

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего
образования
«Кубанский государственный университет»

Факультет компьютерных технологий и прикладной математики
Кафедра вычислительных технологий

УТВЕРЖДАЮ

Проректор по учебной работе,
качеству образования – первый
проректор

Хагуров Т.А.

подпись

«__30__» __мая__ 2025г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
К.М.01.05 «МАШИННОЕ ОБУЧЕНИЕ»

Направление

подготовки/специальность 02.03.02 **Фундаментальная информатика и**
информационные технологии

(код и наименование направления подготовки/специальности)

Направленность (профиль) /специализация

Математическое и программное обеспечение компьютерных технологий

Программа подготовки академический бакалавриат

Форма обучения очная

Квалификация выпускника бакалавр

Краснодар 2025

Рабочая программа дисциплины «Машинное обучение» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.03.02 Фундаментальная информатика и информационные технологии

Программу составил(а):

Приходько Татьяна Александровна, доцент, к. т. н.

Ф.И.О. , должность, ученая степень, ученое звание

подпись

Рабочая программа дисциплины. «Машинное обучение» утверждена на заседании кафедры Вычислительных технологий протокол № 7 «07 » мая 2025 г.

И.о. заведующего кафедрой (разработчика) Еремин.А.А.

(фамилия, инициалы)

подпись

Утверждена на заседании учебно-методической комиссии факультета Компьютерных Технологий и Прикладной Математики протокол № 4 от «23» мая 2024 г

Председатель УМК факультета

Коваленко А.В.

фамилия, инициалы

подпись

Рецензенты:

Гаркуша О.В., доцент кафедры информационных технологий ФБГОУ ВО «Кубанский государственный университет», кандидат физико-математических наук.

Схаляхо Ч.А., доцент КВВУ им.С.М.Штеменко, к.ф.-м.н., доцент

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

1.1 Цель освоения дисциплины

Курс «Машинное обучение» имеет своей целью: формирование у студентов практических навыков работы с данными и решения прикладных задач анализа данных. Достижение планируемых результатов обучения (таблица 2), соотнесенных с индикаторами достижения компетенций и целью реализации ОПОП. В результате изучения дисциплины (модуля) «Машинное обучение» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне.

1.2 Задачи дисциплины

Основные задачи освоения дисциплины:

- Студент должен **знать**
- основные понятия и постановки задач машинного обучения
- теоретические основы линейных и логических методов машинного обучения
- принципы построения композиций моделей
- основные метрики качества для регрессии и классификации

уметь

- Умеет выполнять полный цикл построения модели
- Умеет обучать основные модели машинного обучения, оценивать их качества
- Умеет выполнять кластеризацию и визуализацию данных

владеть математическими методами анализа данных, языками и компьютерными методами машинного обучения.

1.3 Место дисциплины (модуля) в структуре образовательной программы

Курс «Машинное обучение» относится к части блока КМ.01 - Системы искусственного интеллекта основных дисциплин учебного плана.

Для изучения дисциплины студент должен владеть знаниями, умениями и навыками по дисциплинам:

Дискретная математика, Алгебраические структуры, Основы программирования, Алгоритмы вычислительной математики, Конструирование алгоритмов и структур данных, Теория алгоритмов и вычислительных процессов, Основы теории вероятностей и статистических методов, Алгоритмы и структуры данных, Математическая логика и теория алгоритмов, Интеллектуальный анализ данных.

Знания, получаемые при изучении дисциплины «Машинное обучение» используются при изучении профессиональных дисциплин Распределенные задачи и алгоритмы, Программирование в компьютерных сетях, Облачные вычисления, Мультиагентные системы, а также для работ над дипломной и магистерской работой.

1.4 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы.

