

Технологии машинного обучения рабочая программа дисциплины (модуля)

Направление подготовки 09.03.03 Прикладная информатика

Направленность (профиль) Прикладная информатика на железнодорожном транспорте

Квалификация **бакалавр**

Форма обучения **очная**

Общая трудоемкость **10 ЗЕТ**

Виды контроля в семестрах:

экзамены 6

зачеты 5

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	5 (3.1)		6 (3.2)		Итого	
	уп	рп	уп	рп	уп	рп
Неделя	16,5		16,5			
Вид занятий	уп	рп	уп	рп	уп	рп
Лекции	16	16	16	16	32	32
Лабораторные			48	48	48	48
Практические	16	16			16	16
Конт. ч. на аттест. в период ЭС	0,15	0,15	2,3	2,3	2,45	2,45
Итого ауд.	32	32	64	64	96	96
Контактная работа	32,15	32,15	66,3	66,3	98,45	98,45
Сам. работа	103	103	125	125	228	228
Часы на контроль	8,85	8,85	24,7	24,7	33,55	33,55
Итого	144	144	216	216	360	360

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1	овладение студентами основными методами методов машинного обучения, получение навыков программирования алгоритмов в области искусственного интеллекта и анализа полученных результатов.
-----	---

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Цикл (раздел) ОП:	Б1.О.18.02
-------------------	------------

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ОПК-6 Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования;

ОПК-6.2 Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий

В результате освоения дисциплины (модуля) обучающийся должен

3.1 Знать:	
3.1.1	функции и методы библиотеки Pandas: query, tail, sample, head, groupby, agg;
3.1.2	методы фильтрации и агрегации данных;
3.1.3	методы библиотек tensorflow, keras, seaborn, matplotlib;
3.1.4	основные методы машинного обучения;
3.1.5	ансамблевые методы
3.1.6	основные метрики качества моделей машинного обучения: accuracy, precision, recall, F1 score.
3.2 Уметь:	
3.2.1	выполнять фильтрацию данных средствами Python;
3.2.2	агрегировать данные средствами Python;
3.2.3	визуализировать данные средствами библиотек pandas, seaborn, matplotlib;
3.2.4	выбирать тип классификатора в зависимости от поставленной задачи;
3.2.5	строить деревья решений и выполнять по ним предсказание;
3.2.6	строить модель "случайный лес" и выполнять по ней предсказание;
3.2.7	оценивать качество модели обучения;
3.2.8	проводить машинный анализ данных с применением информационных систем и технологий;
3.2.9	решать задачи классификации методами машинного обучения;
3.2.10	решать задачи регрессии методами машинного обучения.
3.3 Владеть:	
3.3.1	навыками построения дерева решений;
3.3.2	реализации алгоритма случайного леса (Random forest);
3.3.3	написания однослойных и многослойных нейронных сетей;
3.3.4	настройки нейронных сетей;
3.3.5	обучения модели.
3.3.6	

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Примечание
	Раздел 1. Введение в обработку данных			
1.1	Основы обработки данных. Открытые данные. Машиночитаемые форматы. /Лек/	5	2	
1.2	Алгоритмы и структуры данных. Датафреймы. Вывод строк и столбцов. Методы loc и iloc. /Лек/	5	2	
1.3	Алгоритмы и структуры данных. Библиотека Pandas: структуры данных: Series и DataFrame. Работа с csv. /Пр/	5	2	
1.4	Датафреймы. Фильтрация данных. Метод query. /Лек/	5	2	
1.5	Датафреймы. Фильтрация данных. Метод query. /Пр/	5	2	

