

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Сибирский государственный индустриальный университет»  
Кафедра прикладных информационных технологий и программирования

УТВЕРЖДАЮ  
Директор института  
информационных технологий и  
автоматизированных систем  
\_\_\_\_\_ Л.Д. Павлова  
подпись  
« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наука о данных и аналитика больших данных

09.04.03 «Прикладная информатика»  
(направленность (профиль): «Прикладная информатика»)

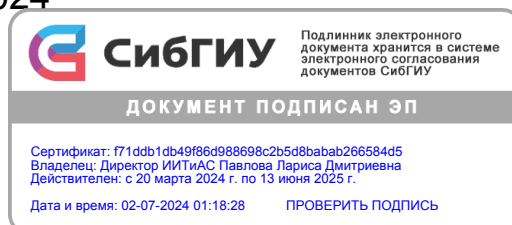
Квалификация выпускника  
Магистр

Форма обучения  
Очная форма

Срок обучения: 2 года

Год начала подготовки 2024

Новокузнецк  
2024



## 1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

- овладение методами анализа и обработки больших объемов данных для принятия обоснованных решений на основе современных технологий.

Задачами учебной дисциплины являются:

- Изучение основных теоретических принципов обработки данных, включая методы сбора, хранения, анализа и визуализации информации;
- овладение практическими навыками работы с различными инструментами и технологиями для обработки и анализа больших объемов данных, такими как Python, SQL, алгоритмы машинного обучения и фреймворки для работы с Big Data.;
- практическое применение полученных знаний и навыков на практике через выполнение проектов по обработке и анализу данных.

## 2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам части, формируемой участниками образовательных отношений **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.04.03 «Прикладная информатика».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:

- Математические и инструментальные методы анализа данных;
- Машинное обучение;
- Программирование глубоких нейронных сетей на Python;
- Анализ текстовых данных.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:

- Методы и инструменты цифровой трансформации.

## 3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

### – Профессиональные компетенции

Наименование категории (группы) ПК	Код и наименование ПК	Код и наименование индикатора достижения ПК	Планируемые результаты обучения
	ПК-1: Способен участвовать в	ПК-1.1 Принимает участие в разработке моделей данных,	– знать: методы моделирования данных, включая

<p>управлении работами по получению, хранению и обработке больших объемов данных</p>	<p>проводит анализ больших объемов данных, строит модели на основе данных</p>	<p>статистическое, машинное обучение и подходы к глубокому обучению. – уметь: разрабатывать модели данных для конкретных приложений, анализировать большие объемы данных, используя соответствующие инструменты и методы.</p>
	<p>ПК-1.2 Принимает участие в разработке методов преобразования данных</p>	<p>– знать: основы различных методов преобразования данных, включая очистку данных, нормализацию, агрегацию и инженерию функций. – уметь: применять инструменты и методы для эффективного преобразования необработанных данных.</p>
	<p>ПК-1.3 Принимает участие в разработке методов обработки больших данных</p>	<p>– знать: методы обработки больших данных, включая распределенные вычисления, параллельную обработку и алгоритмы интеллектуального анализа данных. – уметь: применять алгоритмы и методы программирования для разработки и реализации эффективных и масштабируемых решений.</p>

#### 4 Объем и содержание учебной дисциплины

Учебные занятия по учебной дисциплине проводятся в форме контактной работы и в форме самостоятельной работы обучающихся.

Контактная работа включает в себя занятия лекционного типа (лекции), занятия семинарского типа (семинары, практические занятия,

практикумы), промежуточную аттестацию обучающихся и иные формы взаимодействия обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации ООП на иных условиях, в том числе при проведении промежуточной аттестации обучающихся. Контактная работа может проводиться с применением электронного обучения, дистанционных образовательных технологий.

### Объем учебной дисциплины

Семестр / курс		<b>ИТОГО</b>	<b>6 семестр</b>
Форма промежуточной аттестации			<b>экзамен</b>
Трудоёмкость	<i>академ. час.</i>	<b>144</b>	144
	<i>зачетных единиц</i>	<b>4</b>	4
Лекции, <i>академ. час.</i>		<b>16</b>	16
в форме практической подготовки		<b>0</b>	0
Лабораторные работы, <i>академ. час.</i>		<b>0</b>	0
в форме практической подготовки		<b>0</b>	0
Практические занятия, <i>академ. час.</i>		<b>32</b>	32
в форме практической подготовки		<b>0</b>	0
Курсовая работа / проект, <i>академ. час.</i>		<b>0</b>	0
в форме практической подготовки		<b>0</b>	0
Консультации, <i>академ. час.</i>		<b>0</b>	0
в форме практической подготовки		<b>0</b>	0
Самостоятельная работа, <i>академ. час.</i>		<b>78</b>	78
в форме практической подготовки		<b>0</b>	0
Контроль, <i>академ. час.</i>		<b>18</b>	18
в форме практической подготовки		<b>0</b>	0