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих **профессиональных компетенций и соотнесенных с ними индикаторов достижения компетенций**: УК-1; УК-2; ОПК-3; ОПК-4

Код и наименование индикатора*	Результаты обучения по дисциплине (знает, умеет, владеет (навыки и/или опыт деятельности))
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	
УК-1.1. Знает принципы сбора, отбора и обобщения информации.	Знает принципы сбора, отбора и обобщения информации.
УК-1.2. Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности.	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности.
УК-1.3. Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт научного поиска, опыт библиографического разыскания, создания научных текстов.	Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт научного поиска, опыт библиографического разыскания, создания научных текстов.
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	
УК-2.1. Знает необходимые для осуществления профессиональной деятельности правовые нормы.	Знает необходимые для осуществления профессиональной деятельности правовые нормы.
УК-2.2. Умеет определять круг задач в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов; соотносить главное и второстепенное, решать поставленные задачи в рамках избранных видов профессиональной деятельности.	Умеет определять круг задач в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов; соотносить главное и второстепенное, решать поставленные задачи в рамках избранных видов профессиональной деятельности.
УК-2.3. Имеет практический опыт применения нормативной базы и решения задач в области избранных видов профессиональной деятельности.	Имеет практический опыт применения нормативной базы и решения задач в области избранных видов профессиональной деятельности.
ОПК-3. Способен к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям	
ОПК-3.1. Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей;	Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей;
ОПК-3.2. Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем.	Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем.

Код и наименование индикатора*	Результаты обучения по дисциплине (знает, умеет, владеет (навыки и/или опыт деятельности))
ОПК-3.3. Имеет практический опыт применения разработки программного обеспечения.	Имеет практический опыт разработки программного обеспечения с использованием методов машинного обучения
ОПК-4. Способен участвовать в разработке технической документации программных продуктов и комплексов с использованием стандартов, норм и правил, а также в управлении проектами создания информационных систем на стадиях жизненного цикла	
ОПК-4.1. Знает стандарты разработки технической документации, умеет применить их на практике при разработке ПО;	Знает стандарты разработки технической документации, умеет применить их на практике при разработке ПО;
ОПК-4.2. Знает принципы сбора и анализа информации, создания информационных систем на стадиях жизненного цикла.	Знает принципы сбора и анализа информации, создания информационных систем на стадиях жизненного цикла.
ОПК-4.3. Умеет осуществлять управление проектами информационных систем.	Умеет осуществлять управление проектами информационных систем.
ОПК-4.4. Имеет практический опыт анализа и интерпретации информационных систем.	Имеет практический опыт анализа и интерпретации информационных систем.

2. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 4 зач.ед. (144 часов), их распределение по видам работ представлено в таблице (для студентов ОФО)

Вид учебной работы	Всего часов	Семестры (часы)			
		6			
Контактная работа в том числе:	144	144			
Аудиторные занятия (всего):	78,3	78,3			
В том числе:					
Занятия лекционного типа	32	32			
Занятия семинарского типа (семинары, практ. занятия)					
Лабораторные занятия	32	32			
Иная контрольная работа					
Контроль самостоятельной работы	4	4			
Промежуточная аттестация (ИКР)	0,3	0,3			
Самостоятельная работа, в том числе	40	40			
В том числе:					
Курсовое проектирование	10	10			
Проработка учебного (теоретического) материала	20	20			
Выполнение индивидуальных заданий (подготовка сообщений, презентаций)	16	16			
Реферат					
Подготовка к текущему контролю	4	4			
Контроль:					
Подготовка к экзамену:	35,7	35,7			
Общая трудоемкость час	144	144			

в т.ч. контактная работа	68,3	68,3			
зач. ед.	4	4			

2.2 Структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины. Разделы дисциплины, изучаемые в _6_ семестре (очная форма)

№	Наименование разделов	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	КСР	ЛР	СРС
1.	Введение в машинное обучение.	16	4	2	4	6
2.	Регрессия	12	4		4	4
3.	Классификация Линейные методы. Обработка текстов: bag of words, tf-idf, векторные эмбединги.	26	8		8	10
4.	Прогнозирование на основе рядов	28	8	2	8	10
5.	Отладка моделей машинного обучения	10	4		4	2
6.	Разработка МО проектов	12	4		4	8
	<i>Итого по разделам дисциплины:</i>	108	32	4	32	40
	ИКР	0,3				
	<i>Итого:</i>	108,3				
	<i>Контроль</i>	44,7				
	<i>Итого по дисциплине:</i>	144				

Примечание: Л – лекции, КСР – контрольные и самостоятельные работы, ЛР – лабораторные занятия, СРС – самостоятельная работа студента, Д-доклад, РГЗ – расчетно-графическое задание.

