2
<b>Самостоятельная работа:</b>	<b>70,2</b>	<b>70,2</b>
Проработка материалов по учебной литературе (подготовка к собеседованию, тестированию)	20	20
Изучение материалов, изложенных в лекциях (подготовка к собеседованию, тестированию)	22,2	22,2
Домашняя контрольная работа	10	10
Подготовка отчета по практическим занятиям	19	19
Подготовка к экзамену	33,8	33,8

**5 Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий**

**5.1 Содержание разделов дисциплины**

№ п/п	Наименование раздела дисциплины	Содержание раздела	Трудоемкость, ак. ч.
1	Большие данные	Понятие о Big Data. Причины появления больших данных. Примеры возможностей для промышленности. Понятие жизненного цикла аналитики данных. Когнитивный анализ данных. Парадигма Map Reduce. Проекты Apache Hadoop, Apache Spark.	46
2	Машинное обучение	Классификация задач обработки данных. Информативность и выбор признаков. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация.	46
3	Работа с данными	Основы языка R: основные функции, синтаксис, типы данных. Data Mining: инструменты, основные библиотеки. Технологии хранения больших данных. Парадигма NoSQL.	51,2
		Консультации текущие	1,8
		Консультации перед экзаменом	2
		Вид аттестации - экзамен	0,2
		Экзамен - контроль	33,8

**5.2 Разделы дисциплины и виды занятий**

№ п/п	Наименование раздела дисциплины	Лекции, ак. ч.	Практические занятия, ак. ч.	Лабораторные занятия, ак.ч	СРО, ак. ч.
1	Большие данные	12	12		22
2	Машинное обучение	12	12		22
3	Работа с данными	12	12		27,2
	Консультации текущие			1,8	
	Консультации перед экзаменом			2	
	Вид аттестации - экзамен			0,2	
	Экзамен - контроль			33,8	

**5.2.1 Лекции**

№ п/п	Наименование раздела дисциплины	Тематика лекционных занятий	Трудоемкость, ак. ч.
1	Большие данные	Понятие о Big Data. Причины появления больших данных. Примеры возможностей для промышленности. Понятие жизненного цикла аналитики данных. Когнитивный анализ данных. Парадигма Map Reduce. Проекты Apache Hadoop, Apache Spark.	12
2	Машинное обучение	Классификация задач обработки данных. Информативность и выбор признаков. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация.	12
3	Работа с данными	Основы языка R: основные функции, синтаксис, типы данных. Data Mining: инструменты, основные библиотеки. Технологии хранения больших данных. Парадигма NoSQL.	12

\*в форме практической подготовки

**5.2.2 Практические занятия**

№ п/п	Наименование раздела дисциплины	Тематика практических занятий	Трудоемкость, ак. ч.
1	Большие данные	Жизненный цикл аналитики данных. Когнитивный анализ данных. Map Reduce. Apache Hadoop, Apache Spark.	12
2	Машинное обучение	Задачи классификации и кластеризации. K-means и C-means кластеризация, классификация.	12
3	Работа с данными	Основы языка R: основные функции, синтаксис, типы данных языка R. Инструменты, основные библиотеки для работы с данными.	12

\*в форме практической подготовки

### 5.2.3 Лабораторные занятия

Не предусмотрены учебным планом

### 5.2.4 Самостоятельная работа обучающихся (СРО)

№ п/п	Наименование раздела дисциплины	Вид СРО	Трудоемкость, ак. ч.
1	Большие данные	Проработка материалов по учебной литературе (подготовка к собеседованию, тестированию)	6
		Изучение материалов, изложенных в лекциях (подготовка к собеседованию, тестированию)	7
		Домашняя контрольная работа	3
		Подготовка отчета по практическим занятиям	6
2	Машинное обучение	Проработка материалов по учебной литературе (подготовка к собеседованию, тестированию)	6
		Изучение материалов, изложенных в лекциях (подготовка к собеседованию, тестированию)	7
		Домашняя контрольная работа	3
		Подготовка отчета по практическим занятиям	6
3	Работа с данными	Проработка материалов по учебной литературе (подготовка к собеседованию, тестированию)	8
		Изучение материалов, изложенных в лекциях (подготовка к собеседованию, тестированию)	8,2
		Домашняя контрольная работа	4
		Подготовка отчета по практическим занятиям	7

## 6 Учебно-методическое и информационное обеспечение дисциплины

Для освоения дисциплины обучающийся может использовать:

### 6.1 Основная литература

Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие. — Москва : МИСИ – МГСУ, 2020 <https://e.lanbook.com/book/145102>

Юре, Л. Анализ больших наборов данных. — Москва : ДМК Пресс, 2016 <https://e.lanbook.com/book/93571>

MySQL 8 для больших данных / Ш. Чаллавала, Д. Лакхатария, Ч. Мехта, К. Патель ; перевод с английского А. В. Логунова. — Москва : ДМК Пресс, 2018 <https://e.lanbook.com/book/131684>

## 6.2 Дополнительная литература:

1 Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 188 с. — ISBN 978-5-8114-9834-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/199514>

Нурматова, Е. В. Управление большими базами данных и высоконагруженными системами : учебное пособие / Е. В. Нурматова, Р. Ф. Халабия, Л. В. Бунина. — Москва : РТУ МИРЭА, 2019. — 120 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/171496>

методы и модели исследования сложных систем и обработки больших данных : монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ; под редакцией В. А. Смагина и А. Д. Хомоненко. — Санкт-Петербург : Лань, 2020. — 236 с. — ISBN 978-5-8114-4006-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/126938>

## 6.3 Перечень учебно-методического обеспечения для самостоятельной работы студентов

1. Денисенко В.В., Стукало О.Г., Саввина Е.А. Большие данные [Электронный ресурс] : методические указания для СРО для студентов, обучающихся по направлению 09.03.01 – Информатика и вычислительная техника, очной и заочной форм обучения / Денисенко В.В., Стукало О.Г., Саввина Е.А.; ВГУИТ, Кафедра корпоративных информационных систем и программирования. - Воронеж : ВГУИТ, 2022. - 32 с. - Электрон. ресурс. - <http://biblos.vsu.ru/ProtectedView/Book/ViewBook/2581>

## 6.4 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

Наименование ресурса сети «Интернет»	Электронный адрес ресурса
«Российское образование» - федеральный портал	<a href="http://www.edu.ru/index.php">http://www.edu.ru/index.php</a>
Научная электронная библиотека	<a href="http://www.elibrary.ru/defaulttx.asp?">http://www.elibrary.ru/defaulttx.asp?</a>
Федеральная университетская компьютерная сеть России	<a href="http://www.runnet.ru/">http://www.runnet.ru/</a>
Информационная система «Единое окно доступа к образовательным ресурсам»	<a href="http://www.window.edu.ru/">http://www.window.edu.ru/</a>
Электронная библиотека ВГУИТ	<a href="http://biblos.vsu.ru/megapro/web">http://biblos.vsu.ru/megapro/web</a>
Сайт Министерства науки и высшего образования РФ	<a href="http://minobrnauki.gow.ru">http://minobrnauki.gow.ru</a>
Портал открытого on-line образования	<a href="http://npoed.ru">http://npoed.ru</a>
Информационно-коммуникационные технологии в образовании. Система федеральных образовательных порталов	<a href="http://www.ict.edu.ru/">http://www.ict.edu.ru/</a>
Электронная образовательная среда ФГБОУ ВО «ВГУИТ»	<a href="http://education.vsu.ru">http://education.vsu.ru</a>

## 6.5 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения 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 ОС Windows;MS Office.

## 7 Материально-техническое обеспечение дисциплины (модуля)

<p>Аудитории для проведения занятий лекционного типа, лабораторных и практических занятий</p>	<p>Ауд. 420: Комплекты мебели для учебного процесса. ПЭВМ-12 (компьютер Core i5-4460), проектор Acer projector X1383WH, экран, стенды – 5 шт., блок управления комплекса радиоконтроля и поиска радиопередающих устройств «ОМЕГА» (переносной), МУ защиты ресурсов сети от внутренних и внешних атак CISCO ASA5505-KB, переносной комплекс для автоматизации измерений при проведении исследований и контроля технических средств ЭВТ «НАВИГА-ТОР-ПЗГ»; средство активной защиты информации от утечки за счет побочных электромагнитных излучений и наводок «СОНАТА-РЗ.1»; система защиты речевой информации «Соната-АВ-4Б» (Центральный блок питания и управления + Размыкатели в составе СВАЗ Соната АВ); профессиональный обнаружитель скрытых видеокамер СОКОЛ-М (переносной); портативный обнаружитель закладок Protect1203 (переносной); устройство активной защиты информации «ВЕТО-М»; электронный замок Samsung SHS-2920</p>	<p>Microsoft Windows 7 (64 разрядная) Профессиональная Лицензия (DreamSpark); Microsoft Office (standart) 2007 Профессиональная Лицензия (DreamSpark );Microsoft Access 2007 Профессиональная Лицензия (DreamSpark ); Microsoft Project 2007 Профессиональная Лицензия ( DreamSpark); Microsoft Share Point 2007 Профессиональная Лицензия (DreamSpark ); Microsoft Visio 2007 Профессиональная Лицензия ( DreamSpark ) Microsoft SQL server 2008 Профессиональная Лицензия ( DreamSpark); 1 С Предприятия Лицензия; 7-Zip File Manager (архиватор)Бесплатное ПО; Adobe Acrobat Reader (Бесплатное ПО); Adobe Flash Player (Бесплатное ПО); FAR file managerБесплатное ПО; Google ChromeБесплатное ПО; Java TM 7 (64-bit)Бесплатное ПО; K-Lite Codec PackБесплатное ПО; Mozilla FirefoxБесплатное ПО; Oracle VM VirtualBoxБесплатное ПО; Sublime TextБесплатное ПО; Symantec Endpoint Protection 12(Заменен на AVP Kaspersky)Бесплатное ПО; VMWare Player (Бесплатное ПО); Антивирус “Зоркий глаз” (Бесплатное ПО); Lazarus (аналог Delphi)Бесплатное ПО; SmathStudio (аналог Mathcad)Бесплатное ПО; NanoCAD (аналог Autocad)Бесплатное ПО; Gimp (графический редактор аналог Photoshop) Бесплатное ПО; Avidemax (видео редактор)Бесплатное ПО; Virtual Dub (видео редактор)Бесплатное ПО; Free Pascal (Бесплатное ПО); Страж NT вер.3.0 Сертификат ФСТЭК No 2145 30.07.2013 г.; Ревизор 1XP Сертификат ФСТЭК No 989 08.02.2015 г.; Ревизор 2XP Сертификат ФСТЭК No 990 08.02.2015 г.; Фикс 2.0.2 Сертификат ФСТЭК No1548 15.01.2015 г.; Ревизор сети вер.3.0 Сертификат ФСТЭК No3413 02.06.2015 г.; СЗИ Панцирь К Сертификат ФСТЭК No1973 09.12.2015 г.; СЗИ Dallas Lock 8.0 К Сертификат ФСТЭК No2720 25.09.2015; СЗИ Dallas Lock 8.0 С Сертификат ФСТЭК No2945 16.08.2013</p>
<p>Аудитории для проведения занятий лекционного типа, лабораторных и практических занятий</p>	<p>Ауд. 332а: Комплекты мебели для учебного процесса. ПЭВМ – 12 (компьютер Core i5-4570),средство активной защиты информации изделие «Салют 2000С» с регулятором выходного уровня шума, стенды – 5 шт. Ауд. 424: Комплекты мебели для учебного процесса. ПЭВМ – 12: рабочая станция CPU Core 2Duo E6300 – 1.86 – 10 шт, Celeron D2.8 – 2шт.; стенды – 3 Ауд. 420: Комплекты мебели для учебного процесса. ПЭВМ-12 (компьютер Core i5-4460),</p>	<p>Microsoft Windows 7 (64 разрядная) Профессиональная Лицензия ( DreamSpark ); Microsoft Windows 2003 Профессиональная Лицензия ( DreamSpark ); Microsoft Office (standart) 2007 Профессиональная Лицензия ( DreamSpark );Microsoft Access 2007 Профессиональная Лицензия ( DreamSpark ); Microsoft Project 2007 Профессиональная Лицензия ( DreamSpark ); Microsoft Share Point 2007 Профессиональная Лицензия ( DreamSpark ); Microsoft Visio 2007 Профессиональная</p>

