

УТВЕРЖДАЮ

Проректор по УМР и КО,  
 председатель УМС СГСПУ

 Н.Н. Кислова

# Введение в анализ данных

## рабочая программа дисциплины (модуля)

Закреплена за кафедрой	<b>Информатики, прикладной математики и методики их преподавания</b>		
Учебный план	ФМФИ-623ПИо(4г) Направление подготовки 09.03.03 Прикладная информатика Направленность (профиль): «Корпоративные информационные системы»		
Квалификация	<b>бакалавр</b>		
Форма обучения	<b>очная</b>		
Общая трудоемкость	<b>3 ЗЕТ</b>		
Часов по учебному плану	108	Виды контроля в семестрах:	
в том числе:		зачет с оценкой 5	
аудиторные занятия	42		
самостоятельная работа	66		

**Распределение часов дисциплины по семестрам**

Семестр(Курс.Номер семестра на курсе)	5(3.1)		Итого	
	УП	РПД	УП	РПД
Лекции	16	16	16	16
Семинарские занятия	26	26	26	26
В том числе инт.	10	10	10	10
Итого ауд.	42	42	42	42
Контактная работа	42	42	42	42
Сам. работа	66	66	66	66
Итого	108	108	108	108

Программу составил(и):  
Макарова Елена Леонидовна

При наличии обучающихся из числа лиц с ограниченными возможностями здоровья, которым необходим особый порядок освоения дисциплины (модуля), по их желанию разрабатывается адаптированная к ограничениям их здоровья рабочая программа дисциплины (модуля).

Рабочая программа дисциплины

### **Введение в анализ в данных**

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.03.03 Прикладная информатика (уровень бакалавриата) (приказ Минобрнауки России от 19.09.2017 г. № 922)

составлена на основании учебного плана:

Направление подготовки 09.03.03 Прикладная информатика  
Направленность (профиль): «Корпоративные информационные системы»

утвержденного Учёным советом СГСПУ от 28.10.2022 протокол №4

Рабочая программа одобрена на заседании кафедры

**Информатики, прикладной математики и методики их преподавания**

Протокол от 25.10.2022 г. №3

Зав. кафедрой Добудько Т.В.

Начальник УОП



Н.А. Доманина

**1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)**

**Цель изучения дисциплины:** освоение обучающимися основных методов, технологий и инструментальных средств современного анализа данных.

**Задачи изучения дисциплины:**

- сформировать систематизированное представление о данных текущего учета, статистических данных, больших данных, требованиях к ним, методам их сбора, хранения и очистки;
- сформировать навыки работы с одним из востребованных инструментов анализа данных;
- дать представление о проблемах формализации задачи, интерпретации полученных результатов.

**Область профессиональной деятельности:** 06 Связь, информационные и коммуникационные технологии

**2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

Цикл (раздел) ОП: Б1.В.02

**2.1 Требования к предварительной подготовке обучающегося:**

Содержание дисциплины базируется на материале:

Математика для анализа данных

Базы данных

Объектно-ориентированное программирование

**2.2 Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:**

Анализ данных и методы искусственного интеллекта

Машинное обучение и нейронные сети

**3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)**

**УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений**

**УК-2.1. Формулирует совокупность взаимосвязанных задач в рамках поставленной цели работы, обеспечивающих ее достижение. Определяет ожидаемые результаты решения поставленных задач**

Знает: Формулирует совокупность взаимосвязанных задач в рамках поставленной проблемы анализа больших данных и подбирает необходимый инструментарий (библиотеки Python)

**УК-2.2. Проектирует решение конкретной задачи проекта, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений**

Умеет: проектировать решение конкретной задачи анализа данных, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений.

**УК-2.3 Качественно решает конкретные задачи (исследования, проекта, деятельности) за установленное время**

Владеет: навыками качественно решать конкретные задачи анализа данных, в том числе их сбора за установленное время.

