



Технологии машинного обучения

ИУ-5, бакалавриат, 6 семестр, весна
2021 года

Программа дисциплины №91 Технологии машинного обучения кафедры ИУ5

[2016](#)  / [2017](#)  / [2018](#)  / [2019](#)  / [2020](#)  / [2021](#)  года

[Обложка программы](#) [Литература](#)  [Файлы](#) [Компетенции](#) [Программное обеспечение](#)

Автор(ы):

Примечание:

Уровень подготовки: Бакалавр

Тип: Общая

Семестры		З.Е.	Всего	Лек	Сем	Лр	Др	Сам	Аттестация	Баллы за ДМ	
Семестр 1 17 недель	Объем	5	180	34	0	17	0	129	Экзамен (+30 баллов +36 часов) ДЗач за Кур	ДМ 1	35
	Кол-во			17	0	6	0			ДМ 2	35
Итого:		5	180	34	0	17	0	129		Итого	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											М						М			
	КМ										РК						РК				
	Объем										3						3				
	Тип																	Кур			
	Объем																	72			

Объём дисциплины по видам учебных занятий (в часах) – 5 з.е.

Виды учебной работы	Объем по семестрам, ч	
	Всего	1 семестр
Объем дисциплины	180	180
Аудиторная работа	51	51
Лекции (Л)	34	34
Лабораторные работы (ЛР)	17	17
Самостоятельная работа (СР)	129	129
Проработка материала лекций	7	7
Подготовка к лабораторным работам	8	8
Подготовка к рубежным контролям	6	6
Подготовка к экзамену	36	36
Выполнение курсовой работы	72	72
Вид промежуточной аттестации	экзамен, дифф. зачет (КР)	

Содержание дисциплины, структурированное по модулям

№ п/п	Тема (название) модуля	Виды занятий, часы		
		Л	ЛР	СР
1 семестр				
1.	Технологии разведочного анализа и обработки данных	22	9	11
2.	Технологии использования и оценки моделей машинного обучения	12	8	10
3.	Курсовая работа	0	0	72
4.	Подготовка к экзамену	0	0	36
	ИТОГО	34	17	129

Структура курса

- Лекции:
 - СРЕДА, 13.50, 515ю ГЗ
- Лабораторные работы – 6 лр.
 - ПОНЕДЕЛЬНИК, СРЕДА, ЧЕТВЕРГ (по расписанию)
- Курсовая работа:
 - Решение комплексной задачи, объединяющей различные технологии анализа данных и машинного обучения – [https://github.com/Yorko/mlcourse.ai/wiki/Individual-projects-and-tutorials-\(in-Russian\)](https://github.com/Yorko/mlcourse.ai/wiki/Individual-projects-and-tutorials-(in-Russian))
- Репозиторий курса:
 - https://github.com/ugapanyuk/ml_course_2021/wiki/COURSE_TMO
- Бонусы:
 - участие в соревнованиях по машинному обучению, подготовка статей, выступление на студенческой весне.

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/promo/spec-ds-light/>
- <https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie> - краткий курс
- <https://www.coursera.org/specializations/machine-learning-data-analysis> - специализация из 6 курсов
- <https://mlcourse.ai/> - курс по машинному обучению от OpenDataScience. 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/>
- <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

Книги

- Одна из лучших книг по машинному обучению.



Книги

- Хорошая вводная книга непосредственно по Python.



Книги

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

Дж. Вандер Плас

Python

для сложных задач
наука о данных:
и машинное обучение

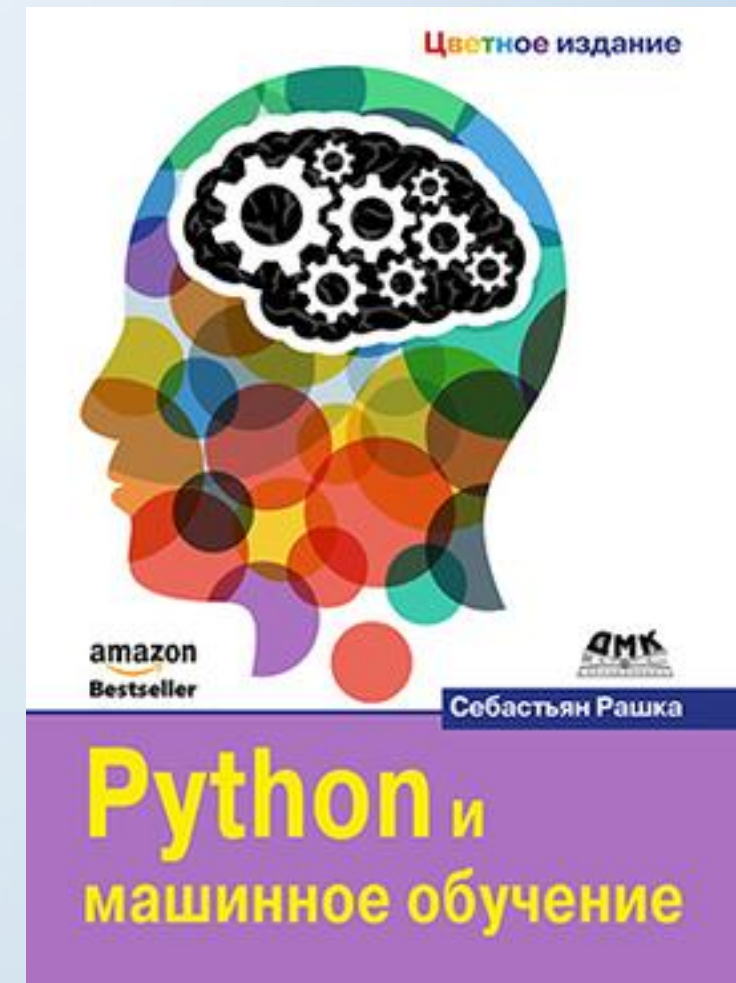


Санкт-Петербург • Москва • Екатеринбург • Воронеж
Нижний Новгород • Ростов-на-Дону • Самара • Минск

2018

Книги

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



Книги

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



Книги

- 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

Книги

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



Книги

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