2.3 Содержание разделов дисциплины:

2.3.1 Занятия лекционного типа

№ раздела	Наименование раздела	Содержание раздела	Форма текущего контроля	Разработ. с участием представителей работодателей
1	2	3	4	5
1	Введение в машинное обучение	▪ Обзор прикладных задач машинного обучения. Библиотеки языка Python для решения задач машинного обучения.	ЛР	

2	Регрессия	Градиентный спуск, методы оценивания градиента. Функции потерь. Метрики качества регрессии. Линейная регрессия, метод наименьших квадратов и максимизация правдоподобия. Теорема Гаусса—Маркова. Явный вид решения в методе наименьших квадратов. Ковариационная матрица для коэффициентов. Практические соображения: что делать с категориальными данными? Вычислительные соображения: точное решение vs градиентный спуск. Регуляризация.	ЛР	
3	Классификация	Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия, кросс-энтропия. Ограничения линейных методов (пример: XOR). Решающие деревья. CART. Ансамбли. Бутстреп. Бэггинг. Случайный лес. Классификация текстов. Пропущенные значения. Обработка текстов: bag of words, tf-idf, векторные эмбединги. Метрики оценки классификации: полнота, точность, F1, ROC, AUC. Валидационная и тестовая выборка. Кросс-валидация. Работа с категориальными признаками. Регрессия. Метрики оценки регрессии: MSE, MAE, R^2 – коэффициент детерминации. Линейная регрессия, полиномиальная регрессия. Переобучение и регуляризация, гребневая регрессия, LASSO, Elastic Net. Линейные модели для классификации. Логистическая регрессия, Регуляризация линейных моделей классификации. Кластеризация. k-means, k-means++, DBSCAN, агломеративная кластеризация. Метрики оценки кластеризации. Метод опорных векторов. Прямая и обратная задача. Определение опорных векторов. Ядерный трюк. Наивный байесовский классификатор. Методы оценки распределения признаков. ЕМ-алгоритм на примере смеси гауссиан.	ЛР РГЗ	
4	Прогнозирование на основе рядов	Временной ряд основные понятия. Разложение временных рядов. Преобразования и корректировки. Компоненты временных рядов. Скользящие средине. Классическое разложение. Признаки временных рядов.		

5	Отладка моделей машинного обучения	Метрики оценки качества на примерах. Предобработка данных. Построение обучающих кривых, переобучение Выбор метрик при обучении моделей	ЛР	
6	Разработка МО проектов	Объединение алгоритмов в цепочки и конвейеры, настройка оптимальных параметров. Построение пайплайнов. Оценка осуществимости модели. Производительность модели, оценка масштаба.	проект	

2.3.2. Занятия семинарского типа

Занятия семинарского типа – не предусмотрены.

2.3.3. Лабораторные занятия

№ работы	№ раздела дисциплины	Наименование лабораторных работ
1	1	Ознакомление с синтаксисом языка R для анализа данных.(4ч)
2	1	Способы подготовки и отображения данных в R (4ч). Возможности ввода/вывода.
4	2	Способы анализа данных в R. Получение первичных элементарных характеристик о наборах данных (элементарные статистики). Способы импорта/экспорта данных(2ч).
5	2	Работа с диаграммами и графиками в R (2ч).
6	3	Проверка статистических гипотез (4ч)
7	3	Решение задач Data Mining. Задачи классификации, кластеризации: деревья решений, RandomForest, k-means. (4ч)
8	3	Корреляционный анализ и регрессионный анализ данных (2ч)
9	4	Решение задач на больших графах (2ч).
10	5	Изучение принципов работы распределенных баз данных
11	1-5	Круглый стол: Совместное обсуждение результатов РГЗ
12	1-5	Обсуждение итогов курса