**УК-2.4. Публично представляет результаты решения задач исследования, проекта, деятельности**

Владеет навыками публичного представления результатов решения задач исследования, проекта, деятельности в области анализа данных.

**4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)**

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Интеракт.
	<b>Раздел 1. Введение в анализ данных</b>			
1.1	Ведение в анализ больших данных. Обзор источников информации /Лек/	5	2	2
1.2	Ведение в анализ больших данных. Обзор источников информации /Сем/	5	4	0
1.3	Ведение в анализ больших данных. Обзор источников информации/Ср/	5	10	0
1.4	Технологии хранения и обработки больших данных /Лек/	5	2	2
1.5	Технологии хранения и обработки больших данных /Сем/	5	4	0
1.6	Технологии хранения и обработки больших данных /Ср/	5	10	0
1.7	Статистические методы анализа данных /Лек/	5	4	0
1.8	Статистические методы анализа данных /Сем/	5	6	2
1.9	Статистические методы анализа данных /Ср/	5	16	0
1.10	Современные программные средства анализа больших данных /Лек/	5	8	0
1.11	Современные программные средства анализа больших данных /Сем/	5	8	2
1.12	Современные программные средства анализа больших данных /Ср/	5	20	0
1.13	Визуализация данных. Дашборды /Сем/	5	4	2
1.14	Визуализация данных. Дашборды /Ср/	5	10	0

**5. Оценочные и методические материалы по дисциплине (модулю)**

**5.1. Содержание аудиторной работы по дисциплине (модулю)**

5 семестр, 8 лекций, 13 семинарских занятий

Раздел 1. Введение в анализ данных

Лекция №1 (2 часа)

Ведение в анализ больших данных. Обзор источников информации

Вопросы и задания

1. Основные определения, термины, задачи анализа больших данных. Вопросы безопасности.

2. Понятие Data Mining. Когнитивный анализ данных.
3. Обзор источников информации для Big Data (открытые источники информации: статистические сборники, опубликованные отчеты и результаты исследований; доступ к закрытой информации).
4. Методики сбора данных.

Лекция №2 (2 часа)

Технологии хранения и обработки больших данных

Вопросы и задания

1. Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.

Лекции №3-4 (4 часа)

Статистические методы анализа данных

Вопросы и задания

1. Основные понятия математической статистики.
2. Методы анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы).

Лекции №5-8 (8 часов)

Современные программные средства анализа больших данных

Вопросы и задания

1. Обзор современных популярных программных средства анализа данных: Statistica, SPSS, Excel, R-Studio и другие; их преимущества и недостатки.
2. Библиотеки Python для анализа данных. Установка и настройка.
3. Поддерживаемые структуры данных и их преобразование.
4. Очистка данных

Семинарские занятия №1-4 (8 часов)

Ведение в анализ больших данных. Обзор источников информации. Технологии хранения и обработки больших данных.

Вопросы и задания

1. Поиск источников информации в сети Интернет: открытые и закрытые источники данных.
2. Портал открытых данных РФ. Сохранение данных, преобразование и первичная обработка данных.

Семинарские занятия №5-13 (18 часов)

Статистические методы анализа данных. Современные программные средства анализа больших данных. Визуализация данных. Дашборды.

Вопросы и задания

1. Представление исходных данных в программах для обработки данных (векторы, массивы, матрицы, списки, таблицы). Статистическая обработка данных в программах табличный процессор, Python, R-Studio: подсчет описательных статистик, графическое представление данных.
2. Группировка данных, обнаружение значимых корреляций, зависимостей и тенденций в результате анализа имеющейся информации, выявления отношений между данными различного типа.
3. Применение различных методов выделения, извлечения и группировки данных, которые позволяют выявить систематизированные структуры данных и вывести из них правила для принятия решений и прогнозирования их последствий (корреляционный, регрессионный, анализы).
4. Возможности графического представления информации в различных программах (: графические функции отображения одномерных и многомерных данных, графический вывод с использованием графических параметров.

