

Документ подписан простой электронной подписью  
Информация о владельце:

ФИО: Кандрашина Елена Александровна

Должность: Врио ректора ФГАОУ ВО «Самарский государственный экономический университет»

Дата подписания: 07.07.2023 14:46:14

Уникальный программный ключ:

b2fd765521f4c570b8c6e8e502a10b4f1de8ae0d

**Министерство науки и высшего образования Российской Федерации**  
**Федеральное государственное автономное образовательное учреждение**  
**высшего образования**  
**«Самарский государственный экономический университет»**

**Институт**      Институт экономики предприятий

**Кафедра**      Информационных систем и технологий (ПГУТИ)

**УТВЕРЖДЕНО**

Ученым советом Университета

(протокол № 11 от 30 мая 2023 г.)

**РАБОЧАЯ ПРОГРАММА**

**Наименование дисциплины**

Б1.В.02 Технологии больших данных

**Основная профессиональная образовательная программа**

09.03.03 Прикладная информатика программа  
Интеллектуальные цифровые системы и сервисы  
в управлении

Квалификация (степень) выпускника Бакалавр

Самара 2023

## Содержание (рабочая программа)

Стр.

- 1 Место дисциплины в структуре ОП
- 2 Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе
- 3 Объем и виды учебной работы
- 4 Содержание дисциплины
- 5 Материально-техническое и учебно-методическое обеспечение дисциплины
- 6 Фонд оценочных средств по дисциплине

Целью изучения дисциплины является формирование результатов обучения, обеспечивающих достижение планируемых результатов освоения образовательной программы.

## 1. Место дисциплины в структуре ОП

Дисциплина Технологии больших данных входит в часть, формируемая участниками образовательных отношений блока Б1. Дисциплины (модули)

Предшествующие дисциплины по связям компетенций: Веб-программирование, Облачные технологии и сервисы, Архитектура интеллектуальных информационных систем, Проектирование пользовательского интерфейса, Разработка интерфейсов и адаптивный Веб-дизайн, Разработка распределенных приложений, Облачные ресурсы для разработки интеллектуальных сервисов

Последующие дисциплины по связям компетенций: Технологии блокчейн, Проектирование и разработка экспертных интеллектуальных систем, Современные технологии разработки приложений, Блокчейн и его приложения, Автоматизированное тестирование программного обеспечения, Корпоративные интеллектуальные системы, Методы и средства проектирования цифровых сервисов в управлении, Управление качеством разработки приложений

## 2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

Изучение дисциплины Технологии больших данных в образовательной программе направлено на формирование у обучающихся следующих компетенций:

### Профессиональные компетенции (ПК):

ПК-1 - Способен осуществлять организационно-технологическое сопровождение работ по созданию интеллектуальных информационных систем и разрабатывать их прототипы

| Планируемые результаты обучения по программе | Планируемые результаты обучения по дисциплине   |  |  |
|--|---|--|--|
| ПК-1   | ПК-1.1: Знать:<br>инструменты и методы организационно-технического сопровождения работ на различных этапах цикла разработки интеллектуальных цифровых систем и сервисов | ПК-1.2: Уметь:<br>выбирать инструменты и методы организационно-технического сопровождения работ на различных этапах цикла разработки интеллектуальных цифровых систем и сервисов | ПК-1.3: Владеть (иметь навыки):<br>навыками выполнения и управления работами по созданию интеллектуальных цифровых систем и сервисов на различных этапах цикла разработки с целью повышения эффективности деятельности организаций |

## 3. Объем и виды учебной работы

Учебным планом предусматриваются следующие виды учебной работы по дисциплине:

### Очная форма обучения

| Виды учебной работы                          | Всего час/ з.е. |
|--|-----------------|
|  | Сем 5           |
| Контактная работа, в том числе:              | 62.3/1.73       |
| Занятия лекционного типа                     | 32/0.89         |
| Лабораторные работы (лабораторный практикум) | 28/0.78         |
| Индивидуальная контактная работа (ИКР)       | 0.3/0.01        |
| Групповая контактная работа (ГКР)            | 2/0.06          |
| Самостоятельная работа:                      | 47.7/1.32       |
| Промежуточная аттестация                     | 34/0.94         |

|  |     |
|--|-----|
| Вид промежуточной аттестации:<br>Экзамен                         | Экз |
| Общая трудоемкость (объем части образовательной программы): Часы | 144 |
| Зачетные единицы   | 4   |

#### 4. Содержание дисциплины

##### 4.1. Разделы, темы дисциплины и виды занятий:

Тематический план дисциплины Технологии больших данных представлен в таблице.

