Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Райхерт Татьяна Николаевна
Должность: Директор
Дата подписания: 07.03.2022 15:09:56
Уникальный программный ключ:
c914df807d771447164c08ee17f8e2f93dde816b

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

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.01.ДВ.04.02 МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ОБЪЕМЫ
ДАНЫХ**

Уровень высшего образования
Направление подготовки

Бакалавриат
09.03.03 Прикладная информатика

Профили

«Прикладная информатика в управлении IT-проектами»

Рабочая программа дисциплины «Машинное обучение и большие объемы данных». Нижнетагильский государственный социально-педагогический институт (филиал) федерального государственного автономного образовательного учреждения высшего образования «Российский государственный профессионально-педагогический университет», Нижний Тагил, 2021. 12 с.

Настоящая программа составлена в соответствии с требованиями ФГОС ВО по направлению 09.03.03 Прикладная информатика (№ 922 от 19.09.2017)

Автор: канд. пед. наук, доцент, доцент кафедры ИТ _____ Н.В. Бужинская

Одобен на заседании кафедры ИТ 24 апреля 2021 г., протокол № 9

Заведующий кафедрой ИТ _____ М.В. Мащенко

Рекомендован к печати методической комиссией ФЕМИ 27 апреля 2021 г., протокол № 6.

Председатель методической комиссии ФЕМИ _____ Н.З. Касимова

© Нижнетагильский государственный социально-педагогический институт (филиал) федерального государственного автономного образовательного учреждения высшего образования «Российский государственный профессионально-педагогический университет», 2021.

© Н.В. Бужинская, 2021.

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ.....	3
1. ЦЕЛЬ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....	4
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	4
3. РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ	4
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ	6
4.1. Объем дисциплины и виды контактной и самостоятельной работы.....	6
4.2. Учебно-тематический план	6
4.3. Содержание дисциплины.....	6
5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ	6
6. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ.....	8
6.1. Организация самостоятельной работы студентов.....	8
6.2. Организация текущего контроля и промежуточной аттестации	9
7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ	10
МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	12

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 в 5 и 6 семестрах.

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:

Наименование категории (группы) универсальных компетенций	Код и наименование универсальной компетенции	Код и наименование индикатора достижения универсальной компетенции
Универсальные	УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИУК 1.1. Знает основные источники и методы поиска информации, необходимой для решения поставленных задач.
		ИУК 1.2. Умеет осуществлять поиск информации для решения поставленных задач, применять методы критического анализа и синтеза информации.

		ИУК 1.3. Грамотно, логично, аргументированно формирует собственные суждения и оценки; отличает факты от мнений, интерпретаций и оценок; применяет методы системного подхода для решения поставленных задач.
Общепрофессиональные	ОПК-1 Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ОПК-1.1. Знает основы математики, физики, вычислительной техники и программирования.
		ОПК-1.2. Умеет решать стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования.
		ОПК-1.3. Организует исследование объектов профессиональной деятельности.
Общепрофессиональные	ОПК-2 Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2.1. Знает современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности.
		ОПК-2.2. Умеет выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности.
		ОПК-2.3. Применяет современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности.
Профессиональные	ПК-2 Способен разрабатывать и адаптировать прикладное программное обеспечение	ПК-2.1. Знает структуру и технологии разработки прикладного ПО
		ПК-2.2. Знает современные языки и среды программирования
		ПК-2.3. Умеет использовать основные технологии разработки программных продуктов
		ПК-2.4. Адаптирует прикладное программное обеспечение под нужды организации

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. Объем дисциплины и виды контактной и самостоятельной работы

Вид работы	Кол-во часов
Общая трудоемкость дисциплины по учебному плану	216
Контактная работа, в том числе:	76
Лекции	24
Практические занятия	52
Самостоятельная работа	140
Подготовка к экзамену, сдача экзамена	36
Подготовка к зачету, сдача зачета	9

4.2. Учебно-тематический план

Очная форма обучения

Наименование разделов и тем дисциплины	Всего часов	Контактная работа		Сам. работа	Формы текущего контроля успеваемости
		Лекции	Практ. занятия		
Тема 1. Основные понятия и примеры прикладных задач	24	4	10	10	Проверка глоссария, интеллект-карты по основным понятиям
Тема 2. Метрические методы классификации	24	4	10	10	Проверка отчетов по лабораторным работам
Тема 3. Логические методы классификации	24	4	6	14	Проверка отчетов по лабораторным работам
Экзамен	36	0	0	36	
Итого за 5 семестр	108	12	26	70	
Тема 4. Линейные методы классификации	34	4	10	20	Проверка отчетов по лабораторным работам
Тема 5. Методы регрессионного анализа	34	4	10	20	Проверка отчетов по лабораторным работам
Тема 6. Байесовские методы классификации	31	4	6	21	Проверка отчетов по лабораторным работам
Зачет с оценкой	9	0	0	9	
Итого за 6 семестр	108	12	26	70	
Итого:	216	24	52	140	