1.6	Датафреймы. Сортировка, группировка и агрегация данных. /Лек/	5	2	
1.7	Датафреймы. Сортировка, группировка и агрегация данных. /Пр/	5	2	
1.8	Визуализация данных. Визуализация данных в Python. Визуализация в Pandas. Библиотеки Seaborn и Matplotlib. /Лек/	5	2	
1.9	Визуализация данных в Python. Методы визуализации в Pandas. Библиотеки Seaborn и Matplotlib. Графики, тепловые карты, диаграммы рассеивания. Анализ и интерпретация результатов визуализации. /Пр/	5	2	
1.10	Фильтрация данных. Библиотека Pandas. Вывод строк и колонок. Методы query, tail, sample, head. Объединение методов. /Пр/	5	2	
1.11	Агрегация данных в Pandas: count, sum, max, min, mean, median. Методы groupby и aggregate. Объединение методов. /Пр/	5	2	
1.12	Отчистка данных. Работа с пропущенными значениями. /Лек/	5	2	
1.13	Предобработка данных. Преобразование данных. /Лек/	5	2	
1.14	Встроенные наборы данных. Работа с пропущенными значениями. Замена NaN. /Пр/	5	2	
1.15	Изменение датафрейма. Добавление строк и столбцов. Комбинирование методов. /Лек/	5	2	
1.16	Изменение датафрейма. Добавление строк и столбцов. Комбинирование методов. /Пр/	5	2	
	Раздел 2. Методы машинного обучения			
2.1	Решающие деревья: основные понятия. Классификация. Деревья решений. Основные параметры дерева. Дерево решений в задаче регрессии. Метод ближайших соседей. /Лек/	6	2	
2.2	Применение ансамблевых методов в машинном обучении. Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. /Лек/	6	2	
2.3	Проблема переобучения. Определение переобучения, методы борьбы с ним. /Лек/	6	2	
2.4	Нейронные сети. Искусственные нейроны. Перцептрон. Обучение нейрона. /Лек/	6	2	
2.5	Градиентный спуск. Дробление шага при градиентном спуске. Стохастический градиентный спуск. Метод наискорейшего спуска. /Лек/	6	2	
2.6	Нейронная сеть. Однослойная модель. /Лек/	6	2	
2.7	Метод обратного распространения ошибки. /Лек/	6	2	
2.8	Метод опорных векторов. /Лек/	6	2	
2.9	Решающие деревья: обучение модели. Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных. Дерево классификации. /Лаб/	6	2	
2.10	Решающие деревья. Выбор оптимальных параметров дерева. Дерево решений в задачах регрессии. /Лаб/	6	2	
2.11	Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных. /Лаб/	6	2	
2.12	Применение ансамблевых методов в машинном обучении. Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. Проблема переобучения. /Лаб/	6	2	
2.13	Типы ошибок: true negative, false positive. Метрики качества: Precision, Recall, F1 score. Тестирование ROC AUC. /Лаб/	6	2	
2.14	Распознавание рукописных цифр. Алгоритмы случайного леса. Оптимальные параметры дерева и леса. /Лаб/	6	2	
2.15	Искусственный нейрон. Вход, функция активации (сигмоида, гиперболический тангенс, SoftMax). Подбор весов. /Лаб/	6	4	

