

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет
им. Н.И. Лобачевского»**

Физический факультет

УТВЕРЖДЕНО

решением Ученого совета ННГУ

протокол № 4 от 26.04.2024 г.

Рабочая программа дисциплины

Машинное обучение

Уровень высшего образования

Магистратура

Направление подготовки / специальность

09.04.02 - Информационные системы и технологии

Направленность образовательной программы

Информационные технологии в системах космической связи и дистанционного
зондирования Земли

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.01.02 Машинное обучение относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-12: Способен осуществлять выбор оптимальных решений, моделирование процессов и объектов профессиональной деятельности при проведении научно-исследовательских и опытно-конструкторских работ	ПК-12.1: Знает: способы выбора и методы математического моделирования процессов ПК-12.2: Умеет: применять методы цифровой обработки данных при решении традиционных задач в области информационных технологий ПК-12.3: Владеет: навыками применения математических моделей и объектов профессиональной деятельности при проведении научно-исследовательских и опытно-конструкторских работ	ПК-12.1: Знает основные модели машинного обучения. ПК-12.2: Умеет применять основные модели машинного обучения. ПК-12.3: Владеет навыками выбора наиболее подходящей модели машинного обучения для поставленной задачи.	Задачи	Экзамен: Дискуссия
ПК-13: Способен понимать и применять в научно-исследовательской и прикладной деятельности современный математический аппарат, основные законы естествознания, современные языки программирования, программное обеспечение, операционные системы, сетевые	ПК-13.1: Знает: современный математический аппарат, основные законы естествознания, современные языки программирования и программное обеспечение ПК-13.2: Умеет: проводить разработку алгоритмического и программного обеспечения в области информационных технологий в научных исследованиях ПК-13.3: Имеет:	ПК-13.1: Знает основные методики оценки качества решения задач машинного обучения. ПК-13.2: Умеет применять методики для повышения качества решения задач машинного обучения. ПК-13.3: Владеет навыком составления стека алгоритмов, наиболее подходящего для решения поставленной задачи.	Задачи	Экзамен: Дискуссия

технологии	практический опыт владения существующими методами и алгоритмами решения задач цифровой обработки сигналов, опыт работы с научными источниками			
------------	---	--	--	--

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	5
Часов по учебному плану	180
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	32
- занятия семинарского типа (практические занятия / лабораторные работы)	32
- КСР	2
самостоятельная работа	69
Промежуточная аттестация	45 Экзамен

3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Введение в дисциплину. Формулировка задач регрессия, классификации, кластеризации	6	2		2	4
Метод наименьших квадратов	6	2		2	4
Полиномиальная регрессия	6	2		2	4
Метод ближайшего соседа	6	2		2	4
Метод опорных векторов	6	2		2	4
Задача выявления аномалий	18	2	8	10	8
Квартильный анализ	8	2	2	4	4

Пороговый классификатор и деревья решений	7	3		3	4
Ансамблирование моделей машинного обучения	7	3		3	4
Байесовский классификатор и логистическая регрессия	19	3	8	11	8
Меры расстояний в задачах кластеризации методы кластеризации, основанные на них	15	3	4	7	8
Методы кластеризации, основанные на плотности объектов	12	3	4	7	5
Задача понижения размерности	9	3	2	5	4
Знакомство с инструментами анализа данных	8		4	4	4
Аттестация	45				
КСР	2			2	
Итого	180	32	32	66	69

Содержание разделов и тем дисциплины

1. Введение. Задачи машинного обучения (регрессия, классификация, кластеризация). Категориальные и количественные признаки входных данных. Кодирование категориальных признаков.
2. Задачи регрессии. Метод наименьших квадратов, линейная регрессия, полиномиальная регрессия нескольких переменных. Обработка пропущенных значений.
3. Задачи классификации. Классификация методом ближайшего соседа (N ближайших соседей). Метод опорных векторов и его нелинейная модификация. Критерии качества классификации. Валидация, кроссвалидация.
4. Задачи выявления аномалий. Z-оценка, правило сигм, критерий Шовене. Квартили распределения, квартильный анализ. Метрические методы поиска аномалий.
5. Пороговый классификатор, дерево решений, построение дерева решений на основе принципа максимума информативности признаков.
6. Применение ансамблей моделей машинного обучения: с последовательным, параллельным и комбинированным обучением (бэггинг, бустинг, стеккинг).
7. Байесовский классификатор. Логистическая регрессия, связь логистической регрессии с нейронными сетями. Нормализация признаков и выхода классификатора. Применение метода градиентного спуска. Проблемы взрывов, исчезновения градиентов, проблема переобучения. Начальная инициализация параметров.
8. Задачи кластеризации. Меры расстояний в задачах кластеризации. Методы кластеризации, основанные на расстоянии между объектами в пространстве признаков.
9. Методы кластеризации, основанные на плотности объектов в пространстве признаков. Количественная оценка качества кластеризации.
10. Задачи понижения размерности. Линейное понижение размерности: сингулярное разложение, метод главных компонент. Нелинейное понижение размерности.
11. Нелинейная модификация метода главных компонент с ядрами. Самоорганизующиеся карты Кохонена.
12. Факультативная лекция. Расстояние Кульбака-Лейблера. Понижение размерности с использованием теоретико-информационного подхода стохастическим вложением соседей с t-распределением (t-SNE).

