



Технологии машинного обучения

ИУ-5, бакалавриат, 6 семестр, весна 2023 года

2023 год

Программа дисциплины №103 <u>Технологии машинного обучения</u> кафедры ИУ5 2019 / 2020 / 2021 / 2022 / 2023 года



Используется в УП:

Читается в 2023 году в следующих УП:

Обложка программы 🖺 Литература Файлы Компетенции Программное обеспечение

Автор(ы): Примечание:

Уровень подготовки: Бакалавр

Тип: Общая

Форма обучения: Первое образование

Семестры		3.E.	Всего	Лек	Сем	Лр	Др	Сам	Свободно	Аттестация	Баллы за ДМ	
Семестр 1 17 недель	Объем		108	34 0	0	17	0		4.75	Экзамен (+30 баллов +36 часов)	ДМ 1	35
	Кол-во	3		17	0	6	0	57			ДМ 2	35
											Итого	100
Итого:		3	108	34	0	17	0	57	40.75		100	

Семестры		Недели																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Семестр 1 17 недель	Модули											M						M			
	KM										PK						PK				
	Объем										3						3				
	Тип																				
	Объем																				

Дисциплина «НИРС» - проект по машинному обучению.

Объём дисциплины по видам учебных занятий (в часах) – 3 з.е.

Виды учебной работы	Объем по	семестрам, ч
виды учесной рассты	Всего	1 семестр
Объем дисциплины	108	108
Аудиторная работа	51	51
Лекции (Л)	34	34
Лабораторные работы (ЛР)	17	17
Самостоятельная работа (СР)	57	57
Проработка материала лекций	9	9
Подготовка к лабораторным работам	12	12
Подготовка к рубежным контролям	6	6
Подготовка к экзамену	30	30
Вид промежуточной аттестации		экзамен

Содержание дисциплины, структурированное по модулям

Nº	To (Виды занятий, часы						
п/п	Тема (название) модуля	\wedge	ΛP	СР				
	1 семестр							
1.	Технологии разведочного анализа и обработки данных	22	9	14				
2.	Технологии использования и оценки моделей машинного обучения	12	8	13				
4.	Подготовка к экзамену	0	0	30				

Структура курса

- Лекции:
 - CPEΔA, 15.40, 524 Γ3
- Лабораторные работы 6 лр.
 - ВТОРНИК, СРЕДА, ПЯТНИЦА (по расписанию)
- НИРС:
 - Решение комплексной задачи, объединяющей различные технологии анализа данных и машинного обучения https://github.com/Yorko/mlcourse.ai/wiki/Individual-projects-and-tutorials-(in-Russian)
- Репозиторий курса:
 - https://github.com/ugapanyuk/courses_current/wiki/COURSE_TMO_SPRING_2023
- Бонусы:
 - участие в соревнованиях по машинному обучению, подготовка статей, выступление на конференциях.

Модуль 1. Технологии разведочного анализа и обработки данных

- 1. Вводная лекция. Цели и задачи курса. Содержание курса. Обзор существующих источников по анализу данных и машинному обучению.
- 2. Понятие набора данных (датасета). Использование библиотеки Pandas для обработки наборов данных. Соединение данных, группировка данных, другие операции с данными. Использование библиотеки Pandasql для выполнения запросов над наборами данных.
- 3. Разведочный анализ данных. Корреляционный анализ данных. Обработка пропусков в данных. Масштабирование данных и его влияние на качество моделей машинного обучения. Основы теории шкалирования. Извлечение признаков для числовых, текстовых, графических данных, даты и времени.
- 4. Визуализация данных с помощью библиотек Matplotlib и Seaborn.
- 5. Библиотека NumPy и векторизация вычислений в Python. Использование технологии разреженных матриц для работы с большими наборами данных.