2.3.3 Примерная тематика курсовых работ (проектов)

Согласно учебному плану по данной дисциплине ведутся курсовые работы. Примерная тематика работ:

Укрупненная тематика:

1. Компьютерные сети и программирование в КС
2. Распределенные алгоритмы
3. Безопасные и надежные вычисления
4. Алгоритмы интеллектуальной обработки данных

Детализированная тематика:

1. Компьютерные сети, программирование в КС
 - 1.1. Программные приложения для сети Интернет

- 1.2. Поиск информации в сетях и ее анализ
2. Распределенные алгоритмы
 - 2.1. Системные распределенные алгоритмы
 - 2.2. Прикладные распределенные алгоритмы
 - 2.3. Мультиагентные системы
 - 2.4. Распределенные хранилища данных
 - 2.5. Распределенный сбор и анализ данных
3. Безопасные и надежные вычисления
 - 3.1. Криптографические преобразования и протоколы
 - 3.2. Надежность вычислений в стационарных и мобильных сетях
 - 3.3. Политики безопасности и прикладные логики
4. Алгоритмы и приложения, включающие интеллектуальную обработку и анализ данных
 - 4.1. Лингвистическая обработка данных.
 - 4.2. Алгоритмы компьютерного зрения и распознавания образов.
 - 4.3. Нейросетевые и нечеткие алгоритмы.

2.3.4 Командный проект

В процессе изучения дисциплины "Машинное обучение" студентами выполняется один командный проект, включающий ролевое выполнение задач. Темы заданий для каждого проекта различны. Задача проекта состоит в проверке умений студентов и проверке эффективности их работы в команде по сбору и анализа данных, построения пайпланов решения задач, оценке эффективности работы ML - проекта.

Темы заданий ежегодно обновляются.

Отчет по выполнению командного проекта должен содержать:

2. постановку задачи;
3. сформированный набор данных;
4. подбор методов машинного обучения, построение конвейера и пайплайна решения
5. тексты скриптов на языке R или Python;
6. ясное и подробное пояснение каждого результата, словесную трактовку графиков;
7. оценка качество модели, выводы;
8. список использованной литературы.

2.4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Введение в машинное обучение	Лекции по дисциплине. Источники основной и дополнительной литературы
2	Регрессия	Лекции по дисциплине. Источники основной и дополнительной литературы
3	Классификация	Лекции по дисциплине. Источники основной и дополнительной литературы

4	Прогнозирование на основе рядов	Лекции по дисциплине. Источники основной и дополнительной литературы
5	Отладка моделей машинного обучения	Лекции по дисциплине. Источники основной и дополнительной литературы
	Разработка МО проектов	Лекции по дисциплине. Источники основной и дополнительной литературы

Учебно-методические материалы для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ) предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом, в форме электронного документа,

Для лиц с нарушениями слуха:

– в печатной форме, в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме, в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

3. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Семестр	Вид занятия (Л, ПР, ЛР)	Используемые интерактивные образовательные технологии	Количество часов
6	Л	Компьютерные презентации и обсуждение	32
	ЛР	Разбор конкретных ситуаций (задач), тренинги по решению задач, компьютерные симуляции (программирование алгоритмов), подготовка и обсуждение докладов.	32
	КРС	Контрольная работа	4
Итого:			68

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

4. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

4.1 Фонд оценочных средств для проведения текущего контроля

№ п/п	Код и наименование индикатора	Результаты обучения	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	УК-1.1. Знает принципы сбора, отбора и обобщения информации.	Знает принципы сбора, отбора и обобщения информации.	Опрос по теме лабораторных работ.	Вопросы 1-34, выносимые на экзамен

