

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Родин Олег Федорович
Должность: И.о. директора
Дата подписания: 26.03.2025 13:52:16
Уникальный программный ключ:
2246bb4b5eca53e35a45d6a91259e790782354e7

Министерство просвещения Российской Федерации
Нижнетагильский государственный социально-педагогический институт (филиал)
Федерального государственного автономного образовательного учреждения
высшего образования
«Российский государственный профессионально-педагогический университет»

Факультет естествознания, математики и информатики
Кафедра информационных технологий и физико-математического образования

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.01.ДВ.04.02 МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ МАССИВЫ
ДАННЫХ**

Направление подготовки	09.03.03 Прикладная информатика
Профили	«Прикладная информатика в управлении IT-проектами»
Автор:	Н.В. Бужинская, к. пед. н., доцент

Одобрена на заседании кафедры информационных технологий и физико-математического образования. Протокол от 6 февраля 2025 г. № 6.

Рекомендована к использованию в образовательной деятельности методической комиссией факультета естествознания, математики и информатики. Протокол от 13 февраля 2025 г. № 5.

1. ЦЕЛЬ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель изучения дисциплины – формирование компетенций в области компьютерной обработки и анализа экономических данных для решения предметно-ориентированных задач на основе использования системного анализа и математического моделирования.

Задачи:

- сформировать целостное представление об осуществлении критического анализа, синтеза и представления информации с использованием системного подхода и статистических методов обработки данных для решения поставленных задач;
- научить применять основы системного подхода и методов математического анализа и моделирования для решения задач в профессиональной деятельности;
- показать возможности современных информационных и коммуникационных технологий для обработки больших массивов статистических данных при решении профессиональных задач;
- научить проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе на основе статистической обработки данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Машинное обучение и большие массивы данных» является дисциплиной по выбору по направлению 09.03.03 Прикладная информатика и включена в блок ДВ.04 «Дисциплины (модули) по выбору». Реализуется кафедрой информационных технологий и физико-математического образования на 4 курсе.

Данная дисциплина является продолжением изучения таких дисциплин как «Высокоуровневые методы информатики и программирования», «Теория систем и системный анализ», «Теория вероятностей и математическая статистика» и др. Полученный при изучении опыт деятельности может быть полезен студентам в выполнении учебных проектов и оформлении выпускных квалификационных работ.

3. РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс изучения дисциплины направлен на формирование и развитие следующих компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Дескрипторы
УК-1. Способен осуществлять поиск,	УК1.2. Применяет логические формы и процедуры, способен к рефлексии по поводу собственной и чужой мыслительной деятельности	Знает логические формы и процедуры, способен к рефлексии по поводу собственной и чужой мыслительной деятельности
		Умеет применять логические формы и процедуры, способен к рефлексии по поводу собственной и чужой мыслительной деятельности
		Владеет технологиями применения логических форм и процедур

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Дескрипторы
критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК1.3. Анализирует источники информации с целью выявления их противоречий и поиска достоверных суждений	Знает подходы к анализу источников информации с целью выявления их противоречий и поиска достоверных суждений
		Умеет анализировать источники информации с целью выявления их противоречий и поиска достоверных суждений
		Владеет технологией анализа источников информации с целью выявления их противоречий и поиска достоверных суждений
ПК-2. Способность разрабатывать и адаптировать прикладное программное обеспечение.	ПК-2.1. Знает структуру и технологии разработки прикладного ПО.	Знает этапы разработки прикладного ПО
		Умеет проектировать структуру ПО
		Владеет навыками проектирования и разработки прикладного ПО
	ПК-2.2. Знает современные языки программирования.	Знает алфавит, синтаксис и семантику современных языков программирования
		Умеет выбирать язык программирования в зависимости от поставленной задачи
		Владеет навыками в области одного из языков программирования
	ПК-2.3. Умеет использовать основные технологии разработки программных продуктов.	Знает технологии разработки программных продуктов
		Умеет выбирать технологии разработки программных продуктов
		Владеет технологией разработки программных продуктов
	ПК-2.4. Адаптирует прикладное программное обеспечение под нужды организации.	Знает способы адаптации прикладного программного обеспечения под нужды организации.
		Умеет выбирать способы адаптации прикладного программного обеспечения под нужды организации.
		Владеет технологией адаптации прикладного программного обеспечения под нужды организации.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Объем дисциплины и виды контактной и самостоятельной работы

Общая трудоемкость дисциплины составляет 6 зач. ед. (216 час.), курс - 4, распределение по видам работ представлено в табл.№1.