#### Разделы, темы дисциплины и виды занятий Очная форма обучения

| № п/п | Наименование темы (раздела) дисциплины                          | Контактная работа |                           |            |          | Самостоятельная работа | Планируемые результаты обучения в соотношении с результатами обучения по образовательной программе |
|-------|---|-------------------|---------------------------|------------|----------|------------------------|--|
|       |   | Лекции            | Занятия семинарского типа | ИКР        | ГКР      |                        |  |
|       |   |                   | Лаборат. работы           |            |          |                        |  |
| 1.    | Введение в анализ больших данных. Методы анализа больших данных | 16                | 14                        |            |          | 20                     | ПК-1.1, ПК-1.2, ПК-1.3   |
| 2.    | Инструментарий больших данных                                   | 16                | 14                        |            |          | 27.7                   | ПК-1.1, ПК-1.2, ПК-1.3   |
|       | Контроль  | 34                |                           |            |          |                        |  |
|       | <b>Итого</b>  | <b>32</b>         | <b>28</b>                 | <b>0.3</b> | <b>2</b> | <b>47.7</b>            |  |

##### 4.2 Содержание разделов и тем

##### 4.2.1 Контактная работа

##### Тематика занятий лекционного типа

| №п/п | Наименование темы (раздела) дисциплины                          | Вид занятия лекционного типа* | Тематика занятия лекционного типа  |
|------|---|-------------------------------|--|
| 1.   | Введение в анализ больших данных. Методы анализа больших данных | лекция                        | Область применения data science и больших данных и их преимущества. Общие сведения о процессе data science.  |
|      |   | лекция                        | Распределенные файловые системы. Инфраструктура распределенного программирования. Инфраструктура интеграции данных. Инфраструктуры машинного обучения. Базы данных NoSQL. Инструменты планирования. Инструменты сравнительного анализа. Развертывание системы. Программирование служб. Безопасность. |
|      |   | лекция                        | Обзор процесса data science. Этапы процесса data science.  |
|      |   | лекция                        | Обзор распространенных методов обработки данных  |
|      |   | лекция                        | Применение машинного обучения в процессе data science. Инструменты Python, используемые в машинном обучении.   |
|      |   | лекция                        | Процесс моделирования. Типы машинного обучения. Частично контролируемое обучение.  |
|      |   | лекция                        | Проблемы при работе с большими объемами данных. Общие методы обработки больших объемов данных.   |

|    |                               |        |   |
|----|-------------------------------|--------|---|
|    |                               | лекция | Общие рекомендации для программистов при работе с большими наборами данных. Примеры реализации работы с большими данными. |
| 2. | Инструментарий больших данных | лекция | Описание, структура и общие особенности работы Hadoop   |
|    |                               | лекция | Описание, структура и общие особенности работы MapReduce.   |
|    |                               | лекция | Описание, структура и общие особенности работы Spark.   |
|    |                               | лекция | Описание, структура и общие особенности работы KNIME.   |
|    |                               | лекция | Введение в NoSQL. Базовые принципы реляционных баз данных. Проблема баз данных, распределенных по многим узлам.           |
|    |                               | лекция | Принципы BASE баз данных NoSQL. Типы баз данных NoSQL.  |
|    |                               | лекция | Связанные данные и графовые базы данных. Neo4j: графовая база данных.   |
|    |                               | лекция | Методы глубокого анализа текста. Выделение основы и лемматизация. Классификатор на базе дерева принятия решений.          |

