

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
имени И.Т. ТРУБИЛИНА»

Факультет прикладной информатики
Системного анализа и обработки информации



УТВЕРЖДЕНО

Декан

Замотайлова Д.А.

Протокол от 25.04.2025 № 7

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)
«СОВРЕМЕННЫЕ МЕТОДЫ АНАЛИЗА ДАННЫХ И МАШИННОГО ОБУЧЕНИЯ»

Уровень высшего образования: бакалавриат

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль) подготовки: Управление цифровой трансформацией бизнеса

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Год набора (приема на обучение): 2025

Срок получения образования: 4 года

Объем:
в зачетных единицах: 2 з.е.
в академических часах: 72 ак.ч.

Разработчики:

Доцент, кафедра системного анализа и обработки информации Павлов Д.А.

Рабочая программа дисциплины (модуля) составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 38.03.05 Бизнес-информатика, утвержденного приказом Минобрнауки от 29.07.2020 № 838, с учетом трудовых функций профессиональных стандартов: "Менеджер по информационным технологиям", утвержден приказом Минтруда России от 30.08.2021 № 588н; "Специалист по информационным системам", утвержден приказом Минтруда России от 13.07.2023 № 586н; "Системный аналитик", утвержден приказом Минтруда России от 27.04.2023 № 367н.

Согласование и утверждение

№	Подразделение или коллегиальный орган	Ответственное лицо	ФИО	Виза	Дата, протокол (при наличии)
1	Факультет прикладной информатики	Председатель методической комиссии/совета	Крамаренко Т.А.	Согласовано	21.04.2025, № 8
2		Руководитель образовательной программы	Вострокнутов А.Е.	Согласовано	21.04.2025, № 8

1. Цель и задачи освоения дисциплины (модуля)

Цель освоения дисциплины - развитие навыков анализа данных и применения машинного обучения для извлечения бизнес-выводов и решения практических задач

Задачи изучения дисциплины:

- Дать представление о фундаментальных понятиях, типах задач (обучение с учителем, без учителя) и парадигмах машинного обучения.;
- Сформировать понимание принципов работы, областей применения, достоинств и недостатков ключевых семейств алгоритмов машинного обучения: линейных моделей, методов ближайших соседей, SVM, деревьев решений, ансамблевых методов (бэггинг, бустинг, случайный лес) и др.;
- Раскрыть основные подходы к предобработке данных (кодирование категориальных переменных, обработка пропусков, масштабирование) и работе с признаками (feature engineering).;
- Объяснить критически важные методики оценки качества моделей (кросс-валидация, метрики классификации и регрессии) .

2. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы

Компетенции, индикаторы и результаты обучения

ОПК-5 Способен организовывать взаимодействие с клиентами и партнерами в процессе решения задач управления жизненным циклом информационных систем и информационно-коммуникационных технологий

ОПК-5.1 Использует современные средства электронной коммуникации при взаимодействии с клиентами и партнерами

Знать:

ОПК-5.1/Зн1 Знает современные средства электронной коммуникации при взаимодействии с клиентами и партнерами

Уметь:

ОПК-5.1/Ум1 Умеет использовать знания о современных средствах электронной коммуникации при взаимодействии с клиентами и партнерами

Владеть:

ОПК-5.1/Нв1 Владеет знаниями использования современных средств электронной коммуникации при взаимодействии с клиентами и партнерами

ОПК-5.2 Выявляет основные потребности клиентов и партнеров в области управления ит-сервисами

Знать:

ОПК-5.2/Зн1 Знает методы выявления основных потребностей клиентов и партнеров в области управления ит-сервисами

Уметь:

ОПК-5.2/Ум1 Умеет выявлять основные потребности клиентов и партнеров в области управления ит-сервисами

Владеть:

ОПК-5.2/Нв1 Владеет знаниями определения основных потребностей клиентов и партнеров в области управления ит-сервисами

ОПК-5.3 Демонстрирует умение взаимодействовать с клиентами и партнерами при заключении договоров на предоставление услуг ит-сервисов

Знать:

ОПК-5.3/Зн1 Знает методы взаимодействия с клиентами и партнерами при заключении договоров на предоставление услуг ит-сервисов

