

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Смирнов Сергей Николаевич  
Должность: врио ректора  
Дата подписания: 05.10.2023 14:33:47  
Уникальный программный ключ: 69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации

ФГБОУ ВО «Тверской государственный университет»

Утверждаю:

Руководитель ООП:

 А.В. Язенин/

2022 г.



**Рабочая программа дисциплины (с аннотацией)**

**Интеллектуальный анализ данных**

02.04.02 Фундаментальная информатика и информационные технологии

Направленность (профиль)

Информационные технологии в управлении и принятии решений

Для студентов 1 курса

Очная форма

Составитель: к.ф.-м.н., доцент Гришина Е.Н.

Тверь, 2022

## **I. Аннотация**

### **1. Цель и задачи дисциплины**

Целью освоения дисциплины является: рассмотрение вопросов обработки и анализа данных.

Задачами освоения дисциплины являются: изучение способов хранения и представления информации, изучение современных методов обработки и анализа данных, в том числе методов и моделей интеллектуального анализа данных.

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

Дисциплина «Интеллектуальный анализ данных» относится к разделу «Математический» обязательной части Блока 1.

Для успешного усвоения курса необходимы знания интегрального и дифференциального исчисления, понятия вероятности, функции плотности, моды, медианы, среднего, математического ожидания, дисперсии, корреляции и других основных терминов теории вероятностей и математической статистики, математического анализа, линейной алгебры и дифференциальных уравнений; а также навыки решения основных задач, рассматриваемых в этих дисциплинах.

Необходимо обладать следующими входными знаниями: хранение и представление информации, проектирование баз данных, проектирование хранилищ данных, владение базовыми методами математической статистики, методами многомерного статистического анализа, подходами к построению и проектированию информационных систем, систем поддержки принятия решений.

Освоение данной дисциплины необходимо как предшествующее для следующих дисциплин: «Анализ нечетких информационных систем», «Применение методов интеллектуального анализа данных».

### **3. Объем дисциплины: 6 зачетных единиц, 216 академических часов, в том числе:**

**контактная аудиторная работа:** практические занятия – 16 часов, в т.ч. практическая подготовка 11 часов; лабораторные работы – 16 часов, в т.ч. практическая подготовка 10 часов;

**самостоятельная работа:** 184 часов, в том числе контроль – 27.

### **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
---	---

<p>ОПК-1 Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий</p>	<p>ОПК-1.1 Обладает фундаментальными знаниями в области математических и естественных наук, теории коммуникаций  ОПК-1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические объекты  ОПК-1.3 Решает актуальные задачи прикладной математики, фундаментальной информатики и информационных технологий</p>
<p>ОПК-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования</p>	<p>ОПК-3.1 Знает и применяет методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей  ОПК-3.2 Соотносит знания в области программирования, интерпретирует прочитанное, определяет и создает информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем  ОПК-3.3 Разрабатывает программное обеспечение и тестирует программные продукты</p>

## 5. Форма промежуточного контроля – экзамен.

## 6. Язык преподавания - русский.

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

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)					Самостоятельная работа, в том числе Контроль (час.)
		Практические занятия		Лабораторные работы		Контроль самостоятельной работы (в том числе курсовая работа)	
		всего	в т.ч. практическая подготовка	всего	в т.ч. практическая подготовка		
<p><b>1. Системы поддержки принятия решений.</b>  1.1. Основные понятия.  1.2. Задачи систем поддержки принятия решений.</p>	23	2	-	1	-	-	20

1.3. База данных как основа систем поддержки принятия решений. 1.4. Архитектура систем поддержки принятия решений.							
<b>2. Хранилище данных.</b> 2.1. Концепция хранилища данных. 2.2. Организация хранилища данных. 2.3. Концепция хранилища данных и анализ.	23	2	-	1	-	-	20
<b>3. OLAP-системы.</b> 3.1. Многомерная модель данных. 3.2. Определение OLAP-систем, основные понятия. 3.3. Концептуальное многомерное представление. 3.4. Архитектура OLAP-систем.	23	1	-	2	2	-	20
<b>4. Интеллектуальный анализ данных.</b> 4.1. Основные понятия интеллектуального анализа данных. 4.2. Задачи интеллектуального анализа данных. 4.3. Область практического применения интеллектуального анализа данных. 4.4. Модели и методы интеллектуального анализа данных.	23	3	3	-	-	-	20
<b>5. Задача классификации и регрессии.</b> 5.1. Постановка задачи классификации и регрессии. 5.2. Методы (алгоритмы) решения задачи классификации и регрессии. 5.3. Представление результатов в задаче классификации и регрессии.	23	2	2	1	1	-	20
<b>6. Задача поиска ассоциативных правил.</b> 6.1. Постановка задачи поиска ассоциативных правил. 6.2. Методы (алгоритмы) решения задачи поиска ассоциативных правил.	25	3	3	2	1	-	20