\*лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации педагогическими работниками организации и (или) лицами, привлекаемыми организацией к реализации образовательных программ на иных условиях, обучающимся

#### Тематика занятий семинарского типа

| №п/п | Наименование темы (раздела) дисциплины                             | Вид занятия семинарского типа** | Тематика занятия семинарского типа   |
|------|--|---------------------------------|--|
| 1.   | Введение в анализ больших данных.<br>Методы анализа больших данных | лабораторные работы             | Подготовка данных с помощью электронных таблиц                             |
|      |  | лабораторные работы             | Построение простого линейного классификатора                               |
|      |  | лабораторные работы             | Классификатор к ближайших соседей  |
|      |  | лабораторные работы             | Наивный байесовский классификатор  |
|      |  | лабораторные работы             | Метод главных компонент  |
|      |  | лабораторные работы             | Алгоритмы кластеризации  |
|      |  | лабораторные работы             | Обработка больших объемов данных на одном компьютере                       |
| 2.   | Инструментарий больших данных                                      | лабораторные работы             | Подготовка данных средствами Python  |
|      |  | лабораторные работы             | Использование PySpark для работы с данными                                 |
|      |  | лабораторные работы             | Знакомство с KNIME   |
|      |  | лабораторные работы             | Предобработка данных в KNIME   |
|      |  | лабораторные работы             | Построение простого классификатора средствами Apache Spark в KNIME         |
|      |  | лабораторные работы             | Построение сложной модели обработки данных средствами Apache Spark в KNIME |
|      |  | лабораторные работы             | Визуализация данных в KNIME  |

\*\* семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия

#### Иная контактная работа

При проведении учебных занятий СГЭУ обеспечивает развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств (включая при необходимости проведение интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализ ситуаций и имитационных моделей, преподавание дисциплин (модулей) в форме курсов, составленных на основе результатов научных исследований, проводимых

организацией, в том числе с учетом региональных особенностей профессиональной деятельности выпускников и потребностей работодателей).

Формы и методы проведения иной контактной работы приведены в Методических указаниях по основной профессиональной образовательной программе.

#### 4.2.2 Самостоятельная работа

| №п/п | Наименование темы (раздела) дисциплины                          | Вид самостоятельной работы *** |
|------|---|--------------------------------|
| 1.   | Введение в анализ больших данных. Методы анализа больших данных | - подготовка к опросу          |
| 2.   | Инструментарий больших данных                                   | - подготовка к опросу          |

\*\*\* самостоятельная работа в семестре, написание курсовых работ, докладов, выполнение контрольных работ

### 5. Материально-техническое и учебно-методическое обеспечение дисциплины

#### 5.1 Литература:

##### Основная литература

1. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов ; под научной редакцией Н. В. Папуловской. — Москва : Издательство Юрайт, 2023. — 121 с. — (Высшее образование). — ISBN 978-5-534-09837-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/514724>

##### Дополнительная литература

1. Гордеев, С. И. Организация баз данных в 2 ч. Часть 1 : учебник для вузов / С. И. Гордеев, В. Н. Волошина. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 310 с. — (Высшее образование). — ISBN 978-5-534-04469-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/513879>

Гордеев, С. И. Организация баз данных в 2 ч. Часть 2 : учебник для вузов / С. И. Гордеев, В. Н. Волошина. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 513 с. — (Высшее образование). — ISBN 978-5-534-04470-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/515097>

##### Литература для самостоятельного изучения

1.

#### 5.2. Перечень лицензионного программного обеспечения

1. RedOS
2. KNIME
3. Python

#### 5.3 Современные профессиональные базы данных, к которым обеспечивается доступ обучающихся

1. Профессиональная база данных «Информационные системы Министерства экономического развития Российской Федерации в сети Интернет» (Портал «Официальная Россия» - <http://www.gov.ru/>)

2. Профессиональная база данных «Финансово-экономические показатели Российской Федерации» (Официальный сайт Министерства финансов РФ - <https://www.minfin.ru/ru/>)

3. Профессиональная база данных «Официальная статистика» (Официальный сайт Федеральной службы государственной статистики - <http://www.gks.ru/>)