Уметь:

ОПК-5.3/Ум1 Умеет применять методы взаимодействия с клиентами и партнерами при заключении договоров на предоставление услуг ит-сервисов

Владеть:

ОПК-5.3/Нв1 Владеет знаниями взаимодействия с клиентами и партнерами при заключении договоров на предоставление услуг ит-сервисов

ОПК-5.4 Разрабатывает индикаторы качества оказания услуг ит-сервисов, проводит их мониторинг и оценку

Знать:

ОПК-5.4/Зн1 Знает методы управления качеством оказания услуг ит-сервисов, проведения их мониторинга и оценки

Уметь:

ОПК-5.4/Ум1 Умеет разрабатывать индикаторы качества оказания услуг ит-сервисов, проводить их мониторинг и оценку

Владеть:

ОПК-5.4/Нв1 Владеет знаниями разработки индикаторов качества оказания услуг ит-сервисов, проведения их мониторинга и оценки

ОПК-5.5 Организует процесс оказания услуг ит-сервиса, проводит эскалацию и закрытие инцидентов

Знать:

ОПК-5.5/Зн1 Знает способы организации процесса оказания услуг ит-сервиса, проведения эскалации и закрытия инцидентов

Уметь:

ОПК-5.5/Ум1 Умеет организовывать процесс оказания услуг ит-сервиса, проводить эскалацию и закрытие инцидентов

Владеть:

ОПК-5.5/Нв1 Владеет знаниями организации процесса оказания услуг ит-сервиса, проведения эскалации и закрытия инцидентов

ПК-П12 Способность использовать знание основных методов искусственного интеллекта в последующей профессиональной деятельности в качестве научных сотрудников, преподавателей образовательных организаций высшего образования, инженеров, технологов

ПК-П12.1 Знает методы разработки оригинальных алгоритмов и программных решений с использованием современных технологий

Знать:

ПК-П12.1/Зн1 Предметная область автоматизации

ПК-П12.1/Зн2 Системы хранения и анализа баз данных

ПК-П12.1/Зн3 Основы программирования

ПК-П12.1/Зн4 Современные объектно-ориентированные языки программирования

ПК-П12.1/Зн5 Современные структурные языки программирования

ПК-П12.1/Зн6 Современные стандарты информационного взаимодействия систем

ПК-П12.1/Зн7 Современный отечественный и зарубежный опыт в профессиональной деятельности

Уметь:

ПК-П12.1/Ум1 Кодировать на языках программирования

ПК-П12.1/Ум2 Тестировать результаты прототипирования

Владеть:

ПК-П12.1/Нв1 Владеть методами кодирования на языках программирования

ПК-П12.2 Умеет использовать методы разработки оригинальных алгоритмов и программных решений с использованием современных технологий

Знать:

ПК-П12.2/Зн1 Предметная область автоматизации

ПК-П12.2/Зн2 Возможности ис

ПК-П12.2/Зн3 Источники информации, необходимой для профессиональной деятельности

Уметь:

ПК-П12.2/Ум1 Анализировать входные данные

ПК-П12.2/Ум2 Планировать работы

Владеть:

ПК-П12.2/Нв1 Владеет знаниями разработки оригинальных алгоритмов и программных решений с использованием современных технологий

ПК-П12.3 Владеть навыками применения методов разработки оригинальных алгоритмов и программных решений с использованием современных технологий

Знать:

ПК-П12.3/Зн1 Языки программирования и работы с базами данных

ПК-П12.3/Зн2 Инструменты и методы верификации структуры программного кода

ПК-П12.3/Зн3 Основы программирования

ПК-П12.3/Зн4 Современные объектно-ориентированные языки программирования

ПК-П12.3/Зн5 Современные структурные языки программирования

ПК-П12.3/Зн6 Современный отечественный и зарубежный опыт в профессиональной деятельности

Уметь:

ПК-П12.3/Ум1 Кодировать на языках программирования

ПК-П12.3/Ум2 Верифицировать структуру программного кода

Владеть:

ПК-П12.3/Нв1 Владеет навыками применения методов разработки оригинальных алгоритмов и программных решений с использованием современных технологий

3. Место дисциплины в структуре ОП

Дисциплина (модуль) «Современные методы анализа данных и машинного обучения» относится к обязательной части образовательной программы и изучается в семестре(ах): 7.

