


МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Тульский государственный университет»

Институт прикладной математики и компьютерных наук
Кафедра «Прикладная математика и информатика»

Утверждено на заседании кафедры
«Прикладная математика и информатика»
24 января 2023 г., протокол № 5

И.о. заведующего кафедрой

 Н.В. Ларин

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
«Методы машинного обучения»

**основной профессиональной образовательной программы
высшего образования – программы бакалавриата**

по направлению подготовки
01.03.02 Прикладная математика и информатика

с направленностью (профилем)
Искусственный интеллект и анализ данных

Форма обучения: очная

Идентификационный номер образовательной программы: 010302-02-23

Тула 2023 год

ЛИСТ СОГЛАСОВАНИЯ
рабочей программы дисциплины (модуля)

Разработчик:

Смирнов О.И., доцент каф. ПМиИ, к.ф.-м.н., доцент

(ФИО, должность, ученая степень, ученое звание)



(подпись)

1 Цель и задачи освоения дисциплины (модуля)

Целью освоения дисциплины (модуля) является формирование знаний в области искусственного интеллекта, а также получения навыков проектирования систем искусственного интеллекта и работы с инструментальными средствами реализации принципов искусственного интеллекта.

Задачами освоения дисциплины (модуля) являются:

- формирование теоретических знаний в области искусственного интеллекта;
- развитие навыков решения прикладных задач в области искусственного интеллекта;
- формирование способностей для самостоятельной разработки алгоритмов решения задач и их анализа.

2 Место дисциплины (модуля) в структуре основной профессиональной образовательной программы

Дисциплина (модуль) относится к части основной профессиональной образовательной программы, формируемой участниками образовательных отношений.

Дисциплина (модуль) изучается в шестом и седьмом семестрах.

3 Перечень планируемых результатов обучения по дисциплине (модулю)

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения основной профессиональной образовательной программы (формируемыми компетенциями) и индикаторами их достижения, установленными в общей характеристике основной профессиональной образовательной программы, приведён ниже.

В результате освоения дисциплины (модуля) обучающийся должен:

Знать:

- 1) анализ требований и определяет необходимые классы задач машинного обучения (код компетенции – ПК-4, код индикатора – ПК-4.1);
- 2) метрики оценки результатов моделирования и критерии качества построенных моделей (код компетенции – ПК-4, код индикатора – ПК-4.2).

Уметь:

- 1) осуществлять оценку и выбор инструментальных средств для решения поставленной задачи (код компетенции – ПК-5, код индикатора – ПК-5.1);
- 2) разрабатывать модели машинного обучения для решения задач (код компетенции – ПК-5, код индикатора – ПК-5.2);

Владеть:

- 1) навыками оценки, выбора и при необходимости разработки методов машинного обучения (код компетенции – ПК-4, код индикатора – ПК-4.3);
- 2) навыками создания, поддержки и использования систем искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения (код компетенции – ПК-5, код индикатора – ПК-5.3).

Полные наименования компетенций и индикаторов их достижения представлены в общей характеристике основной профессиональной образовательной программы.

4 Объем и содержание дисциплины (модуля)

4.1 Объем дисциплины (модуля), объем контактной и самостоятельной работы обучающегося при освоении дисциплины (модуля), формы промежуточной аттестации по дисциплине (модулю)

Номер семестра	Формы промежуточной аттестации	Общий объем в зачетных единицах	Общий объем в академических часах	Объем контактной работы в академических часах						Объем самостоятельной работы в академических часах
				Лекционные занятия	Практические (семинарские) занятия	Лабораторные работы	Клинические практические занятия	Консультации	Промежуточная аттестация	
Очная форма обучения										
6	ДЗ, КП	4	144	32	–	32	–	2,5	0,5	77
7	Э	4	144	28	–	14	–	2	0,25	99,75
Итого	–	8	288	60	–	46	–	4,5	0,75	176,75

Условные сокращения: Э – экзамен, ЗЧ – зачет, ДЗ – дифференцированный зачет (зачет с оценкой), КП – защита курсового проекта, КР – защита курсовой работы.

