

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ  
УПРАВЛЕНИЯ  
И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

**МЕТОДЫ И АЛГОРИТМЫ РАСПОЗНАВАНИЯ И ЦИФРОВОЙ  
ОБРАБОТКИ ДАННЫХ**

Методические указания по организации самостоятельной работы для  
студентов направления

**“Программная инженерия”**

(уровень магистратуры)

Замятин Николай Владимирович

**Методы и алгоритмы распознавания и цифровой обработки данных:** методические указания по организации самостоятельной работы для студентов направления подготовки “Программная инженерия” (уровень магистратуры) / Н.В. Замятин. – Томск, 2018- с 11.

## Содержание

1. Введение .....	4
2. Рекомендации по выполнению самостоятельной работы.....	4
3. Проработка лекционного материалп.....	5
4. Подготовка к лабораторным работам.....	7
5. Подготовка рефератов.....	8
6. Вопросы для самоподготовки.....	9
7. Темы для самостоятельного изучения и подготовки рефератов...	10
8. Подготовка к контрольным работам.....	10
9. Литература.....	11

### **1. Введение**

Дисциплина «Методы и алгоритмы распознавания и цифровой обработки данных» представляет систематическое изложение материала по методам и алгоритмам распознавания и обработки данных, и дает базовые знания, необходимые специалисту по направлению «Программная инженерия» независимо от его специализации. Наряду с изучением методов и алгоритмов распознавания и цифровой обработки данных, как единого целого рассматриваются основные понятия и наиболее важные характеристики программных компонентов, используемых для распознавания и обработки данных, вопросы нейросетевой, нечеткой и визуальной кластеризации и классификации.

Данное методическое пособие должно помочь студенту правильно выбрать тему, выделить проблемные места, сформулировать вопросы, по которым студент может оценить степень усвоения материала, а также указать необходимую литературу для самостоятельного изучения разделов данной дисциплины.

## **2. Рекомендации по выполнению самостоятельной работы**

Самостоятельная работа является внеаудиторной и предназначена для самостоятельного ознакомления студента с определенными разделами дисциплины по рекомендованным преподавателем материалам и подготовки к выполнению рефератов по дисциплине. Целью самостоятельной работы студентов является:

- научить студента осмысленно и самостоятельно работать сначала с учебным материалом, затем с научной информацией, заложить основы самоорганизации и самовоспитания с тем, чтобы привить умение в дальнейшем непрерывно повышать свою квалификацию
- закрепление, расширение и углубление знаний, умений и навыков, полученных студентами на аудиторных занятиях, изучение студентами дополнительных материалов по изучаемым дисциплинам и умение выбирать необходимый материал из различных источников;
- воспитание у студентов самостоятельности, организованности, самодисциплины, творческой активности, потребности развития познавательных способностей и творчества в достижении поставленных целей.

Самостоятельная работа студентов выражается в освоении необходимого объема учебной программы по дисциплине, выработке навыков профессиональной деятельности при изучении вынесенных на самостоятельную работу вопросов. Это выражается в подготовке к лекционным и лабораторным занятиям, подготовке рефератов, подготовки отчетов по лабораторным работам. В связи с большим объемом материала по методам и алгоритмам распознавания и цифровой обработки данных, того минимального времени, отведенного для их изучения учебным планом, явно недостаточно. Поэтому студентам нужно выполнить самостоятельное изучение ряда тем, а результаты выразить в виде рефератов.

Прежде чем приступить к изложению ответов на вопросы задания, студент получает необходимые знания об основных понятиях, терминах, общих вопросах распознавания и обработки данных из лекций. Эти знания также можно получить в процессе самоподготовки по предлагаемым вопросам к теме по дисциплине.

Самостоятельная подготовка состоит в подборке и изучении предлагаемой в настоящем указании учебно-методической литературы, а также использовании дополнительной литературы. В связи с быстрым развитием появлением новых методов и алгоритмов распознавания и обработки данных, литература, которую делают ссылки на момент изучения данной дисциплины, может устареть. Поэтому при выполнении работы целесообразно использовать Интернет.

Темы рефератов, тесты, письменные ответы на которые являются основой для оценки результатов самостоятельной работы, определяются текущей темой лекции и выполняемой лабораторной работой. Приветствуется инициативное предложение варианта темы реферата студентом, в рамках перспективного развития программных средств распознавания и обработки данных, например обработка больших данных.

Если реферат не зачтен, то с учетом замечаний преподавателя подлежит доработке и повторной сдаче. Студенты, не предоставившие письменной работы и не доработавшие ее после замечаний преподавателя, к экзамену по дисциплине не допускаются.