	<p>проектор Acer projector X1383WH, экран, стенды – 5 шт., блок управления комплекса радиоконтроля и поиска радиопередающих устройств «ОМЕГА» (переносной), МУ защиты ресурсов сети от внутренних и внешних атак CISCO ASA5505-KB, переносной комплекс для автоматизации измерений при проведении исследований и контроля технических средств ЭВТ «НАВИГАТОР-ПЗГ»; средство активной защиты информации от утечки за счет побочных электромагнитных излучений и наводок «СОНАТА-РЗ.1»; система защиты речевой информации «Соната-АВ-4Б» (Центральный блок питания и управления + Размыкатели в составе СВАЗ Соната АВ); профессиональный обнаружитель скрытых видеокамер СОКОЛ-М (переносной); портативный обнаружитель закладок Protect1203 (переносной); устройство активной защиты информации «ВЕТО-М»; электронный замок Samsung SHS-2920</p>	<p>Лицензия (DreamSpark ) Microsoft SQL server 2008 Профессиональная Лицензия ( DreamSpark ); 1 С Предприятие Лицензия; 7-Zip File Manager (архиватор) Бесплатное ПО; Adobe Acrobat Reader Бесплатное ПО; Adobe Flash Player Бесплатное ПО; FAR file manager Бесплатное ПО; Google Chrome Бесплатное ПО; Java TM 7 (64-bit) Бесплатное ПО; K-Lite Codec Pack Бесплатное ПО; Mozilla Firefox Бесплатное ПО; Oracle VM VirtualBox Бесплатное ПО; Sublime Text Бесплатное ПО; Symantec Endpoint Protection 12 (Заменен на AVP Kaspersky) Бесплатное ПО; VMWare Player Бесплатное ПО; Антивирус “Зоркий глаз” Бесплатное ПО; Lazarus (аналог Delphi) Бесплатное ПО; Smath Studio (аналог Mathcad) Бесплатное ПО; NanoCAD (аналог Autocad) Бесплатное ПО; Gimp (графический редактор аналог Photoshop) Бесплатное ПО; Avidemux (видео редактор) Бесплатное ПО; Virtual Dub (видео редактор) Бесплатное ПО; Free Pascal Бесплатное ПО (ауд.420) Страж NT вер.3.0 Сертификат ФСТЭК No 2145 30.07.2013 г.; Ревизор 1XP Сертификат ФСТЭК No 989 08.02.2015 г.; Ревизор 2XP Сертификат ФСТЭК No 990 08.02.2015 г.; Фикс 2.0.2 Сертификат ФСТЭК No1548 15.01.2015 г.; Ревизор сети вер.3.0 Сертификат ФСТЭК No3413 02.06.2015 г.; СЗИ Панцирь К Сертификат ФСТЭК No1973 09.12.2015 г.; СЗИ Dallas Lock 8.0 К Сертификат ФСТЭК No2720 25.09.2015; СЗИ Dallas Lock 8.0 С Сертификат ФСТЭК No2945 16.08.2013</p>
<p>Аудитории для самостоятельной работы, курсового и дипломного проектирования</p>	<p>Читальные залы библиотеки: Компьютеры со свободным доступом в сеть Интернет и Электронными библиотечными и информационно справочными системами; Ауд.424: Комплекты мебели для учебного процесса. Количество ПЭВМ – 12 (рабочая станция CPU Core 2Duo E6300 – 1.86 – 10 шт, Celeron D2.8 – 2 шт.), стенды – 3</p>	

## 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.**

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

## ПРИЛОЖЕНИЕ к рабочей программе

1. Организационно-методические данные дисциплины для заочной форм обучения

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

Общая трудоёмкость дисциплины составляет 5 зачётных единицы.

Виды учебной работы	Всего ак. ч.	Распределение трудоёмкости по семестрам, ак. ч.
		6 семестр
<b>Общая трудоёмкость дисциплины</b>	<b>180</b>	<b>180</b>
<b>Контактная работа, в т.ч. аудиторные занятия:</b>	<b>11,6</b>	<b>11,6</b>
Лекции	4	4
<i>в том числе в форме практической подготовки</i>	-	-
Практические занятия (ПЗ)	4	4
<i>в том числе в форме практической подготовки</i>	-	-
Консультации текущие	1,4	1,4
Консультации перед экзаменом	2	2
Виды аттестации (экзамен)	0,2	0,2
<b>Самостоятельная работа:</b>	<b>161,6</b>	<b>161,6</b>
Проработка материалов по конспекту лекций (собеседование, тестирование, решение кейс-заданий)	2	2
Проработка материалов по учебникам (собеседование, тестирование, решение кейс-заданий)	145,6	145,6
Выполнение контрольной работы	10	10
Оформление текста контрольной работы	4	4
Подготовка к экзамену	6,8	6,8

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ  
для промежуточной аттестации**

по дисциплине

**Большие данные**

## 1 Перечень компетенций с указанием этапов их формирования

№ п/п	Компетенции	Формулировка компетенции	Код и наименование индикатора достижения компетенции
1	ОПК-2	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности	ИД1 <sub>ОПК-2</sub> – Демонстрирует знания принципов работы современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности ИД2 <sub>ОПК-2</sub> – Применяет знания принципов работы современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности ИД3 <sub>ОПК-2</sub> – Решает стандартные задачи с применением современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности

Код и наименование индикатора достижения компетенции	Результаты обучения (показатели оценивания)
ИД1 <sub>ОПК-2</sub> – Демонстрирует знания принципов работы современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	Знает: современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	Умеет: выбирать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности
	Владеет: навыками применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности

## 2 Паспорт оценочных материалов по дисциплине

№ п/п	Разделы дисциплины	Код и наименование индикатора достижения компетенции	Оценочные средства		Технология/процедура оценивания (способ контроля)
			наименование	№№ заданий	
1	Большие данные	ОПК-2	Банк тестовых заданий	1-19, 49, 54	Бланочное или компьютерное тестирование (уровневая шкала)
			Домашняя работа	71-74	Проверка преподавателем (оценка в системе «зачтено/не зачтено»)
			Подготовка отчета по практическим занятиям	85-86	Проверка преподавателем (оценка в системе «зачтено/не зачтено»)
			Собеседование (вопросы для	57-59	Проверка преподавателем

			экзамена)		(уровневая шкала)
2	Машинное обучение	ОПК-2	Банк тестовых заданий	20-35, 50-51, 55	Бланочное или компьютерное тестирование (уровневая шкала)
			Домашняя работа	75-80	Проверка преподавателем (оценка в системе «зачтено/не зачтено»)
			Подготовка отчета по практическим занятиям	87-88	Проверка преподавателем (оценка в системе «зачтено/не зачтено»)
			Собеседование (вопросы для экзамена)	60-65	Проверка преподавателем (уровневая шкала)
3	Работа с данными	ОПК-2	Банк тестовых заданий	36-48, 52-53, 56	Бланочное или компьютерное тестирование (уровневая шкала)
			Домашняя работа	81-84	Проверка преподавателем (оценка в системе «зачтено/не зачтено»)
			Подготовка отчета по практическим занятиям	89	Проверка преподавателем (оценка в системе «зачтено/не зачтено»)
			Собеседование (вопросы для экзамена)	66-70	Проверка преподавателем (уровневая шкала)

### 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**

Аттестация обучающегося по дисциплине проводится в форме тестирования и предусматривает возможность последующего собеседования (экзамен). Каждый вариант теста включает 25 контрольных заданий, из них:

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

#### 3.1 Тесты (тестовые задания)

**ОПК 2** - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

№	Тестовое задание
---	------------------

задания	
	<i>Выбрать один ответ</i>
1.	<p><b>Выберите верный ответ</b></p> <p>a. большие данные – это обработка или хранение более 1 Тб информации</p> <p><b>b. проблема больших данных – это такая проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна</b></p> <p>c. большие данные – это огромная PR-акция крупных вендоров и не более того</p> <p>d. большие данные – это явление, когда цифровые данные наиболее полно представляют изучаемый объект</p>
2.	<p><b>Выберите НЕверный ответ:</b></p> <p>a. <b>большие данные – это данные объёма свыше 1 Тб</b></p> <p>b. проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна</p> <p>c. большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров</p> <p>d. большие данные как правило не структурированы</p>
3.	<p><b>Hadoop – это:</b></p> <p>a. <b>набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах</b></p> <p>b. распределённая СУБД, позволяющая обрабатывать большие данные</p> <p>c. язык выполнения заданий в парадигме MapReduce</p> <p>d. распределённая файловая система, предназначенная для хранения файлов большого объёма</p>
4.	<p><b>Выберите верное определение</b></p> <p><b>ТХТ-файлы это:</b></p> <p>a. упорядоченный набор пар «ключ: значение»</p> <p><b>b. Последовательность символов, обычно сгруппированных в строки</b></p> <p>c. язык разметки, в котором существуют теги для определения элементов, что затрудняет его чтения</p> <p>d. Текстовый файл, разделитель — точка с запятой. Каждая строка файла — запись данных, запятые разделяют запись на разные поля</p>
5.	<p><b>Выберите верное определение</b></p> <p><b>JSON-файл это-</b></p> <p>a. <b>упорядоченный набор пар «ключ: значение»</b></p> <p>b. Последовательность символов, обычно сгруппированных в строки</p> <p>c. язык разметки, в котором существуют теги для определения элементов, что затрудняет его чтения</p> <p>d. Текстовый файл, разделитель — точка с запятой. Каждая строка файла — запись данных, запятые разделяют запись на разные поля</p>
6.	<p><b>Выберите верное определение</b></p> <p><b>XML формат это -</b></p> <p>a. упорядоченный набор пар «ключ: значение»</p> <p>b. Последовательность символов, обычно сгруппированных в строки</p> <p><b>c. язык разметки, в котором существуют теги для определения элементов, что затрудняет его чтения</b></p> <p>d. Текстовый файл, разделитель — точка с запятой. Каждая строка файла — запись данных, запятые разделяют запись на разные поля</p>
7.	<p><b>Выберите верное определение</b></p> <p><b>CSV-файл это-</b></p> <p>a. упорядоченный набор пар «ключ: значение»</p> <p>b. Последовательность символов, обычно сгруппированных в строки</p> <p>c. язык разметки, в котором существуют теги для определения элементов, что затрудняет его чтения</p> <p><b>d. Текстовый файл, разделитель — точка с запятой.</b></p>
8.	<p><b>Как называется совокупность условий и факторов, создающих потенциальную или реально существующую опасность нарушения безопасности информации?</b></p> <p>a. уязвимость</p> <p>b. слабое место системы</p> <p><b>c. угроза</b></p> <p><b>d. атака</b></p>
9.	<b>В отношении информации, доступ к которой ограничен федеральными законами,</b>