4.2 Содержание лекционных занятий

Очная форма обучения

№ п/п	Темы лекционных занятий
6 семестр	
1	Введение. Терминология машинного обучения
2	Постановка основных задач
3	Математика в машинном обучении: краткий обзор
4	Оптимизация в машинном обучении
5	Метрические алгоритмы
6	Линейные модели
7 семестр	
7	Деревья решений
8	Контроль качества и выбор модели
9	Ансамблирование в машинном обучении
10	Методы, основанные на деревьях: случайный лес, бустинг
11	Введение в рекомендательные системы
12	Сложность алгоритмов, переобучение, смещение и разброс

4.3 Содержание практических (семинарских) занятий

Занятия указанного типа не предусмотрены основной профессиональной образовательной программой.

4.4 Содержание лабораторных работ

Очная форма обучения

№ п/п	Наименования лабораторных работ
6 семестр	
1	Обучение с учителем (с размеченными данными / метками) целевая функция объект метка классификация. Прогнозирование. Пространство объектов признаковое пространство. Извлечение признаков. Визуализация задач функции ошибки эмпирический риск обучающая выборка. Задачи оптимизации в обучении. Модель алгоритмов. Алгоритм. Обучение. Обобщающая способность. Схема решения задачи машинного обучения. Как решаются задачи. Обучение без учителя /с размеченными данными. Обучение с частично размеченными данными трансдуктивное обучение. Обучение с подкреплением. Структурный вывод. Активное обучение. Онлайн-обучение. Transfer Learning. Multitask Learning. Feature Learning. Проблемы в машинном обучении. Примеры модельных задач
2	Бритва Оккама. Теорема о бесплатном сыре. Футбольный оракул. Сведения из теории вероятностей и математической статистики. Задание распределений. Средние и отклонения. Условная плотность, маргинализация и обуславливание. Точечное оценивание. Оценка максимального правдоподобия. Дивергенция Кульбака-Лейблера ковариация и корреляция. Оценка плотности гистограммного подхода. Парзенковский подход. Нормальное распределение. Центральная предельная теорема. Теория информации. Проклятие размерности. Сингулярное разложение матрицы (SVD) матричное дифференцирование
3	Методы безусловной оптимизации. Методы нулевого порядка. Методы первого порядка. Методы второго порядка. Градиентный спуск. Наискорейший градиентный спуск. Стохастический градиентный спуск. Обучение: Пакетное, онлайн, по минибатчам. Метод градиентного спуска в машинном обучении. Стационарные точки. Метод Ньютона. Квази-ньютоновские методы. Оптимизация с ограничениями
4	Метрические алгоритмы. Ближайший центроид. Подход, основанный на близости. kNN в задаче классификации. kNN в задаче регрессии. Обоснование 1NN. Ленивые и нетерпеливые алгоритмы. Весовые обобщения kNN. Различные метрики: Минковского, Евклидова, Манхэттенская, Махаланобиса, Canberra distance, Хэмминга, косинусное, расстояние Джаккарда, DTW, Левенштейна. Приложения метрического подхода: нечёткий матчнинг таблиц, Ленкор, в DL, классификация текстов. Эффективные методы поиска ближайших соседей. Регрессия Надарая-Ватсона
5	Линейная регрессия. Обобщённая линейная регрессия. Проблема вырожденности матрицы. Регуляризация. Основные виды регуляризации. Гребневая регрессия. LASSO (Least Absolute Selection and Shrinkage Operator). Elastic Net. Селекция признаков. Ошибка с весами. Устойчивая регрессия. Линейные скоринговые модели в задаче бинарной классификации. Логистическая регрессия. Probit-регрессия. Многоклассовая логистическая регрессия. Линейный классификатор. Персептрон. Оценка функции ошибок через гладкую функцию
7 семестр	
6	Деревья решений (CART). Предикаты / ветвления. Ответы дерева. Критерии расщепления в задачах классификации: Missclassification criteria, энтропийный, Джини. Критерии останова при построении деревьев. Проблема переобучения для деревьев. Подрезка (post-pruning). Классические алгоритмы построения деревьев решений: ID3, C5.0. Важности признаков. Проблема пропусков. Категориальные признаки. Сравнение: деревья и линейные модели

