

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Смирнов Сергей Николаевич  
Должность: врио ректора  
Дата подписания: 08.11.2023 10:11:22  
Уникальный программный ключ:  
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации  
ФГБОУ ВО «Тверской государственный университет»

Утверждаю:  
Руководитель ООП  
Н.А. Семькина  
  
« 4 » 09  


Рабочая программа дисциплины (с аннотацией)

**Аналитика больших данных**

Специальность

**10.05.01 Компьютерная безопасность**

Специализация

**«Математические методы защиты информации»**

Для студентов очной и заочной форм обучения

**СПЕЦИАЛИТЕТ**

Для студентов 4 курса ОФО

Составитель:

Царькова Е.Г.



Тверь 2023

## **I. Аннотация**

### **1. Цель и задачи дисциплины**

Целью освоения дисциплины является овладение студентами моделями и методами интеллектуального анализа данных и машинного обучения в задачах поиска информации, обработки и анализа данных, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа больших данных (Big Data).

**Задачами** освоения дисциплины являются:

- ознакомление с целями и задачами интеллектуального анализа данных;
- изучение основных моделей, методов, технологий анализа данных и машинного обучения;
- формирование навыков адекватного применения указанных моделей и методов, а также программных средств, в которых они реализованы;
- формирование навыков (приобретение опыта) анализа реальных данных с помощью изученных методов.

### **2. Место дисциплины в структуре ООП**

Данная дисциплина относится к части, формируемой участниками образовательных отношений, связана с другими дисциплинами образовательной программы: «Информатика», «Математическая логика и теория алгоритмов».

Дисциплины, для которых освоение данной дисциплины необходимо как предшествующее: «Производственная практика «Проектно-технологическая практика», «Производственная практика «Научно-исследовательская работа», «Подготовка к процедуре защиты и защита выпускной квалификационной работы»

**3. Объем дисциплины:** 3 зачетные единицы, 108 академических часов, в том числе:

- контактная аудиторная работа: лекции – 34 часов, в т.ч. практическая подготовка – 0 часов;
- лабораторные занятия – 34 час, в т.ч. практическая подготовка – 4 часа;
- самостоятельная работа: 40 часов, в т.ч. контроль 27 часов.

### **4. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы**

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
<b>УК-1.</b> Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	<b>УК-1.1.</b> Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	<b>УК-1.2.</b> Определяет пробелы в информации, необходимой для решения проблемной ситуации, и

	проектирует процессы по их устранению
	<b>УК-1.3.</b> Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников
	<b>УК-1.4.</b> Разрабатывает и содержательно аргументирует стратегию решения проблемной ситуации на основе системного и междисциплинарного подходов
	<b>УК-1.5.</b> Строит сценарии реализации стратегии, определяя возможные риски и предлагая пути их устранения
<b>ПК-1.</b> Способен участвовать в теоретических и экспериментальных научно-исследовательских работах по оценке защищенности информации в компьютерных системах, составлять научные отчеты, обзоры по результатам выполнения исследований	<b>ПК-1.1.</b> Разрабатывает методики выполнения аналитических работ
<b>ПК-3.</b> Способен применять методы и методики оценивания безопасности компьютерных систем при проведении контрольного анализа системы защиты	<b>ПК-3.3.</b> Проводит анализ безопасности компьютерных систем

**5. Форма промежуточной аттестации и семестр прохождения** – экзамен в 10 семестре.

**6. Язык преподавания русский.**

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

#### Очная форма обучения

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)			Самостоятельная работа, в том числе Контроль (час.)
		Лекции	Практические занятия		
			всего	в т.ч. практическая подготовка	
Тема 1. Понятие интеллектуального анализа данных. Data Mining как часть рынка информационных технологий.	6	2	2	0	2