2	УК-1.2. Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности.	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности.	Опрос по теме лабораторных работ.	Вопросы 1-34, выносимые на экзамен
3	УК-1.3. Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт научного поиска, опыт библиографического разыскания, создания научных текстов.	Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт научного поиска, опыт библиографического разыскания, создания научных текстов.	Опрос по теме лабораторных работ.	Вопросы 1-34, выносимые на экзамен
4	УК-2.1. Знает необходимые для осуществления профессиональной деятельности правовые нормы.	Знает необходимые для осуществления профессиональной деятельности правовые нормы.	Опрос по теме лабораторных работ.	Вопросы 1-34, выносимые на экзамен
5	УК-2.2. Умеет определять круг задач в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов; соотносить главное и второстепенное, решать поставленные задачи в рамках избранных видов профессиональной деятельности.	Умеет определять круг задач в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов; соотносить главное и второстепенное, решать поставленные задачи в рамках избранных видов профессиональной деятельности.	Опрос по теме лабораторных работ.	Вопросы 1-34, выносимые на экзамен
6	УК-2.3. Имеет практический опыт применения нормативной базы и решения задач в области избранных видов профессиональной деятельности.	Имеет практический опыт применения нормативной базы и решения задач в области избранных видов профессиональной деятельности.	Опрос по теме лабораторных работ.	Вопросы 20-34, выносимые на экзамен
7	ОПК-3.1. Знает методы теории алгоритмов, методы системного и	Знает методы теории алгоритмов, методы системного и прикладного	Опрос по теме	Вопросы 27-34, выносимые на экзамен

	прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей;	программирования, основные положения и концепции в области математических, информационных и имитационных моделей;	лабораторных работ.	
8	ОПК-3.2. Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем.	Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем.	Опрос по теме лабораторных работ.	Вопросы 27-51, выносимые на экзамен
9	ОПК-3.3. Имеет практический опыт применения разработки программного обеспечения.	Имеет практический опыт применения разработки программного обеспечения.	Опрос по теме лабораторных работ.	Вопросы 27-51, выносимые на экзамен
10	ОПК-4.1. Знает стандарты разработки технической документации, умеет применить их на практике при разработке ПО;	Знает стандарты разработки технической документации, умеет применить их на практике при разработке ПО;	Опрос по теме лабораторных работ.	Вопросы 27-51, выносимые на экзамен
11	ОПК-4.2. Знает принципы сбора и анализа информации, создания информационных систем на стадиях жизненного цикла.	Знает принципы сбора и анализа информации, создания информационных систем на стадиях жизненного цикла.	Опрос по теме лабораторных работ.	Вопросы 27-51, выносимые на экзамен
12	ОПК-4.3. Умеет осуществлять управление проектами информационных систем.	Умеет осуществлять управление проектами информационных систем.	Опрос по теме лабораторных работ.	Вопросы 27-51, выносимые на экзамен
13	ОПК-4.4. Имеет практический опыт анализа и интерпретации информационных систем.	Имеет практический опыт анализа и интерпретации информационных систем.		Вопросы 27-51, выносимые на экзамен

Фонд оценочных средств дисциплины состоит из средств текущего контроля (вопросы при защите ЛР, контрольной работы) лабораторных работ, средств итоговой аттестации (экзамен в 7 семестре).

Оценка успеваемости осуществляется по результатам:

- выполнения лабораторных работ;
- выполнения лабораторных работ;
- ответа на экзамене (для выявления знания и понимания теоретического материала дисциплины).

Текущий контроль включает контрольную работу по итогам первой половины курса.

Пример экзаменационного билета:

1. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
2. Свойства описательных статистик (Дескриптивный анализ). Меры изменчивости.
3. Что такое наивный байесовский алгоритм? Какую задачу обработки данных он выполняет? Приведите и поясните формулу теорема Байеса.
4. Индивидуальное задание.

Пример задания к билету:

Для чтения файла воспользуйтесь командой

```
chem <- read.csv(file = file.choose(), header = TRUE, sep = ",")
```

Дан датасет: **SmokeBan.csv**

Сокращают Ли Запреты На Курение На Рабочем Месте Курение?

Описание

Оценка влияния запрета на курение на рабочем месте на курение работников, работающих в помещениях.

Формат

Фрейм данных, содержащий 500 наблюдений по 7 переменным.

курильщик фактор. Является ли данный человек в настоящее время курильщиком?

бан фактор. Существует ли запрет на курение в рабочей зоне?