	<p><b>необходимо соблюдать следующее требование:</b></p> <ul style="list-style-type: none"> <li>a. обеспечение доступности</li> <li>b. обеспечение неотказуемости</li> <li><b>c. обеспечение конфиденциальности</b></li> <li>d. обеспечение целостности</li> </ul>
10.	<p><b>Как называется гражданин или юридическое лицо, осуществляющие деятельность по эксплуатации информационной системы, в том числе по обработке информации, содержащейся в ее базах данных?</b></p> <ul style="list-style-type: none"> <li><b>a. оператор информационной системы</b></li> <li>b. обладатель информации</li> <li>c. субъект информации</li> <li>d. обладатель информационной системы</li> </ul>
11.	<p><b>Информация, к которой нельзя ограничить доступ:</b></p> <ul style="list-style-type: none"> <li>a. информация о работе государственных органов</li> <li>b. информация об окружающей среде</li> <li>c. персональные данные субъекта</li> <li>d. информация о здоровье субъекта</li> </ul>
12.	<p><b>Для полнотекстового интеллектуального поиска и аналитики по полуструктурированным данным в формате JSON отлично подходит СУБД</b></p> <ul style="list-style-type: none"> <li>a. HBase</li> <li>b. Cassandra</li> <li>c. Hive</li> <li><b>d. Elasticsearch</b></li> </ul>
13.	<p><b>Для реализации микросервисной архитектуры и интеграции разрозненных систем подходит</b></p> <ul style="list-style-type: none"> <li><b>a. Apache Kafka</b></li> <li>b. Apache Spark</li> <li>c. Apache AirFlow</li> <li>d. Apache <a href="#">Hadoop</a></li> </ul>
14.	<p><b>Apache NiFi используется для</b></p> <ul style="list-style-type: none"> <li>a. визуализации результатов аналитики</li> <li>b. эффективного хранения больших данных</li> <li><b>c. маршрутизации потоков Big Data и построения ETL-конвейеров</b></li> <li>d. оптимизации SQL-запросов к DWH</li> </ul>
15.	<p><b>Повысить производительность Apache Kafka можно с помощью:</b></p> <ul style="list-style-type: none"> <li>a. Увеличения плотности разделов на каждом брокере</li> <li>b. Повышения коэффициента репликации</li> <li>c. Увеличения размера сообщений</li> <li><b>d. Замены HDD-дисков на SSD</b></li> </ul>
16.	<p><b>Автоматизировать запуск пакетных задач в рамках конвейера обработки больших данных по расписанию можно с помощью</b></p> <ul style="list-style-type: none"> <li>a. Apache Hive</li> <li>b. Apache Hadoop</li> <li>c. Apache Kafka</li> <li><b>d. Apache AirFlow</b></li> </ul>
17.	<p><b>Какие вероятные разочарования тренда больших данных?</b></p> <ul style="list-style-type: none"> <li><b>a. из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных</b></li> <li>b. из-за угрозы безопасности личной жизни (privacy) граждан будут упрощены процедуры сбора данных, что приведёт к падению ценности больших данных</li> <li>c. выпуск баз данных с обработкой данных в памяти</li> <li>d. нет</li> </ul>
18.	<p><b>Выберите одно неверное высказывание про MapReduce:</b></p> <ul style="list-style-type: none"> <li>a. интерфейс для массово-параллельной обработки данных, где вычисления производятся на узлах, где информация изначально была сохранена</li> <li>b. MapReduce – это две операции: распределения и сборки данных</li> <li><b>c. MapReduce был придуман разработчиками Hadoop</b></li> <li>d. MapReduce был анонсирован разработчиками Google</li> </ul>
19.	<p><b>Какие из следующих технологий СУБД не используют принцип MapReduce</b></p> <ul style="list-style-type: none"> <li>a. Hadoop</li> <li>b. Cassandra</li> <li>c. HDInsight</li> </ul>

	d. <b>Redis</b>
20.	<b>Какие СУБД полностью полагаются на оперативную память при хранении информации:</b> a. <b>Oracle Exalytics</b> b. <b>SAP HANA</b> c. BigTable d. HBase
21.	<b>В чём преимущество колоночно-ориентированных СУБД?</b> a. они позволяют выполнять более сложные SQL-запросы по сравнению с реляционными СУБД b. <b>они позволяют динамически дополнять содержание записей новыми полями</b> c. они имеют более гибкие возможности аналитики они позволяют эффективно делать межколоночные сравнения
22.	<b>Начиная с каких размеров данных обоснованно применение кластера Hadoop для хранения данных?</b> a. 100Гб b. 1Тб c. <b>100Тб</b> d. 10 Тб
23.	<b>Взаимодействия пользователя с функционалом веб-приложения начинается с</b> a. <b>Ввод адреса в браузер</b> b. Отправка запроса к серверу c. Генерация ответа сервером d. Отображение HTTP ответа
24.	<b>Задаёт параметры, требуемые для конкретного соединения</b> a. expires b. last-modified c. cache-control d. <b>connection</b>
25.	<b>Какой уровень даёт возможность общения устройствам из разных сетей</b> a. канальный b. физический c. транспортный d. <b>сетевой</b>
26.	<b>Для определения новых типов данных используется оператор ...</b> a. <b>typedef</b> b. switch; c. char d. default
27.	<b>Недостатком, какого приложения являются уязвимости в безопасности</b> a. MPA приложения b. PWA приложения c. <b>SPA приложения</b> d. FTPприложения
28.	<b>Модель сети TCP/IP на самом низком имеет уровень</b> a. канальный b. <b>физический</b> c. транспортный d. сетевой
29.	<b>Плотный индекс - это</b> a. <b>запись создается для каждого поискового ключа, оцененного в базе данных.</b> b. запись, которая отображается только для некоторых значений в файле c. индекс который формируется только для архивов d. индекс который архивируется для уменьшения объема
30.	<b>Неплотный (разреженный) индекс - это</b> a. запись создается для каждого поискового ключа b. <b>запись, которая отображается только для некоторых значений в файле</b> c. индекс который формируется только для архивов d. индекс который архивируется для уменьшения объема
31.	<b>Ниже приведена последовательность этапов проекта аналитики в соответствии с CRISP-DM, укажите первый этап.</b> a. моделирование (Modeling) b. внедрение (Deployment)

	<p>c. подготовка данных (Data Preparation)  d. <b>понимание бизнеса (Business understanding)</b></p>
32.	<p><b>На каком из этапов процесса CRISP-DM происходит проверка гипотез?</b>  a. понимание бизнеса (Business understanding)  b. понимание данных (Data Understanding)  c. <b>моделирование (Modeling)</b>  d. оценка (Evaluation)</p>
33.	<p><b>Вы являетесь владельцем и аналитиком в компании из 10 человек, в которой требуется проанализировать продажи за 1 год (1 млн. продаж). Какие из этапов CRISP-DM можно опустить:</b>  a. <b>понимание бизнеса (Business understanding)</b>  b. подготовка данных (Data Preparation)  c. моделирование (Modeling)  d. оценка (Evaluation)</p>
34.	<p><b>Клиент покупает билет на самолет через интернет. В момент покупки, он хочет знать насколько может упасть стоимость этого билета в ближайшем будущем и когда. К какому типу относится эта задача анализа данных?</b>  a. <b>прогнозирование</b>  b. кластеризация  c. классификация  d. цензурирование</p>
35.	<p><b>Ниже приведена последовательность этапов проекта аналитики в соответствии с CRISP-DM, укажите первый этап.</b>  a. моделирование (Modeling)  b. внедрение (Deployment)  c. подготовка данных (Data Preparation)  d. <b>понимание бизнеса (Business understanding)</b></p>
36.	<p><b>Для машинного обучения подходят данные</b>  a. Любых форматов в цифровом виде  b. Числовые типа int  c. Бинарные  d. <b>Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов</b></p>
37.	<p><b>Для распределенного глубокого машинного обучения (Deep Learning) больше подходит фреймворк</b>  a. TensorFlow  b. Flask  c. <b>PyTorch</b>  d. Scikit-learn</p>
	<i>Выбрать несколько ответов</i>
38.	<p><i>Множественный выбор (3 верных варианта ответов)</i>  <b>Свойства больших данных:</b>  a. <b>объем</b>  b. <b>скорость</b>  c. <b>разнообразие</b>  d. структурированность  e. доступность</p>
39.	<p><i>Выберите несколько вариантов ответов. Множественный выбор (4 верных варианта ответов)</i>  <b>Наиболее часто используемые форматы для представления больших данных</b>  a. csv,  b. <b>xlsx</b>  c. <b>txt</b>  d. <b>json</b>  e. <b>XML</b>  f. HTML  g. jpg, gif  h. docx</p>
40.	<p><i>Множественный выбор (3 верных варианта ответов)</i>  <b>Какие из следующих средств разумно использовать для анализа данных, представленных единственным csv-файлом размера более 100Гб:</b>  a. <b>Hadoop</b></p>