6.3. Представление результатов в задаче поиска ассоциативных правил.							
<b>7. Задача кластеризации.</b> 7.1. Постановка задачи кластеризации. 7.2. Методы (алгоритмы) решения задачи кластеризации. 7.3. Представление результатов в задаче кластеризации.	26	3	3	3	1	-	20
<b>8. Визуальный анализ данных.</b> 8.1. Выполнение визуального анализа данных. 8.2. Характеристика средств визуализации данных. 8.3. Методы визуализации.	23	-	-	3	3	-	20
<b>9. Анализ текстовой информации.</b> 9.1. Задача анализа текстов. 9.2. Извлечение ключевых понятий из текста. 9.3. Классификация текстовых документов. 9.4. Методы кластеризации текстовых документов. 9.5. Средства анализа текстовой информации.	27	-	-	3	2	-	24
<b>ИТОГО</b>	216	16	10	16	11	--	184

### III. Образовательные технологии

Учебная программа – наименование разделов и тем (в строгом соответствии с разделом II РПД)	Вид занятия	Образовательные технологии
Системы поддержки принятия решений	Практические занятия, лабораторные работы	Компьютерные: показ презентаций
Хранилище данных	Практические занятия, лабораторные работы	Компьютерные: показ презентаций
OLAP-системы	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций
Интеллектуальный анализ данных	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций

Задача классификации и регрессии	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций
Задача поиска ассоциативных правил	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций
Задача кластеризации	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций
Визуальный анализ данных	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций
Анализ текстовой информации	Практические занятия, лабораторные работы	Компьютерные: выполнение компьютерных лабораторных работ, показ презентаций

Преподавание учебной дисциплины строится на сочетании практических занятий, лабораторных работ и различных форм самостоятельной работы студентов. В процессе освоения дисциплины используются следующие образовательные технологии, способы и методы формирования компетенций: практические занятия в диалоговом режиме, выполнение индивидуальных заданий в рамках лабораторных и самостоятельных работ, подготовка и проведение докладов.

Дисциплина предусматривает выполнение домашних заданий.

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

Для проведения текущей и промежуточной аттестации:

ОПК-1 Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий

ОПК-1.1 Обладает фундаментальными знаниями в области математических и естественных наук, теории коммуникаций

1. Опишите алгоритм Apriori.
2. Опишите алгоритмы 1-R и Naive Bayes для построения правил классификации.

Способ проведения – письменный.

Критерии оценивания:

Дан правильный развернутый ответ – 12 баллов;

Ответ содержит неточности – 5 баллов.

ОПК-1.2 Осуществляет первичный сбор и анализ материала, интерпретирует различные математические объекты

1. Приведите постановку задачи кластеризации.
2. Определите средства анализа текстовой информации.

Способ проведения – письменный.

Критерии оценивания:

Дан правильный развернутый ответ – 12 баллов;

Ответ содержит неточности – 5 баллов.

ОПК-1.3 Решает актуальные задачи прикладной математики, фундаментальной информатики и информационных технологий

1. Произведите сравнение 2 известных Вам алгоритмов кластеризации и сделайте вывод об эффективности их работы.

2. Произведите сравнение 2 известных Вам алгоритмов классификации и сделайте вывод об эффективности их работы.

Способ проведения – письменный.

Критерии оценивания:

Дан правильный развернутый ответ – 12 баллов;

Ответ содержит неточности – 5 баллов.