Возраст- возраст в годах.

образование, указывающий на наивысший достигнутый уровень образования: окончание средней школы (hs), выпускник средней школы, какой-либо колледж, выпускник колледжа, степень магистра (или выше).

afam фактор. Является ли индивид афроамериканцем?

hispanic (латиноамериканец) фактор. Является ли индивид латиноамериканцем?

пол фактор, указывающий на пол.

Подробности SmokeBank - это набор данных с наблюдениями за 500 работников, работающих в помещении, который является подмножеством набора данных, собранных в рамках Национального опроса по вопросам здравоохранения в 1991 году, а затем снова (с разными респондентами) в 1993 году. Набор данных содержит информацию о том, подпадали или не подпадали под действие запрета на курение на рабочем месте.

1. Найти средний возраст курильщиков (не используя цикл).
2. Подсчитать число курильщиков афроамериканцев моложе 35 лет. И отдельно курильщиков латиноамериканцев старше 30 лет.
3. Проверить влияние образования и то, является ли человек в настоящее время курильщиком, а также влияние пола на курение.

4. Построить боксплоты, сопоставляющие сообщества курильщиков афроамериканцев, латиноамериканцев и всех остальных.
 5. Проверить гипотезу о среднем возрасте курильщиков всех трех групп.
 6. Построить гистограммы по трем группам людей.
 7. Полученные результаты показать и прокомментировать устно.
- *) Все графики должны сопровождаться заголовками и подписями по осям.*

Перечень вопросов, для подготовки к экзамену

1. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
2. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
3. Метрики качества алгоритм регрессии и классификации.
4. Оценивание качества алгоритмов. Отложенная выборка, ее недостатки. Оценка полного скользящего контроля. Кросс-валидация. Leave-one-out.
5. Деревья решений. Методы построения деревьев. Их регуляризация.
6. Композиции алгоритмов. Разложение ошибки на смещение и разброс.
7. Случайный лес, его особенности.
8. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.
9. Кластеризация. Алгоритм K-Means.
10. Типы задач машинного обучения
11. Типы данных
12. Градиентный спуск
13. Перекрестная проверка. Оптимизация гиперпараметров моделей
14. Предварительная обработка данных
15. Инженерия Признаков
16. Pipeline в Машинном Обучении
17. Кластеризация k-средних
18. Иерархическая кластеризация
19. DBSCAN
20. Метод главных компонент
21. Сингулярное Разложение Матриц
22. t-SNE
23. Neighborhood Components Analysis
24. Линейная регрессия
25. Регуляризация линейной регрессии
26. Метод k-ближайших соседей
27. Метод опорных векторов
28. AdaBoost
29. Gradient Boosting
30. Логистическая регрессия
31. Random Forest
32. Наивный байесовский классификатор
33. Дискриминантный анализ
34. LMS-платформа - не предусмотрена

Критерии оценивания к экзамену:

- 84-100 баллов (оценка «отлично») - изложенный материал фактически верен, наличие глубоких исчерпывающих знаний в объеме пройденной программы дисциплины в соответствии с поставленными программой курса целями и задачами обучения; правильные, уверенные действия по применению полученных знаний на практике, грамотное и логически стройное изложение материала при ответе, усвоение основной и знакомство с дополнительной литературой; Практические задания выполнены в срок в полном объеме.

- 67-83 баллов (оценка «хорошо») - наличие твердых и достаточно полных знаний в объеме пройденной программы дисциплины в соответствии с целями обучения, правильные действия по применению знаний на практике, четкое изложение материала, допускаются отдельные логические и стилистические погрешности. Практические задания выполнены в срок в объеме не менее 80%.

- 50-66 баллов (оценка удовлетворительно) - наличие твердых знаний в объеме пройденного курса в соответствии с целями обучения, изложение ответов с отдельными ошибками, уверенно исправленными после дополнительных вопросов; правильные в целом действия по применению знаний на практике; Практические задания выполнены в объеме не менее 60%.

