

**Федеральное государственное образовательное бюджетное
учреждение высшего образования
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»
(Финансовый университет)**

Краснодарский филиал Финуниверситета

Кафедра «Математика и информатика»

СОГЛАСОВАНО

ООО «Портал-Юг»
Генеральный директор



Е.В. Мостовой

«20» февраля 2024 г.

УТВЕРЖДАЮ

Краснодарский филиал
Финуниверситета

Директор

Э.В.Соболев
«20» февраля 2024 г.

Калайдин Е. Н.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
МАШИННОЕ ОБУЧЕНИЕ**

для студентов, обучающихся по направлению подготовки
01.04.02 Прикладная математика и информатика,

Направленность программы магистратуры:
Анализ больших данных и машинное обучение в экономике и финансах

*Рекомендовано Ученым советом Краснодарского филиала Финуниверситета
(протокол № 12 от 20.02.2024)*

*Одобрено кафедрой «Математика и информатика»
(протокол № 13 от 13.02.2024)*

Краснодар 2024

УДК: 681.14(076)
ББК: 32.973.2я73
К17

Рецензенты: Кирий В.А., доцент кафедры «Математика и информатика», канд. физ.-мат. наук, Коренева О.В., доцент кафедры «Математика и информатика», канд. техн. наук.

Калайдин Е. Н. «Машинное обучение». Рабочая программа дисциплины для обучающихся магистратуры по направлению 01.04.02 «Прикладная математика и информатика», направленность программы магистратуры «Анализ больших данных и машинное обучение в экономике и финансах». – Краснодар: Краснодарский филиал Финуниверситета, кафедра «Математика и информатика», 2024 г.

В рабочей программе дисциплины определены ее цель, требования к результатам освоения дисциплины, содержание программы, тематика аудиторных занятий, формы самостоятельной работы, оценочные средства для текущего контроля и промежуточной аттестации, учебно-методическое и информационное обеспечение.

Учебное издание

Калайдин Е. Н.

Машинное обучение

Рабочая программа дисциплины

Компьютерный набор и верстка Е. Н. Калайдин
Формат 60х90/16. Гарнитура Times New Roman
Усл.п.л. 1,75. Изд. № - 2024. Тираж экз.

Заказ №

Отпечатано в Краснодарском филиале Финансового университета

© Калайдин Е. Н., 2024
© КФ Финансового университета, 2024

СОДЕРЖАНИЕ

1. Наименование дисциплины	4
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	4
3. Место дисциплины в структуре образовательной программы	7
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся.....	7
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий.....	8
6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	13
7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	17
8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	18
9. Методические указания для обучающихся по освоению дисциплины	19
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	21
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.	21

1. Наименование дисциплины

Дисциплина «Машинное обучение»

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Дисциплина «Машинное обучение» обеспечивает формирование компетенций: ПКН-4; ПК-2; ПК-4; ПК-5 направления подготовки 01.04.02 «Прикладная математика и информатика» направленности программы магистратуры «Анализ больших данных и машинное обучение в экономике и финансах».

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-4	Способность разрабатывать информационные системы и алгоритмы на основе математических методов и моделей, в том числе из области искусственного интеллекта, в решении профессиональных задач	1. Демонстрирует знания в области оригинальных алгоритмов и программной реализации.	<u>Знать:</u> оригинальные алгоритмы и программную реализацию <u>Уметь:</u> владеть оригинальными алгоритмами и программной реализацией
		2. Создает оригинальные алгоритмические и программные средства в решении профессиональных задач.	<u>Знать:</u> оригинальные алгоритмические и программные средства в решении профессиональных задач <u>Уметь:</u> создавать оригинальные алгоритмические и программные средства в решении профессиональных задач