ОПК-3 Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования

ОПК-3.1 Знает и применяет методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей

1. Опишите алгоритм k-means и сравните уровень сложности вычислений с другими известными Вам алгоритмами кластеризации.

2. Реализуйте алгоритм k-means.

Способ проведения – письменный.

Критерии оценивания:

Дан правильный развернутый ответ – 12 баллов;

Ответ содержит неточности – 5 баллов.

ОПК-3.2 Соотносит знания в области программирования, интерпретирует прочитанное, определяет и создает информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем

1. Дайте понятие OLAP-системы и приведите характерные особенности таких систем.

2. Определите практическое применение алгоритма ID3.

Способ проведения – письменный.

Критерии оценивания:

Дан правильный развернутый ответ – 12 баллов;

Ответ содержит неточности – 5 баллов.

ОПК-3.3 Разрабатывает программное обеспечение и тестирует программные продукты

1. Реализуйте алгоритм Apriori.

2. Реализуйте один из известных вам алгоритмов кластеризации.

3. Реализуйте алгоритм покрытия.  
Способ проведения – письменный.  
Критерии оценивания:  
Дан правильный развернутый ответ – 12 баллов;  
Ответ содержит неточности – 5 баллов.

## V. Учебно-методическое и информационное обеспечение дисциплины

### 1) Рекомендуемая литература

#### а) Основная литература

1. Нестеров, С.А. Интеллектуальный анализ данных средствами MS SQL Server 2008 / С.А. Нестеров. - 2-е изд., испр. - М.: Национальный Открытый Университет «ИНТУИТ», 2016. - 338 с.: ил. - Библиогр. в кн.; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=429083>
2. Интеллектуальные системы: учебное пособие / А. Семенов, Н. Соловьев, Е. Чернопрудова, А. Цыганков; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Оренбургский государственный университет». - Оренбург: ОГУ, 2013. - 236 с.; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=259148>
3. Ярушкина Н. Г. Интеллектуальный анализ временных рядов: учебное пособие / Н.Г. Ярушкина, И.Г. Перфильева, Т.В. Афанасьева. - Москва; Москва: Издательский Дом "ФОРУМ": ООО "Научно-издательский центр ИНФРА-М", 2012. - 160 с. - ISBN 9785819904961. -Режим доступа: <http://znanium.com/go.php?id=249314>
4. Ясницкий, Л.Н. Интеллектуальные системы [Электронный ресурс]: учеб. — Электрон. дан. — Москва: Издательство "Лаборатория знаний", 2016. — 224 с. — Режим доступа: <https://e.lanbook.com/book/90254>

#### б) Дополнительная литература:

1. Богданова, С.В. Информационные технологии: учебное пособие для студентов высших учебных заведений / С.В. Богданова, А.Н. Ермакова; Министерство сельского хозяйства РФ, ФГБОУ ВПО Ставропольский государственный аграрный университет. - Ставрополь: Сервисшкола, 2014. - 211 с.: ил. - Библиогр. в кн.; [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=277476>
2. Грошев, А.С. Информационные технологии: лабораторный практикум / А.С. Грошев. - 2-е изд. - М.; Берлин: Директ-Медиа, 2015. - 285 с.: ил., табл. - Библиогр. в кн. - ISBN 978-5-4475-5065-3; То же [Электронный ресурс]. - Режим доступа: <http://biblioclub.ru/index.php?page=book&id=434666>

### 2) Программное обеспечение

Помещение для самостоятельной работы обучающихся: Компьютерный класс факультета ПМиК № 46 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)	Перечень программного обеспечения (со свободными лицензиями): Adobe Acrobat Reader DC, Anaconda3 2019.07 (Python 3.7.3 64-bit), Apache Tomcat 8.0.27, Cadence SPB/OrCAD 16.6, GlassFish Server Open Source Edition 4.1.1, Google Chrome, IntelliJ IDEA, IIS 10.0 Express, Java SE Development Kit 8 Update 191 (64-bit), JetBrains PyCharm Community Edition 2019.2.1, Kaspersky Endpoint Security для Windows, Lazarus 2.0.12, MiKTeX, NetBeans IDE 8.2, Notepad++ (64-bit x64), ONLYOFFICE Desktop Editors 7.1 (x64), Origin 8.1 Sr2, Python 3.10.7, R for Windows 3.6.1, RStudio Desktop, Visual Studio Community 2022, VLC media player, WinDjView 2.1, Unreal Commander v3.57x64
---	---

3) Современные профессиональные базы данных и информационные справочные системы

1. ЭБС «ZNANIUM.COM» [www.znanium.com](http://www.znanium.com);
2. ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/>;
3. ЭБС «Лань» <http://e.lanbook.com>.