Тема 2. Понятие данных. Набор данных и их атрибуты. Измерения. Типы наборов данных. Форматы хранения данных. Классификация видов данных. Метаданные.	12	4	4	2	4
Тема 3. Методы и стадии Data Mining. Задачи Data Mining: информация и знания. Сопоставление и сравнение понятий «информация», «данные», «знание».	10	2	2	0	6
Тема 4. Задачи Data Mining: классификация и кластеризация. Задачи и процесс классификации. Методы и точность классификации. Оценивание классификационных методов. Задача, процесс и оценка качества кластеризации. Применение кластерного анализа.	12	4	4	0	4
Тема 5. Задачи Data Mining: прогнозирование и визуализация. Задача прогнозирования. Прогнозирование и временные ряды. Сравнение задач прогнозирования и классификации. Задача визуализации.	6	2	2	0	2
Тема 6. Сферы применения Data Mining.	10	4	4	0	2
Тема 7. Анализ данных. Описательная статистика. Корреляционный анализ. Регрессионный анализ.	10	2	2	0	6
Тема 8. Деревья решений. Метод опорных векторов. Метод «ближайшего соседа», байесовская классификация.	12	4	4	2	4
Тема 9. Нейронные сети: методы классификации и прогнозирования.	10	2	2	0	6
Тема 10. Нейронные сети: самоорганизующиеся карты Кохонена.	10	4	4	0	2
Тема 11. Инструментальные средства Data Mining, Big Data (Hadoop и аналоги).	10	4	4	0	2

ИТОГО	108	34	34	4	40
-------	-----	----	----	---	----

### III. Образовательные технологии

Учебная программа – наименование разделов и тем	Вид занятия	Образовательные технологии
Тема 1. Понятие интеллектуального анализа данных. Data Mining как часть рынка информационных технологий.	лекция лабораторные	Дискуссионные технологии, дистанционные образовательные технологии, проблемная лекция.
Тема 2. Понятие данных. Набор данных и их атрибуты. Измерения. Типы наборов данных. Форматы хранения данных. Классификация видов данных. Метаданные.		
Тема 3. Методы и стадии Data Mining. Задачи Data Mining: информация и знания. Сопоставление и сравнение понятий «информация», «данные», «знание».		
Тема 4. Задачи Data Mining: классификация и кластеризация. Задачи и процесс классификации. Методы и точность классификации. Оценивание классификационных методов. Задача, процесс и оценка качества кластеризации. Применение кластерного анализа.		
Тема 5. Задачи Data Mining: прогнозирование и визуализация. Задача прогнозирования. Прогнозирование и временные ряды. Сравнение задач прогнозирования и классификации. Задача визуализации.	лекция лабораторные	Дискуссионные технологии, дистанционные образовательные технологии, проблемная лекция, кейс-технология, технология развития креативного мышления
Тема 6. Сферы применения Data Mining.		
Тема 7. Анализ данных. Описательная статистика. Корреляционный анализ. Регрессионный анализ.		
Тема 8. Деревья решений. Метод опорных векторов. Метод «ближайшего соседа», байесовская классификация.		

Тема 9. Нейронные сети: методы классификации и прогнозирования.		
Тема 10. Нейронные сети: самоорганизующиеся карты Кохонена.		
Тема 11. Инструментальные средства Data Mining, Big Data (Hadoop и аналоги).		

#### **IV. Оценочные материалы для проведения текущей и промежуточной аттестации**

##### ***Оценочные материалы для проведения текущей аттестации***

##### Задания для практических (семинарских) занятий

###### **Тема 1.**

**Задание 1 (УК-1.1, УК-1.3, УК-1.4.):** Раскройте понятие Data Mining. Дайте классификацию систем Business Intelligence и аргументируйте выбор платформ для анализа данных.

**Задание 2 (УК-1.4, ПК-1.1):** Приведите примеры, демонстрирующие сложность разработки и эксплуатации приложений Data Mining.

###### **Тема 2.**

**Задание 1 (УК-1.1, УК-1.3):** Что такое большие данные? Раскройте понятия «измерения», «шкалы», «генеральная совокупность», «большие данные».