Таблица 1. Распределение трудоемкости дисциплин по видам

Вид работы	Форма обучения
------------	----------------

	очная
Общая трудоемкость дисциплины по учебному плану	216
Контактная работа, в том числе:	22
Лекции	8
Лабораторные работы	12
Самостоятельная работа	183
Зачет с оценкой + экзамен	13

4.2. Содержание и тематическое планирование дисциплины

Таблица 2. Тематический план дисциплины (заочная форма обучения)

Наименование разделов и тем дисциплины	Всего часов	Контактная работа		Сам. работа	Оценочные средства текущего контроля	Оценочные средства промежуточного контроля
		Лекции	Практ. занятия			
Тема 1. Основные понятия и примеры прикладных задач	34	2	2	30	Проверка глоссария, интеллект-карты по основным понятиям	Устный ответ на экзамене Решение задач
Тема 2. Метрические методы классификации	34	2	2	30	Проверка отчетов по лабораторным работам	Устный ответ на экзамене Решение задач
Тема 3. Логические методы классификации	31	0	2	29	Проверка отчетов по лабораторным работам	Устный ответ на экзамене Решение задач
Экзамен	9	0	0	9		
Итого за зимнюю сессию	108	4	6	98		
Тема 4. Линейные методы классификации	34	2	2	30	Проверка отчетов по лабораторным работам	Устный ответ на экзамене Решение задач
Тема 5. Методы регрессионного анализа	34	2	2	30	Проверка отчетов по лабораторным работам	Устный ответ на экзамене Решение задач
Тема 6. Байесовские методы классификации	36	0	2	34	Проверка отчетов по лабораторным работам	Устный ответ на экзамене Решение задач
Зачет с оценкой	4	0	0	4		
Итого за летнюю сессию	108	4	6	98		
Итого:	216	8	12	196		

4.3. Содержание тем дисциплины

Тема 1. Основные понятия и примеры прикладных задач.

Машинное обучение. Большие массивы данных. Постановка задач обучения по прецедентам. Объекты и признаки. Типы шкал: бинарные, номинальные, порядковые,

количественные. Типы задач: классификация, регрессия, прогнозирование, кластеризация. Примеры прикладных задач. Основные понятия: модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль. Методика экспериментального исследования и сравнения алгоритмов на модельных и реальных данных. Полигон алгоритмов классификации. CRISP-DM — межотраслевой стандарт ведения проектов интеллектуального анализа данных.

Тема 2. Метрические методы классификации.

Метод ближайших соседей и его обобщения. Метод ближайших соседей (kNN) и его обобщения. Подбор числа k по критерию скользящего контроля. Обобщённый метрический классификатор, понятие отступа. Метод потенциальных функций, градиентный алгоритм. Отбор эталонов и оптимизация метрики. Отбор эталонных объектов. Псевдокод: алгоритм СТОЛП. Функция конкурентного сходства, алгоритм FRiS-СТОЛП. Функционал полного скользящего контроля, формула быстрого вычисления для метода 1NN. Профиль компактности.

Тема 3. Логические методы классификации.

Понятия закономерности и информативности. Понятие логической закономерности. Эвристическое, статистическое, энтропийное определение информативности. Асимптотическая эквивалентность статистического и энтропийного определения. Сравнение областей эвристических и статистических закономерностей. Разновидности закономерностей: конъюнкции пороговых предикатов (гиперпараллелепипеды), синдромные правила, шары, гиперплоскости. Бинаризация признаков. Алгоритм разбиения области значений признака на информативные зоны. Решающие списки и деревья. Решающий список. Жадный алгоритм синтеза списка. Решающее дерево. Псевдокод: жадный алгоритм ID3. Недостатки алгоритма и способы их устранения. Проблема переобучения. Редукция решающих деревьев: предредукция и постредукция. Преобразование решающего дерева в решающий список. Небрежные решающие деревья (oblivious decision tree).

Тема 4. Линейные методы классификации.

Градиентные методы. Линейный классификатор, непрерывные аппроксимации пороговой функции потерь. Связь с методом максимума правдоподобия. Метод стохастического градиента и частные случаи: адаптивный линейный элемент ADALINE, персептрон Розенблатта, правило Хэбба. Теорема Новикова о сходимости. Доказательство теоремы Новикова. Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, «выбывание» из локальных минимумов. Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь. Задача квадратичного программирования и двойственная задача. Понятие опорных векторов.

Тема 5. Методы регрессионного анализа.

Многомерная линейная регрессия. Задача регрессии, многомерная линейная регрессия. Метод наименьших квадратов, его вероятностный смысл и геометрический смысл. Сингулярное разложение. Нелинейная параметрическая регрессия. Метод Ньютона-Рафсона, метод Ньютона-Гаусса. Одномерные нелинейные преобразования признаков: метод настройки с возвращениями (backfitting) Хасты-Тибширани. Непараметрическая регрессия. Сглаживание. Локально взвешенный метод наименьших квадратов и оценка Надарая-Ватсона. Выбор функции ядра. Выбор ширины окна сглаживания. Сглаживание с переменной шириной окна. Неквадратичные функции потерь. Прогнозирование временных рядов.

Тема 6. Байесовские методы классификации.

Оптимальный байесовский классификатор. Принцип максимума апостериорной вероятности. Функционал среднего риска. Ошибки I и II рода. Теорема об оптимальности байесовского классификатора. Оценивание плотности распределения: три основных подхода. Наивный байесовский классификатор. Непараметрическое оценивание плотности.