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

<http://www.intuit.ru/department/database/datamining/>

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

Практические и лабораторные занятия осуществляются по следующему плану:

1. Проверка выполнения заданий, выданных для домашней работы.
2. Обзор пройденного материала.
3. Решение задач по очередной теме.
4. Обсуждение результатов решения задачи.
5. Выдача задач для домашней работы.

Примерные варианты тем докладов

1. Практическое применение интеллектуального анализа данных.
2. Информационное хранилище (витрины данных, информационное хранилище двухуровневой и трехуровневой архитектуры).
3. Модели данных (реляционная, сетевая, иерархическая модели данных).
4. Концепция многомерного представления данных.
5. Алгоритм ID3 и его практическое применение.
6. Алгоритм k-means и его практическое применение.

Тестовые задания

1. Какие задачи решаются с помощью систем поддержки принятия решений:
  - а) ввод данных;

- б) хранение данных;
  - в) анализ данных.
2. Какая задача, решаемая с помощью систем поддержки принятия решений, является основной:
- а) ввод данных;
  - б) хранение данных;
  - в) анализ данных.
3. С помощью какого класса задач анализа осуществляется группирование и обобщение необходимых аналитику данных:
- а) информационно-поисковый;
  - б) оперативно-аналитический;
  - в) интеллектуальный.
4. С помощью какого класса задач анализа осуществляется поиск данных на основе заранее определенных запросов:
- а) информационно-поисковый;
  - б) оперативно-аналитический;
  - в) интеллектуальный.
5. С помощью какого класса задач анализа осуществляется поиск функциональных закономерностей в накопленных данных, построение моделей и правил:
- а) информационно-поисковый;
  - б) оперативно-аналитический;
  - в) интеллектуальный.
6. Какие подсистемы входят в обобщенную архитектуру системы поддержки принятия решений:
- а) информационно-поисковая система;
  - б) подсистема ввода данных;
  - в) подсистема хранения;
  - г) подсистема анализа;
  - д) подсистема отчетов.
7. В какой подсистеме осуществляется операционная обработка данных на основе использования СУБД:
- а) информационно-поисковая система;
  - б) подсистема ввода данных;
  - в) подсистема хранения;
  - г) подсистема анализа;
  - д) подсистема отчетов.
8. В какой подсистеме подсистемы анализа реализуются методы и алгоритмы добычи данных:
- а) подсистема информационно-поискового анализа;
  - б) подсистема оперативного анализа;
  - в) подсистема интеллектуального анализа.
9. В каких системах присутствие человеческого фактора при вводе данных повышает вероятность ошибочных данных и может создать локальные проблемы в системах:
- а) система поддержки принятия решений;
  - б) OLTP – система;
  - в) система управления базами данных;
  - г) экономическая информационная система;
  - д) информационно-поисковая система.
10. К каким системам предъявляется требование обеспечения формирования произвольных запросов к базе данных:
- а) система поддержки принятия решений;
  - б) OLTP – система;

- в) система управления базами данных;
- г) экономическая информационная система;
- д) информационно-поисковая система.

11. Для каких систем является приоритетных высокая производительность и доступность данных:

- а) система поддержки принятия решений;
- б) OLTP – система;
- в) система управления базами данных;
- г) экономическая информационная система;
- д) информационно-поисковая система.

12. Хранилище данных – это

а) предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений;

б) модель некоторой предметной области, состоящая из связанных между собой данных об объектах, их свойствах и характеристиках;

в) система, обладающая свойствами ввода, хранения и анализа данных, относящихся к некоторой предметной области, с целью поиска решений.

13. Выделите свойства хранилища данных:

- а) предметная ориентация;
- б) интеграция;
- в) поддержка хронологии;
- г) неизменяемость;
- д) допущение избыточности данных;
- е) приоритетность характеристик системы.