Письменная работа подписывается лицом ее выполнившим, с указанием фамилии, инициалов, даты и сдается для проверки и рецензирования преподавателю.

### **3. Проработка лекционного материала**

Проработка материала лекций дает наибольший эффект при овладении знаниями из дисциплины, если это выполняется систематически и сразу же после лекции или перед следующей лекцией. Для этого необходимо иметь конспект лекций и необходимые материалы в виде указанной в рабочей программе основной и дополнительной литературы. Желательно при изучении материала по конспекту, выделять основные смысловые единицы и отвечать на вопросы, которые также приведены в рабочей программе. Для лучшего усвоения материала нужно использовать интернет.

**Раздел 1.** Задачи распознавания и обработки данных. Классификация, группировка, прогнозирование, нахождение ассоциаций и зависимостей, визуализация. Виды данных. Данные в виде сигналов и изображений. Основные разделы, на которых базируется анализ данных: статистика, базы данных и знаний, распознавание образов, искусственный интеллект, машинное обучение.

Классификация методов обработки данных. Этапы анализа данных: выявление закономерностей, прогнозирование, анализ исключений. Сферы

применения анализа данных: финансы и банковское дело, маркетинг, медицина, генетика, биоинформатика, интернет.

При изучении этой темы студенту необходимо вспомнить основные положения теории информации, понятия системы, данных и знаний, баз и хранилищ данных.

Студент должен понимать, как развивались методы обработки данных с точки зрения диалектического материализма, почему возникла необходимость учитывать качество данных и потребность в их обработке.

Особое внимание следует уделить классификации методов и алгоритмов обработки данных, кластеризации и классификации. Также понять, в чем различие методов обработки данных, так как на сегодняшний день эти понятия размыты. Уметь отличать методы кластеризации и классификации друг от друга и уметь приводить примеры применения таких методов для обработки больших данных, в том числе и текстов.

Поскольку дисциплина называется “Методы и алгоритмы распознавания и цифровой обработки данных”, то также основное внимание нужно уделить рассмотрению необходимости применения математических методов для обработки и распознавания данных.

**Раздел 2.** Методы и алгоритмы обработки данных. Стационарность, нормальность, независимость, однородность, функции распределения.

При изучении этой темы студенту необходимо вспомнить основные процедуры статистической обработки данных, гипотезы о вероятностной природе данных. Существующих законах распределения.

Студент должен уяснить основные требования, которые предъявляются к качеству данных и качеству кластеризации и классификации.

Студент должен понимать основные основные методы и алгоритмы распознавания и цифровой обработки данных.

Студент должен знать этапы обработки данных и их последовательность при распознавании образов..

Особое внимание следует уделить понятиям индуктивного и дедуктивного обучения.

**Раздел 3.** Методы и алгоритмы распознавания данных. Дискриминантный анализ, кластерный анализ, анализ главных компонент, факторный анализ. Динамические модели и прогноз на основе временных рядов. Методы анализа данных, основанные на использовании метрики: метод опорных векторов, метод ближайших соседей. Распознавание и классификация данных.

При изучении этой темы студенту необходимо вспомнить, как формируются случайные числа, законы распределения, условные и безусловные вероятности.

Студент должен уяснить принципы кластеризации и классификации.

Студент должен понимать, как выполняются процедуры обработки данных для повышения их качества.

Студент должен знать основные принципы кластеризации и классификации, требования к качеству данных и применение методов и алгоритмов распознавания и обработки больших данных.

Особое внимание следует уделить понятию качества данных, обучению с учителем и без учителя.

#### **4. Подготовка к лабораторным работам**

Лекции закладывает основы знаний по предмету в обобщенной форме, а лабораторные занятия направлены на расширение и детализацию этих знаний, на выработку и закрепление навыков профессиональной деятельности. Подготовка к лабораторным работам предполагает предварительную самостоятельную работу студентов в соответствии с методическими указаниями по каждой лабораторной работе.

Цель выполнения лабораторных работ заключается в закреплении теоретического материала путем систематического контроля за самостоятельной работой студентов;

- Формирование умений использования теоретических знаний в процессе выполнения лабораторных работ;
- развитие аналитического мышления путем обобщения результатов лабораторных работ;
- формирование навыков оформления результатов лабораторных/практических работ в виде таблиц, графиков, выводов.