В процессе изучения дисциплины студент готовится к решению типов задач профессиональной деятельности, предусмотренных ФГОС ВО и образовательной программой.

4. Объем дисциплины (модуля) и виды учебной работы

Период	/доимость сы)	/доимость ЭТ)	ая работа всего)	ая контактная (часы)	(часы)	ые занятия сы)	ые занятия сы)	ьяная работа сы)	тая аттестация сы)
--------	------------------	------------------	---------------------	-------------------------	--------	-------------------	-------------------	---------------------	-----------------------

обучения	Общая труд (час)	Общая труд (ЗП)	Контактн (часы,	Внеаудиторная работа	Зачет	Лабораторн (ча	Лекционн (ча	Самостоятел (ча	Промежуточ (ча
Седьмой семестр	72	2	47	1		30	16	25	Зачет
Всего	72	2	47	1		30	16	25	

5. Содержание дисциплины (модуля)

5.1. Разделы, темы дисциплины и виды занятий

(часы промежуточной аттестации не указываются)

Наименование раздела, темы	Всего	Внеаудиторная контактная работа	Лабораторные занятия	Лекционные занятия	Самостоятельная работа	Планируемые результаты обучения, соответствующие результатам освоения программы
Раздел 1. Введение в машинное обучение	14		6	4	4	ОПК-5.1 ОПК-5.2
Тема 1.1. Основные понятия и задачи в машинном обучении	6		2	2	2	
Тема 1.2. Работа с данными и признаками	8		4	2	2	
Раздел 2. Линейные модели	26		12	6	8	ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-5.4 ОПК-5.5
Тема 2.1. Линейная регрессия	14		6	4	4	
Тема 2.2. Линейная классификация	12		6	2	4	ПК-П12.1 ПК-П12.2 ПК-П12.3
Раздел 3. Ансамблевые методы	22	1	8	4	9	ПК-П12.1
Тема 3.1. Решающие деревья	10		4	2	4	ПК-П12.2
Тема 3.2. Бэггинг, случайные леса и разложение ошибки на смещение и разброс	12	1	4	2	5	ПК-П12.3
Раздел 4. Снижение размерности	10		4	2	4	ПК-П12.1
Тема 4.1. Снижение размерности	10		4	2	4	ПК-П12.2 ПК-П12.3
Итого	72	1	30	16	25	

5.2. Содержание разделов, тем дисциплин

Раздел 1. Введение в машинное обучение

(Лабораторные занятия - 6ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 4ч.)

Тема 1.1. Основные понятия и задачи в машинном обучении

(Лабораторные занятия - 2ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 2ч.)

Введение в машинное обучение.

Виды задач и методов машинного обучения.

Основные термины, постановки задач и примеры применения.

Тема 1.2. Работа с данными и признаками

(Лабораторные занятия - 4ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 2ч.)

Работа с данными и признаками.

Метод k ближайших соседей.

Раздел 2. Линейные модели

(Лабораторные занятия - 12ч.; Лекционные занятия - 6ч.; Самостоятельная работа - 8ч.)

Тема 2.1. Линейная регрессия

(Лабораторные занятия - 6ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 4ч.)

Область применимости линейных моделей

Измерение ошибки в задачах регрессии

Переобучение

Оценивание качества моделей

Обучение линейной регрессии

Градиентный спуск и оценивание градиента

Модификация градиентного спуска

Регуляризация

Преобразования признаков

Тема 2.2. Линейная классификация

(Лабораторные занятия - 6ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 4ч.)

Линейные модели классификации

Метрики качества классификации

Логистическая регрессия

Метод опорных векторов

Многоклассовая классификация

Раздел 3. Ансамблевые методы

(Внеаудиторная контактная работа - 1ч.; Лабораторные занятия - 8ч.; Лекционные занятия - 4ч.; Самостоятельная работа - 9ч.)