###### **Тема 3.**

**Задание 1 (УК-1.3):** Перечислите свойства информации. Какие требования предъявляются к информации? Сопоставьте и сравните понятия «информация», «данные», «знания».

###### **Тема 4.**

**Задание 1 (УК-1.4, УК-1.5, ПК-1.1):** Имеется выборка размера 40x100. Объекты выборки – технические средства. 20 технических средств признаны неисправными, оставшиеся 20 – исправны. Признаки – значения 20 параметров. Необходимо определить параметры, свидетельствующие о неисправности технического средства. Решить эту задачу средствами пакета анализа данных, предварительно сформулировав ответы на следующие вопросы:

Какой тип имеет данная задача анализа данных? Почему?

Какие базовые предположения, по Вашему мнению, нужно использовать для успешного решения этой задачи? Почему?

Какой алгоритм (несколько алгоритмов) Вы бы использовали для решения этой задачи? Почему?

**Задание 2 (УК-1.1, УК-1.4, УК-1.5, ПК-1.1):** Имеется выборка размера 40x100. Объекты выборки – технические средства. 20 технических средств признаны неисправными, оставшиеся 20 – исправны. Признаки – значения 20 параметров. Необходимо определить параметры, свидетельствующие о

неисправности технического средства. Решить эту задачу средствами пакета анализа данных, предварительно сформулировав ответы на следующие вопросы:

Какой тип имеет данная задача анализа данных? Почему?

Какие базовые предположения, по Вашему мнению, нужно использовать для успешного решения этой задачи? Почему?

Какой алгоритм (несколько алгоритмов) Вы бы использовали для решения этой задачи? Почему?

#### **Тема 5.**

**Задание 1 (УК-1.3, ): Перечислите методы прогнозирования временных рядов.**

**Задание 2 (УК-1.2, УК-1.3 ): Что входит в понятия «тренд», «сезонность» и «цикл»? С применением пакета аналитического пакета на основе ретроспективных данных постройте прогноз.**

#### **Тема 6.**

**Задание 1 (УК-1.3, УК-1.4, ПК-1.1, ПК-3.3): Приведите пример применения интеллектуального анализа данных в сфере информационной безопасности, в банковской сфере, в сфере страхования, электронной коммерции, в промышленности.**

#### **Тема 7.**

**Задание 1 (УК-1.4, УК-1.5): По набору данных с применением аналитического пакета сформируйте описательную статистику (среднее, стандартная ошибка, медиана, мода, стандартное отклонение, эксцесс, интервал, минимум, максимум). Подготовьте отчет «Описательная статистика».**

**Задание 2 (УК-1.4, УК-1.5, ПК-1.1): По набору данных с применением аналитического пакета проведите регрессионный анализ данных, постройте диаграмму рассеивания, сформируйте регрессионную статистику. Подготовьте отчет «Регрессионная статистика».**

#### **Тема 8.**

**Задание 1 (УК-1.2, УК-1.5): Постройте дерево решений «Эффективность системы физической защиты объекта информатизации».**

#### **Тема 9.**

**Задание 1 (УК-1.4, УК-1.5): На основе ретроспективных данных с применением аналитической платформы Longinom выполните решение задачи прогнозирования, предварительно обучив искусственную нейронную сеть.**

#### **Тема 10.**

**Задание 1 (УК-1.4, ПК-1.1): Имеется база данных организации с показателями деятельности за текущий период (количество показателей равно 21). Необходимо провести их классификацию в аналитическом пакете Longinom с использованием самоорганизующихся карт Кохонена.**

#### **Тема 11.**

**Задание 1 (УК-1.2, ПК-1.1): Опишите основные стадии анализа данных с применением аналитической платформы Longinom. Опишите основные функциональные возможности платформы Longinom.**

**Задание 2 (УК-1.2, УК-1.3, УК-1.4):** Сравните известные Вам современные VI-системы. В чем их достоинства и недостатки? Как согласуется их использование с требованиями законодательства РФ в сфере импортозамещения?