Подготовка к лабораторным работам заключается в следующем:

- открыть методические указания по лабораторным работам к данной дисциплине
- ознакомиться с целью лабораторной работы
- просмотреть необходимый теоретический материал из методических указаний
- просмотреть материал лекций по данной теме
- ознакомиться с вариантами заданий для данной лабораторной работы
- ознакомиться и подготовить ответы на контрольные вопросы
- подготовить черновик отчета по лабораторной работе.

Для выполнения предлагаются следующие лабораторные работы:

**Лабораторная работа 1.** Знакомство с Аналитической Платформой “Deductor”. Цель работы:

- получение первоначальных сведений о возможностях аналитической платформы;
- изучение основных модулей, работа с мастерами импорта, экспорта, обработки и визуализации данных.

**Лабораторная работа 2.** Хранилища данных (Организация структуры). Цель работы: изучить программную среду хранилища данных в Deductor Warehouse, ознакомиться с архитектурой хранилищ данных, научиться создавать, и наполнять информацию из хранилища данных

**Лабораторная работа 3.** Поиск ассоциативных правил. Цель работы: изучить возможность поиска ассоциативных правил используя аналитическую платформу DEDUCTOR

**Лабораторная работа 4.** Распознавание образов данных (НС Хемминга). Цель работы: изучение функционирования нейроподобных элементов в виде сети Хемминга. Разработка программы для распознавания образов при преобразовании информации.

**Лабораторная работа 5.** Кластерная обработка данных (НС Кохонена). Цель работы: научиться использовать метод кластерной обработки данных в виде самоорганизующихся карт Кохонена».

**Лабораторная работа 6.** Классификация данных (НС с ВР). Цель работы: исследование способности нейронной сети решать задачи классификации. Сеть необходимо обучить классификации по пяти классам по 10-20 числовым признакам. Используемая модель: одномерная сеть Кохонена.

**Лабораторная работа 7.** Алгоритмы распознавания прецедентов. Цель работы: изучить существующие алгоритмы распознавания образов в виде прецедентов.

**Лабораторная работа 8.** Фильтрация данных (Фильтр Калмана). Цель работы: ознакомиться с методом фильтрации данных фильтром Калмана. Оценить возможность применения фильтра Калмана на практике для обработки данных.

**Лабораторная работа 9.** Парциальная обработка данных. Цель работы: изучить возможности аналитической платформы процедур обработки данных и выполнить обработку данных выбранной предметной области.

Краткие теоретические сведения, задания на выполнение лабораторной работа, содержание отчета и контрольные вопросы – все это изложено в методических указания к выполнению лабораторных работ.

## **5. Подготовка реферата**

Подготовка реферата необходима, чтобы закрепить свои знания при изучении тем, отведенных для самостоятельной работы. Письменная работа студента в виде реферата, являющаяся основой для оценки результатов его самостоятельной работы, состоит из ответа на заданную тему и решения теста. Она выполняется студентом самостоятельно по каждой теме, определенной рабочей программой, отдельно и лично им предоставляется преподавателю или на кафедру в указанный преподавателем срок, но не позднее 15 дней до экзамена.

Реферат должен включать следующие обязательные структурные части: титульный лист, содержание (оглавление), основная часть ответов по каждому вопросу, список использованной литературы, приложения. Возможна краткая оценка современного состояния исследуемой проблемы (введение) по всем или в отдельности по каждому вопросу задания. Введение располагается после содержания.

При непосредственном написании рефератов недопустимо дословное копирование материала из Интернета, а также с использованием сканеров фрагментов текста из учебников, учебных пособий и брошюр, за исключением

отдельных научных и практически положений. В этом случае необходимо сделать ссылку на первоисточник.

При подготовке реферата должны соблюдаться такие требования, как логическая последовательность изложения ответа по каждому вопросу, убедительность аргументации, если она присутствует, краткость, конкретность и точность формулировок, исключающих возможность неоднозначного толкования. Объем реферата 10-12 страниц 14 шрифтом.

При наличии затруднений в подборе учебной и нормативной литературы для выполнения задания, в изучении отдельных вопросов по дисциплине «Математическое моделирование» студенты могут получить необходимую консультацию на кафедре АОИ в специально определенные для этого дни.

В тексте необходимо применять научные и технические термины, обозначения и определения, установленные соответствующими стандартами, а при их отсутствии – общепринятые в научной и технической литературе. Не допускается применение для одного и того же понятия без соответствующих пояснений, различных научных терминов, близких по смыслу (синонимов) и иностранных слов, сокращений слов, обозначений, кроме установленных правилами русской орфографии или соответствующими государственными стандартами.