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Самостоятельная работа студентов включает в себя:

- активное изучение учебных и учебно-методических пособий, лекционного материала по соответствующим разделам дисциплины, в том числе с использованием систем компьютерной графики и электронных образовательных ресурсов;
- использование профессиональных прикладных программ моделирования физических процессов и методов обработки данных;
- использование профессиональных прикладных программ моделирования физических процессов и методов обработки данных.

Студенты на основе лекционного материала разрабатывают, тестируют и отлаживают программные реализации современных методов и алгоритмов обработки экспериментальных

5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)

5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:

5.1.1 Типовые задания (оценочное средство - Задачи) для оценки сформированности компетенции ПК-12:

Поиск и визуализация аномалий в базе данных «ирисы Фишера». Решение задачи «ирисы Фишера» (классификация подвида ириса по параметрам цветка)

5.1.2 Типовые задания (оценочное средство - Задачи) для оценки сформированности компетенции ПК-13:

Решение задачи «Титаник» с использованием логистической регрессии (прогноз выживания пассажира при крушении по полу, возрасту, каюте, классу обслуживания). Обработать пропущенные значения. Сравнить результаты с использованием и без использования кодирования категориальных признаков, нормализации признаков, с учетом обработки и без учета обработки пропущенных данных

Критерии оценивания (оценочное средство - Задачи)

Оценка	Критерии оценивания
зачтено	Все задачи полностью выполнены
не зачтено	Задачи не выполнены

5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено		зачтено				
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи с отдельными и несущественными недочетами, выполнены все задания в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продemonстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой

	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

5.3.1 Типовые задания (оценочное средство - Дискуссия) для оценки сформированности компетенции ПК-12

Пороговый классификатор и деревья решений

5.3.2 Типовые задания (оценочное средство - Дискуссия) для оценки сформированности компетенции ПК-13

Полиномиальная регрессия

Критерии оценивания (оценочное средство - Дискуссия)

Оценка	Критерии оценивания
превосходно	Знания превышают программу. На дополнительные вопросы отвечает, в том числе и за пределами программы курса
отлично	Ответ полный. Грамотно отвечает на большинство дополнительных вопросов. Свободно владеет материалом.
очень хорошо	Ответ полный. Отвечает на большинство дополнительных вопросов.
хорошо	Ответ с незначительными ошибками. Отвечает без наводящих вопросов
удовлетворительно	Не отвечает на половину вопросов билета. Отвечает только с наводящими вопросами
неудовлетворительно	Не отвечает на вопросы билета. Слабо ориентируется в терминологии.

Оценка	Критерии оценивания
плохо	Отвечать затрудняется

6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - Москва : ДМК-пресс, 2017., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=659331&idb=0>.

Дополнительная литература:

1. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / Флах П. - Москва : ДМК-пресс, 2015., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=647134&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

Jupyter Lab <https://jupyter.org/>

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами, специализированным оборудованием: в том числе, – высокотехнологичным оборудованием: серверная вычислительная техника, включая сервера и АРМы Гравитон, серверные шкафы, программно-аппаратные комплексы, сетевое оборудование;

– вычислительными ресурсами: терминал-классы с 26 стационарными и 3 мобильными рабочими местами на базе современных ПК с лицензионным программным обеспечением; – офисным и мультимедийным оборудованием, включая проектор, экран и ТВ-панель, специализированная мебель.

Перечисленное выше оборудование входит в состав Учебно-лабораторного интерактивного комплекса "Распределенные вычисления" для проведения занятий для студентов с использованием современной вычислительной техники при обучении моделированию, проектированию и разработке распределенных вычислительных комплексов и для проведения практических занятий по дисциплинам, предусмотренных программой.

Специальное образовательное пространство (СОП) Учебно-лабораторный интерактивный комплекс «Распределенные вычисления» (уч. корп.3, ауд. 513, 528, 520) создано научно-образовательным отделением космической связи ПИШ ННГУ и утверждено приказом ННГУ №06.49-04-0669/23 от 29.12.2023 г. для реализации образовательных программ (ОП) ПИШ ННГУ, в том числе, ОП «Информационные технологии в системах космической связи и дистанционного зондирования Земли» направления подготовки 09.04.02 «Информационные системы и технологии», разработанной с целью исполнения Программы развития ПИШ ННГУ в рамках федерального проекта Минобрнауки России "Передовые инженерные школы"

государственной программы Российской Федерации "Научно-технологическое развитие Российской Федерации" (<https://analytics.engineers2030.ru/schools/unn>).

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 09.04.02 - Информационные системы и технологии.

Автор(ы): Чуманкин Юрий Евгеньевич.

Заведующий кафедрой: Морозов Олег Александрович, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 15.04.2024, протокол № б/н.