#### 5.4. Информационно-справочные системы, к которым обеспечивается доступ обучающихся

1. Справочно-правовая система «Консультант Плюс»
2. Справочно-правовая система «ГАРАНТ-Максимум»

### 5.5. Специальные помещения

|   |   |
|---|---|
| Учебные аудитории для проведения занятий лекционного типа                         | Комплекты ученической мебели<br>Мультимедийный проектор<br>Доска<br>Экран   |
| Учебные аудитории для проведения практических занятий (занятий семинарского типа) | Комплекты ученической мебели<br>Мультимедийный проектор<br>Доска<br>Экран<br>Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ |
| Учебные аудитории для групповых и индивидуальных консультаций                     | Комплекты ученической мебели<br>Мультимедийный проектор<br>Доска<br>Экран<br>Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ |
| Учебные аудитории для текущего контроля и промежуточной аттестации                | Комплекты ученической мебели<br>Мультимедийный проектор<br>Доска<br>Экран<br>Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ |
| Помещения для самостоятельной работы  | Комплекты ученической мебели<br>Мультимедийный проектор<br>Доска<br>Экран<br>Компьютеры с выходом в сеть «Интернет» и ЭИОС СГЭУ |
| Помещения для хранения и профилактического обслуживания оборудования              | Комплекты специализированной мебели для хранения оборудования   |

### 5.6 Лаборатории и лабораторное оборудование

|                                    |   |
|------------------------------------|---|
| Аудитории для лабораторных занятий | Количество посадочных мест по количеству обучающихся.<br>Компьютеры с выходом в сеть «Интернет» |
|------------------------------------|---|

### 6. Фонд оценочных средств по дисциплине Технологии больших данных:

#### 6.1. Контрольные мероприятия по дисциплине

| Вид контроля           | Форма контроля  | Отметить<br>нужное<br>знаком<br>« + » |
|------------------------|---|---------------------------------------|
| Текущий контроль       | Оценка докладов                                       |                                       |
|                        | Устный/письменный опрос                               | +                                     |
|                        | Тестирование  |                                       |
|                        | Практические задачи                                   |                                       |
|                        | Оценка контрольных работ (для заочной формы обучения) |                                       |
| Промежуточный контроль | Экзамен   | +                                     |

Порядок проведения мероприятий текущего и промежуточного контроля определяется Методическими указаниями по основной профессиональной образовательной

программе высшего образования; Положением о балльно-рейтинговой системе оценки успеваемости обучающихся по основным образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры в федеральном государственном автономном образовательном учреждении высшего образования «Самарский государственный экономический университет».

## 6.2. Планируемые результаты обучения по дисциплине, обеспечивающие достижение планируемых результатов обучения по программе

### Профессиональные компетенции (ПК):

ПК-1 - Способен осуществлять организационно-технологическое сопровождение работ по созданию интеллектуальных информационных систем и разрабатывать их прототипы

| Планируемые результаты обучения по программе         | Планируемые результаты обучения по дисциплине   |  |   |
|--|---|--|---|
|  | ПК-1.1: Знать:  | ПК-1.2: Уметь:   | ПК-1.3: Владеть (иметь навыки):   |
|  | инструменты и методы организационно-технического сопровождения работ на различных этапах цикла разработки интеллектуальных цифровых систем и сервисов | выбирать инструменты и методы организационно-технического сопровождения работ на различных этапах цикла разработки интеллектуальных цифровых систем и сервисов | навыками выполнения и управления работами по созданию интеллектуальных цифровых систем и сервисов на различных этапах цикла разработки с целью повышения эффективности деятельности организаций |
| Пороговый  | Методы, используемые при создании интеллектуальных цифровых систем  | Выбирать методы, используемые при создании интеллектуальных цифровых систем  | Навыками выполнения работ по созданию интеллектуальных цифровых систем и сервисов на этапе подготовки данных  |
| Стандартный (в дополнение к пороговому)              | Методы и инструменты, используемые при создании интеллектуальных цифровых систем  | Выбирать методы и инструменты, используемые при создании интеллектуальных цифровых систем  | Навыками выполнения работ по созданию интеллектуальных цифровых систем и сервисов на этапе подготовки данных и выбора математической модели.  |
| Повышенный (в дополнение к пороговому, стандартному) | Методы и инструменты, используемые при создании интеллектуальных цифровых систем и сервисов   | Выбирать методы и инструменты, используемые при создании интеллектуальных цифровых систем и сервисов   | Навыками выполнения работ по созданию интеллектуальных цифровых систем и сервисов на этапе подготовки данных, выбора и тестирования математической модели.                                      |

## 6.3. Паспорт оценочных материалов

| № п/п | Наименование темы (раздела) дисциплины                      | Контролируемые планируемые результаты обучения в соотношении с результатами обучения по программе | Вид контроля/используемые оценочные средства |               |
|-------|---|---|--|---------------|
|       |   |   | Текущий                                      | Промежуточный |
| 1.    | Введение в анализ больших данных.<br>Методы анализа больших | ПК-1.1, ПК-1.2, ПК-1.3  | Устный/письменный опрос                      | Экзамен       |



|    |                               |                        |                         |         |
|----|-------------------------------|------------------------|-------------------------|---------|
|    | данных                        |                        |                         |         |
| 2. | Инструментарий больших данных | ПК-1.1, ПК-1.2, ПК-1.3 | Устный/письменный опрос | Экзамен |

#### 6.4.Оценочные материалы для текущего контроля

##### Вопросы для устного/письменного опроса

| Раздел дисциплины  | Вопросы   |
|--|---|
| Введение в анализ больших данных.<br>Методы анализа больших данных | <ol style="list-style-type: none"> <li>1. Что такое структурированные данные и неструктурированные данные?</li> <li>2. Что собой представляют данные на естественном языке и машинные данные?</li> <li>3. Расскажите про графовые данные.</li> <li>4. Расскажите про аудио, видео и графические данные.</li> <li>5. Дайте понятие потоковых данных.</li> <li>6. Как выполняется назначение цели исследования?</li> <li>7. Как осуществляется сбор данных?</li> <li>8. Как осуществляется подготовка данных?</li> <li>9. Как выполняется исследование данных?</li> <li>10. Поясните процесс построения модели данных.</li> <li>11. Что такое распределённые файловые системы?</li> <li>12. Поясните инфраструктуру распределенного программирования.</li> <li>13. Поясните инфраструктуру интеграции данных.</li> <li>14. Поясните инфраструктуру машинного обучения.</li> <li>15. Дайте понятие базы данных NoSQL.</li> <li>16. Что собой представляют инструменты планирования?</li> <li>17. Что собой представляют собой инструменты сравнительного анализа?</li> <li>18. В чем заключается развертывание системы, программирование служб.</li> <li>19. Что такое data science?</li> <li>20. Каковы цели и задачи data science?</li> <li>21. Перечислите этапы data science</li> <li>22. Расскажите об этапе data science, указанном преподавателем.</li> <li>23. Каковы области применения data science?</li> <li>24. Какие методы анализа данных вы знаете?</li> <li>25. Расскажите про методы математической статистики.</li> <li>26. Расскажите про методы искусственного интеллекта.</li> <li>27. Расскажите про методы имитационного моделирования.</li> <li>28. Расскажите про методы интеллектуального анализа данных</li> <li>29. Расскажите о применении машинного обучения в процессе data science.</li> <li>30. Какие инструменты Python могут быть использованы в машинном обучении?</li> <li>31. Пояснит процесс моделирования.</li> <li>32. Какие типы машинного обучения вы знаете?</li> <li>33. Что представляет собой частично контролируемое обучение?</li> <li>34. Какие существуют проблемы при работе с большими объемами данных?</li> <li>35. Расскажите об общих методах обработки больших объемов данных.</li> <li>36. Каковы общие рекомендации для программистов при работе с большими наборами данных?</li> <li>37. Каким образом технология Spark может помочь реализации работе с большими данными на одном компьютере?</li> <li>38. Какие преимущества позволяет реализовать работа с большими данными на одном компьютере?</li> </ol> |