		3. Демонстрирует знания в области современных интеллектуальных технологий.	<p><u>Знать:</u> теоретические основы современных интеллектуальных технологий</p> <p><u>Уметь:</u> использовать современные интеллектуальные технологии</p>
		4. Выбирает интеллектуальные технологии и решает профессиональные задачи с их использованием.	<p><u>Знать:</u> теоретические основы интеллектуальных технологий для решения профессиональных задач с их использованием</p> <p><u>Уметь:</u> выбирать интеллектуальные технологии и решать профессиональные задачи с их использованием</p>
ПК-2	Способность применять лингвистические и сетевые модели машинного обучения, предобученные на больших наборах данных в экономике и финансах.	1. Демонстрирует знание современных предобученных лингвистических моделей или инструментов создания предобучения сетевых моделей.	<p><u>Знать:</u> современные предобученные лингвистические модели или инструменты создания предобучения сетевых моделей</p> <p><u>Уметь:</u> использовать современные предобученные лингвистические модели или инструменты создания предобучения сетевых моделей</p>
		2. Владеет методологией и инструментарием использования предобученных моделей для создания решения прикладных задач, разрабатывает на базе известных решений технологию решения прикладной задачи с помощью предобученных моделей.	<p><u>Знать:</u> методологию и инструментарий использования предобученных моделей для создания решения прикладных задач, теоретические основы для разработки на базе известных решений технологии решения прикладной задачи с помощью предобученных моделей</p> <p><u>Уметь:</u> владеть методологией и инструментарием использования предобученных моделей для создания решения прикладных задач, разрабатывать на базе известных решений технологию решения прикладной задачи с помощью предобученных моделей</p>

		3 . Владеет практическим навыком применения технологии использования предобученных моделей для решения прикладных задач.	<p><u>Знать:</u> теоретические основы технологий использования предобученных моделей для решения прикладных задач</p> <p><u>Уметь:</u> владеть практическим навыком применения технологии использования предобученных моделей для решения прикладных задач</p>
ПК-4	Способность самостоятельно разрабатывать прикладные средства лингвистического анализа с использованием методов машинного обучения, в том числе методов глубокого обучения в экономике и финансах.	1 . Демонстрирует знание актуальных методов экономического анализа с использованием методов машинного обучения.	<p><u>Знать:</u> актуальные методы экономического анализа с использованием методов машинного обучения</p> <p><u>Уметь:</u> применять актуальные методы лингвистического анализа с использованием методов машинного обучения</p>
		2 . Владеет современными методами и инструментами машинного обучения для создания и обучения прикладных моделей экономического анализа	<p><u>Знать:</u> теоретические основы современных методов и инструментов машинного обучения для создания и обучения прикладных моделей экономического анализа</p> <p><u>Уметь:</u> владеть современными методами и инструментами машинного обучения для создания и обучения прикладных моделей экономического анализа</p>
		3 . Владеет современными методами и инструментами глубокого обучения для создания и обучения прикладных моделей экономического анализа.	<p><u>Знать:</u> теоретические основы современных методов и инструментов глубокого обучения для создания и обучения прикладных моделей экономического анализа</p> <p><u>Уметь:</u> владеть современными методами и инструментами глубокого обучения для создания и обучения прикладных моделей экономического анализа</p>

ПК-5	Способность самостоятельно разрабатывать прикладные средства анализа сетевых структур с использованием методов машинного обучения, в том числе методов глубокого обучения, в экономике и финансах	1. Демонстрирует знание актуальных методов экономического анализа с использованием методов машинного обучения.	<u>Знать:</u> актуальные методы экономического анализа с использованием методов машинного обучения <u>Уметь:</u> применять актуальные методы экономического анализа с использованием методов машинного обучения
		2. Использует современные методы и инструменты машинного обучения для создания, обучения и оценки качества прикладных моделей сетевых структур.	<u>Знать:</u> теоретические основы современных методов и инструментов машинного обучения для создания, обучения и оценки качества прикладных моделей сетевых структур <u>Уметь:</u> применять современные методы и инструменты машинного обучения для создания, обучения и оценки качества прикладных моделей сетевых структур
		3. Применяет современные методы и инструменты глубокого обучения для создания, обучения и оценки качества прикладных нейросетевых моделей сетевых структур.	<u>Знать:</u> теоретические основы современных методов и инструментов глубокого обучения для создания, обучения и оценки качества прикладных нейросетевых моделей сетевых структур <u>Уметь:</u> применять современные методы и инструменты глубокого обучения для создания, обучения и оценки качества прикладных нейросетевых моделей сетевых структур