Ответы на вопросы реферата по текущей теме должны быть даны в электронном и письменном виде. Электронный вариант высылается преподавателю по электронной почте в течении недели после получения текущей темы, письменный вариант приносится на лекцию или контрольную работу. Отсчет страниц (листов) начинается с титульного. Номера страниц (листов) проставляются внизу страницы посередине. Нумерация страниц (листов) текста работы и приложений должна быть сквозной.

## **6. Вопросы для самоподготовки**

1. Необходимость обработки данных
2. Классификация методов обработки данных
3. Процедуры очистки данных
4. Консолидация данных
5. Профилирование данных
6. Трансформация данных
7. Фильтрация данных
8. Обогащение данных
9. Визуализация данных
10. Понятие классификации
11. Постановка задачи классификации
12. Понятие класса
13. Понятие признака классификации
14. Понятие о прецедентах
15. Методы классификации
16. Метрические методы классификации

17. Алгоритм ближайших соседей
18. Алгоритм окна Парзена
19. Алгоритм потенциальных функций
20. Логические методы классификации
21. Алгоритм решающего дерева
22. Линейные методы классификации
23. Алгоритм SVM
24. Перцептрон
25. Нелинейные методы классификации
26. Многослойные нейронные сети
27. Статистические методы классификации
28. Алгоритм Байесовского классификатора
29. Примеры применения классификации в различных предметных областях.

## **7. Темы для самостоятельного изучения и подготовки рефератов**

1. Методики анализа больших данных
2. Алгоритмы распознавания текстов
3. Управление качеством данных
4. Пакеты прикладных программ ETL
5. Обработка данных при машинном обучении
6. Модели классификации данных.

## **8. Подготовка к контрольным работам**

Контрольная работа – письменная работа небольшого объема, для проверки знаний материала изучаемой дисциплины и навыков его практического применения. Контрольные работы могут состоять из одного или нескольких теоретических вопросов (обычно из 5).

С помощью контрольной работы студент постигает наиболее сложные проблемы дисциплины, систематизирует знания, учится кратко и последовательно излагать свои мысли, правильно оформлять работу.

Контрольная работа обычно проводится перед контрольной точкой с целью определить уровень знаний студента.

Для подготовки к контрольной работе по материалу, изложенному в лекциях или самостоятельно изученному, необходимо изучить содержание конспектов лекций и по непонятным вопросам использовать интернет или указанную в рабочей программе литературу.

Темы контрольных работ:

- Алгоритмы кластеризации данных для различных предметных областей
- Алгоритмы классификации данных для различных предметных областей
- SVM метод

- Классификация на нейронных сетях
- Кластеризация на нейронных сетях

## 9. Литература

1. Замятин, Н. В. Модели и алгоритмы распознавания и обработки данных: Комплекс лабораторных работ [Электронный ресурс] / Н. В. Замятин. — Томск: ТУСУР, 2017. — 67 с. — Режим доступа: <https://edu.tusur.ru/publications/7041>
2. Методы и модели анализа данных: OLAP и Data Mining. / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод – СПб.: БХВ-Петербург, 2004.- 336 с.: ил.
3. Официальный сайт компании «BaseGroup Labs» [Электрон. ресурс]. Рязань, 1995-2010.- Режим доступа: <http://www.basegroup.ru/>
4. Data Mining и аналитическая платформа Deductor [Электрон. ресурс] : [статья]. М., 2008.- Режим доступа: [http://sttc.ru/index.php?option=com\\_content&task=view&id=56&Itemid=90](http://sttc.ru/index.php?option=com_content&task=view&id=56&Itemid=90)
5. Г.И. Просветов (МГУ им. М.В. Ломоносова). Дерево решений [Электрон. ресурс] : [статья] / Г.И. Просветов.- СПб, 2008.- Режим доступа: [http://www.elitarium.ru/2008/04/09/derevo\\_reshenijj.html](http://www.elitarium.ru/2008/04/09/derevo_reshenijj.html)
6. Деревянко В.А. Поиск ассоциативных правил при интеллектуальном анализе данных [Электрон. ресурс] : [статья] / В.А. Деревянко.- 2009.- Режим доступа: [http://www.rammus.ru/products/arda/article\\_lam\\_translation](http://www.rammus.ru/products/arda/article_lam_translation)
2. Замятин, Н. В. Нечеткая логика и нейронные сети: Учебное пособие [Электронный ресурс] / Замятин Н. В. — Томск: ТУСУР, 2014. — 289 с. — Режим доступа: <https://edu.tusur.ru/publications/7020>.