Модуль 2. Технологии использования и оценки моделей машинного обучения

- 1. Основные задачи машинного обучения: обучение с учителем, обучение без учителя, обучение с подкреплением, другие виды обучения. Задачи классификации и регрессии.
- 2. Основные модели машинного обучения, предназначенные для решения задач классификации и регрессии: метод ближайших соседей, методы на основе градиентного спуска с регуляризацией, машина опорных векторов, деревья решений. Ансамбли моделей: случайный лес, градиентный бустинг. Использование библиотеки scikit-learn для построения моделей.
- 3. Оценка качества моделей классификации и регрессии. Метрики качества и их выбор для оценки моделей.

Курсы по машинному обучению

- https://netology.ru/programs/data-scientist
- http://digitaltech.school/courses/data-science-and-python
- https://skillbox.ru/course/profession-machine-learning/
- https://geekbrains.ru/geek_university/data-science
- https://otus.ru/categories/data-science/
- https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie краткий курс
- https://www.coursera.org/specializations/machine-learning-dataanalysis - специализация из 6 курсов
- https://mlcourse.ai/ курс по машинному обучению от ОрепDataScience. Русская версия курса - https://ods.ai/tracks/open-ml-course Статьи на хабре: https://habr.com/ru/company/ods/blog/322626/

Интересные ссылки:

- http://www.machinelearning.ru основной ресурс по машинному обучению на русском
- https://neurohive.io/ru/ статьи на русском по ИИ
- https://www.kaggle.com/ (соревнования, датасеты, курсы, ноутбуки)
- https://towardsdatascience.com/
- https://machinelearningmastery.com/
- https://paperswithcode.com/sota публикации, сгруппированные по основным направлениям ИИ
- https://arxiv.org/ открытая библиотека научных публикаций Корнеллского университета

• Одна из лучших книг по машинному обучению.

O'REILLY"

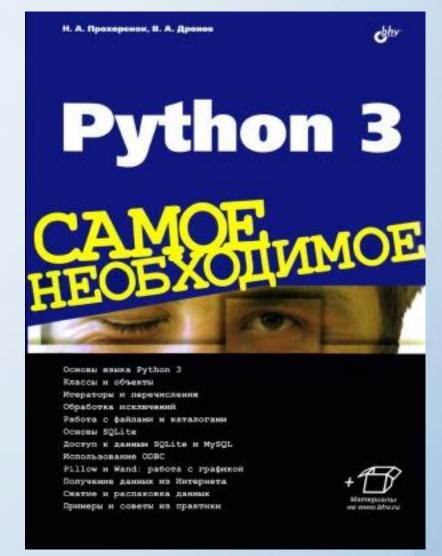
ASHOPINA

Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow

Концепции, инструменты и техники для создания интеллектуальных систем



• Хорошая вводная книга непосредственно по Python.



Дж. Вандер Плас

• Хорошая практическая книга с описанием библиотек Python. Меньше примеров задач, больше описания библиотек.

Python для сложных задач наука о данных: и машинное обучение



Санкт-Петербург - Москва - Екатеринбург - Воронеж Нижний Новгород - Ростов-на-Дону - Самара - Минск

2018

- Рассматривается весь жизненный цикл анализа данных выделение признаков, оценка качества моделей.
- Рассматриваются много различных задач машинного обучения, в том числе довольно специфических.

мнение экспертов

Python и машинное обучение

Машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2

3-е издание — охватывает TensorFlow 2, порождающие состязательные сети и обучение с подкреплением

Себастьян Рашка Вахид Мирджалили



- Хорошая практическая книга. Исторически была переведена первой.
- Содержит в основном примеры решения задач. В меньшей степени содержит описание библиотек.



- Рассматриваются как задачи обучения с учителем, так и задачи обучения без учителя.
- Рассматривается весь жизненный цикл анализа данных выделение признаков, оценка качества моделей.

Андреас Мюллер, Сара Гвидо

Введение в машинное обучение с помощью Python

Руководство для специалистов по работе с данными



Москва 2016-2017

- Хорошая вводная книга в основном по методам машинного обучения. Методы разбираются достаточно детально.
- Примеров кода относительно немного.



- Теоретический учебник. Разбираются теоретические основы машинного обучения на основе большого количества примеров.
- Не привязан к конкретному языку программирования.

