

МИНПРОСВЕЩЕНИЯ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
"Тульский государственный педагогический университет им. Л.Н. Толстого"
(ФГБОУ ВО "ТГПУ им. Л.Н. Толстого")

Машинное обучение

рабочая программа дисциплины (модуля)

Закреплена за кафедрой	кафедра алгебры, математического анализа и геометрии
ОПОП	Направление 02.03.02 Фундаментальная информатика и информационные технологии направленность (профиль) Инженерия программного обеспечения
Квалификация	Бакалавр
Год начала подготовки	2022
Форма обучения	очная
Общая трудоемкость	4 з.е.

Виды контроля по семестрам:
экзамен 8

Семестр(Курс.Номер семестра на курсе)	8(4.2)		Итого	
	УП	РПД	УП	РПД
Лекции	22	22	22	22
Лабораторные	24	24	24	24
Итого ауд.	46	46	46	46
КСР	2	2	2	2
Контактная работа	48	48	48	48
Сам. работа	60	60	60	60
Часы на контроль	36	36	36	36
Практическая подготовка	0	0	0	0
Семинары	0	0	0	0
Консультации	0	0	0	0
Итого трудоемкость в часах	144	144	144	144

Программу составил(и):

Рабочая программа дисциплины

Машинное обучение

разработана в соответствии с ФГОС:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии (приказ Минобрнауки России от 23.08.2017 г. № 808)

составлена на основании учебного плана:

Направление 02.03.02 Фундаментальная информатика и информационные технологии
направленность (профиль) Инженерия программного обеспечения
утвержденного Учёным советом вуза от 28.02.2022 протокол № 3.

РПД утверждена Учёным советом университета
протокол от 28.2.2022 г. № 3

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

формирование представления о современных математических моделях и методах, применяемых к автоматизации обработки больших объёмов данных, с элементами обучения и получения навыков практического применения полученных знаний при реализации программных комплексов

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП

Цикл (раздел) ООП:	Б1.В
2.1	Требования к предварительной подготовке обучающегося:
1.	Программирование
2.	Объектно-ориентированное программирование
3.	Методы оптимизации
4.	Компьютерное моделирование
5.	Аналитика больших данных
6.	Исследование операций
7.	эксплуатационная практика
8.	Применение нейронных сетей
9.	Технологии виртуальной реальности
10.	Экономико-правовые основы рынка программного обеспечения
11.	Компьютерное моделирование
12.	Методы оптимизации
13.	Технологии параллельного программирования
14.	Экономические информационные системы
15.	Основы проектирования веб-интерфейсов
16.	технологическая (проектно-технологическая) практика
17.	Веб-программирование
18.	Операционные системы
19.	Объектно-ориентированное программирование
20.	практика по получению первичных навыков профессиональной деятельности
21.	Программирование
2.2	Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:

3. СООТНЕСЕНИЕ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ) С ИНДИКАТОРАМИ ДОСТИЖЕНИЯ КОМПЕТЕНЦИЙ

3.1 Компетенции обучающегося и индикаторы их достижения:	
ОПК-2: Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности	
ОПК-2.1	Знает основные положения и концепции в области программирования, архитектуру языков программирования, теории коммуникации, знает основную терминологию, знаком с содержанием Единого Реестра Российских программ
	основные понятия и алгоритмы машинного обучения и анализа данных и основные понятия, алгоритмы и структуры данных необходимых для их программной реализации
ОПК-2.2	Умеет анализировать типовые языки программирования, составлять программы
	формулировать задачи машинного обучения и анализа данных в виде алгоритмов и структур данных
ОПК-2.3	Имеет практический опыт решения задач анализа, интеграции различных типов программного обеспечения, анализа типов коммуникаций
	Владеть практической реализацией методов машинного обучения и анализа данных на языке python
ПК-3: Способность к реализации всех этапов жизненного цикла программного обеспечения на основе современных парадигм, методологий, инструментальных и вычислительных средств	
ПК-3.1	Знает основы разработки и реализации процессов жизненного цикла программного обеспечения
	основные понятия и машинного обучения и анализа данных
ПК-3.2	Умеет разрабатывать и документировать программные средства на всех этапах их жизненного цикла
	формулировать задачи в терминах машинного обучения и анализа данных, подбирать необходимую математическую модель и алгоритмы
ПК-3.3	Имеет практический опыт разработки программных средств и документации
	практическое использование методов машинного обучения и анализа данных для решения прикладных задач