№ п/п	Наименования лабораторных работ
7	Проблема контроля качества. Выбор модели в широком смысле. Правила разбиения выборки. Отложенный контроль (held-out data, hold-out set). Скользящий контроль. Бутстреп. Контроль по времени. Локальный контроль. Кривые обучения. Перебор параметров
8	Ансамбли алгоритмов: примеры и обоснование. комитеты (голосование) / усреднение. Бэггинг. Кодировки / перекодировки ответов, ECOC. Стекинг и блендинг. Бустинг: AdaBoost, Forward stagewise additive modeling (FSAM). «Ручные методы». Однородные ансамбли
9	Случайный лес, его параметры, их настройка. Бэггинг и OOB (out of bag). Важность признаков. Близость с помощью RF. Extreme Random Trees. Градиентный бустинг над деревьями, его параметры, современные реализации, Продвинутое методы оптимизации.
10	Рекомендательные системы. Персонализация, онлайн и оффлайн рекомендации. Рекомендация по контенту. Коллаборативная фильтрация: GroupLens-алгоритм, SVD, SVD++, timeSVD++, адаптация SVD под социальные связи. One-class recommendation. Факторизационная машина, факторизационная машина с полями (FFM – field-aware factorization machine). Простые методы рекомендаций: FPM – Frequent Pattern Mining. Deep Semantic Similarity Model (DSSM). Контекст рекомендации. Knowledge-based Recommendations. Важность объяснений (explanations). Использование дополнительной информации. Современные тренды в практике построения рекомендательных систем
11	Проблема обобщения. Переобучение. Недообучение. Сложность алгоритмов. Смещение и разброс. Способы борьбы с переобучением

4.5 Содержание клинических практических занятий

Занятия указанного типа не предусмотрены основной профессиональной образовательной программой.

4.6 Содержание самостоятельной работы обучающегося

Очная форма обучения

№ п/п	Виды и формы самостоятельной работы
6 семестр	
1	Подготовка к лабораторным работам
3	Выполнение курсового проекта
3	Подготовка к промежуточной аттестации и ее прохождение
7 семестр	
4	Подготовка к лабораторным работам
5	Подготовка к промежуточной аттестации и ее прохождение

5 Система формирования оценки результатов обучения по дисциплине (модулю) в рамках текущего контроля успеваемости и промежуточной аттестации обучающегося

Очная форма обучения

Мероприятия текущего контроля успеваемости и промежуточной аттестации обучающегося			Максимальное количество баллов
6 семестр			
Текущий контроль успеваемости	Первый рубежный контроль	Оцениваемая учебная деятельность обучающегося:	
		Посещение лекционных занятий.	5
		Выполнение лабораторных работ.	25
		Итого	30
	Второй рубежный контроль	Оцениваемая учебная деятельность обучающегося:	
		Посещение лекционных занятий	5
		Выполнение лабораторных работ.	25
		Итого	30
Промежуточная аттестация	Дифференцированный зачет		40 (100*)
	Защита курсового проекта		100
7 семестр			
Текущий контроль успеваемости	Первый рубежный контроль	Оцениваемая учебная деятельность обучающегося:	
		Посещение лекционных занятий.	5
		Выполнение лабораторных работ.	25
		Итого	30
	Второй рубежный контроль	Оцениваемая учебная деятельность обучающегося:	
		Посещение лекционных занятий	5
		Выполнение лабораторных работ.	25
		Итого	30
Промежуточная аттестация	Экзамен		40 (100*)