**Оценочные материалы для проведения промежуточной аттестации**

Проверяемые индикаторы достижения компетенций: УК-1.1; УК-1.2; УК-1.3; УК-1.4; УК-1.5; ПК-1.1; ПК-1.2; ПК-3.2; ПК-3.3.

Каждый студент отвечает на вопросы теста и дает развернутый ответ на теоретический вопрос.

**Примерные вопросы к зачету**

1. Основные понятия интеллектуального анализа данных. Основные постановки задач. Примеры прикладных задач.

2. Методы прогнозирования временных рядов.

3. Случайные величины. Дискретные и непрерывные распределения. Примеры.

4. Оценивание параметров распределений, метод максимального правдоподобия. Бутстрэппинг.

5. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.

6. Метрики качества алгоритм регрессии и классификации.

7. Корреляционный анализ данных.

8. Деревья решений. Методы построения деревьев. Их регуляризация.

9. Композиции алгоритмов. Разложение ошибки на смещение и разброс.

10. Случайный лес, его особенности.

11. Методы поиска выбросов в данных. Методы восстановления пропусков в данных. Работа с несбалансированными выборками.

12. Задача анализа потребительской корзины. Поддержка и достоверность. Частые, замкнутые и максимальные частые множества.

13. Ассоциативные правила.

14. Алгоритм Априори.

15. Задача кластеризации. Алгоритм K-Means. Оценки качества кластеризации.

**Вид и способ** проведения промежуточной аттестации: индивидуальный устный опрос сочетается с самостоятельной практической работой студента.

**Критерии** оценивания и шкала оценивания:

Максимально возможное количество баллов – 3 балла. Для получения зачета необходимо ответить на вопросы теста и дать ответ на теоретический вопрос с суммарной оценкой не менее 2-х баллов.

**3 балла:**

Ответ на вопрос демонстрирует знание и корректное использование терминологии. Факты и примеры в полном объеме обосновывают выводы. Имеется решение теста верное от 85 – 100% всех заданий.



## **2 балла:**

Ответ на вопрос демонстрирует знание и корректное использование терминологии. Ответ не содержит фактических ошибок. Верно даны ответы на 70-84% тестовых заданий.

## **1 балл:**

Ответ демонстрирует знание и корректное использование терминологии. Правильные решения тестовых заданий составляют от 41-69%.

## **0 баллов:**

В ответе преобладают рассуждения общего характера И/ИЛИ содержит существенные фактические ошибки, искажающие смысл. Правильные тестовые ответы составляют менее 40%.

## **V. Учебно-методическое и информационное обеспечение дисциплины**

### **1) Рекомендуемая литература**

#### **а) Основная литература**

1. Андрейчиков А. В. Интеллектуальные информационные системы и методы искусственного интеллекта : учебник / А. В. Андрейчиков, О. Н. Андрейчикова. - 1. - Москва : ООО "Научно-издательский центр ИНФРА-М", 2023. - 530 с. - (Высшее образование: Магистратура). - ВО - Магистратура. – Режим доступа: <https://znanium.com/catalog/document?id=417737>

2. Голицына О. Л. Информационные системы и технологии : учебное пособие / О. Л. Голицына, Н. В. Максимов, И. И. Попов; Национальный исследовательский ядерный университет "МИФИ"; Российский экономический университет им. Г.В. Плеханова. - Москва : Издательство "ФОРУМ", 2023. - 400 с. - (Высшее образование). - Профессиональное образование. – Режим доступа: <https://znanium.com/catalog/document?id=427018>

3. Милютин Е. М. Интеллектуальные информационные системы: курс лекций. Ч. 1 : Интеллектуальные информационные системы / Е. М. Милютин. - Брянск : Брянский ГАУ, 2022. - 40 с. – Режим доступа : <https://e.lanbook.com/book/304886>

4. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.]; под ред. Мхитаряна В.С. - Электрон. дан. - Москва : Юрайт, 2021. - 490 с. - (Высшее образование). – URL.- Режим доступа: <https://urait.ru/bcode/469022>