	b. Data Warehouse c. «Песочница» d. Python								
41.	<b>Множественный выбор (2 верных варианта ответов)</b> <b>Для чего аналитику необходима «песочница»?</b> a. для высокопроизводительной аналитики за счёт использования оперативной памяти и inDB операций b. для хранения всех полученных от заказчика данных c. для построения отчётов о результатах анализа d. для снижения затрат, связанных с репликацией данных								
42.	<b>Выберите верные утверждения. Множественный выбор. (3 верных варианта ответов)</b> <b>Преимущества индексации:</b> a. помогает сократить общее количество операций ввода-вывода, необходимых для извлечения этих данных b. предлагает более быстрый поиск и поиск данных для пользователей. c. помогает сократить табличное пространство d. вы не можете выполнять какие-либо другие индексы для проиндексированных данных. вам не разрешено разбивать организованную по индексу таблицу								
43.	<b>Принцип MapReduce состоит в том, чтобы</b> a. производить вычисления на узлах, где информация изначально была сохранена b. использовать вычислительные мощности систем хранения c. использовать функциональное программирование для решения задач массивно-параллельной обработки d. нет правильного ответа								
	<i>Вопрос на сопоставление</i>								
44.	Расставьте названия основных технологических этапов анализа данных (Постановка задачи, Ввод данных в обработку, Качественный анализ, Количественное описание данных, Интерпретация результатов): <b>Этап 1. - Постановка задачи</b> <b>Этап 2. - Ввод данных в обработку</b> <b>Этап 3. - Качественный анализ</b> <b>Этап 4. - Количественное описание данных</b> <b>Этап 5. - Интерпретация результатов</b>								
45.	Расставьте верное определение типов шкал измерения и их характеристики: <b>Интервальная шкала - Наряду с классификацией и упорядочиванием данных позволяет количественно оценить различие между классами. При проведении сравнений вводится единица измерения и произвольное начало отсчета («нуль-пункт»).</b> <b>Шкала отношений - Отличается от интервальной шкалы тем, что в ней задано абсолютное начало отсчета. Можно определить, во сколько раз одно измерение превосходит другое.</b>								
46.	Распределите задачи при использовании методов OLAP и Data Mining: <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 50%; text-align: center;">OLAP</th> <th style="width: 50%; text-align: center;">Data Mining</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">Каковы средние показатели травматизма для курящих и некурящих?</td> <td style="text-align: center;">Встречаются ли точные шаблоны в описаниях людей, подверженных повышенному травматизму?</td> </tr> <tr> <td style="text-align: center;">Каковы средние размеры телефонных счетов существующих клиентов в сравнении со счетами бывших клиентов отказавшихся от услуг телефонной компании?</td> <td style="text-align: center;">Имеются ли характерные портреты клиентов, которые, по всей вероятности, собираются отказаться от услуг телефонной компании?</td> </tr> <tr> <td style="text-align: center;">Какова средняя величина ежедневных покупок по украденной и не украденной кредитной карточке?</td> <td style="text-align: center;">Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?</td> </tr> </tbody> </table>	OLAP	Data Mining	Каковы средние показатели травматизма для курящих и некурящих?	Встречаются ли точные шаблоны в описаниях людей, подверженных повышенному травматизму?	Каковы средние размеры телефонных счетов существующих клиентов в сравнении со счетами бывших клиентов отказавшихся от услуг телефонной компании?	Имеются ли характерные портреты клиентов, которые, по всей вероятности, собираются отказаться от услуг телефонной компании?	Какова средняя величина ежедневных покупок по украденной и не украденной кредитной карточке?	Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?
OLAP	Data Mining								
Каковы средние показатели травматизма для курящих и некурящих?	Встречаются ли точные шаблоны в описаниях людей, подверженных повышенному травматизму?								
Каковы средние размеры телефонных счетов существующих клиентов в сравнении со счетами бывших клиентов отказавшихся от услуг телефонной компании?	Имеются ли характерные портреты клиентов, которые, по всей вероятности, собираются отказаться от услуг телефонной компании?								
Какова средняя величина ежедневных покупок по украденной и не украденной кредитной карточке?	Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?								
	<i>Вставить пропущенное слово или число</i>								
47.	<b>Вставьте пропущенное слово большими буквами в нужном падеже</b> _____ данные это разнообразные данные, поступающие с более высокой скоростью, объем которых постоянно растет. Таким образом, три основных свойства больших данных – это разнообразие, высокая скорость поступления и большой объем. <b>Ответ: БОЛЬШИЕ</b>								
48.	<b>Вставьте пропущенное слово большими буквами в нужном падеже</b>								

— вспомогательная структура, связанная с файлом и предназначенная для поиска информации по тому же принципу, что и в книге с предметным указателем  
**Ответ: ИНДЕКС**

*Задачи на 1-2 действия*

49. Рассмотрим результаты тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров направления «Менеджмент» 2 курса ЗФО [31], [33] (см. таблицу 2). В таблице 2.2 для каждого объекта (студенты) представлены значения свойств: «Пред. тестов», «Вып. тестов», «Затрачено (мин.)», «Попытки», «Дата тестирования», «Оценка» и т.п. В таблицах из 3-4 строк объекты можно перечислять в произвольном порядке, но если объектов много, то располагать их необходимо в осмысленном порядке (т.е. согласно некоторому правилу). Например, в таблице 1 можно перечислить в алфавитном порядке ФИО бакалавров, по возрастанию или убыванию оценок, времени или попыток тестирования и т.п. Таблица 2.3 представляет «поворот на бок» таблицы 2.

Таблица 1 - Результаты тестирования по дисциплине «Теория вероятностей и математическая статистика» БМ 2 курс ЗФО

№ п/п	Код	ФИО	Пред. тестов	Вып. тестов	Затрачено (мин.)	Попытки	Дата тестирования	Оценка
1.	11123	Смык Инна Владимировна	15	13	34	3	11.04.2013	отл
2.	11124	...	0	0	0	0	...	...
3.	11128	Иванов Петр Сергеевич	15	9	38	1	13.04.2013	уд
...	...	...	...	...	...	...	...	...

Таблица 2 - Незульаты тестирования по дисциплине «Теория вероятностей и математическая статистика» БМ 2 курс ЗФО

Код	11123	11124	11128
ФИО	Смык Инна Владимировна	...	Иванов Петр Сергее
Предъявлено тестов	15	0	15
Выполнено тестов	13	0	9
Затрачено (мин.)	34	0	38
Попытки	3	0	1
Дата тестирования	11.04.2013	...	13.04.2013
Оценка	отл.	...	уд

Таблица типа «**объекты - объекты - один**», представляет собой таблицу, содержащую информацию о некотором одном свойстве пар объектов, чаще всего принадлежащих разным классам (см. таблицу 3).

Таблица 3- Шаблон таблиц типа «объекты - объекты - один»

Имя 1-го класса объектов	Имя 2-го класса объектов			
	Имя 1-го объекта 2-го класса	Имя 2-го объекта 2-го класса	...	Имя i-го объекта класса
Имя 1-го объекта 1-го класса	*	*	...	*
Имя 2-го объекта 1-го класса	*	*	...	*
...	...	...	...	...
Имя n-го объекта 1-го класса	*	*	...	*

\* - поля заполняются значениями свойства пары объектов

В таблице 3 верхний заголовок (головка) имеет сложную (двухъярусную) структуру. Данная таблица может быть «повернута на бок» - строки превращены в столбцы, а столбцы - в строки. В таблице данного типа фиксируется одно свойство пары объектов, поэтому в ее ячейках всегда содержатся значения одного типа: или числа, или слова, или графические изображения.

50. Рассмотрим итоги семестрового тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров 2 курса ЗФО. В таблице 1 приведены итоги данного тестирования головка таблицы представлена двухъярусной структурой: первый ярус - имя 2-го класса объектов «количество оценок», второй - наименование самих оценок (имена объектов 2-го класса). Боковик таблицы представлен объектами 1-го класса «Направление подготовки». Поля таблицы заполнены - количеством баллов по паре объектов.

	<p>Таблица 1 - Итоги тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров 2 курса ЗФО</p> <table border="1"> <thead> <tr> <th rowspan="2">Направление обучения</th> <th colspan="5">Количество оценок</th> </tr> <tr> <th>неуд.</th> <th>уд.</th> <th>хорошо</th> <th>отлично</th> <th>не</th> </tr> </thead> <tbody> <tr> <td>Бакалавр экономика</td> <td>0</td> <td>21</td> <td>36</td> <td>10</td> <td></td> </tr> <tr> <td>Бакалавр менеджмент</td> <td>0</td> <td>12</td> <td>18</td> <td>7</td> <td></td> </tr> <tr> <td>Бакалавр бизнес-информатика</td> <td>0</td> <td>4</td> <td>7</td> <td>4</td> <td></td> </tr> </tbody> </table> <p>Таблица 2 представляет «поворот на бок» таблицы 2, в которой приведены итоги тестирования по дисциплине «Теория вероятностей и математическая статистика» головка таблицы представлена двухъярусной структурой: первый ярус - имя 2-го класса объектов «направление обучения», второй - наименование самих направлений (имена объектов 2-го класса).</p> <p>Таблица 2 - Итоги тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров 2 курса ЗФО</p> <table border="1"> <thead> <tr> <th rowspan="2">Количество оценок</th> <th colspan="3">Направление обучения</th> </tr> <tr> <th>Бакалавр экономика</th> <th>Бакалавр менеджмент</th> <th>Бакалавр бизнес-информа</th> </tr> </thead> <tbody> <tr> <td>неуд.</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>уд.</td> <td>21</td> <td>12</td> <td>4</td> </tr> <tr> <td>хорошо</td> <td>36</td> <td>18</td> <td>7</td> </tr> <tr> <td>отлично</td> <td>10</td> <td>7</td> <td>4</td> </tr> <tr> <td>неявка</td> <td>1</td> <td>1</td> <td>0</td> </tr> </tbody> </table> <p>Боковик таблицы представлен объектами 1-го класса «Количество оценок». Поля таблицы заполнены - количеством баллов по паре объектов.</p>	Направление обучения	Количество оценок					неуд.	уд.	хорошо	отлично	не	Бакалавр экономика	0	21	36	10		Бакалавр менеджмент	0	12	18	7		Бакалавр бизнес-информатика	0	4	7	4		Количество оценок	Направление обучения			Бакалавр экономика	Бакалавр менеджмент	Бакалавр бизнес-информа	неуд.	0	0	0	уд.	21	12	4	хорошо	36	18	7	отлично	10	7	4	неявка	1	1	0
Направление обучения	Количество оценок																																																								
	неуд.	уд.	хорошо	отлично	не																																																				
Бакалавр экономика	0	21	36	10																																																					
Бакалавр менеджмент	0	12	18	7																																																					
Бакалавр бизнес-информатика	0	4	7	4																																																					
Количество оценок	Направление обучения																																																								
	Бакалавр экономика	Бакалавр менеджмент	Бакалавр бизнес-информа																																																						
неуд.	0	0	0																																																						
уд.	21	12	4																																																						
хорошо	36	18	7																																																						
отлично	10	7	4																																																						
неявка	1	1	0																																																						
51.	<p>Рассмотрим итоги семестрового тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров 2 курса ЗФО. В таблице 1 приведены итоги данного тестирования головка таблицы представлена двухъярусной структурой: первый ярус - имя 2-го класса объектов «количество оценок», второй - наименование самих оценок (имена объектов 2-го класса). Боковик таблицы представлен объектами 1-го класса «Направление подготовки». Поля таблицы заполнены - количеством баллов по паре объектов</p> <p style="text-align: right;">Таблица 1</p> <p style="text-align: center;">Итоги тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров 2 курса ЗФО</p> <table border="1"> <thead> <tr> <th rowspan="2">Направление обучения</th> <th colspan="5">Количество оценок</th> </tr> <tr> <th>неуд.</th> <th>уд.</th> <th>хорошо</th> <th>отлично</th> <th>не</th> </tr> </thead> <tbody> <tr> <td>Бакалавр экономика</td> <td>0</td> <td>21</td> <td>36</td> <td>10</td> <td></td> </tr> <tr> <td>Бакалавр менеджмент</td> <td>0</td> <td>12</td> <td>18</td> <td>7</td> <td></td> </tr> <tr> <td>Бакалавр бизнес-информатика</td> <td>0</td> <td>4</td> <td>7</td> <td>4</td> <td></td> </tr> </tbody> </table> <p>Таблица 2 представляет «поворот на бок» таблицы 1, в которой приведены итоги тестирования по дисциплине «Теория вероятностей и математическая статистика» головка таблицы представлена двухъярусной структурой: первый ярус - имя 2-го класса объектов «направление обучения», второй - наименование самих направлений (имена объектов 2-го класса).</p> <p>Таблица 2 - Итоги тестирования по дисциплине «Теория вероятностей и математическая статистика» бакалавров 2 курса ЗФО</p> <table border="1"> <thead> <tr> <th rowspan="2">Количество оценок</th> <th colspan="3">Направление обучения</th> </tr> <tr> <th>Бакалавр экономика</th> <th>Бакалавр менеджмент</th> <th>Бакалавр бизнес-информа</th> </tr> </thead> <tbody> <tr> <td>неуд.</td> <td>0</td> <td>0</td> <td>0</td> </tr> <tr> <td>уд.</td> <td>21</td> <td>12</td> <td>4</td> </tr> <tr> <td>хорошо</td> <td>36</td> <td>18</td> <td>7</td> </tr> <tr> <td>отлично</td> <td>10</td> <td>7</td> <td>4</td> </tr> <tr> <td>неявка</td> <td>1</td> <td>1</td> <td>0</td> </tr> </tbody> </table> <p>Боковик таблицы представлен объектами 1-го класса «Количество оценок». Поля таблицы заполнены - количеством баллов по паре объектов.</p>	Направление обучения	Количество оценок					неуд.	уд.	хорошо	отлично	не	Бакалавр экономика	0	21	36	10		Бакалавр менеджмент	0	12	18	7		Бакалавр бизнес-информатика	0	4	7	4		Количество оценок	Направление обучения			Бакалавр экономика	Бакалавр менеджмент	Бакалавр бизнес-информа	неуд.	0	0	0	уд.	21	12	4	хорошо	36	18	7	отлично	10	7	4	неявка	1	1	0
Направление обучения	Количество оценок																																																								
	неуд.	уд.	хорошо	отлично	не																																																				
Бакалавр экономика	0	21	36	10																																																					
Бакалавр менеджмент	0	12	18	7																																																					
Бакалавр бизнес-информатика	0	4	7	4																																																					
Количество оценок	Направление обучения																																																								
	Бакалавр экономика	Бакалавр менеджмент	Бакалавр бизнес-информа																																																						
неуд.	0	0	0																																																						
уд.	21	12	4																																																						
хорошо	36	18	7																																																						
отлично	10	7	4																																																						
неявка	1	1	0																																																						
52.	<p>По отдельным показателям социально - экономического положения Воронежской области за период с 2018г. по 2022г. (см. таблицу 1), необходимо: 1) при использовании процедур кластерного анализа, сформировать группы показателей с аналогичной динамикой за рассматриваемый период; 2) дать содержательную интерпретацию полученных групп; 3) на основе результатов кластеризации и корреляционного анализа, осуществить выбор факторов, наиболее подходящих для анализа среднемесячных денежных доходов населения (руб. на душу населения); 4) подобрать уравнение регрессии, позволяющее получить достоверный прогноз среднемесячных денежных доходов населения; 5) оценить прогнозные значения</p>																																																								