* В случае отказа обучающегося от результатов текущего контроля успеваемости

Шкала соответствия оценок в стобалльной и академической системах оценивания результатов обучения по дисциплине (модулю)

Система оценивания результатов обучения	Оценки			
Стобалльная система оценивания	0 – 39	40 – 60	61 – 80	81 – 100
Академическая система оценивания (экзамен, дифференцированный зачет, защита курсового проекта, защита курсовой работы)	Неудовлетворительно	Удовлетворительно	Хорошо	Отлично
Академическая система оценивания (зачет)	Не зачтено	Зачтено		

6 Описание материально-технической базы (включая оборудование и технические средства обучения), необходимой для осуществления образовательного процесса по дисциплине (модулю)

Для осуществления образовательного процесса по дисциплине (модулю) требуется учебная аудитория, оборудованная доской для написания мелом. Для проведения лаборатор-

ных работ требуется аудитория, оснащенная компьютерами с подключением к сети «Интернет» и обеспечением доступа в электронную-образовательную среду.

7 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

7.1 Основная литература

1. Платонов, А. В. Машинное обучение: учебное пособие для вузов / А. В. Платонов. — Москва: Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/520544>.

2. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Москва: Издательство Юрайт, 2023. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511121>.

7.2 Дополнительная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511020>.

2. Бессмертный, И. А. Системы искусственного интеллекта: учебное пособие для вузов / И. А. Бессмертный. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2023. — 157 с. — (Высшее образование). — ISBN 978-5-534-07467-3. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/512657>.

3. Федоров, Д. Ю. Программирование на языке высокого уровня Python: учебное пособие для вузов / Д. Ю. Федоров. — 5-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2023. — 227 с. — (Высшее образование). — ISBN 978-5-534-17323-9. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/532868>.

4. Целых А.Н. Извлечение знаний методами машинного обучения [Электронный ресурс]: учебное пособие по курсам «Модели и методы инженерии знаний», «Методы машинного обучения»/ Целых А.Н., Котов Э.М.— Электрон. текстовые данные.— Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2022.— 105 с.— Режим доступа: <https://ipr-smart.ru/131448>.— IPR SMART, по паролю.

5. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/45418.html>. — Режим доступа: для авторизир. пользователей.

6. Сараев, П. В. Методы машинного обучения : методические указания и задания к лабораторным работам по курсу / П. В. Сараев. — Липецк : Липецкий государственный технический университет, ЭБС АСВ, 2017. — 48 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/83183.html>. — Режим доступа: для авторизир. пользователей.

7. Ракитский, А. А. Методы машинного обучения : учебно-методическое пособие / А. А. Ракитский. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/90591.html>. — Режим доступа: для авторизир. пользователей.

8 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

1. <https://e.lanbook.com/> – ЭБС «Лань», доступ авторизованный.
2. <https://urait.ru/> – Образовательная платформа «Юрайт», доступ авторизованный.
3. <https://www.iprbookshop.ru/> – Цифровой образовательный ресурс IPR SMART, доступ авторизованный.
4. <https://tsutula.bookonline.ru/> – ЭБС ТулГУ «BookOnLime» учебные издания ТулГУ по всем дисциплинам, доступ авторизованный.
5. <https://dlib.eastview.com/browse/udb/12> - Политематическая база данных периодических изданий East View, доступ авторизованный.
6. <https://cyberleninka.ru/> - Научная электронная библиотека «КиберЛенинка», доступ свободный.
7. <https://www.elibrary.ru/> - Научная электронная библиотека eLibrary.ru, доступ свободный.

9 Перечень информационных технологий, необходимых для осуществления образовательного процесса по дисциплине (модулю)

9.1 Перечень необходимого ежегодно обновляемого лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

1. Пакет офисных приложений «МойОфис».

9.2 Перечень необходимых современных профессиональных баз данных и информационных справочных систем

Современные профессиональные базы данных и информационные справочные системы не требуются.