#### **б) Дополнительная литература:**

1. Новикова О. А. Анализ данных : учебное пособие. Ч. 1 / О. А. Новикова, Е. Г. Андрианова. - Москва : РТУ МИРЭА, 2020. - 162 с. – Режим доступа: <https://e.lanbook.com/book/167597>

2. Нестеров С. А. Интеллектуальный анализ данных с использованием SQL Server [Электронный ресурс] : учебник для вузов / С. А. Нестеров. - Санкт-Петербург : Лань, 2023. - 160 с. – Режим доступа: <https://e.lanbook.com/book/311861>

3. Демидова Л. А. Интеллектуальный анализ данных на языке Python [Электронный ресурс] : учебно-методическое пособие / Л. А. Демидова. - Москва : РТУ МИРЭА, 2021. - 92 с. – Режим доступа: <https://e.lanbook.com/book/218693>

4. Ризаев И. С. Интеллектуальный анализ данных [Электронный ресурс] : учебное пособие / И. С. Ризаев, Э. Г. Тахавова. - Казань : КНИТУ-КАИ, 2020. - 116 с. – Режим доступа: <https://e.lanbook.com/book/264896>

5. Шелухин О. И. Искусственный интеллект и машинное обучение в кибербезопасности [Электронный ресурс] : учебно-методическое пособие для выполнения лабораторных работ. направление подготовки: 10.03.01 информационная безопасность. профили: «безопасность компьютерных систем», «безопасность автоматизированных систем» / О. И. Шелухин, А. В. Осин, Д. И. Раковский. - Москва : МТУСИ, 2022. - 52 с. – Режим доступа: <https://e.lanbook.com/book/333755>

Официальный сайт компании Loginom Company [Электронный ресурс]. – Режим доступа: <https://loginom.ru/>

Энциклопедия по Бизнес-анализу [Электронный ресурс] режим <https://wiki.loginom.ru>

Loginom – Руководство пользователя [Электронный ресурс]. – Режим доступа: <https://help.loginom.ru/userguide/>.

## **2) Программное обеспечение**

Google Chrome	бесплатно
Kaspersky Endpoint Security 10 для Windows	Акт на передачу прав ПК545 от 16.12.2022
Lazarus	бесплатно
OpenOffice	бесплатно
Многофункциональный редактор ONLYOFFICE	бесплатное ПО
ОС Linux Ubuntu	бесплатное ПО

## **3) Современные профессиональные базы данных и информационные справочные системы**

1. ЭБС Лань <https://e.lanbook.com/> Договор № 4-е/23 от 02.08.2023 г.
2. ЭБС Znanium.com <https://znanium.com/> Договор № 1106 эбс от 02.08.2023 г.
3. ЭБС Университетская библиотека online <https://biblioclub.ru> Договор № 02-06/2023 от 02.08.2023 г.
4. ЭБС ЮРАЙТ <https://urait.ru/> Договор № 5-е/23 от 02.08.2023 г.
5. ЭБС IPR SMART <https://www.iprbookshop.ru/> Договор № 3-е/23К от 02.08.2023 г.
6. <https://cyberleninka.ru/> научная электронная библиотека «Киберленинка».
7. Научная электронная библиотека eLIBRARY.RU (подписка на журналы) [https://elibrary.ru/projects/subscription/rus\\_titles\\_open.asp](https://elibrary.ru/projects/subscription/rus_titles_open.asp);
8. Репозиторий ТвГУ <http://eprints.tversu.ru>

**4) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:**

<https://cyberleninka.ru/> научная электронная библиотека «Киберленинка».