4.3. Содержание 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, «выбивание» из локальных минимумов. Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случаи линейной делимости и отсутствия линейной делимости. 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) Хасти-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.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Обучение по дисциплине «Машинное обучение и большие объемы данных» целесообразно построить с использованием компетентностного подхода, в рамках которого образовательный процесс строится с учетом специфики будущей профессиональной деятельности студентов. Лекционные занятия должны стимулировать познавательную активность студентов, поэтому преподавателю необходимо обращаться к примерам, взятым из практики, включать проблемные вопросы.

Основными методами, используемыми на практических занятиях, будут: метод демонстрационных примеров, мастер-класс, практикум с использованием практико-ориентированных задач и проектная технология.

6. УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ

6.1. Организация самостоятельной работы студентов

Самостоятельная работа студентов включает изучение вопросов, вынесенных за рамки аудиторных занятий, расширение и углубление знаний по темам, рассмотренным на лекционных занятиях. При подготовке к практическим занятиям студенты изучают технологии программирования, решают задачи.

Проверка качества усвоения знаний в течение семестра осуществляется на каждом лабораторном занятии как в устной (обсуждение изученной литературы), так и в письменной (проверка отчетов) форме.

Подобное разнообразие видов текущего контроля дает основания для объективной оценки уровня подготовки каждого студента.

Тематика занятий

№ п.п.	Наименование лабораторных работ	Кол-во ауд. часов
5 семестр		
1	Язык Python как средство работы с данными. Установка библиотек	2
2	Классификация данных в языке Python	2
3	Основы работы с Pandas	2
4	Анализа данных в Python	2
5	Решение задач	2
6	Построение графиков и диаграмм	2
7	Линейная регрессия	2
8	Деревья решений	2
9	Метод случайного леса	2
10	Метод опорных векторов	2
11	Машинное зрение и распознавание лиц в реальном времени	2
12	Выполнение проекта	2
13	Защита проекта	2
6 семестр		
14	Работа с массивами данных в C#	2
15	Нейронные сети на C#	2
16	Решение задач	2
17	Создание приложения UWP	2
18	Выполнение и защита проекта	2
19	Возможности языка R в машинном обучении. Установка.	2

20	Синтаксис языка R	2
21	Методы работы с данными на языке R	2
22	Классификация данных	2
23	Кластеризация данных	2
24	Математическое и нейросетевое моделирование	2
25	Выполнение проекта	2
26	Защита проекта	2
	Итого:	52 часа

6.2. Организация текущего контроля и промежуточной аттестации

Текущий контроль усвоения знаний ведется по итогам представления выполненных самостоятельных заданий и защиты отчетов по лабораторным работам; участия в дискуссиях на лекционных занятиях, проверки составленного глоссария и результатов тестирования.

Текущий контроль учебных достижений студентов может быть проведен с использованием накопительной балльно-рейтинговой системы оценки в соответствии с Положением о НБРС.

Промежуточная аттестация по данной дисциплине проводится в форме зачета с оценкой в шестом семестре и экзамена в пятом, на которых теоретические знания студентов проверяются в ходе устного ответа на вопрос, а практические по итогам выполнения и презентации практического задания.

Примерные вопросы к экзамену (5 семестр)

1. Цели и задачи машинного обучения.
2. Особенности работы с большими массивами данных.
3. Постановка задач обучения по прецедентам. Объекты и признаки.
4. Типы шкал: бинарные, номинальные, порядковые, количественные.
5. Типы задач: классификация, регрессия, прогнозирование, кластеризация.
6. Метрические методы классификации.
7. Метод ближайших соседей и его обобщения.
8. Подбор числа k по критерию скользящего контроля.
9. Обобщённый метрический классификатор, понятие отступа.
10. Логические методы классификации.
11. Понятия закономерности и информативности.
12. Понятие логической закономерности.
13. Эвристическое, статистическое, энтропийное определение информативности.
14. Решающие списки и деревья.

Примерное практическое задание

С помощью Python решите задачу регрессии. Для сотрудников какой-нибудь компании определите, как значение зарплаты зависит от других данных, таких как опыт работы, уровень образования, роль, город, в котором они работают, и так далее.

Критерии оценки на экзамене

Не удовлетворительно – низкий уровень знаний и умений по курсу (студент в деятельности совершает ошибки и не способен к их исправлению, не может ответить на теоретический вопрос или выполнить практическое задание).