### Содержание учебной дисциплины

Раздел 1 Основы аналитики больших данных;

Тема 1.1 Введение в большие данные (Определение больших данных и причины их появления. Примеры возможностей для бизнеса. Различие между Business Intelligence и Big Data);

Тема 1.2 Жизненный цикл аналитики данных (Понятие жизненного цикла аналитики данных. Роли, необходимые для успешного создания проекта по аналитике данных);

Тема 1.3 Высокопроизводительные вычисления (Распределенные вычисления на нескольких серверах, вычислительная парадигма MapReduce. Проект Apache Hadoop и его экосистема. Apache Spark и его компоненты. Вычисления в реальном времени, Apache Storm, Flink);

Раздел 2 Расширенные приложения аналитики больших данных;

Тема 2.1 Масштабирование и многоуровневое хранение данных (Теорема CAP. Парадигма NoSQL. Классификация NoSQL баз данных);

Тема 2.2 Визуализация данных и результатов анализа (Техники визуализации данных, введение в язык R. Визуализация данных в R);

Тема 2.3 Сложные методы аналитики (Классификация задач анализа: Text, Data, Web, Social Mining. Применение машинного обучения в аналитике. K-means и C-means кластеризация, классификация. Логистическая регрессия, ассоциации, алгоритм Априори.);

Тема 2.4 Анализ текста (Поисковые механизмы: Lucene, Solr, ElasticSearch..Алгоритмы Work2Vec и Glove).

### 5 Перечень тем лекций

№ раздела / темы дисциплины	Темы лекций	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
Раздел 1.	Основы аналитики больших данных	8	
Раздел 2.	Расширенные приложения аналитики больших данных	8	
<b>Итого:</b>		<b>16</b>	<b>0</b>

### 6 Перечень тем практических занятий (семинаров)

№ раздела / темы дисциплины	Темы практических занятий (семинаров)	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
Тема 1.1.	Введение в большие данные	4	
Тема 1.2.	Жизненный цикл аналитики данных	4	
Тема 1.3.	Высокопроизводительные вычисления	4	
Тема 2.1.	Масштабирование и многоуровневое хранение данных	4	
Тема 2.2.	Визуализация данных и результатов анализа	4	
Тема 2.3.	Сложные методы аналитики	6	
Тема 2.4.	Анализ текста	6	
<b>Итого:</b>		<b>32</b>	<b>0</b>

### 7 Перечень тем лабораторных работ

№ раздела / темы дисциплины	Темы лабораторных работ	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
	<i>Отсутствуют</i>		
<b>Итого:</b>		<b>0</b>	<b>0</b>

## 8 Перечень тем курсовых работ (проектов)

№ раздела / темы дисциплины	Темы курсовых работ (проектов)	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
	<i>Отсутствуют</i>		
<b>Итого:</b>		<b>0</b>	<b>0</b>

## 9 Виды самостоятельной работы

№ раздела / темы дисциплины	Виды самостоятельной работы	Трудоемкость, <i>академ. час</i>	
		всего	в форме практической подготовки
Раздел 1.	1. Изучение лекционного материала; 2. Подготовка к практическому занятию; 3. Прохождение тестирования.	39	
Раздел 2.	1. Изучение лекционного материала; 2. Подготовка к практическому занятию; 3. Прохождение тестирования.	39	
<i>Контроль</i>	<i>Подготовка к экзамену</i>	18	
<b>Итого:</b>		<b>96</b>	<b>0</b>

## 10 Учебно-методическое и информационное обеспечение учебной дисциплины

### а) литература:

1 Рашка, С. Python и машинное обучение / С. Рашка; пер. с англ. А. В. Логунова. - Москва : ДМК Пресс, 2017. - 418 с. - ISBN 978-5-97060-409-0. – URL: <https://www.studentlibrary.ru/book/ISBN9785970604090.html> (дата обращения: 01.03.2024);

2 Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах; пер. с англ. А. А. Слинкина. - 2-е изд. - Москва : ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. – URL: <https://www.studentlibrary.ru/ru/book/ISBN9785898183004.html> (дата обращения: 27.06.2024);

3 Мастицкий, С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. - 2-е изд. - Москва : ДМК Пресс, 2023. - 497 с. - ISBN 978-5-89818-601-2. – URL: <https://www.studentlibrary.ru/ru/book/ISBN9785898186012.html> (дата обращения: 27.06.2024).