2.16	Определение тональности текста с помощью нейронных сетей. Библиотеки tensorflow, Keras, набор данных IMDb. /Лаб/	6	2	
2.17	Библиотека Keras. Полносвязный и сглаживающий слой, слой Dropout. Решение проблем переобучения. /Лаб/	6	2	
2.18	Классификация изображений по нескольким классам (более двух). Набор данных Fashion Mnist. /Лаб/	6	4	
2.19	Внешние источники данных для анализа. Kuggle. Подготовка данных. Работа с изображениями разного формата. /Лаб/	6	2	
2.20	Распознавание образов. Классификация изображений кошек и собак. Библиотека keras, ImageDataGenerator. Достижение заданной точности модели на валидационной выборке. /Лаб/	6	4	
2.21	Написание собственных классов для описания нейронной сети. Нейрон. Слой. Сеть. Обучение модели. Выполнение предсказания. /Лаб/	6	4	
2.22	Предсказание пола по росту и весу с помощью нейронной сети. Выполнение предсказания. Минимизация значения функции потерь. Обучение: стохастический градиентный спуск. /Лаб/	6	2	
2.23	Метод опорных векторов. /Лаб/	6	4	
2.24	Метод обратного распространения ошибки /Лаб/	6	4	
2.25	Работа с неструктурированными данными. /Лаб/	6	4	
	Раздел 3. Самостоятельная работа			
3.1	Библиотека NumPy. Работа с матрицами. /Ср/	5	6	
3.2	Подготовка к лекциям /Ср/	5	8	
3.3	Подготовка к практическим работам /Ср/	5	7	
3.4	Алгоритм обратного распространения ошибки. Функции активации. Оценка работы сети. /Ср/	6	18	
3.5	Задачи распознавания образов. Классификация объектов. Поиск изображения по образцу. /Ср/	6	18	
3.6	Компьютерное зрение. Библиотека OpenCV. Сегментация изображений. Детектирование объектов. /Ср/	6	18	
3.7	Компьютерное зрение. Библиотека OpenCV. Отслеживание движущихся объектов во времени. Распознавание лиц. /Ср/	6	20	
3.8	Подготовка к лекциям /Ср/	6	5	
3.9	Подготовка к лабораторным занятиям /Ср/	6	30	
3.10	Обработка естественного языка. Классификация длинных текстов. Векторная модель текста. /Ср/	5	10	
3.11	Обработка естественного языка. Дистрибутивная семантика. Векторные представления слов. /Ср/	5	10	
3.12	Математические основы искусственного интеллекта /Ср/	5	40	
3.13	Биологический нейрон. Принцип работы. /Ср/	5	10	
3.14	Лингвистический анализ. Извлечение признаков. /Ср/	5	12	
3.15	Языковые модели. Генерация текста. Рекуррентные нейронные сети. /Ср/	6	16	
	Раздел 4. Контактные часы на аттестацию			
4.1	Зачёт /КЭ/	5	0,25	
4.2	Экзамен /КЭ/	6	2,35	

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Оценочные материалы для проведения промежуточной аттестации обучающихся приведены в приложении к рабочей программе дисциплины.

Формы и виды текущего контроля по дисциплине (модулю), виды заданий, критерии их оценивания,

распределение баллов по видам текущего контроля разрабатываются преподавателем дисциплины с учетом ее специфики и доводятся до сведения обучающихся на первом учебном занятии.

Текущий контроль успеваемости осуществляется преподавателем дисциплины (модуля), как правило, с использованием ЭИОС или путем проверки письменных работ, предусмотренных рабочими программами дисциплин в рамках контактной работы и самостоятельной работы обучающихся. Для фиксации результатов текущего контроля может использоваться ЭИОС.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л1.1	Никольский С. Н.	Автоматизация информационного поведения и искусственный интеллект: Учебное пособие	Москва: МИРЭА, 2020	https://e.lanbook.com/book/163824

6.2 Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине (модулю)

6.2.1 Перечень лицензионного и свободно распространяемого программного обеспечения

6.2.1.1 Python

6.2.2 Перечень профессиональных баз данных и информационных справочных систем

6.2.2.1 Информационная справочная система Техэксперт <https://tech.company-dis.ru>

6.2.2.2 Информационная справочная система "Гарант" <http://www.garant.ru>

6.2.2.3 База данных Государственных стандартов <http://gostexpert.ru/>

6.2.2.4 База данных «Железнодорожные перевозки» <https://cargo-report.info/>

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

- 7.1 Учебные аудитории для проведения занятий лекционного типа, укомплектованные специализированной мебелью и техническими средствами обучения: мультимедийное оборудование для предоставления учебной информации большой аудитории и/или звукоусиливающее оборудование (стационарное или переносное).
- 7.2 Учебные аудитории для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения: мультимедийное оборудование и/или звукоусиливающее оборудование (стационарное или переносное).
- 7.3 Помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.
- 7.4 Помещения для хранения и профилактического обслуживания учебного оборудования.
- 7.5 Лаборатории, оснащенные специальным лабораторным оборудованием: компьютерной техникой с установленным ПО: Python, а также с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета