Приложение 4

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Институт информационных технологий, математики и механики

УТВЕРЖДЕНО

решением Ученого совета ННГУ

протокол № 10 от 02.12.2024 г.

Рабочая программа дисциплины

Машинное обучение и анализ данных

Уровень высшего образования
Бакалавриат

Направление подготовки / специальность
02.03.02 - Фундаментальная информатика и информационные технологии

Направленность образовательной программы
Сопряженная разработка программного и аппаратного обеспечения

Форма обучения
очная

г. Нижний Новгород

2025 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.12.01 Машинное обучение и анализ данных 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-2: Способен к применению общенаучных базовых знаний математических и естественных наук, *фундаментальной информатики и информационных технологий; применять в профессиональной деятельности современные языки программирования и методы параллельной обработки данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии*	*ПК-2.1: Знает основы научно-исследовательской деятельности в области информационных технологий, имеет научные знания в теории информационных систем* *ПК-2.2: Умеет применять полученные знания в области фундаментальных научных основ теории информации и решать стандартные задачи в собственной научно-исследовательской деятельности* *ПК-2.3: Имеет практический опыт научно-исследовательской деятельности в области информационных технологий*	*ПК-2.1:* *ЗНАЕТ: основные концептуальные и теоретические модели машинного обучения, основные алгоритмы и методы машинного обучения основы языка Python или среды вычислений R* *ПК-2.2:* *УМЕЕТ: использовать методы машинного обучения на практике, оценивать качество методов работать с библиотекой Scikit-Learn или средой для статистических вычислений R* *ПК-2.3:* *ИМЕЕТ: практический опыт разработки и анализа концептуальных и теоретических моделей машинного обучения*	*Тест*	*Зачёт: Контрольные вопросы*

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	**2**

Часов по учебному плану	**72**
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	**32**
- занятия семинарского типа (практические занятия / лабораторные работы)	**0**
- КСР	**1**
самостоятельная работа	**39**
Промежуточная аттестация	**0** **Зачёт**

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	
	ооо	ооо	ооо	ооо	ооо
Тема 1. Постановка задач машинного обучения	23	10		10	13
Тема 2. Обучение с учителем	23	10		10	13
Тема 3. Обучение без учителя	25	12		12	13
Аттестация	0				
КСР	1			1	
Итого	72	32	0	33	39

Содержание разделов и тем дисциплины

Тема 1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор. Экспериментальные методы оценки качества обучения. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков). Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа. Наивный байесовский классификатор. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия. Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска.

Тема 2. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях. Машина опорных векторов. Ядра и спрямляющие пространства. Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосования. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (случайный лес.).
Тема 3. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. ЕМалгоритм. Иерархическая кластеризация. Агломеративные и разделяющие методы. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска. Размерность Вапника– Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.

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 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 при изучении дисциплины проводится с целью углубления знаний по дисциплине и предусматривает:
- повторение пройденного учебного материала, чтение рекомендованной литературы;
- выполнение общих и индивидуальных домашних заданий;
- работу с электронными источниками;
- подготовку к сдаче формы промежуточной аттестации.

Цель самостоятельной работы - формирование навыков непрерывного самообразования и профессионального совершенствования.
Самостоятельная работа способствует формированию аналитического и творческого мышления, совершенствует способы организации исследовательской деятельности, воспитывает целеустремленность, системность и последовательность в работе студентов, развивает у них навык завершать начатую работу.
Основные виды самостоятельной работы студентов:
- работа с основной и дополнительной литературой;
- изучение категориального аппарата дисциплины;
- самостоятельное изучение тем дисциплины;
- подготовка к зачёту;
- работа в библиотеке;
- изучение сайтов по темам дисциплины в сети Интернет.

Работа с основной и дополнительной литературой
Изучение рекомендованной литературы следует начинать с учебников и учебных пособий,

затем переходить к научным монографиям и материалам периодических изданий. Работа с литературой предусматривает конспектирование наиболее актуальных и познавательных материалов. Это не только мобилизует внимание, но и способствует более глубокому осмыслению материала, его лучшему запоминанию, а также позволяет студентам проводить систематизацию и сравнительный анализ изучаемой информации. Таким образом, конспектирование – одна из основных форм самостоятельного труда, которая требует от студента активно работать с учебной литературой и не ограничиваться конспектом лекций. Студент должен уметь самостоятельно подбирать необходимую литературу для учебной и научной работы, уметь обращаться с предметными каталогами и библиографическим справочником библиотеки.

Изучение категориального аппарата дисциплины

Изучение и осмысление экономических категорий требует проработки лекционного материала, выполнения практических заданий, изучение словарей, энциклопедий, справочников. Индивидуальная самостоятельная работа студента направлена на овладение и грамотное применение экономической терминологии в области компьютерного моделирования.

Самостоятельное изучение тем дисциплины

Особое место отводится самостоятельной проработке студентами отдельных разделов и тем изучаемой дисциплины. Такой подход вырабатывает у студентов инициативу, стремление к увеличению объема знаний, умений и навыков, всестороннего овладения способами и приемами профессиональной деятельности.

Изучение вопросов определенной темы направлено на более глубокое усвоение основных категорий экономической теории, понимание экономических процессов, происходящих в обществе, совершенствование навыка анализа теоретического и эмпирического материала.

Подготовка к зачёту

Промежуточная аттестация студентов по дисциплине проходит в виде зачёта. Условием успешного прохождения промежуточной аттестации является систематическая работа студента в течение семестра. В этом случае подготовка к зачёту является систематизацией всех полученных знаний по данной дисциплине.

Рекомендуется внимательно изучить перечень вопросов к зачёту, а также использовать в процессе обучения программу, учебно-методический комплекс, другие методические материалы.

Желательно спланировать троекратный просмотр материала перед зачётом. Во-первых, внимательное чтение с осмыслением, подчеркиванием и составлением краткого плана ответа. Во-вторых, повторная проработка наиболее сложных вопросов. В-третьих, быстрый просмотр материала или планов ответов для его систематизации в памяти.

Изучение сайтов по темам дисциплины в сети Интернет

Ресурсы Интернет являются одним из альтернативных источников быстрого поиска требуемой информации. Их использование возможно для получения основных и дополнительных сведений по изучаемым материалам. Необходимо помнить об оформлении ссылок на Интернет-источники.

Для повышения эффективности самостоятельной работы студентов преподавателю целесообразно использовать следующие виды деятельности:

- консультации,
- выдача заданий на самостоятельную работу,
- информационное обеспечение обучения,
- контроль качества самостоятельной работы студентов.

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 - Тест) для оценки сформированности компетенции ПК-2:

1. Какая из следующих задач является задачей обучения без учителя?

a. восстановление регрессии

b. классификация

c. кластеризация

2. Какая из следующих задач является задачей обучения с учителем?

a. классификация

b. кластеризация

c. восстановление функции распределения

3. Переобучением называется

a. явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много больше ошибки на обучающей выборке

b. явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много меньше ошибки на обучающей выборке

c. явление, возникающее при решении задач обучения с учителем, при котором решающее правило допускает большую ошибку на тестовой выборке

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	дано 51-100% правильных ответов
не зачтено	дано 0-50% правильных ответов

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**				
Знания	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.
Умения	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продемонстрированы все основные умения. Решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Навыки	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	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован 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 - Контрольные вопросы) для оценки сформированности компетенции ПК-2

1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор.

2. Экспериментальные методы оценки качества обучения.

3. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков).

4. Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.

5. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.

6. Наивный байесовский классификатор.

7. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.

8. Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях.

9. Машина опорных векторов. Ядра и спрямляющие пространства.

10. Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенныхзначений.

11. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосования. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (случайный лес).

12. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Мемедиан. Метод нечетких множеств. EM-алгоритм.

13. Иерархическая кластеризация. Агломеративные и разделяющие методы.

14. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска.

15. Размерность Вапника– Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического рик ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Владение основным и дополнительным материалом достаточное или с незначительными ошибками и погрешностями
не зачтено	Владение материалом, необходимым по данному предмету, недостаточно. Работу за время семестра можно оценить как неудовлетворительную.

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

Основная литература:

1. Котельников Е. В. Введение в машинное обучение и анализ данных : учеб. пособие / Котельников Е. В.,Котельникова А. В. - Киров : ВятГУ, 2023. - 68 с. - Книга из коллекции ВятГУ - Информатика. - Текст : электронный., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=894096&idb=0.
2. Платонов Алексей Владимирович. Машинное обучение : учебное пособие для вузов / А. В.

Платонов. - 2-е изд. - Москва : Юрайт, 2026. - 89 с. - (Высшее образование). - URL: https://urait.ru/bcode/589132 (дата обращения: 24.01.2026). - ISBN 978-5-534-20732-3 : 409.00. - Текст : электронный., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=996216&idb=0.
3. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : монография / Флах П. - Москва : ДМК-пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=878863&idb=0.

Дополнительная литература:

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. - Москва : Юрайт, 2026. - 448 с. - (Высшее образование). - URL: https://urait.ru/bcode/583032 (дата обращения: 24.01.2026). - ISBN 978-5-534-19964-2 : 2159.00. - Текст : электронный., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=994954&idb=0.
2. Баюк О.А. Практикум по анализу данных на языках Python и R : учебное пособие / Баюк О.А., Исаева М.Р., Самсонкин М.О. - Москва : Прометей, 2023. - 100 с. - ISBN 978-5-00172-356-1. - Текст : электронный., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=936025&idb=0.
3. Основы анализа данных и интеллектуальные системы : учебное пособие для вузов / Косников С. Н.,Золкин А. Л.,Ахмадуллин Ф. Р.,Урусова А. Б.,Малова Н. Н.,Поскряков И. А.,Вербицкий Р. А.; Золкин А. Л., Ахмадуллин Ф. Р., Урусова А. Б., Малова Н. Н., Поскряков И. А., Вербицкий Р. А. - Санкт-Петербург : Лань, 2025. - 176 с. - Книга из коллекции Лань - Информатика. - ISBN 978-5-507-50239-4. - Текст : электронный., https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=929007&idb=0.

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

1. The R Project for Statistical Computing https://www.r-project
2. Welcome to Python.org https://www.python.org/
3. scikit-learn: machine learning in Python scikit-learn.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.

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 02.03.02 - Фундаментальная информатика и информационные технологии.

Автор(ы): Золотых Николай Юрьевич, доктор физико-математических наук, доцент.

Заведующий кафедрой: Мееров Иосиф Борисович, кандидат технических наук.

Программа одобрена на заседании методической комиссии от 02.12.2024, протокол № 5.