Тема 3.1. Решающие деревья

(Лабораторные занятия - 4ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 4ч.)

Построение деревьев

Критерии информативности

Критерии останова

Методы стрижки деревьев

Тема 3.2. Бэггинг, случайные леса и разложение ошибки

на смещение и разброс

(Внеаудиторная контактная работа - 1ч.; Лабораторные занятия - 4ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 5ч.)

Бутстрап
Смещение и разброс
Бэггинг
Случайные леса
Бустинг
Современные методы градиентного бустинга
Стекинг

Раздел 4. Снижение размерности

(Лабораторные занятия - 4ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 4ч.)

Тема 4.1. Снижение размерности

(Лабораторные занятия - 4ч.; Лекционные занятия - 2ч.; Самостоятельная работа - 4ч.)

PCA
tSNE
Кластеризация

6. Оценочные материалы текущего контроля

Раздел 1. Введение в машинное обучение

Форма контроля/оценочное средство: Расчетно-графическая работа

Вопросы/Задания:

1. Выполнить задания из интерактивного блокнота

Задание состоит из двух разделов, посвященных работе с табличными данными с помощью библиотеки `pandas` и визуализации с помощью `matplotlib`. В первом разделе вам предстоит выполнить базовые задания с помощью вышеуказанных библиотек, а во втором распределить студентов по курсам. Баллы даются за выполнение отдельных пунктов. Задачи в рамках одного раздела рекомендуется решать в том порядке, в котором они даны в задании.

Задание направлено на освоение `jupyter notebook` (будет использоваться в дальнейших заданиях), библиотек `pandas` и `matplotlib`.

Раздел 2. Линейные модели

Форма контроля/оценочное средство: Кейс-задание

Вопросы/Задания:

1. Используя интерактивный блокнот выполнить задания к теме

Задание 1. Мы будем работать с данными из соревнования New York City Taxi Trip Duration, в котором нужно было предсказать длительность поездки на такси. Скачайте обучающую выборку из этого соревнования и загрузите ее.

Задание 2. Для начала давайте посмотрим, сколько всего было поездок в каждый из дней. Постройте график зависимости количества поездок от дня в году (например, можно воспользоваться `sns.countplot`).

Задание 3. Нарисуйте на одном графике зависимости количества поездок от часа в сутках для разных месяцев (разные кривые, соответствующие разным месяцам, окрашивайте в разные цвета, воспользуйтесь `hue` в `sns.relplot`). Аналогично нарисуйте зависимости количества поездок от часа в сутках для разных дней недели.

Задание 4. Разбейте выборку на обучающую и тестовую в отношении 7:3 (используйте `train_test_split` из `sklearn`). По обучающей выборке нарисуйте график зависимости среднего логарифма времени поездки от дня недели. Затем сделайте то же самое, но для часа в сутках и дня в году.

Задание 5. Обучите Ridge-регрессию с параметрами по умолчанию, закодировав все

категориальные признаки с помощью OneHotEncoder. Численные признаки отмасштабируйте с помощью StandardScaler. Используйте только признаки, которые мы выделили в этой части задания.

Форма контроля/оценочное средство: Расчетно-графическая работа

Вопросы/Задания:

1. В данном задании необходимо реализовать обучение линейной регрессии с помощью различных вариантов градиентного спуска.

Задание 1. Реализация градиентного спуска

В этом задании вам предстоит написать собственные реализации различных подходов к градиентному спуску с опорой на подготовленные шаблоны в файле descents.py.

Все реализуемые методы должны быть векторизованы!

Задание 2. Реализация линейной регрессии

В этом задании вам предстоит написать свою реализацию линейной регрессии, обучаемой с использованием градиентного спуска, с опорой на подготовленные шаблоны в файле linear_regression.py - LinearRegression. По сути линейная регрессия будет оберткой, которая запускает обучение

2. Работа с интерактивным блокнотом

В этом задании вы:

ознакомитесь с тем, что происходит "внутри" метода опорных векторов и логистической регрессии

познакомитесь с калибровкой вероятности

изучите методы трансформации переменных и методы отбора признаков

попытаетесь оценить экономический эффект модели

Скачайте интерактивный блокнот с заданиями. В пустые ячейки впишите свой программный код

Для сдачи задания переименуйте получившийся файл *.ipynb в соответствии со следующим форматом: homework-practice-04-linclass-Username.ipynb, где Username — ваша фамилия и имя на латинице именно в таком порядке (например, homework-practice-04-linclass-IvanovIvan.ipynb).