среднемесячных денежных доходов населения на 2023-2024 гг.

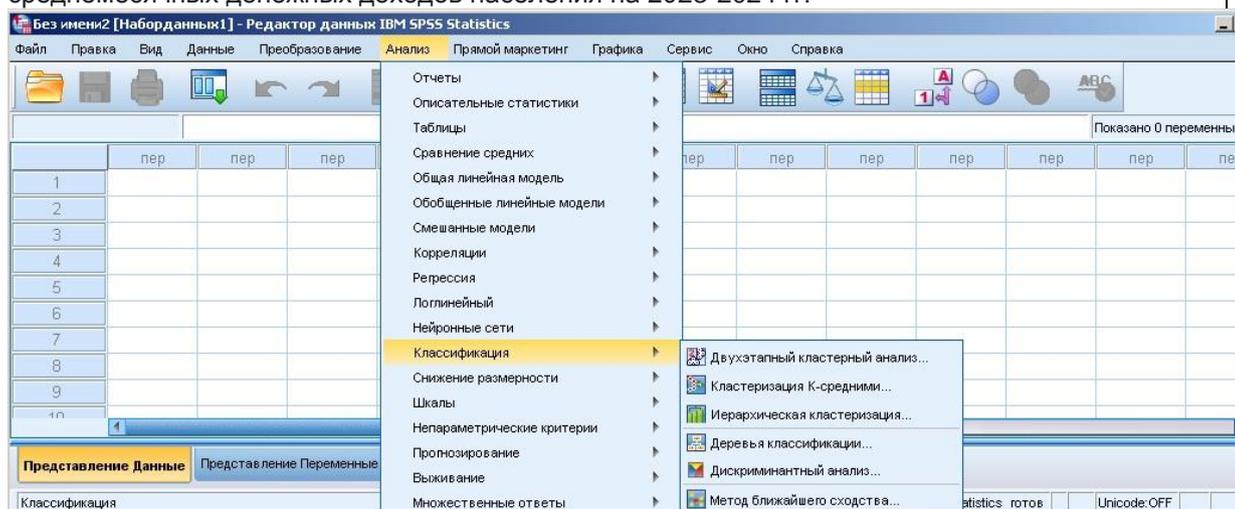


Рис. 1. Меню SPSS «Анализ» подгруппа «Классификация»

Таблица 1

Показатели социально-экономического развития Воронежской области за период 2018-2022 гг.

№ п/п	Показатель развития	2018	2019	2020	2021	2022
1	2	3	4	5	6	7
1	Объем промышленного производства (ОПП), млн. руб.	82820,4	86547,3	93384,5	120839,6	128573,3
2	Инвестиции в основной капитал (ИвОК), млн. руб.	94800	123000	143000	170000	216775,1
3	Оборот розничной торговли (ОРТ), млн. руб.	168700	193700	247200	323000	368600
4	Объем выпуска сельскохозяйственной продукции (ОВСХП), млн. руб.	35549,95	25838,4	42659,2	44664,18	48728,62
5	Доходы консолидированного бюджета (ДКБ), млн. руб.	59451,3	71235,3	77324,1	87607,5	91499,7
6	Расходы консолидированного бюджета (РКБ), млн. руб.	61507,8	72948,6	79522,7	89179,6	100209,5
7	Количество крупных и средних предприятий и организаций (ККСПиО)	1028	929	839	827	824
8	Сальдированный финансовый результат деятельности предприятий и организаций (СФРДПиО), млн. руб.	7120,7	5932,1	15027,7	25594	17672,1
9	Численность официально зарегистрированных безработных (ЧОЗБ), тыс. чел.	28,1	19,8	18	13	11,1
10	Среднемесячные денежные доходы населения (СрДДН), руб. на душу населения	11652,6	13111,3	14837,5	18876,4	21861,1

Решение. 1) **Описание исходных данных в SPSS:** В редакторе данных SPSS (рабочий лист «Представление Данные» см. рисунок 2), каждый столбец соответствует отдельной переменной, а каждая строка - отдельному наблюдению (т.е. в каждой ячейке хранится одно наблюдаемое значение переменной).

	Показатель	r2009	r2010	r2011	r2012	r2013	пер	пер	пер	пер	пер	пер
1	ОПП	82820,40	86547,30	93384,50	120839,60	128573,30						
2	ИвОК	94800,00	123000,00	143000,00	170000,00	216775,10						
3	ОРТ	168700,00	193700,00	247200,00	323000,00	368600,00						
4	ОВСХП	35549,95	25838,40	42659,20	44664,18	48728,62						
5	ДКБ	59451,30	71235,30	77324,10	87607,50	91499,70						
6	РКБ	61507,80	72948,60	79522,70	89179,60	100209,50						
7	ККСПиО	1028,00	929,00	839,00	827,00	824,00						
8	СФРДПиО	7120,70	5932,10	15027,70	25594,00	17672,10						
9	ЧОЗБ	28,10	19,80	18,00	13,00	11,10						
10	СрдДН	11652,60	13111,30	14837,50	18876,40	21861,10						
11												

Рис. 2. Представление исходных данных в SPSS Statistics версии 22

Переменные в SPSS определяются во вкладке «Представление Переменные» в левом нижнем углу рабочего листа «Представление Данные» (см. рисунок 3).

	Имя	Тип	Ширина	Знаков ...	Метка	Значения	Пропущенн...	Столбцы	Выравнивание	Мера	Роль
1	Показатель	Текстовая	10	0		Нет	Нет	12	По левом...	Номинальная	Входная
2	r2009	Числовой	8	2		Нет	Нет	8	По право...	Нет данных	Входная
3	r2010	Числовой	8	2		Нет	Нет	8	По право...	Нет данных	Входная
4	r2011	Числовой	8	2		Нет	Нет	8	По право...	Нет данных	Входная
5	r2012	Числовой	8	2		Нет	Нет	8	По право...	Нет данных	Входная
6	r2013	Числовой	8	2		Нет	Нет	8	По право...	Нет данных	Входная

Рис. 3. Окно работы с переменными в SPSS Statistics версии 22

В данной вкладке последовательно, построчно описываются переменные, при выборе имени которых используются следующие правила: имена переменных должны содержать символы (буквы) латинского алфавита, цифры и некоторые специальные символы (без пробелов); длина имени не должна превышать 8 символов; имя переменной обязательно должно начинаться с буквы; последний символ имени не должен быть точкой или знаком подчеркивания.

53. По имеющимся данным деятельности крупнейших банков России на 01.11.2023. (см. таблицу 3) необходимо: 1) выявить в структуре исходных данных скрытые факторы и дать их содержательную интерпретацию; 2) представить результаты вычислений в графическом виде; 3) упорядочить структуру данных в соответствии со значениями выявленных факторов.

Таблица 3

Характеристика крупнейших банков России на 01.11.2023 (млрд. руб.)

Название банка	Активы нетто	Высоколик в. активы	Депозитный портфель		Кредитный портфель		Чистая прибыль	Пр за д к п
			Депозиты физлиц.	Депозиты юрлиц.	Кредиты физлиц.	Кредиты юрлиц.		
1	2	3	4	5	6	7	8	
Сбербанк	18697,50	1235,86	7917,92	3930,85	3918,32	9153,27	286,69	
Альфа-Банк	1852,17	131,78	386,42	575,98	274,85	1025,86	42,35	
Райффайзенбанк	728,06	34,84	262,70	185,48	196,64	344,61	16,06	
Газпромбанк	4008,23	243,40	381,18	2311,58	291,14	2323,97	29,33	
ВТБ 24	2426,47	99,75	1409,57	273,50	1363,39	256,92	16,43	
Фольксваген Банк РУС	37,42	1,88	0,00	7,02	23,58	9,04	2,43	
Транскапиталбанк	155,52	12,15	36,89	56,92	27,68	88,36	1,77	
Совкомбанк	179,89	8,70	88,81	10,67	70,30	3,25	2,71	

Юникредит Банк	1032,90	32,02	81,77	436,89	158,54	526,27	8,96	
ХКФ Банк (Хоум Кредит Банк)	349,98	9,77	197,66	37,44	251,93	21,29	1,61	
Ситибанк	339,77	35,50	80,12	165,51	51,10	86,65	4,85	
Горбанк	3,92	0,22	0,95	0,27	0,01	3,05	0,61	
ФК Открытие (бывший Номос-Банк)	1131,77	31,30	104,57	463,29	23,04	548,50	7,80	
ВТБ	6543,38	138,31	18,11	2442,62	0,19	2761,68	73,19	1.
Московский Кредитный Банк	504,24	22,10	146,88	178,99	120,72	242,22	5,82	
Национальный Клиринговый Центр	967,92	447,25	0,00	111,12	0,00	101,56	7,03	
Банк Москвы	2053,89	72,20	242,96	609,40	169,70	892,03	8,65	2.
Нордеа Банк	332,01	36,14	9,35	47,54	23,52	211,99	3,90	
Тинькофф. Кредитные Системы	121,27	6,40	40,88	19,91	94,26	2,09	2,51	
Ханты-Мансийский Банк Открытие	353,19	36,41	69,37	120,55	90,99	127,49	0,43	

Решение. 1) **Выявление скрытых факторов и проведение классификации.** Воспользуемся факторным анализом, для выявления скрытых факторов и проведения классификации банков. Формирование матрицы наблюдений, описывающий данные измерений 20 объектов по 8 соответствующим переменным  $x_1, \dots, x_8$ , поводится при группировке показателей по колонкам. Для описания данных используется редактор данных SPSS Statistics версии 22 (см. рисунок 4).