3.2 Результаты обучения по дисциплине:**В результате освоения дисциплины обучающийся должен:**

	Знать:
3.1	основные понятия и алгоритмы машинного обучения и анализа данных и основные понятия, алгоритмы и структуры данных необходимых для их программной реализации;
3.2	основные понятия и машинного обучения и анализа данных
	Уметь:
У.1	формулировать задачи машинного обучения и анализа данных в виде алгоритмов и структур данных
У.2	формулировать задачи в терминах машинного обучения и анализа данных, подбирать необходимую математическую модель и алгоритмы
	Владеть:
В.1	Владеть: практической реализацией методов машинного обучения и анализа данных на языке python
В.2	Владеть: практическим использованием методов машинного обучения и анализа данных для решения прикладных задач

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Литература	Содержание
	Машинное обучение				
1.1	Введение в предметную область. Основы Python. Специализированные библиотеки. Методы предобработки данных /Лек/	8	6		Введение в предметную область. Примеры использования методов машинного обучения для решения прикладных задач. Повторение основ программирования на языке Python. Знакомство со специализированными библиотеками языка программирования Python для научных расчетов и анализа данных. NumPy, SciPy, pandas. Знакомство с различными методами предобработки данных, описательными статистиками и основными способами визуализации данных, методами снижения размерности. Метод главных компонент. Важность нормировки данных. Предобработка данных. Работа с пропущенными значениями
1.2	Введение в предметную область. Основы Python. Специализированные библиотеки. Методы предобработки данных /Лаб/	8	4		Введение в предметную область. Примеры использования методов машинного обучения для решения прикладных задач. Повторение основ программирования на языке Python. Знакомство со специализированными библиотеками языка программирования Python для научных расчетов и анализа данных. NumPy, SciPy, pandas. Знакомство с различными методами предобработки данных, описательными статистиками и основными способами визуализации данных, методами снижения размерности. Метод главных компонент. Важность нормировки данных. Предобработка данных. Работа с пропущенными значениями

1.3	Введение в предметную область. Основы Python. Специализированные библиотеки. Методы предобработки данных /Ср/	8	12		Введение в предметную область. Примеры использования методов машинного обучения для решения прикладных задач. Повторение основ программирования на языке Python. Знакомство со специализированными библиотеками языка программирования Python для научных расчетов и анализа данных. NumPy, SciPy, pandas. Знакомство с различными методами предобработки данных, описательными статистиками и основными способами визуализации данных, методами снижения размерности. Метод главных компонент. Важность нормировки данных. Предобработка данных. Работа с пропущенными значениями
1.4	Основы машинного обучения. Обучение на неразмеченных данных /Лек/	8	4		Основы машинного обучения и основные типы задач. Классификация задач машинного обучения. Обучение на неразмеченных данных. Кластеризация. Иерархическая кластеризация. Метод К-средних, DBSCAN и др. Обзор методов кластеризации, реализованных в библиотеке sklearn.
1.5	Основы машинного обучения. Обучение на неразмеченных данных /Лаб/	8	6		Основы машинного обучения и основные типы задач. Классификация задач машинного обучения. Обучение на неразмеченных данных. Кластеризация. Иерархическая кластеризация. Метод К-средних, DBSCAN и др. Обзор методов кластеризации, реализованных в библиотеке sklearn.
1.6	Основы машинного обучения. Обучение на неразмеченных данных /Ср/	8	12		Основы машинного обучения и основные типы задач. Классификация задач машинного обучения. Обучение на неразмеченных данных. Кластеризация. Иерархическая кластеризация. Метод К-средних, DBSCAN и др. Обзор методов кластеризации, реализованных в библиотеке sklearn.
1.7	Задача обучения с учителем /Лек/	8	4		Задачи обучения с учителем. Разделение данных на обучающие и тестовые. Определение переобученности модели. Критерии оценки качества полученных моделей.
1.8	Задача обучения с учителем /Лаб/	8	4		Задачи обучения с учителем. Разделение данных на обучающие и тестовые. Определение переобученности модели. Критерии оценки качества полученных моделей.
1.9	Задача обучения с учителем /Ср/	8	12		Задачи обучения с учителем. Разделение данных на обучающие и тестовые. Определение переобученности модели. Критерии оценки качества полученных моделей.
1.10	Задачи регрессии. Задачи классификации /Лек/	8	4		Постановка задачи регрессии. Линейный регрессионный анализ. Отбор признаков, коллинеарность, влиятельные наблюдения, анализ остатков. Непараметрическая регрессия (ядерное сглаживание). L1 и L2 регуляризация. Метрики качества Постановка задачи классификации, обзор основных методов ее решения. Бинарная и многоклассовая классификация. Логистическая регрессия. Решающие деревья. Метрики качества классификации (точность/специфичность, ROC-кривая, площадь под кривой).