**б) ресурсы информационно-телекоммуникационной сети «Интернет»:**

1 Консультант студента : электронно-библиотечная система / ООО «КОНСУЛЬТАНТ СТУДЕНТА». – Москва, [200 – ]. – URL: <http://www.studentlibrary.ru>. – Режим доступа: для авторизир. пользователей;

2 ЛАНЬ : электронно-библиотечная система : [коллекция «Инженерно-технические науки»] / ООО «Издательство ЛАНЬ». – Санкт-Петербург, [200 – ]. – URL: <http://e.lanbook.com>. – Режим доступа: для авторизир. пользователей;

3 НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU : база данных / ООО «НЭБ». – Москва, [200 – ]. – URL: <http://elibrary.ru>. – Режим доступа: по подписке;

4 Образовательная платформа ЮРАЙТ / ООО «Электронное издательство ЮРАЙТ». – Москва, [200 – ]. – URL: <https://urait.ru>. – Режим доступа: для авторизир. пользователей;

5 Университетская библиотека онлайн : электронно-библиотечная система / ООО «Директ-Медиа». – Москва, [200 – ]. – URL: <https://biblioclub.ru>. – Режим доступа: для авторизир. пользователей. – URL: <http://www.biblioclub.ru>;

6 Электронная библиотека // Научно-техническая библиотека СибГИУ : сайт. – Новокузнецк, [200 – ]. – URL: <http://library.sibsiu.ru/LibrELibraryFullText.asp>. – Режим доступа: для авторизир. пользователей. – URL: <https://library.sibsiu.ru/LibrELibraryFullText.asp>;

7 Электронные периодические издания ИВИС : универсальная база данных / ООО «ИВИС». – Москва, [200 – ]. – URL: <http://eivis.ru>. – Режим доступа: по подписке;

8 Электронный каталог : сайт / Научно-техническая библиотека СибГИУ. – Новокузнецк, [199 – ]. – URL: <http://libr.sibsiu.ru>. – URL: <https://libr.sibsiu.ru>.

**в) лицензионное и свободно распространяемое программное обеспечение:**

- 7-Zip;
- Adobe Acrobat Reader;
- Astra Linux Special Edition;
- Kaspersky Endpoint Security;
- Microsoft Office;
- Microsoft Windows;
- OnlyOffice;
- Р7-Офис.

**г) базы данных и информационно-справочные системы:**

1 ГАРАНТ : справочно-правовая система / ООО «Правовой центр «Гарант». – Кемерово, [200 – ]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.;

2 КонсультантПлюс : справочно-правовая система / ООО «Информационный центр АНВИК». – Новокузнецк, [199 – ]. – Режим доступа: компьютерная сеть библиотеки Сиб. гос. индустр. ун-та.;

3 Техэксперт : информационно-справочная система / ООО «Группа компаний «Кодекс». – Кемерово, [200 – ]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.

## **11 Материально-техническое обеспечение учебной дисциплины**

Материально-техническое обеспечение учебной дисциплины включает учебные аудитории, оснащенные оборудованием, компьютерной техникой, и техническими средствами обучения, в том числе:

- учебную аудиторию для проведения занятий лекционного типа, оборудованную учебной доской, экраном и мультимедийным проектором;
- учебную аудиторию для проведения занятий семинарского типа (практических занятий), оснащенную учебной доской, экраном и мультимедийным проектором;
- учебную аудиторию (помещения) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду, научно-техническую библиотеку СибГИУ.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению подготовки (специальности) 09.04.03 «Прикладная информатика».

Составитель(и):

доцент Маслова Елена Владимировна (кафедра прикладных информационных технологий и программирования).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.



## Приложение

### Аннотация

#### рабочей программы дисциплины «Наука о данных и аналитика больших данных»

по направлению подготовки (специальности)

**09.04.03 «Прикладная информатика»**

(направленность (профиль): «Прикладная информатика»)

форма обучения – Очная форма

#### **1 Цели и задачи освоения учебной дисциплины**

Целями учебной дисциплины являются:

- овладение методами анализа и обработки больших объемов данных для принятия обоснованных решений на основе современных технологий.

Задачами учебной дисциплины являются:

- Изучение основных теоретических принципов обработки данных, включая методы сбора, хранения, анализа и визуализации информации;
- овладение практическими навыками работы с различными инструментами и технологиями для обработки и анализа больших объемов данных, такими как Python, SQL, алгоритмы машинного обучения и фреймворки для работы с Big Data.;
- практическое применение полученных знаний и навыков на практике через выполнение проектов по обработке и анализу данных.