## 5.2. Содержание самостоятельной работы по дисциплине (модулю)

### Содержание обязательной самостоятельной работы по дисциплине

№ п/п	Темы дисциплины	Содержание самостоятельной работы	Продукты деятельности
1	Введение в анализ больших данных. Обзор источников информации	Решение задач повышенной сложности по индивидуальному заданию	Выполненное задание
2	Технологии хранения и обработки больших данных	Решение задач повышенной сложности по индивидуальному заданию	Выполненное задание
3	Статистические методы анализа	Решение задач повышенной сложности по индивидуальному заданию	Выполненное задание
4	Современные программные средства анализа больших данных	Решение задач повышенной сложности по индивидуальному заданию	Выполненное задание
5	Визуализация данных. Дашборды	Решение задач повышенной сложности по индивидуальному заданию	Выполненное задание

### Содержание самостоятельной работы по дисциплине на выбор

№ п/п	Темы дисциплины	Содержание самостоятельной работы	Продукты деятельности
1	Визуализация данных. Дашборды	Создание дашборда по открытым данным государственных органов	Дашборд
2	Все темы дисциплины	Участие в олимпиадах, конкурсах	Грамота, сертификат

### 5.3. Образовательные технологии

При организации изучения дисциплины будут использованы следующие образовательные технологии: информационно-коммуникационные технологии, технология организации самостоятельной работы, технология рефлексивного обучения, технология модульного обучения, технология игрового обучения, технологии групповой дискуссии, интерактивные технологии, технология проблемного обучения, технология организации учебно-исследовательской деятельности, технология проектного обучения, технология развития критического мышления.

### 5.4. Текущий контроль, промежуточный контроль и промежуточная аттестация

Балльно-рейтинговая карта дисциплины оформлена как приложение к рабочей программе дисциплины. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине оформлен отдельным документом.

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

### 6.1. Рекомендуемая литература

#### 6.1.1. Основная литература

	Авторы, составители	Заглавие, ссылка на электронную библиотечную систему	Издательство, год
Л1.1	Агалаков, С. А.	Анализ данных в среде R: практикум URL: <a href="https://biblioclub.ru/index.php?page=book&amp;id=614033">https://biblioclub.ru/index.php?page=book&amp;id=614033</a>	Омск: Омский государственный университет им. Ф.М. Достоевского (ОмГУ), 2020
Л1.2	Шелудько, В. М.	Основы программирования на языке высокого уровня Python: учебное пособие URL: <a href="https://biblioclub.ru/index.php?page=book&amp;id=500056">https://biblioclub.ru/index.php?page=book&amp;id=500056</a>	Ростов-на-Дону; Таганрог: Южный федеральный университет, 2017
Л1.3	Целых, А. Н.	Современные методы прикладной информатики в задачах анализа данных: учебное пособие по курсу «Методы интеллектуального анализа данных» URL: <a href="https://biblioclub.ru/index.php?page=book&amp;id=683920">https://biblioclub.ru/index.php?page=book&amp;id=683920</a>	Ростов-на-Дону; Таганрог: Южный федеральный университет, 2021

#### 6.1.2. Дополнительная литература

	Авторы, составители	Заглавие, ссылка на электронную библиотечную систему	Издательство, год
Л2.1	Горелов, В. И.	Анализ статистических данных: практикум URL: <a href="https://biblioclub.ru/index.php?page=book&amp;id=574944">https://biblioclub.ru/index.php?page=book&amp;id=574944</a>	Москва: Университетская книга, 2015
Л2.2	Кругиков, В. Н.	Анализ данных: учебное пособие URL: <a href="https://biblioclub.ru/index.php?page=book&amp;id=278426">https://biblioclub.ru/index.php?page=book&amp;id=278426</a>	Кемерово: Кемеровский государственный университет, 2014