Параметрическое оценивание плотности. Разделение смеси распределений. Логистическая регрессия.

Лабораторные работы для заочной формы обучения

Тематика занятий		
№ п.п.	Наименование лабораторных работ	Кол-во ауд. часов
Зимняя сессия		
1	Язык Python как средство работы с данными. Установка библиотек	2
2	Основы работы с Pandas	2
3	Анализа данных в Python	2
Летняя сессия		
4	Работа с массивами данных	2
5	Нейронные сети	2
6	Кластеризация данных	2
	Итого:	12 часов

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

5.1. Перечень основной и дополнительной литературы

Основная литература

1. Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск : УлГТУ, 2017. — 290 с. — ISBN 978-5-9795-1712-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/165053> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.

2. Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук ; перевод с английского А. Б. Огурцова. — Москва : ДМК Пресс, 2018. — 250 с. — ISBN 978-5-97060-508-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/97353> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.

3. Ракитский А.А. Методы машинного обучения : учебно-методическое пособие / Ракитский А.А.. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/90591.html> (дата обращения: 27.01.2022). — Режим доступа: для авторизир. пользователей.

Дополнительная литература

1. Боровков, А. А. Математическая статистика : учебник для вузов / А. А. Боровков. — 5-е изд., стер. — Санкт-Петербург : Лань, 2021. — 704 с. — ISBN 978-5-8114-7677-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/164711> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.

2. Остроух, А. В. Интеллектуальные информационные системы и технологии : монография / А. В. Остроух, А. Б. Николаев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 308 с. — ISBN 978-5-8114-8578-9. — Текст : электронный // Лань : электронно-

библиотечная система. — URL: <https://e.lanbook.com/book/177839> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.

5.2. Электронные образовательные ресурсы, в т.ч. профессиональные базы данных и информационные справочные системы

https://www.ntspi.ru/library/directories_and_files/web_res/systems/	Электронно-библиотечные системы НТГСПИ
https://www.ntspi.ru/library/directories_and_files/web_res/systems/libraris/	Электронные базы данных НТГСПИ
https://www.ntspi.ru/library/periodika/	Периодика НТГСПИ
https://iprmedia.ru	ЭБС «Ай Пи Эр Медиа»
https://ibooks.ru	ЭБС «Айбукс»
https://urait.ru	ЭБС Юрайт
http://e.lanbook.com	ЭБС издательства «ЛАНЬ»
http://elibrary.ru	Научная электронная библиотека eLIBRARY.RU
http://www.consultant.ru	«КонсультантПлюс»
http://cyberleninka.ru	НЭБ «КиберЛенинка»
https://polpred.ru	ООО «Полпред-Справочники» (база данных)
https://eivis.ru	ООО «ИВИС»
www.delpress.ru	«Деловая пресса»

Интернет-ресурсы:

1. «Российское образование: федеральный портал [сайт]. — URL: <https://www.edu.ru/>
2. eLIBRARY.RU : научная электронная библиотека [сайт]. — URL: <https://elibrary.ru>
3. INTUIT.ru : Учебный курс — Intel. Обучение для будущего [сайт]. — URL: <http://www.intuit.ru/department/education/intelteach/>
4. LEARNINGAPPS: сервис для разработки электронных дидактических материалов [сайт]. — URL: <https://learningapps.org/>.
5. Единое окно доступа к образовательным ресурсам : Федеральный портал [сайт]. — URL: <http://window.edu.ru/window/library>.

5.3. Комплект программного обеспечения

1. Среда электронного обучения «Русский Moodle» (<https://do.ntspi.ru/>).
2. Электронная информационно-образовательная среда РГППУ (<https://eios.rsvpu.ru/>).
3. Платформа для организации и проведения вебинаров «Mirapolis Virtual Room».
4. Microsoft Office /LibreOffice /P-Офис.
5. Kaspersky Endpoint Security.
6. Adobe Reader.
7. Браузеры Firefox, Google Chrome, Яндекс.Браузер.
8. GIMP, Inkscape, Paint Net
9. Movavi / Windows Movie Maker/ Free Video Editor.
- Net Beans IDE

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Помещения

Помещение для проведения занятий лекционного типа, компьютерный класс (не менее 10 рабочих мест с установленным программным обеспечением и доступом в сеть «Интернет», кабинет для индивидуальных консультаций, самостоятельной работы, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации. Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечены доступом в электронную информационно-образовательную среду университета.

6.2. Оборудование и технические средства обучения

6.2.1. Оборудование, в т.ч. специализированное

Стационарный компьютер или ноутбук, проекционное оборудование, кликер, акустические колонки.

6.2.2. Технические средства обучения

Документ-камера, интерактивная доска (панель).

Персональные компьютеры/ ноутбуки, веб-камера, наушники.

6.2.3. Учебные и наглядные пособия

Печатные и электронные учебные пособия, и наглядный материал: графические изображения, схемы, таблицы, раздаточный материал.

Презентации лекций, видео-презентации, видео-лекции.