Удовлетворительно – достаточный уровень знаний и умений по курсу (деятельность студента не носит целенаправленный характер, однако отличается стремлением самостоятельно ставить перед собой цели; может определить основные понятия дисциплины);

Хорошо – средний уровень знаний и умений по курсу (деятельность студента носит целенаправленный характер, отличается стремлением самостоятельно ставить перед собой различные цели: знает методологию в данной области);

Отлично – высокий уровень знаний и умений (в деятельности студентов отмечается четкость, действенность и реальность целей самообразования; самообразование становится устойчивым, целенаправленным, систематичным).

Примерные вопросы к зачету (6 семестр)

1. Линейные методы классификации.
2. Градиентные методы.
3. Линейный классификатор, непрерывные аппроксимации пороговой функции потерь. Теорема Новикова о сходимости.
4. Эвристики: инициализация весов, порядок предъявления объектов, выбор величины градиентного шага, «выбивание» из локальных минимумов.
5. Метод опорных векторов.
6. Задача квадратичного программирования и двойственная задача.
7. Методы регрессионного анализа.
8. Многомерная линейная регрессия.
9. Задача регрессии, многомерная линейная регрессия.
10. Метод наименьших квадратов, его вероятностный смысл и геометрический смысл. Сглаживание.
11. Неквадратичные функции потерь.
12. Прогнозирование временных рядов
13. Байесовские методы классификации.
14. Оптимальный байесовский классификатор
15. Непараметрическое оценивание плотности.
16. Параметрическое оценивание плотности.
17. Логистическая регрессия.

Примерное практическое задание

Постройте кластеризацию данных для определения вида материнских плат. Данные возьмите из соответствующей базы на основе четырех параметров. Постройте соответствующие графики.

Критерии оценки устного ответа на вопрос

- полнота ответа;
- лаконичность ответа и умение выделить главное;
- соответствие современным достижениям науки;
- логичность ответа и умение построить завершённую монологическую речь;
- научно-популярный (деловой) стиль изложения;
- наличие практических примеров из жизни или профессиональной деятельности.

Критерии оценки практического задания

- работоспособность предложенных алгоритмов;
- умение объяснить принцип решения задачи.

Критерии оценки

- Зачтено – студент отвечает на теоретический вопрос, может привести примеры, программа работает без ошибок.
- Не зачтено – студент не может ответить на теоретический вопрос, или допускает существенные ошибки, практическая задача решена неправильно.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

Основная литература: указывается до 5 наименований не старше 5 лет

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: *указывается до 5 наименований не старше 5 лет*

4. 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.

5. 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.

Сетевые ресурсы (*указываются при необходимости обращения обучающихся при выполнении практических заданий*):

1. Введение в машинное обучение. [Электронный ресурс]. URL: <https://ru.coursera.org/learn/vvedenie-mashinnoe-obuchenie> (дата обращения: 18.02.2020).

2. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными / А. Мюллер, С. Гвидо. [Электронный ресурс]. URL: https://codernet.ru/books/python/vvedenie_v_mashinnoe_obuchenie_s_pomoshhyu_python/ (дата обращения: 18.02.2020).

3. Единое окно доступа к образовательным ресурсам: Федеральный портал. URL: <http://window.edu.ru/window/library>. (дата обращения: 09.02.2020). Режим доступа: свободный. Текст: электронный.

Программное обеспечение общего и профессионального назначения:

LibreOffice

Microsoft Visual Studio,

Expressions и

Embedded.

Microsoft Visio,

OneNote,

Project.

Серверы Microsoft SQL,

BizTalk

SharePoint

Сублицензионный договор № Tr000142285 от 16.02.2017 г., продление 02.08.2018 г.
№ счета 5024818829

Бесплатное ПО:

GIMP, Inkscape, Paint Net

7-Zip

Blender
Ramus Educational
Python, Dev C++
Net Beans IDE
Python 3.6.

Информационные системы и платформы:

1. Среда электронного обучения «Русский Moodle» (<https://do.ntspi.ru/>).
2. Интернет-платформа онлайн-курсов со свободным кодом «Open edX» (<https://www.edx.org/>).
3. Интернет-платформа онлайн-курсов «Открытое образование» (<https://openedu.ru/>).
4. Электронная информационно-образовательная среда РГППУ (<https://eios.rsvpu.ru/>).
5. Платформа для организации и проведения вебинаров «Mirapolis Virtual Room».

МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

1. Учебная аудитория для проведения занятий лекционного типа.
2. Учебная аудитория для проведения занятий практического типа, проведения групповых и индивидуальных консультаций, проведения текущего контроля и промежуточной аттестации.
3. Помещения для самостоятельной работы, оснащенные персональными компьютерами с доступом в интернет, доступом в электронную информационно-образовательную среду, программное обеспечение общего и профессионального назначения.