### 6.2 Перечень программного обеспечения

- Acrobat Reader DC
- Dr.Web Desktop Security Suite, Dr.Web Server Security Suite
- GIMP
- Microsoft Office 365 Pro Plus - subscription license (12 month) (Пакет программ Word, Excel, Access, PowerPoint, Outlook, OneNote, Publisher, Teams, OneDrive, Yammer, Stream, SharePoint Online).
- Microsoft Windows 10 Education
- XnView
- Архиватор 7-Zip

### 6.3 Перечень информационных справочных систем, профессиональных баз данных

- ЭБС «Университетская библиотека онлайн»
- Базы данных Springer eBooks

## 7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Наименование специального помещения: учебная аудитория для проведения занятий лекционного типа, лабораторных занятий, групповых консультаций, индивидуальных консультаций, текущего контроля, промежуточной аттестации, Учебная аудитория. Оснащенность: Меловая доска-1шт., Комплект учебной мебели
7.2	Наименование специального помещения: помещение для самостоятельной работы, Читальный зал. Оснащенность: ПК-4шт. с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду СГСПУ, Письменный стол-4 шт., Парта-2 шт.

## 8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Работа над теоретическим материалом происходит кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удается разобраться в материале, необходимо сформулировать вопрос и задать преподавателю.

Проработка рабочей программы дисциплины, уделяя особое внимание целям и задачам, структуре и содержанию дисциплины. Конспектирование источников, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы, работа с информационными источниками в разных форматах.

Также в процессе изучения дисциплины методические рекомендации могут быть изданы отдельным документом.

Курс 3 Семестр 5

Вид контроля		Минимальное количество баллов	Максимальное количество баллов
<b>Наименование раздела</b>			
Текущий контроль по разделу:			
1	Аудиторная работа	13	26
2	Самостоятельная работа (специальные обязательные формы)	2	8
3	Самостоятельная работа (специальные формы на выбор студента)	5	16
Контрольное мероприятие по разделу			
Промежуточный контроль		20	40
Промежуточная аттестация		36	60
<b>Итого:</b>		<b>56</b>	<b>100</b>

Вид контроля	Перечень или примеры заданий, критерии оценки и количество баллов	Темы для изучения и образовательные результаты
Текущий контроль по разделу		
Аудиторная работа	<p>Семинарское занятие №1 Поиск источников информации в сети Интернет: открытые и закрытые источники данных.                      Вопросы для семинарского занятия</p> <ol style="list-style-type: none"> <li>1. Дайте определение понятия «анализ данных».</li> <li>2. Перечислите основные задачи анализа данных.</li> <li>3. Приведите примеры применения методов анализа данных.</li> <li>4. Приведите пример актуального направления в области анализа данных.</li> <li>5. Приведите алгоритм построения системы анализа данных.</li> </ol> <p>Практические задания</p> <ol style="list-style-type: none"> <li>1. Найдите в сети Интернет два сайта, на которых используются системы прогнозирования.</li> <li>2. Найдите в сети Интернет два сайта, на которых используются рекомендательные системы.</li> <li>3. Пользуясь системой SCOPUS, проанализируйте динамику количества публикаций за пять лет по направлениям Deep Learning, Big Data, Recommender Systems, Social Network Analysis.</li> <li>4. Пользуясь системой SCOPUS, найдите</li> </ol>	<p>Тема</p> <p>Введение в анализ больших данных. Обзор источников информации</p> <p>Тема</p> <p>Технологии хранения и обработки больших данных</p> <p>Тема</p> <p>Статистические методы анализа</p> <p>Тема</p> <p>Современные программные средства анализа больших данных</p> <p>Тема</p> <p>Библиотеки Python для анализа данных. Установка и настройка.</p> <p>Тема</p> <p>Визуализация данных. Дашборды</p> <p>Образовательные результаты:                      Знает: Формулирует совокупность взаимосвязанных задач в рамках поставленной проблемы анализа</p>

	<p>пять публикаций с наибольшей цитируемостью публикаций за последние десять лет по направлениям Deep Learning, Big Data, Recommender Systems, Social Network Analysis. 5.Пользуясь системами SCOPUS, Web of Science, E-library (РИНЦ), выявите нескольких ведущих ученых в сфере анализа данных</p> <p>Критерий оценивания:                  1 балл – выполнена теоретическая часть занятия                  2 балла – выполнена практическая часть занятия                  Итого – 13х2=26 баллов</p>	<p>больших данных и подбирает необходимый инструментарий (библиотеки Python)</p> <p>Умеет: проектировать решение конкретной задачи анализа данных, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений.</p> <p>Владеет: навыками качественно решать конкретные задачи анализа данных, в том числе их сбора за установленное время; навыками публичного представления результатов решения задач исследования, проекта, деятельности в области анализа данных.</p>
<p>Самостоятельная работа (обяз.)</p>	<p>Решение задач повышенной сложности по индивидуальному заданию                  Каждый пример оценивается в 0-2 балла.                  Итого – 4х2= 8баллов</p>	<p>Тема</p> <p>Ведение в анализ больших данных. Обзор источников информации</p> <p>Тема</p> <p>Технологии хранения и обработки больших данных</p> <p>Тема</p> <p>Статистические методы анализа</p> <p>Тема</p> <p>Современные программные средства анализа больших данных</p> <p>Тема</p> <p>Библиотеки Python для анализа данных. Установка и настройка.</p> <p>Тема</p> <p>Визуализация данных. Дашборды</p> <p>Образовательные результаты:                  Знает: Формулирует совокупность взаимосвязанных задач в рамках поставленной проблемы анализа больших данных и подбирает необходимый инструментарий (библиотеки Python)</p> <p>Умеет: проектировать решение конкретной задачи анализа данных, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений.</p> <p>Владеет: навыками качественно решать конкретные задачи анализа данных, в том числе их сбора за установленное время; навыками публичного представления результатов решения задач исследования, проекта, деятельности в области анализа данных.</p>

Направление подготовки 09.03.03 Прикладная информатика  
 Направленность (профиль): «Корпоративные информационные системы»  
 Рабочая программа дисциплины «Введение в анализ данных»

Самостоятельная работа (на выбор)	<p>Создание дашборда по открытым данным государственных органов.                  дашборд                  Итого 4 балла</p> <p>Участие в олимпиадах, конкурсах</p> <p>Критерии оценивания:                  I или II место – 12 баллов                  III место (приз зрительских симпатий / другое) – 10 баллов                  участие – 5 баллов</p>	<p>Тема</p> <p>Ведение в анализ больших данных. Обзор источников информации</p> <p>Тема</p> <p>Технологии хранения и обработки больших данных</p> <p>Тема</p> <p>Статистические методы анализа</p> <p>Тема</p> <p>Современные программные средства анализа больших данных</p> <p>Тема</p> <p>Библиотеки Python для анализа данных. Установка и настройка.</p> <p>Тема</p> <p>Визуализация данных. Дашборды</p> <p>Образовательные результаты:                  Знает: Формулирует совокупность взаимосвязанных задач в рамках поставленной проблемы анализа больших данных и подбирает необходимый инструментарий (библиотеки Python)</p> <p>Умеет: проектировать решение конкретной задачи анализа данных, выбирая оптимальный способ ее решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений.</p> <p>Владеет: навыками качественно решать конкретные задачи анализа данных, в том числе их сбора за установленное время; навыками публичного представления результатов решения задач исследования, проекта, деятельности в области анализа данных.</p>
Контрольное мероприятие по модулю	–	
Промежуточный контроль (кол-во баллов)	Минимальное количество баллов – 20, максимальное – 40	
Промежуточная аттестация	Представлены в фонде оценочных средств для промежуточной аттестации по дисциплине	