14. Основное назначение OLAP-систем:

- а) обнаружение скрытых знаний и закономерностей;
- б) поддержка аналитической деятельности;
- в) поддержка запросов пользователей-аналитиков;
- г) ввод и хранение данных.

15. Интеллектуальный анализ данных – это

а) исследование и обнаружение в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны;

б) технология оперативной аналитической обработки данных, использующая методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений;

в) предметно-ориентированный, интегрированный, изменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений;

г) модель некоторой предметной области, состоящая из связанных между собой данных об объектах, их свойствах и характеристиках.

## Контрольные вопросы

1. Перечислите и опишите основные этапы интеллектуального анализа данных.
2. Какие методы интеллектуального анализа данных вы знаете, опишите перечисленные методы.
3. Приведите постановку задачи классификации и регрессии.
4. Какими способами может быть произведено представление обнаруженных функциональных зависимостей в задачах классификации и регрессии?
5. Опишите алгоритмы 1-R и Naive Bayes для построения правил классификации. Реализуйте алгоритмы для задачи с 4 независимыми переменными и 1 зависимой переменной.
6. Опишите и реализуйте алгоритм покрытия.

7. Приведите постановку задачи поиска ассоциативных правил.
8. Опишите и реализуйте алгоритм Apriori.
9. Приведите постановку задачи кластеризации.
10. Опишите один из известных вам алгоритмов кластеризации и реализуйте его на примере.

#### Вопросы для подготовки к экзамену

1. Системы поддержки принятия решений. Основные понятия. Задачи СППР.
2. База данных как основа СППР.
3. Архитектура ССПР.
4. Хранилище данных. Концепция хранилища данных.
5. Организация хранилища данных. Концепция хранилища данных и анализ.
6. OLAP-системы. Основные понятия.
7. Концептуальное многомерное представление.
8. Архитектура OLAP-систем.
9. Интеллектуальный анализ данных. Основные понятия ИАД.
10. Задачи ИАД. Область практического применения ИАД.
11. Модели и методы ИАД.
12. Задача классификации и регрессии. Постановка задачи.
13. Методы (алгоритмы) решения задачи классификации и регрессии.
14. Представление результатов в задаче классификации и регрессии.
15. Задача поиска ассоциативных правил. Постановка задачи.
16. Методы (алгоритмы) решения задачи поиска ассоциативных правил.
17. Представление результатов в задаче поиска ассоциативных правил.
18. Задача кластеризации. Постановка задачи.
19. Методы (алгоритмы) решения задачи кластеризации.
20. Представление результатов в задаче кластеризации.
21. Визуальный анализ данных.
22. Методы визуализации.
23. Анализ текстов, классификация текстовых документов.
24. Методы кластеризации текстовых документов.
25. Средства анализа текстовой информации.

### **VII. Материально-техническое обеспечение**

#### Для аудиторной работы

Учебная аудитория № 212 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)	Набор учебной мебели, мультимедийный комплекс (доска, проектор, панель управления).
Учебная аудитория № 205 (170002, Тверская обл., г.Тверь, Садовый переулок, д.35)	Набор учебной мебели, экран, проектор.

### Для самостоятельной работы

Помещение для самостоятельной работы обучающихся: Компьютерный класс факультета ПМиК № 46 (170002, Тверская обл., г. Тверь, Садовый переулок, д.35)	Компьютер, экран, проектор, кондиционер.
---	--

### Х. Сведения об обновлении рабочей программы дисциплины

№ п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Дата и протокол заседания кафедры, утвердившего изменения
1.	3. Объем дисциплины	Выделение часов на практическую подготовку	От 29.10.2020 года, протокол № 3 ученого совета факультета
2.	II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	Выделение часов на практическую подготовку по темам	От 29.10.2020 года, протокол № 3 ученого совета факультета
3.	V. Учебно-методическое и информационное обеспечение дисциплины 2) Программное обеспечение	Внесены изменения в программное обеспечение	От 29.09.2022 года, протокол № 2 ученого совета факультета
4.	VII. Материально-техническое обеспечение	Внесены изменения в материально-техническое обеспечение аудиторий	От 29.09.2022 года, протокол № 2 ученого совета факультета
5.			