<http://www.intuit.ru/> Национальный Открытый Университете «ИНТУИТ»

[http://www.cisco.com/c/ru\\_ru/index.html](http://www.cisco.com/c/ru_ru/index.html) Сетевой Академии Cisco

**VI. Методические материалы для обучающихся по освоению дисциплины**

***Методические рекомендации по организации самостоятельной работы студентов***

На лекциях будет представлен необходимый теоретически материал по темам и представлены практические задания для решения на занятиях в аудитории под руководством преподавателя и самостоятельно. Многие задачи являются стандартными и имеют уже готовые шаблоны (алгоритмы) решения, тем не менее, для получения большего познавательного и учебного эффекта, рекомендуется написание собственного оригинального кода.

Самостоятельная работа студентов в рамках данной дисциплины в основном состоит в подготовке к практическим занятиям и работе с разными источниками. Освоению учебного материала большую помощь окажет личный творческий подход, связанный с дополнительным просмотром материала по отдельным темам.

Самостоятельная работа является необходимой на всей стадиях и при всех формах изучения предмета. Важно помнить, что часы для самостоятельной работы, из всего объема времени затраченного на дисциплину, будут превосходить иные виды работ. Важно продумать стиль фиксации нового и важного материала.

Рекомендуется немедленно обсуждать любые возникшие в процессе обучения вопросы, проблемы и неясности с преподавателем, не откладывая это обсуждение до контрольной точки. Проконсультироваться с преподавателем можно во время и после практических занятий, во время консультаций, а также по электронной почте и в личном кабинете электронной образовательной среды (LMS).

**Требования к рейтинг-контролю для студентов очной формы обучения.**

Текущая работа студентов очной формы обучения оценивается в 100 баллов, которые распределяются между двумя модулями (периодами обучения) следующим образом:

Модуль (период обучения)	Максимальная сумма баллов в модуле	Максимальная сумма баллов за работу на практических занятиях	Реферирование, представление научной статьи, создание и отладка кода	Максимальный балл за рейтинговую контрольную работу
1	50	18	12	20
2	50	18	12	20

Правила формирования рейтинговой оценки и шкалу пересчета рейтинговых баллов в оценку на экзамене см. в «Положении о рейтинговой системе обучения в ТвГУ»:

<https://tversu.ru/sveden/files/204->

[R Pologhenie o reytingovoy sisteme obucheniya v TvGU.pdf](#)

## VII. Материально-техническое обеспечение

Учебный процесс по данной дисциплине проводится в аудиториях, оснащенных мультимедийными средствами обучения. Для организации самостоятельной работы студентов необходимо наличие персональных компьютеров с доступом в Интернет.

Наименование специальных* помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
<p>Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, Учебная аудитория № 314 (Корпус 3, 170002, Тверская обл., г.Тверь, пер. Садовый, дом 35)</p>	<p>Набор учебной мебели, меловая доска, Мультимедийный комплект учебного класса</p>	<p>Google Chrome-бесплатно; Kaspersky Endpoint Security 10 для Windows-Акт на передачу прав ПК545 от 16.12.2022; Lazarus –бесплатно; OpenOffice –бесплатно; Многофункциональный редактор ONLYOFFICE бесплатное ПО- бесплатно; ОС Linux Ubuntu бесплатное ПО-бесплатно</p>
<p>Учебная аудитория для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, учебная аудитория 203, 224, 170002, г.Тверь, Садовый пер-к, д. 35</p>	<p>Стол, стулья, переносной ноутбук, проектор</p>	<p>Google Chrome-бесплатно; Kaspersky Endpoint Security 10 для Windows-Акт на передачу прав ПК545 от 16.12.2022; Lazarus –бесплатно; OpenOffice –бесплатно; Многофункциональный редактор ONLYOFFICE бесплатное ПО- бесплатно; ОС Linux Ubuntu бесплатное ПО-бесплатно</p>

Наличие учебно-наглядных пособий, презентаций для проведения занятий лекционного и семинарского типа, обеспечивающих тематические иллюстрации.

## VIII. Сведения об обновлении рабочей программы дисциплины

№п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Реквизиты документа, утвердившего изменения
1	V. Учебно-методическое и информационное обеспечение дисциплины	Обновление списков ПО. Обновление ссылок из ЭБС.	Протокол № 1 от 1.09.2023