\*Без имени2 [Набор данных:1] - Редактор данных IBM SPSS Statistics

Показано 9 переменных

	banki	aktiv_net	vlic_aktiv	depoz_fl	depoz_yl	kred_fl	kred_yl	ch_prib	pr_zad_
1	Сбербанк России	18697,50	1235,86	7917,92	3930,85	3918,32	9153,27	286,69	3.
2	Альфа-Банк	1852,17	131,78	386,42	575,98	274,85	1025,86	42,35	6
3	Райффайзенбанк	728,06	34,84	262,70	185,48	196,64	344,61	16,06	1.
4	Газпромбанк	4008,23	243,40	381,18	2311,58	291,14	2323,97	29,33	2.
5	ВТБ 24	2426,47	99,75	1409,57	273,50	1363,39	256,92	16,43	11.
6	Фольксваген Банк РУС	37,42	1,88	,00	7,02	23,58	9,04	2,43	
7	Транскапиталбанк	155,52	12,15	36,89	56,92	27,68	88,36	1,77	
8	Совкомбанк	179,89	8,70	88,81	10,67	70,30	3,25	2,71	
9	Юникредит Банк	1032,90	32,02	81,77	436,89	158,54	526,27	8,96	1.
10	ХКФ Банк (Хоум Кредит Банк)	349,98	9,77	197,66	37,44	251,93	21,29	1,61	4.
11	Ситибанк	339,77	35,50	80,12	165,51	51,10	86,65	4,85	
12	Горбанк	3,92	,22	,95	,27	,01	3,05	,61	
13	ФК Открытие (бывший Номос-Банк)	1131,77	31,30	104,57	463,29	23,04	548,50	7,80	2.
14	Московский Кредитный Банк	504,24	22,10	146,88	178,99	120,72	242,22	5,82	
15	Национальный Клиринговый Центр	967,92	447,25	,00	111,12	,00	101,56	7,03	
16	ВТБ	6543,38	138,31	18,11	2442,62	,19	2761,68	73,19	12.
17	Банк Москвы	2053,89	72,20	242,96	609,40	169,70	892,03	8,65	22.
18	Нордеа Банк	332,01	36,14	9,35	47,54	23,52	211,99	3,90	
19	ТКС Банк (Тинькофф Кредитные Системы)	121,27	6,40	40,88	19,91	94,26	2,09	2,51	1.
20	Ханты-Мансийский Банк Открытие	353,19	36,41	69,37	120,55	90,99	127,49	,43	1.

Представление Данные Представление Переменные

Процессор IBM SPSS Statistics готов Unicode: OFF

Рис. 4. Исходные данные по банкам в SPSS Statistics версии 22

Описание переменных в SPSS представлено на рисунке 5 (имя и тип переменной, форматы ввода и вывода переменных).

	Имя	Тип	Ширина	Знаков ...	Метка	Значения	Пропущен...	Столбцы	Выравнивание	Мера
1	banki	Текстовая	37	0	Название банка	Нет	Нет	28	По левом...	Номинальная
2	aktiv_net	Числовой	8	2	Активы нетто	Нет	Нет	8	По право...	Шкалы
3	vlic_aktiv	Числовой	8	2	Высоколиквидные акти...	Нет	Нет	8	По право...	Шкалы
4	depoz_fl	Числовой	8	2	Депозиты физлиц	Нет	Нет	8	По право...	Шкалы
5	depoz_yl	Числовой	8	2	Депозиты юрлиц	Нет	Нет	8	По право...	Шкалы
6	kred_fl	Числовой	8	2	Кредиты физлиц	Нет	Нет	8	По право...	Шкалы
7	kred_yl	Числовой	8	2	Кредиты юрлиц	Нет	Нет	8	По право...	Шкалы
8	ch_prib	Числовой	8	2	Чистая прибыль	Нет	Нет	8	По право...	Шкалы
9	pr_zad_krp	Числовой	8	2	Просроченная задолжн...	Нет	Нет	8	По право...	Шкалы

Рис. 5. Описание переменных по банкам в SPSS Statistics версии 22

Для проведения вычислений в меню «Анализ» - «Снижение размерности» выбирается «Факторный анализ» (см. рисунок 6).

21 : banki

	banki	fl	depoz_yl	kred_fl	kred_yl	ch_prib	vlog_v_cb	пер
1	Сбербанк России	7,92	3930,85	3918,32	9153,27	286,69	1843,84	
2	Альфа-Банк	6,42	575,98	274,85	1025,86	42,35	219,14	
3	Райффайзенбанк	2,70	185,48	196,64	344,61	16,06	46,73	
4	Газпромбанк	1,18	2311,58	291,14	2323,97	29,33	427,99	
5	ВТБ 24	9,57	273,60	1363,39	256,92	16,43	162,07	
6	Фольксваген Банк РУС	,00	7,02	23,58	9,04	2,43	,00	
7	Транскапиталбанк	6,89	56,92	27,68	88,36	1,77	17,18	
8	Совкомбанк	9,04	19,67	70,30	3,25	2,71	66,42	

Рис. 6. Меню «Анализ» - «Снижение размерности»

В открывшемся диалоговом окне «Факторный анализ» (см. рисунок 7), отмечаются все переменные слева и переносятся в окно справа «Переменные».

Факторный анализ

Переменные:

Название банка [banki]

Высоколиквидные акт...  
 Депозиты физлиц. [depo...  
 Депозиты юрлиц. [depo...  
 Кредиты физлиц. [kred...  
 Кредиты юрлиц. [kred\_yl]  
 Чистая прибыль [ch\_prib]  
 Просроченная задолжн...

Описательные...  
 Извлечение...  
 Вращение...  
 Значения факторов...  
 Параметры...

Переменная отбора наблюдений:

Значение...

OK Вставка Сброс Отмена Справка

Рис. 7. Окно «Факторный анализ»

Далее щелчком по кнопке «Описательные...», открывается окно «Факторный анализ: Описательные» (см. рисунок 8), в котором остается по умолчанию вывод первичных результатов, включающий: первичные относительные дисперсии простых факторов, собственные значения и процентные доли объясненной дисперсии. Дополнительно при выборе в пункте «начальное решение» можно вывести одномерные статистики, а также корреляционную матрицу (см. рисунок 8).

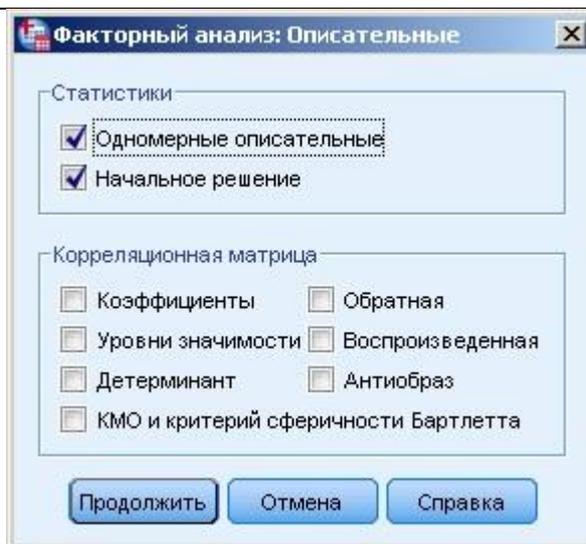


Рис. 8. Окно «Факторный анализ: Описательные»

При использовании кнопки «Извлечение...» (см. рисунок 7) в появившемся окне «**Факторный анализ: Выделение факторов**» (см. рисунок 9) можно выбрать различные методы отбора факторов. Наиболее распространенным является предлагаемый по умолчанию метод главных компонент (см. рисунок 3.28), при котором количество отобранных факторов в этом случае приравнивается к числу собственных значений, превосходящих единицу.

Также дополнительно можно указать нужное количество факторов, вывести график собственных значений. Так повернутое факторное решение предоставляет малозначимую информацию и его можно предотвратить, выбрав опцию «**Неповернутое решение**».

Кнопкой «Вращение...» (см. рисунок 7), вызывается окно «**Факторный анализ: Вращение**», позволяющее выбрать метод вращения (см. рисунок 10). **Ортогональным** называется вращение, при котором сохраняется прямоугольная система координат. Существуют следующие основные **методы вращения**: варимакс; квартимакс; эквимакс; прямой облимин; промакс.

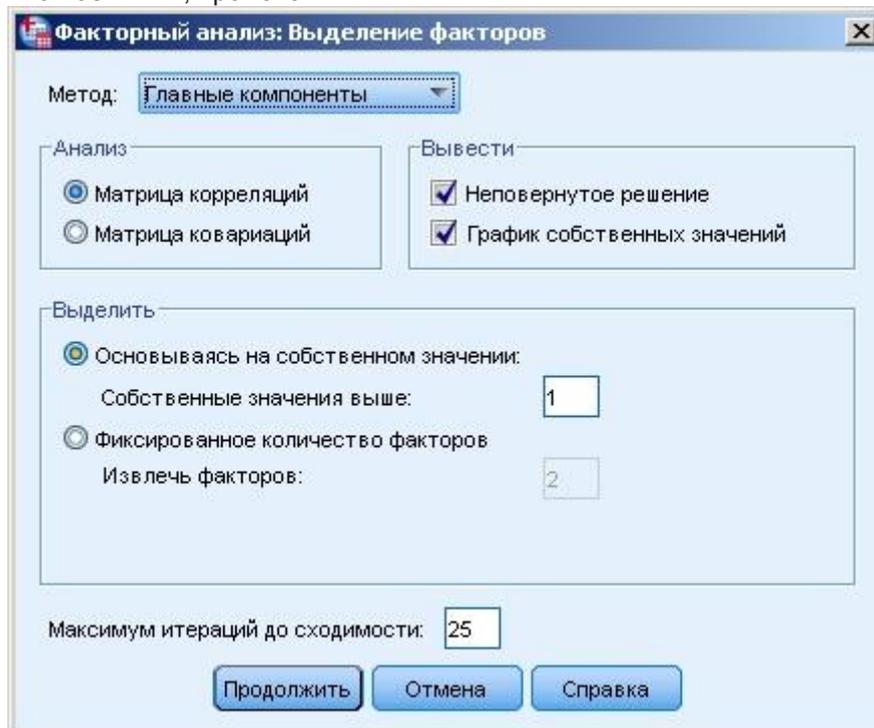


Рис. 9. Окно «Факторный анализ: Выделение факторов»

В нашем случае активизируется метод варимакс (см. рисунок 10) и остается активизированным вывод повернутой матрицы факторов.

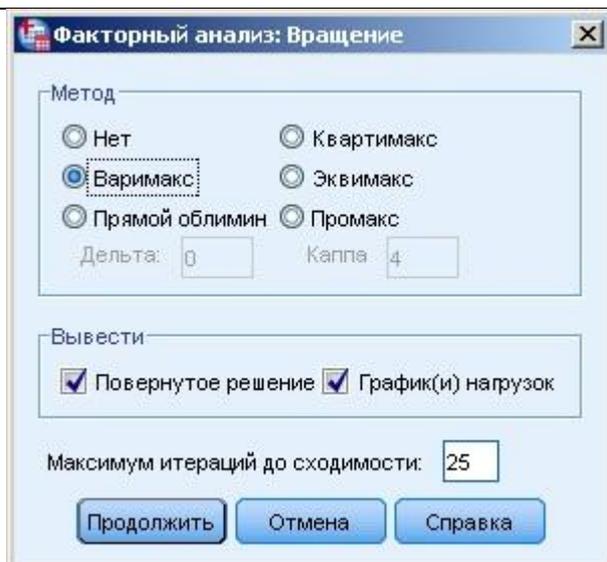


Рис. 10. Окно «Факторный анализ: Вращение»

Найденные значения факторов в виде дополнительных переменных можно сохранить при нажатии на кнопку **«Значения факторов...»** (см. рисунок 7). В появившемся окне **«Факторный анализ: Значение факторов»** отмечается пункт **«Сохранить как переменные»** (см. рисунок 11). По умолчанию обычно устанавливается регрессионный метод.