1.11	Задачи регрессии. Задачи классификации /Лаб/	8	6		Постановка задачи регрессии. Линейный регрессионный анализ. Отбор признаков, коллинеарность, влиятельные наблюдения, анализ остатков. Непараметрическая регрессия (ядерное сглаживание). L1 и L2 регуляризация. Метрики качества Постановка задачи классификации, обзор основных методов ее решения. Бинарная и многоклассовая классификация. Логистическая регрессия. Решающие деревья. Метрики качества классификации (точность/специфичность, ROC-кривая, площадь под кривой).
1.12	Задачи регрессии. Задачи классификации /Ср/	8	12		Постановка задачи регрессии. Линейный регрессионный анализ. Отбор признаков, коллинеарность, влиятельные наблюдения, анализ остатков. Непараметрическая регрессия (ядерное сглаживание). L1 и L2 регуляризация. Метрики качества Постановка задачи классификации, обзор основных методов ее решения. Бинарная и многоклассовая классификация. Логистическая регрессия. Решающие деревья. Метрики качества классификации (точность/специфичность, ROC-кривая, площадь под кривой).
1.13	Ансамбль алгоритмов машинного обучения /Лек/	8	4		Ансамбли алгоритмов машинного обучения. Агрегирование моделей. Ансамбли решающих деревьев. Метод случайного леса. Градиентный бустинг.
1.14	Ансамбль алгоритмов машинного обучения /Лаб/	8	4		Ансамбли алгоритмов машинного обучения. Агрегирование моделей. Ансамбли решающих деревьев. Метод случайного леса. Градиентный бустинг.
1.15	Ансамбль алгоритмов машинного обучения /Ср/	8	12		Ансамбли алгоритмов машинного обучения. Агрегирование моделей. Ансамбли решающих деревьев. Метод случайного леса. Градиентный бустинг.
1.16	Задачи обучения с учителем /КСР/	8	2		Задачи обучения с учителем. Разделение данных на обучающие и тестовые. Определение переобученности модели. Критерии оценки качества полученных моделей.

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

5.1. Типовые задания для проведения текущего контроля

1. Постановка задачи восстановления логистической регрессии
 2. Алгоритм отыскания оптимальных параметров
 3. Алгоритм Ньютона-Рафсона.
 4. Линейная регрессия (пример).
 5. Полиномиальная регрессия.
 6. Криволинейная регрессия.
 7. Применение линейной постановки задачи для моделирования кривых второго порядка.
 8. Сравнительный анализ различных определений обучаемости
- Задача 1. Реализуйте алгоритм kNN классификации по k ближайшим соседям, используя простое евклидовое расстояние.
- Задача 2. Реализуйте алгоритм k-means для кластеризации на 2-4 кластера.
- Задача 3. Реализуйте алгоритм DBSCAN, найдите параметры для кластеризации на 4 кластера.
- Задача 4. Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10).
- Задача 5. Примените метод SVM (например, из библиотеки sklearn) для датасета blobs2. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF).
- Задача 6. Реализуйте алгоритм логистической регрессии со стохастическим градиентным спуском, обучите его на датасете spmbase_old (train) и проверьте на датасете

spmbase_new (val). Получите ROC кривые для вариантов без нормировки и с нормировкой признаков.
 Задача 7. Модифицируйте модель из задачи 3, заменив последний нейрон на 10 нейронов, и реализовав мультиклассовую классификацию с softmax в качестве решающей функции и кросс-энтропией в качестве функции потерь и обучите на подготовленном датасете mnist.