- 0-49 баллов (оценка неудовлетворительно) - ответы не связаны с вопросами, наличие грубых ошибок в ответе, непонимание сущности излагаемого вопроса, неумение применять знания на практике, неуверенность и неточность ответов на дополнительные и наводящие вопросы». Практические задания не выполнены либо предоставлены не в срок в объеме менее 50%.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на экзамене;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в печатной форме увеличенным шрифтом, в форме электронного документа.

Для лиц с нарушениями слуха:

– в печатной форме, в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

– в печатной форме, в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

5. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

5.1 Основная литература:

1. Машинное обучение : учебник : [16+] / Е. Ю. Бутырский, В. В. Цехановский, Н. А. Жукова [и др.]. – Москва : Директ-Медиа, 2023. – 368 с. : ил., табл., схем., граф. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=701807> (дата обращения: 07.06.2025). – Библиогр. в кн. – ISBN 978-5-4499-3778-0. – DOI 10.23681/701807. – Текст : электронный.
2. Целых, А. Н. Выявление инцидентов информационной безопасности и мошеннических транзакций методами машинного обучения : учебное пособие : [16+] / А. Н. Целых, Э. М. Котов ; Южный федеральный университет, Южный федеральный университет, Инженерно-технологическая академия. – Ростов-на-Дону : Южный федеральный университет, 2023. – 118 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=713464> (дата обращения: 07.06.2025). – Библиогр. в кн. – ISBN 978-5-9275-4515-5. – Текст : электронный.
3. Лимановская, О. В. Основы машинного обучения : учебное пособие / О. В. Лимановская, Т. И. Алферьева ; науч. ред. И. . Обабков ; Уральский федеральный университет им. первого Президента России Б. Н. Ельцина. – Екатеринбург : Издательство Уральского университета, 2020. – 91 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=699059> (дата обращения: 07.06.2025). – ISBN 978-5-7996-3015-7. – Текст : электронный.
4. Кревецкий, А. В. Основы технологий искусственного интеллекта : учебное пособие : [16+] / А. В. Кревецкий, Ю. А. Ипатов, Н. И. Роженцова ; под общ. ред. А. В. Кревецкого ; Поволжский государственный технологический университет. – Йошкар-Ола : Поволжский государственный технологический университет, 2023. – 272 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=714624> (дата обращения: 07.06.2025). – Библиогр.: с. 264-267. – ISBN 978-5-8158-2358-7. – Текст : электронный.
5. Целых, А. Н. Применение временных рядов для анализа больших данных : учебное пособие по курсу «Математические методы анализа больших данных» : [16+] / А. Н. Целых, В. С. Васильев, Э. М. Котов ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2021. – 86 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=691448> (дата обращения: 07.06.2025). – Библиогр. в кн. – ISBN 978-5-9275-3983-3. – Текст : электронный.

5.2. Дополнительная литература:

1. Сахарова, Л. В. Методы оптимизации для машинного обучения : учебное пособие : [16+] / Л. В. Сахарова, Г. В. Лукьянова ; Ростовский государственный экономический университет (РИНХ). – Ростов-на-Дону : Издательско-полиграфический комплекс РГЭУ (РИНХ), 2023. – 87 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=711220> (дата обращения: 07.06.2025). – Библиогр. в кн. – ISBN 978-5-7972-3139-4. – Текст : электронный.
2. Лисьев, Г. А. Технологии поддержки принятия решений : учебное пособие : [16+] / Г. А. Лисьев, И. В. Попова. – 4-е изд., стер. – Москва : ФЛИНТА, 2022. – 133 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=103806> (дата обращения: 06.06.2025). – ISBN 978-5-9765-1300-6. – Текст : электронный.