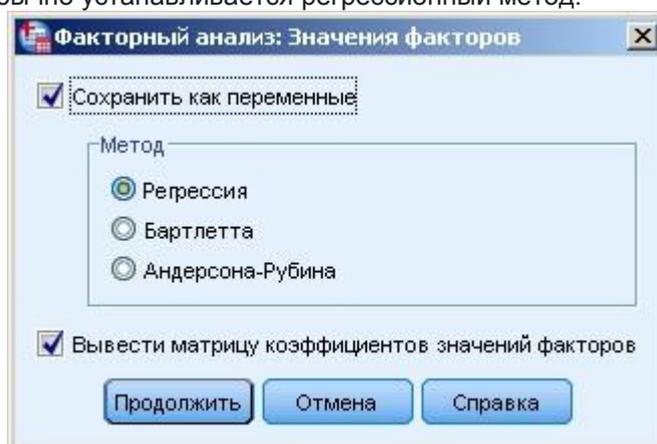


Рис. 11. Окно «Факторный анализ: Значение факторов»

Кнопка **«Параметры ...»** (см. рисунок 7) предназначена для обработки пропущенных значений. В окне **«Факторный анализ: Параметры»** обеспечивается возможность замены пропущенных значений (см. рисунок 12).

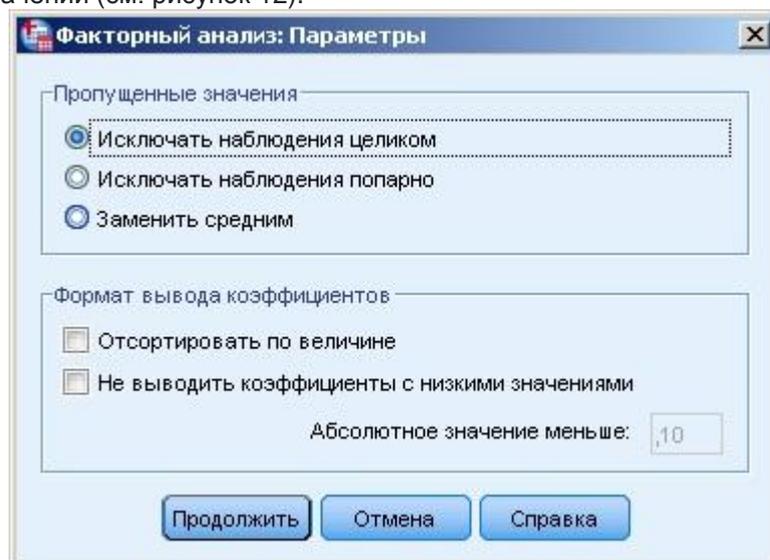


Рис. 12. Окно «Факторный анализ: Параметры»

Выполнения расчетов начнется после нажатия кнопки **«Ок»** окна **«Факторный анализ»** (см. рисунок 7), в окне обзора **«Вывод1 [Документ1]»** появятся результаты (см. рисунок 13).

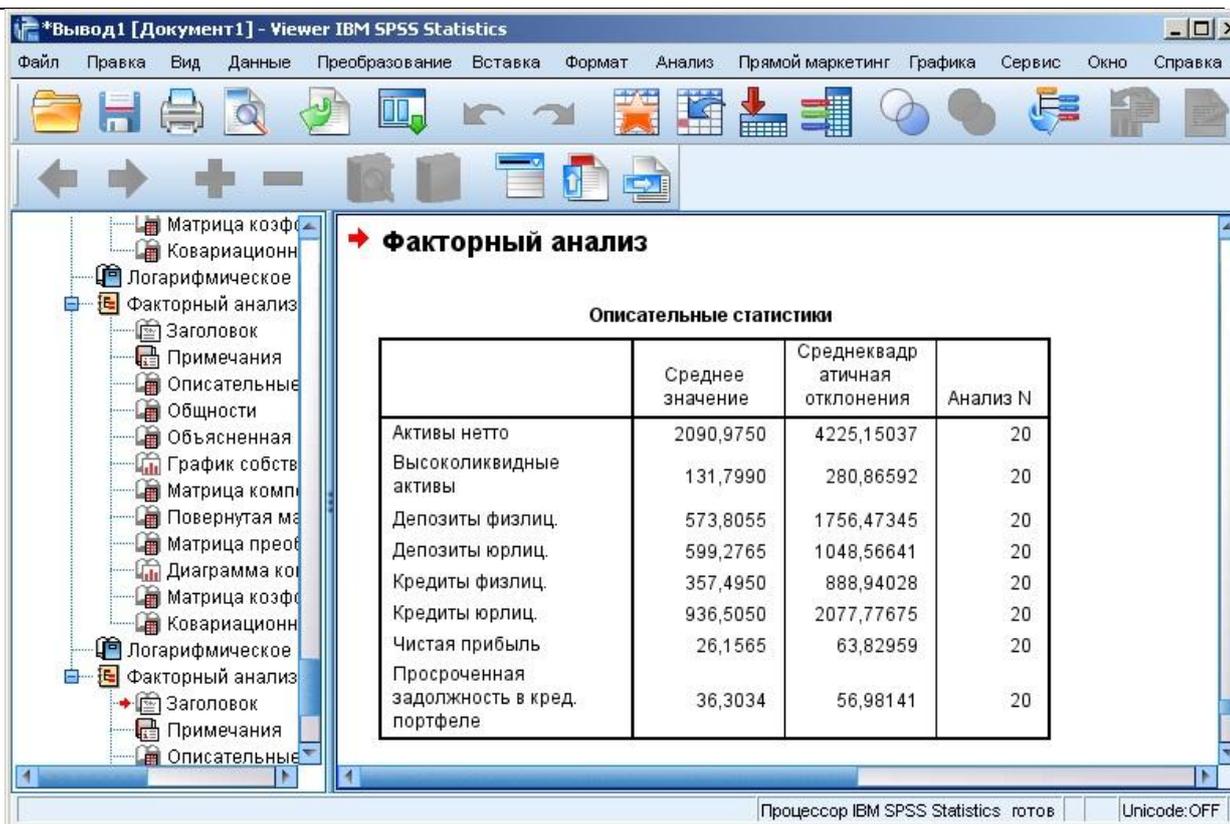


Рис. 13. Окно обзора «Вывод1 [Документ1]»

В данном окне можно увидеть рассчитанные описательные статистики, таблицу полной объясненной дисперсии (см. таблицу 11).

Согласно таблице 3 видно, что только два фактора имеют собственные значения больше единицы и два отобранных фактора (компоненты 1 и 2) объясняют 93,65% совокупной дисперсии.

#### Кейс задание

54. Искусственный интеллект и большие данные на фондовом рынке. По примеру многих своих знакомых трейдер Руслан решил начать торговать на фондовом рынке. Его однокурсница Людмила поделилась с ним своим положительным опытом торговли акциями при помощи торгового робота – специальной программы, предлагаемой брокерской компанией «Инвест-чемпион» (компания также осуществляет деятельность инвестиционного советника). Руслан проанализировал сайты нескольких банков и брокеров и также остановился на «Инвест-чемпионе», поскольку эта компания позволяла своим клиентам пользоваться её торговыми роботами бесплатно. Как подчеркивала компания, эти роботы «являются не просто алгоритмическими программами, а основываются на искусственном интеллекте. Они способны к самообучению на основе анализа рыночных данных, к самостоятельному сбору информации из новостных лент, социальных сетей, аналитических сайтов и телеграм-каналов, к использованию фундаментального и технического анализа, и даже к учету поведенческих стереотипов других участников рынка. При этом инвестирование через таких роботов не означает передачу средств клиента в доверительное управление: и деньги, и приобретаемые ценные бумаги находятся на счетах клиента. Руслан заключил с «Инвест-чемпионом» договор о комплексном обслуживании на рынке ценных бумаг, один из разделов которого предусматривал право клиента на использование любого торгового робота из нескольких сотен имеющихся у компании без дополнительной платы (при этом комиссии за совершаемые сделки компания удерживала). Правда, вначале наш герой попытался торговать самостоятельно, но быстро разочаровался в этом: его доходность за месяц составила чуть больше 1% без учета налогов. Тогда Руслан выбрал в личном кабинете робота с гордым названием «Эйнштейн» и доверил ему свои средства. Инвестиционная стратегия «Эйнштейна» предполагала вложения в широкий круг лучших российских и зарубежных акций. Руслан вздохнул с облегчением и занялся написанием курсовой работы и поиском подходящей стажировки в зарубежном вузе, позабыв о фондовом рынке. Ежедневно приходившие ему отчеты от робота он довольно быстро перестал читать, поскольку они были слишком объемными. Когда прошло полгода, Руслан решил сравнить результаты своего робота со среднерыночными. К его большому огорчению, оказалось, что его портфель упал на 2%, хотя за это же время индекс МосБиржи увеличился на 15%, а американский индекс S&P