5.2. Типовые задания для проведения промежуточной аттестации

Вопросы к экзамену

1. Примеры использования методов машинного обучения для решения прикладных задач.
2. Основы программирования на языке Python.
3. Специализированные библиотеки языка программирования Python для научных расчетов и анализа данных. NumPy, SciPy, pandas.
4. Методы предобработки данных.
5. Описательные статистики
6. Основные способы визуализации данных.
7. Методы снижения размерности.
8. Метод главных компонент. Важность нормировки данных.
9. Предобработка данных.
10. Работа с пропущенными значениями.
11. Основы машинного обучения и основные типы задач.
12. Классификация задач машинного обучения.
13. Обучение на неразмеченных данных.
14. Кластеризация. Иерархическая кластеризация.
15. Метод K-средних, DBSCAN и др. Обзор методов кластеризации, реализованных в библиотеке sklearn.
16. Задачи обучения с учителем. Разделение данных на обучающие и тестовые.
17. Определение переобученности модели.
18. Критерии оценки качества полученных моделей.
19. Постановка задачи регрессии. Линейный регрессионный анализ.
20. Отбор признаков, коллинеарность, влиятельные наблюдения, анализ остатков.
21. Непараметрическая регрессия (ядерное сглаживание). L1 и L2 регуляризация.
22. Метрики качества.
23. Постановка задачи классификации, обзор основных методов ее решения.
24. Бинарная и многоклассовая классификация.
25. Логистическая регрессия.
26. Решающие деревья.
27. Метрики качества классификации (точность/специфичность, ROC-кривая, площадь под кривой).
28. Ансамбли алгоритмов машинного обучения. Агрегирование моделей.
29. Ансамбли решающих деревьев. Метод случайного леса. Градиентный бустинг.

5.3. Перечень видов оценочных средств

Контроль и оценка результатов освоения учебной дисциплины осуществляется преподавателем в процессе проведения практических занятий, а также выполнения двух индивидуальных заданий. Для формирования итоговой оценки знаний, умений используется балльно-рейтинговая система, учитывающая значительную долю практических и индивидуальных занятий. Итоговая рейтинговая оценка по дисциплине рассчитывается из 100 баллов, которые складываются из следующих составляющих: 1) За каждую из двух индивидуальных работ студент может максимально получить по 20 баллов. 2) Студентам, желающим повысить свой рейтинг, предлагаются задания повышенной сложности, которые максимально могут быть оценены в 10 баллов. 3) На экзамене ответ студента может быть максимально оценен в 30 баллов. Отметка «отлично» выставляется, если студент в целом за семестр набрал от 81 до 100 баллов (с учетом баллов, набранных на промежуточной аттестации). Отметка «хорошо» выставляется, если студент в целом за семестр набрал от 61 до 80 баллов (с учетом баллов, набранных на промежуточной аттестации). Отметка «удовлетворительно» выставляется, если студент в целом за семестр набрал от 41 до 60 баллов (с учетом баллов, набранных на промежуточной аттестации). Отметка «неудовлетворительно» выставляется, если студент в целом за семестр набрал менее 41 балла (с учетом баллов, набранных на промежуточной аттестации).

5.4. Процедура применения оценочных материалов

Промежуточная аттестация может проводиться с применением электронного обучения и (или) дистанционных образовательных технологий в соответствии с «Порядком проведения промежуточной аттестации с применением электронного обучения и /или дистанционных образовательных технологий». Проведение экзамена с применением дистанционных образовательных технологий может проходить по следующим процедурам: в форме устного собеседования преподавателя со студентом по предложенным вопросам к экзамену (без предварительной подготовки к конкретному вопросу в период проведения экзамена), в виде решения обучающимся уникального кейс-задания, в виде защиты индивидуального учебного проекта.