|                               |  |
|-------------------------------|--|
| Инструментарий больших данных | <ol style="list-style-type: none"> <li>1. Расскажите про Hadoop</li> <li>2. Расскажите про MapReduce</li> <li>3. Расскажите про Spark</li> <li>4. Расскажите про KNIME.</li> <li>5. Каковы возможности ограниченного использования «стандартных» инструментов интеллектуального анализа данных при работе с big data.</li> <li>6. Что такое NoSQL?</li> <li>7. Расскажите о базовых принципах реляционных баз данных.</li> <li>8. Каковы проблемы баз данных, распределенных по многим узлам?</li> <li>9. Расскажите о принципах BASE баз данных NoSQL.</li> <li>10. Какие типы баз данных NoSQL вы знаете?</li> <li>11. Что такое связанные данные?</li> <li>12. Что такое граф?</li> <li>13. Что такое графовая база данных?</li> <li>14. Когда и почему используются графовые базы данных?</li> <li>15. Расскажите про особенности Neo4j.</li> <li>16. Расскажите про язык Cypher.</li> <li>17. Что такое глубокий анализ текста?</li> <li>18. Что такое обработка текста на естественном языке?</li> <li>19. Практические примеры использования глубокого анализа текста.</li> <li>20. Расскажите про методы глубокого анализа текста</li> <li>21. Что такое TF-IDF?</li> <li>22. Для чего применяется визуализация данных для конечного пользователя?</li> <li>23. Что такое одноразовое представление результатов?</li> <li>24. Что такое новая точка зрения на данные?</li> <li>25. Что такое информационная панель для отображения данных в реальном времени?</li> <li>26. Какие способы визуализации данных вы знаете?</li> </ol> |
|-------------------------------|--|

## 6.5. Оценочные материалы для промежуточной аттестации

### Фонд вопросов для проведения промежуточного контроля в форме экзамена

| Раздел дисциплины  | Вопросы   |
|--|---|
| Введение в анализ больших данных.<br>Методы анализа больших данных | <ol style="list-style-type: none"> <li>1. Структурированные и неструктурированные данные. Типы данных.</li> <li>2. Работа с данными: сбор, подготовка, исследование.</li> <li>3. Работа с данными: построение модели данных, отображение, автоматизация</li> <li>4. Распределенные файловые системы.</li> <li>5. Инфраструктуры: интеграции данных, машинного обучения, распределенного программирования.</li> <li>6. Процесс data science: обзор, основные этапы</li> <li>7. Машинное обучение в data science</li> <li>8. Типы машинного обучения.</li> <li>9. Python в data science</li> <li>10. Работа с большими данными: проблемы методы, рекомендации.</li> </ol> |
| Инструментарий больших данных                                      | <ol style="list-style-type: none"> <li>11. Работа с Hadoop</li> <li>12. Работа с Map Reduce</li> <li>13. Работа со Spark</li> <li>14. Работа с KNIME</li> <li>15. Особенности использования стандартных методов ИИ при работе с</li> </ol>  |

|  |   |
|--|---|
|  | большими данными<br>16. Парадигма NoSQL: описание, особенности, принципы<br>17. Базовые принципы реляционных баз данных<br>18. Графовая база данных Neo4j<br>19. Глубокий анализ текста<br>20 Способы визуализации данных |
|--|---|

#### 6.6. Шкалы и критерии оценивания по формам текущего контроля и промежуточной аттестации

##### Шкала и критерии оценивания

| Оценка                | Критерии оценивания для мероприятий контроля с применением 4-х балльной системы |
|-----------------------|---|
| «отлично»             | Повышенный ПК-1.1, ПК-1.2, ПК-1.3   |
| «хорошо»              | Стандартный ПК-1.1, ПК-1.2, ПК-1.3  |
| «удовлетворительно»   | Пороговый ПК-1.1, ПК-1.2, ПК-1.3  |
| «неудовлетворительно» | Результаты обучения не сформированы на пороговом уровне                         |