	<p>500 – на 12%. Большинство друзей Руслана, игравшие на рынке акций самостоятельно, получили доходности в диапазоне от 10% до 20% за полгода. Руслан попытался выяснить у своего брокера, какие именно сделки совершал робот и как ему удалось получить убытки, и ему был предъявлен огромный список более чем в пять тысяч сделок. Проанализировать каждую из них на предмет ее целесообразности Руслан, разумеется, не смог. А когда он поинтересовался о возмещении понесенного ущерба и упущенной выгоды, сотрудники «Инвест-Чемпиона» лишь развели руками: сделки с использованием робота осуществляются в рамках брокерского договора, а значит, юридически их совершает сам клиент. И в договоре было четко написано, что компания не гарантирует какого-либо конкретного результата в случае использования торгового робота, а клиент несет все риски неблагоприятного изменения стоимости активов.</p> <p>Вопросы:</p> <p>Вопрос 1 (для команды «позиции регулятора»): Проанализируйте с позиции регулятора, нужны ли какие-либо меры по регулированию использования искусственного интеллекта на фондовом рынке, и если да, то какие. Ответ обоснуйте.</p> <p>Вопрос 1 (для команды «позиции бизнеса»): Проанализируйте с позиции брокерской компании, нужны ли какие-либо меры по регулированию использования искусственного интеллекта на фондовом рынке, и если да, то какие. Ответ обоснуйте.</p> <p>Вопрос 2 (для обеих команд): Какие возможности и риски, на ваш взгляд, влечет за собой распространение использования искусственного интеллекта на фондовом рынке для личных финансов?</p>
55.	<p>Искусственный интеллект, большие данные и безопасность расчетов и платежей Антон – студент 4 курса одного из ведущих технических вузов страны, в который поступил благодаря участию в олимпиаде «Высокий уровень». На первых курсах он занимался только учебой, чтобы получать повышенную стипендию. С третьего курса Антон начал работать в службе технической поддержки одного из операторов мобильной связи. Теперь работа и учеба занимают практически всё время Антона, а его путешествия ограничиваются поездками к родителям в соседний областной центр раз в пару месяцев. К своим финансам Антон относится очень бережно, ведет личный бюджет. Очень основательно выбирал свою первую платежную карту. Остановился на банке «Безопасный», который позиционировался на рынке банковских услуг как лидер по внедрению решений, обеспечивающих раннее выявление аномальных активностей в IT-инфраструктуре, за что неоднократно получал награду в номинации «Технология года» в категории «Финансы и экономика: цифровые решения для безопасности». И вот однажды, подсчитав свои накопления, Антон решил сделать себе подарок – слетать на концерт любимой группы в одной из европейских столиц. Каково же было его удивление, когда при покупке долгожданных билетов на сайте организатора мероприятия банк «Безопасный» заблокировал операцию, да и карту в целом! Обращение в банк и выяснение причин блокировки заняло немало времени и сил: робот-консультант не смог сказать ничего конкретного, люди-менеджеры долго пересылали Антона от одного сотрудника к другому. Наконец, начальник отдела предотвращения мошеннических операций объяснил Антону, что система безопасности банка, основанная на искусственном интеллекте, определила данную транзакцию как аномальную, абсолютно не типичную для Антона. Следовательно, существовал высокий риск, что на деньги Антона посягают неизвестные злоумышленники, и блокировка карты – наилучший способ сохранить средства клиента, который должен быть только благодарен банку за такую заботу. Карту в итоге разблокировали через несколько дней, но нервные клетки и потерянное время Антона вернуть не удалось, да и билеты на концерт успели закончиться. Но еще больше Антон, никуда не уехавший из России, удивился через месяц, когда обнаружил в личном кабинете мобильного банка уведомление о транзакции на сумму 399 долларов США, совершённую неделю назад в магазине в самом центре Нью-Йорка. Кто-то приобрел за счет средств с его банковского счета новенькую модель умных часов известного бренда. К большому огорчению Антона, оспорить этот платеж не получилось, попытки обратиться в банк за возмещением понесенного ущерба также не увенчались успехом. А куда же смотрел искусственный интеллект, который должен был «выявить аномалии с помощью эвристики, графических алгоритмов и анализа потоков транзакций» как обещал банк «Безопасный» на своем сайте? Почему он вдруг решил, что такие операции нормальны и обычны для Антона? Эти вопросы так и остались без ответа.</p> <p>Вопросы:</p> <p>Вопрос 1 (для команды «позиции регулятора»): Проанализируйте с позиции регулятора, нужны ли какие-либо меры по регулированию использования искусственного интеллекта на рынке платежных услуг, и если да, то какие. Ответ обоснуйте.</p> <p>Вопрос 1 (для команды «позиции бизнеса»): Проанализируйте с позиции кредитной организации, нужны ли какие-либо меры по регулированию использования искусственного интеллекта на рынке платежных услуг, и если да, то какие. Ответ обоснуйте.</p>

	Вопрос 2 (для обеих команд): Какие возможности и риски, на ваш взгляд, влечет за собой распространение использования искусственного интеллекта на рынке банковских платежных услуг для личных финансов?
56.	<p>Искусственный интеллект и большие данные в скоринговых моделях Игорь, молодой и талантливый сотрудник консалтинговой компании, налил себе кофе и в очередной раз заглянул на сайт, посвященный дизайнерской отделке квартир. Накануне он подал в один из ведущих банков заявку на ипотечный кредит и сейчас был абсолютно уверен, что её одобряют: ведь у него высокие официальные доходы и ни разу не было просроченных или непогашенных долгов. Как долго он выбирал квартиру, общался с риэлторами и продавцами, и вот наконец остался последний шаг! Лаконичное смс-сообщение от банка «Ваша заявка на получение кредита отклонена» было для Игоря подобно грому с ясного неба, причем о причинах отказа в нем не было ни слова. Однако Игорь считал себя человеком упорным, целеустремленным и не собирался быстро отказываться от мечты, поэтому подал заявку в другой банк. Но и там решение было отрицательным без объяснения причин. На раздраженные расспросы Игоря менеджер банка заявил, что не может объяснить, почему именно заявка была отклонена, так как все решения об одобрении кредитов принимаются искусственным интеллектом на основании кредитного скоринга клиентов. Игорь возмутился такому ответу и попросил, чтобы его заявку рассматривал человек, так как сам он не видит ни одной причины, почему в кредите ему может быть отказано. А если эти причины есть, то он хотел бы их выяснить, чтобы иметь возможность исправить ситуацию со своим риск-профилем. Однако менеджер, хоть и посочувствовал Игорю, но ответил, что это невозможно: политика банка в области выдачи кредитов не предусматривает рассмотрение заявки людьми. Оказывается, искусственный интеллект при оценке кредитных рисков работает с очень высокой точностью, и с момента его внедрения процент невозвратных кредитов существенно снизился, а прибыль банка выросла. Поэтому банк не считает целесообразным пересмотр результатов кредитного скоринга человеком. Факторы, на основании которых искусственный интеллект принимает свои решения, также не раскрываются. В итоге Игорь остался в полном недоумении о том, что ему делать дальше: ведь он так и не узнал, в чем именно состоит проблема, из-за которой искусственный интеллект присваивает ему высокий кредитный риск, и, соответственно, не понимает, что сделать, чтобы её решить. Возмущенный таким отношением к клиентам, Игорь решил подать жалобу на последнюю кредитную организацию в Банк России.</p> <p>Вопросы:</p> <p>Вопрос 1 (для команды «позиции регулятора»): Проанализируйте с позиции регулятора, нужны ли какие-либо меры по регулированию использования искусственного интеллекта в кредитном скоринге, и если да, то какие. Ответ обоснуйте.</p> <p>Вопрос 1 (для команды «позиции бизнеса»): Проанализируйте с позиции кредитной организации, нужны ли какие-либо меры по регулированию использования искусственного интеллекта в кредитном скоринге, и если да, то какие. Ответ обоснуйте.</p> <p>Вопрос 2 (для обеих команд): Какие возможности и риски, на ваш взгляд, влечет за собой распространение использования искусственного интеллекта в кредитном скоринге для личных финансов?</p>

### 3.2 Собеседование (вопросы для зачета)

**ОПК 2** - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

Номер вопроса	Текст вопроса
57.	Что включает в себя понятие «Большие данные»
58.	Форматы хранения больших данных
59.	Программы обработки больших данных
60.	Источники больших данных
61.	Способы получения больших данных
62.	Методы и инструменты анализа больших данных
63.	Архитектура систем хранения и обработки больших данных.
64.	Особенности использования внешних и внутренних каналов связи при передаче больших объемов данных.
65.	Состав и назначение модулей обработки, подсистем индексирования и хранения данных
66.	Стандартные средства разработки для организации процессов поточной обработки

	больших объемов данных
67.	Зачем нужен VPN?
68.	Основные элементы системы построения виртуализированных сетевых инфраструктур.
69.	Как организовано облачное хранение
70.	Где лучше хранить большие данные: собственный сервер, аренда стойки, облако.

### 3.3 Домашняя работа

**ОПК 2** - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

Номер вопроса	Текст вопроса
71	Понятие виртуализированный контейнер
72	Состояние ИБ. Основные принципы ИБ
73	Риски при обработке больших данных
74	Степень участия различных категорий персонала в обработке защищаемой информации
75	Мероприятия по обеспечению защиты данных
76	Основные этапы пакетной обработки данных.
77	Методология анализа данных Crisp DM
78	Методология анализа данных SEMMA
79	Методология анализа данных KDD
80	Python – пакеты для анализа данных
81	Этапы исследования по Crisp DM
82	Оценка рисков и условий реализации проекта
83	Разведочный анализ данных
84	Как построить красивые диаграммы в Python

### 3.4. Подготовка к отчету по практическим работам

**ОПК 2** - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности

Номер вопроса	Текст задания
85	Построение диаграмм размаха. Что показывает ящик с усами?
86.	Оценка параметров. Средняя арифметическая. Средняя геометрическая.
87.	Оценка параметров Средняя гармоническая. Средняя квадратическая
88	Оценка параметров мода и медиана.
89	Перцентили. Квартили. Межквартильное расстояние.

### 44. 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, а также методическими указаниями для самостоятельной работы обучающихся:

Стукало О.Г. Большие данные [Электронный ресурс] : методические указания к самостоятельной работе обучающихся по направлению подготовки 09.03.02 очной, и заочной форм обучения / О.Г. Стукало; ВГУИТ. - Воронеж, 2022. - 8 с. - Электрон. ресурс. – Режим доступа: <http://education.vsu.ru>

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.

**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
<b>ОПК 2 - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности</b>					
<b>Знает</b>	современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	Результаты тестирования	Обучающимся даны правильные ответы менее чем на 59,99 % всех тестовых вопросов	Неудовлетворительно	Не освоена / недостаточный
			Обучающимся даны правильные ответы на 60-74,99% всех тестовых вопросов	Удовлетворительно	Освоена / базовый
			Обучающимся даны правильные ответы на 75-84,99% всех тестовых вопросов	Хорошо	Освоена / повышенный
			Обучающимся даны правильные ответы на 85-100% всех тестовых вопросов	Отлично	Освоена / повышенный
		Собеседование ( экзамен)	Обучающийся обладает частичными и разрозненными знаниями, только некоторые из которых может связывать между собой	Неудовлетворительно	Не освоена / недостаточный
			Обучающийся обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Удовлетворительно	Освоена / базовый
			Обучающийся обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Хорошо	Освоена / повышенный
			Обучающийся обладает системным взглядом на изучаемый объект	Отлично	Освоена / повышенный
<b>Умеет</b>	выбирать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	Подготовка к отчету по практическим работам	Обучающийся не владеет умениями выполнения заданий; не демонстрирует умений, предусмотренных планируемыми результатами обучения	Неудовлетворительно	Не освоена / недостаточный
			Обучающийся испытывает затруднения при выполнении заданий по алгоритму; демонстрирует минимальный набор умений, предусмотренных планируемыми результатами обучения	Удовлетворительно	Освоена / базовый
			Обучающийся выполняет задания с использованием алгоритма решения, при выполнении допускает незначительные ошибки и неточности, формулирует выводы; демонстрирует умения,	Хорошо	Освоена / повышенный

			предусмотренные планируемые результаты обучения		
			Обучающийся выполняет задания, формируя алгоритм решения, при выполнении не допускает ошибок и неточностей, формулирует выводы; демонстрирует умения, предусмотренные планируемые результаты обучения	Отлично	Освоена / повышенный
<b>Владеет</b>	навыками применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	Домашняя контрольная работа	Обучающийся не владеет умениями выполнения заданий; не демонстрирует умений, предусмотренных планируемые результаты обучения	Неудовлетворительно	Не освоена / недостаточный
			Обучающийся испытывает затруднения при выполнении заданий по алгоритму; демонстрирует минимальный набор умений, предусмотренных планируемые результаты обучения	Удовлетворительно	Освоена / базовый
			Обучающийся выполняет задания с использованием алгоритма решения, при выполнении допускает незначительные ошибки и неточности, формулирует выводы; демонстрирует умения, предусмотренные планируемые результаты обучения	Хорошо	Освоена / повышенный
			Обучающийся выполняет задания, формируя алгоритм решения, при выполнении не допускает ошибок и неточностей, формулирует выводы; демонстрирует умения, предусмотренные планируемые результаты обучения	Отлично	Освоена / повышенный