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

6.1. Рекомендуемая литература

6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

- | | |
|----|--|
| Э1 | Полнотекстовый архив ведущих западных научных журналов на российской платформе Национального электронно-информационного консорциума (НЭИКОН) |
|----|--|

Э2	Базы данных издательства Springer
Э3	Web of Science Core Collection – политематическая реферативно-библиографическая и наукометрическая (библиометрическая) база данных
Э4	Портал «Информационно-коммуникационные технологии в образовании»
Э5	Большая коллекция материалов по машинному обучению на русском языке.

6.3. Информационные технологии

6.3.1 Перечень лицензионного и свободно распространяемого программного обеспечения

1.	Операционная система ROSA Enterprise Linux Desktop № RL00450-1-110518-01. RL00450-1-110518-17 от 11 мая 2018 г.
2.	Операционная система Microsoft Windows 10 Professional Russian. Контракт № ПР/ФЕН/15/18 от 23.10.2015 г., договор № Пр/16/6 от 05 апреля 2016 г.
3.	Программное обеспечение Microsoft Office Enterprise 2007 Russian. Лицензия №46138962 от 16.11.2009
4.	Программа для распознавания текста ABBYY FineReader 9.0 Corporate Edition. Лицензионный сертификат - код позиции AF90-3U1V25-102, ABBYY FineReader 9.0 Corporate Edition Volume License Concurrent от 28 июля 2009 г.
5.	Электронный словарь ABBYY Lingvo X3 Европейская версия - Код позиции AL14-2U1V05-102, ABBYY Lingvo x3 Европейская версия. Именная лицензия Concurrent от 28 июля 2009 г.
6.	Комплексная система антивирусной защиты Kaspersky Endpoint Security для бизнеса – стандартный Russian Edition. 500-999 Node 2 year Educational Renewal License. Лицензия № 13C8-190514-084943-783-1256 от 15.05.2019
7.	Файловый архиватор 7z. Свободно распространяемое ПО
8.	Браузеры Google Chrome, Mozilla, Opera. Свободно распространяемое ПО
9.	Инструмент для очистки и оптимизации операционных систем Microsoft Windows С Cleaner. Свободно распространяемое ПО
10.	Программа просмотра файлов формата RPD Adobe Acrobat Reader DC. Свободно распространяемое ПО
11.	Среда выполнения Adobe Flash Player. Свободно распространяемое ПО
12.	Файловый менеджер Far manager. Свободно распространяемое ПО
13.	Система Интернет-телефонии Skype. Свободно распространяемое ПО
14.	Редактор диаграмм, схем, блок-схем, UML-схем Dia 0.97.2. Свободно распространяемое ПО

6.3.2 Перечень информационных справочных систем и профессиональных баз данных

1.	Библиотека федерального портала «Российское образование» (http://www.edu.ru)
2.	Полнотекстовый архив ведущих западных научных журналов на российской платформе Национального электронно-информационного консорциума (НЭИКОН)(http://neicon.ru)
3.	Web of Science Core Collection – политематическая реферативно-библиографическая и наукометрическая (библиометрическая) база данных (http://webofscience.com)
4.	Портал «Информационно-коммуникационные технологии в образовании» (http://www.ict.edu.ru)
5.	Портал Федеральных государственных образовательных стандартов высшего образования (http://fgosvo.ru)

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Ауд.	Назначение	Оборудование и технические средства обучения	Вид
4-128	Компьютерная лаборатория экономико-математического моделирования	доска учебная, компьютерные столы, компьютеры, мультимедийная установка Mitsubishi Electric, ноутбук, принтер, программное обеспечение (1С: Бухгалтерия; 1С: Предприятие. Управление торговлей; 1С: Зарплата и управление Персоналом; 1С предприятие. Управление производственным предприятием), сканер	

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Дисциплина «Введение в машинное обучение и анализ данных» направлена на формирование у студентов представления о современных математических моделях и методах, применяемых к автоматизации обработки больших объемов данных, с элементами обучения. Получения навыков практического применения полученных знаний при реализации программных комплексов.