5.3. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

Электронно-библиотечные системы (ЭБС):

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН» <http://www.biblioclub.ru/>
3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» www.znanium.com
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

Профессиональные базы данных

1. Scopus <http://www.scopus.com/>
2. ScienceDirect <https://www.sciencedirect.com/>
3. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
4. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
5. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
6. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ)) <https://rusneb.ru/>
7. Президентская библиотека им. Б.Н. Ельцина <https://www.prilib.ru/>
8. База данных CSD Кембриджского центра кристаллографических данных (CCDC) <https://www.ccdc.cam.ac.uk/structures/>
9. Springer Journals: <https://link.springer.com/>
10. Springer Journals Archive: <https://link.springer.com/>
11. Nature Journals: <https://www.nature.com/>
12. Springer Nature Protocols and Methods: <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials: <http://materials.springer.com/>
14. Nano Database: <https://nano.nature.com/>
15. Springer eBooks (i.e. 2020 eBook collections): <https://link.springer.com/>
16. "Лекториум ТВ" <http://www.lektorium.tv/>
17. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

Информационные справочные системы

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

Ресурсы свободного доступа

1. КиберЛенинка <http://cyberleninka.ru/>;
2. Американская патентная база данных <http://www.uspto.gov/patft/>
3. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
4. Федеральный портал "Российское образование" <http://www.edu.ru/>;
5. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
6. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
7. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
8. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
9. Служба тематических толковых словарей <http://www.glossary.ru/>;
10. Словари и энциклопедии <http://dic.academic.ru/>;
11. Образовательный портал "Учеба" <http://www.ucheba.com/>;
12. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы http://xn--273--84d1f.xn--p1ai/voprosy_i_otvety

Собственные электронные образовательные и информационные ресурсы КубГУ

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>

4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий [http://mschool.kubsu.ru/](http://mschool.kubsu.ru;);
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

6. Методические указания для обучающихся по освоению дисциплины

По курсу предусмотрено проведение лекционных занятий, на которых дается основной систематизированный материал, лабораторных работ, контрольной работы, зачета и экзамена.

Важнейшим этапом курса является самостоятельная работа по дисциплине с использованием указанных литературных источников и методических указаний автора курса.

Виды и формы СР, сроки выполнения, формы контроля приведены выше в данном документе.

Для лучшего освоения дисциплины при защите ЛР студент должен ответить на несколько вопросов из лекционной части курса.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

7. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

7.1 Перечень информационных технологий

- Проверка домашних заданий и консультирование посредством электронной почты.
- Использование электронных презентаций при проведении лекций и практических занятий.

7.2 Перечень необходимого программного обеспечения

1. Python,
2. R, R Studio.
3. Программы для демонстрации и создания презентаций («Microsoft Power Point»).

8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Наименование специальных помещений	Оснащенность специальных помещений	Перечень лицензионного программного обеспечения
------------------------------------	------------------------------------	---

Учебные аудитории для проведения занятий лекционного типа (ауд. 129, 131, А305.)	Мебель: учебная мебель Технические средства обучения: экран, проектор, компьютер	PowerPoint.
Учебные аудитории для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации (ауд. 147,148)	Мебель: учебная мебель Технические средства обучения: экран, проектор, компьютер	Аудитория, (кабинет) – компьютерный класс
Учебные аудитории для проведения лабораторных работ. Лаборатория 102,105,106	Мебель: учебная мебель Технические средства обучения: компьютер	Лаборатория, укомплектованная специализированными техническими средствами обучения – компьютерный класс, с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета

Для самостоятельной работы обучающихся предусмотрены помещения, укомплектованные специализированной мебелью, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование помещений для самостоятельной работы обучающихся	Оснащенность помещений для самостоятельной работы обучающихся	Перечень лицензионного программного обеспечения
Помещение для самостоятельной работы обучающихся (читальный зал Научной библиотеки)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	1. OS Windows, MS Office 2. R, R Studio. 3. Антивирус.

Помещение для самостоятельной работы обучающихся (ауд. 105, 148,150)	Мебель: учебная мебель Комплект специализированной мебели: компьютерные столы Оборудование: компьютерная техника с подключением к информационно-коммуникационной сети «Интернет» и доступом в электронную информационно-образовательную среду образовательной организации, веб-камеры, коммуникационное оборудование, обеспечивающее доступ к сети интернет (проводное соединение и беспроводное соединение по технологии Wi-Fi)	1.OS Windows, MS Office 2.R, R Studio. 3.Антивирус.
---	---	--