Раздел 3. Ансамблевые методы

Форма контроля/оценочное средство: Расчетно-графическая работа

Вопросы/Задания:

1. Необходимо воспользоваться возможностями бутстрапа для оценки смещения и разброса алгоритмов машинного обучения.

В этом задании вам предстоит воспользоваться возможностями бутстрапа для оценки смещения и разброса алгоритмов машинного обучения.

Раздел 4. Снижение размерности

Форма контроля/оценочное средство: Задача

Вопросы/Задания:

1. В задании необходимо решить предложенные задачи по программированию.

В задании необходимо решить предложенные задачи по программированию – вписать свой код в ячейки после условий задач вместо комментария `### YOUR CODE HERE ###` и сохранить изменения. Для решения использовать приложенный файл.

7. Оценочные материалы промежуточной аттестации

Седьмой семестр, Зачет

Контролируемые ИДК: ОПК-5.1 ОПК-5.2 ОПК-5.3 ОПК-5.4 ОПК-5.5 ПК-П12.1 ПК-П12.2 ПК-П12.3

Вопросы/Задания:

1. Основные этапы решения задач машинного обучения
Основные этапы решения задач машинного обучения
2. Бинаризация числовых признаков
Бинаризация числовых признаков
3. Измерение ошибки в задачах регрессии
Измерение ошибки в задачах регрессии
4. Переобучение линейной регрессии
Переобучение линейной регрессии
5. Оценивание качества линейной регрессии
Оценивание качества линейной регрессии
6. Обучение линейной регрессии
Обучение линейной регрессии
7. Градиентный спуск и оценивание градиента
Градиентный спуск и оценивание градиента
8. Стохастический градиентный спуск
Стохастический градиентный спуск
9. Модификации градиентного спуска
Модификации градиентного спуска
10. Регуляризация
Регуляризация
11. Преобразование признаков
Преобразование признаков
12. Линейные модели классификации
Линейные модели классификации
13. Метрики качества классифик
Метрики качества классифик

8. Материально-техническое и учебно-методическое обеспечение дисциплины

8.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Сапрыкин О. Н. Интеллектуальный анализ данных / Сапрыкин О. Н.. - Самара: Самарский университет, 2020. - 80 с. - 978-5-7883-1563-8. - Текст: электронный. // RuSpLAN: [сайт]. - URL: <https://e.lanbook.com/img/cover/book/188906.jpg> (дата обращения: 19.06.2025). - Режим доступа: по подписке

Дополнительная литература

1. ЛУЦЕНКО Е. В. Инженерия знаний и интеллектуальные системы: метод. рекомендации / ЛУЦЕНКО Е. В., Аршинов Г. А.. - Краснодар: КубГАУ, 2020. - 109 с - Текст: электронный. // : [сайт]. - URL: <https://edu.kubsau.ru/mod/resource/view.php?id=7101> (дата обращения: 08.09.2025). - Режим доступа: по подписке

8.2. Профессиональные базы данных и ресурсы «Интернет», к которым обеспечивается доступ обучающихся

Профессиональные базы данных

Не используются.

Ресурсы «Интернет»

1. <http://www.iprbookshop.ru/> - IPRbook
2. <https://edu.kubsau.ru/> - Образовательный портал КубГАУ
3. <https://znanium.com/> - Znanium.com

8.3. Программное обеспечение и информационно-справочные системы, используемые при осуществлении образовательного процесса по дисциплине

Перечень программного обеспечения

(обновление производится по мере появления новых версий программы)

Не используется.

Перечень информационно-справочных систем

(обновление выполняется еженедельно)

Не используется.

8.4. Специальные помещения, лаборатории и лабораторное оборудование

9. Методические указания по освоению дисциплины (модуля)

10. Методические рекомендации по освоению дисциплины (модуля)