#### **2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)**

Учебная дисциплина относится к учебным дисциплинам части, формируемой участниками образовательных отношений **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 09.04.03 «Прикладная информатика».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:

- Математические и инструментальные методы анализа данных;
- Машинное обучение;
- Программирование глубоких нейронных сетей на Python;
- Анализ текстовых данных.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:

- Методы и инструменты цифровой трансформации.

### 3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

#### – Профессиональные компетенции

Наименование категории (группы) ПК	Код и наименование ПК	Код и наименование индикатора достижения ПК	Планируемые результаты обучения
	ПК-1: Способен участвовать в управлении работами по получению, хранению и обработке больших объемов данных	ПК-1.1 Принимает участие в разработке моделей данных, проводит анализ больших объемов данных, строит модели на основе данных	– знать: методы моделирования данных, включая статистическое, машинное обучение и подходы к глубокому обучению. – уметь: разрабатывать модели данных для конкретных приложений, анализировать большие объемы данных, используя соответствующие инструменты и методы.
		ПК-1.2 Принимает участие в разработке методов преобразования данных	– знать: основы различных методов преобразования данных, включая очистку данных, нормализацию, агрегацию и инженерию функций. – уметь: применять инструменты и методы для эффективного преобразования необработанных данных.
		ПК-1.3 Принимает участие в разработке методов обработки больших данных	– знать: методы обработки больших данных, включая распределенные вычисления, параллельную обработку и алгоритмы интеллектуального анализа данных. – уметь: применять алгоритмы и методы программирования для разработки и

			реализации эффективных и масштабируемых решений.
--	--	--	--

#### 4 Объем учебной дисциплины

Семестр / курс		<b>ИТОГО</b>	<b>6 семестр</b>
Форма промежуточной аттестации			экзамен
Трудоёмкость	<i>академ. час.</i>	<b>144</b>	144
	<i>зачетных единиц</i>	<b>4</b>	4
Лекции, <i>академ. час.</i>		<b>16</b>	16
в форме практической подготовки		<b>0</b>	0
Лабораторные работы, <i>академ. час.</i>		<b>0</b>	0
в форме практической подготовки		<b>0</b>	0
Практические занятия, <i>академ. час.</i>		<b>32</b>	32
в форме практической подготовки		<b>0</b>	0
Курсовая работа / проект, <i>академ. час.</i>		<b>0</b>	0
в форме практической подготовки		<b>0</b>	0
Консультации, <i>академ. час.</i>		<b>0</b>	0
в форме практической подготовки		<b>0</b>	0
Самостоятельная работа, <i>академ. час.</i>		<b>78</b>	78
в форме практической подготовки		<b>0</b>	0
Контроль, <i>академ. час.</i>		<b>18</b>	18
в форме практической подготовки		<b>0</b>	0

#### 5 Краткое содержание учебной дисциплины

В структуре учебной дисциплины выделяются следующие основные разделы (темы):

Раздел 1 Основы аналитики больших данных;

Тема 1.1 Введение в большие данные (Определение больших данных и причины их появления. Примеры возможностей для бизнеса. Различие между Business Intelligence и Big Data);

Тема 1.2 Жизненный цикл аналитики данных (Понятие жизненного цикла аналитики данных. Роли, необходимые для успешного создания проекта по аналитике данных);

Тема 1.3 Высокопроизводительные вычисления (Распределенные вычисления на нескольких серверах, вычислительная парадигма MapReduce. Проект Apache Hadoop и его экосистема. Apache Spark и его компоненты. Вычисления в реальном времени, Apache Storm, Flink);

Раздел 2 Расширенные приложения аналитики больших данных;

Тема 2.1 Масштабирование и многоуровневое хранение данных (Теорема CAP. Парадигма NoSQL. Классификация NoSQL баз данных);

Тема 2.2 Визуализация данных и результатов анализа (Техники визуализации данных, введение в язык R. Визуализация данных в R);

Тема 2.3 Сложные методы аналитики (Классификация задач анализа: Text, Data, Web, Social Mining. Применение машинного обучения в аналитике. K-means и C-means кластеризация,

классификация. Логистическая регрессия, ассоциации, алгоритм Априори.);

Тема 2.4 Анализ текста (Поисковые механизмы: Lucene, Solr, ElasticSearch..Алгоритмы Work2Vec и Glove).

**6 Составитель(и):**

доцент Маслова Елена Владимировна (кафедра прикладных информационных технологий и программирования).