3. Место дисциплины в структуре образовательной программы

Дисциплина «Машинное обучение» относится к циклу дисциплин направленности программы магистратуры, направления подготовки 01.04.02 - Прикладная математика и информатика, ОП «Анализ больших данных и машинное обучение в экономике и финансах».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Модуль 5 (в часах)
----------------------------------	------------------------	-----------------------

Общая трудоёмкость дисциплины (в том числе курсовой проект)	3/108	108
Контактная работа- Аудиторные занятия	40	40
Лекции	10	10
Семинары, практические занятия	30	30
Самостоятельная работа	68	68
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Введение в машинное обучение

Основные понятия машинного обучения. Связь с другими дисциплинами. Контекст машинного обучения как дисциплины - анализ данных и искусственный интеллект. Сферы применения машинного обучения. Типы задач машинного обучения - обучение с учителем и без учителя. Структура данных для машинного обучения. Инструментальные средства машинного обучения. Понятие модели машинного обучения.

Тема 2. Классификация. Регрессия

Постановка задачи классификации как задачи машинного обучения. Отличия от задачи регрессии. Структура данных для классификации. Логистическая регрессия - функция гипотезы, граница принятия решений, функция ошибки логистической регрессии, градиентный спуск для логистической регрессии, многоклассовая классификация, алгоритм “один против всех”.

Постановка задачи регрессии. Линейная регрессия с одной переменной - функция гипотезы, функция ошибки, метод градиентного спуска. Регрессия с несколькими переменными - множественная линейная регрессия, нормализация признаков, полиномиальная регрессия. Практическое построение регрессии - загрузка и представление данных, реализация метода градиентного спуска, оценка качества регрессии, подбор скорости обучения, знакомство с библиотекой sklearn.

Тема 3. Методы обучения с учителем. Задачи обучения без учителя

Универсальность методов обучения с учителем, общая постановка задачи. Линейные модели - линейная и логистическая регрессии - как единая модель. Полиномиальные модели. Метод опорных векторов, ядра - линейное, гауссово, другие. Перцептрон. Деревья решений. К ближайших соседей.

Наивная байесовская модель. Достоинства и недостатки разных типов моделей, их сравнительная характеристика, применимость. Применение этих моделей для решения задач классификации и регрессии.

Задача кластеризации - постановка задачи, структура датасета, результат и интерпретация. Метод К средних - формализация, гиперпараметры, применимость. Другие методы кластеризации - DBSCAN, иерархическая, агломеративная кластеризация. Задача обнаружения аномалий. Задача понижения размерности - метод главных компонент, метод независимых компонент. Обучение с подкреплением.

Тема 4. Диагностика систем машинного обучения

Метрики эффективности машинного обучения - сравнение с функциями ошибки. Типичные метрики эффективности для моделей регрессии - MAE, MSE, RMSE, MSLE, MAPE и другие. Метрики эффективности для моделей классификации - accuracy, precision, recall, F1, ROC, PR и другие. Недообучение и переобучение. Проблема bias-variance. Оценка сложности моделей. Обобщающая способность моделей, тестовый набор, кривые обучения. Методы борьбы с недо- и переобучением. Регуляризация. Задача выбора модели - кросс-валидация, гиперпараметры моделей, поиск по сетке, валидационный набор.

Тема 5. Предварительный анализ и обработка данных

Сбор данных для моделей обучения с учителем - реляционная форма данных, понятие чистых данных, оценка источников и объемов данных. Описательный (предварительный) анализ данных (EDA) - анализ репрезентативности, шкалы и типы, визуализация, проблема несбалансированности, обнаружение корреляций, аномалий в данных. Очистка и преобразование данных - удаление лишних признаков, удаление непоказательных объектов, заполнение отсутствующих значений, создание суррогатных признаков, преобразование шкал, воспроизводимость преобразования данных.

Тема 6. Практическое использование моделей машинного обучения

Стохастический и пакетный градиентный спуск. Отбор признаков. Частичное обучение с учителем. Ансамблирование моделей - беггинг, бустинг, стакинг, случайный лес, XGBoost, CatBoost. Конвейеризация моделей машинного обучения. Основные этапы проекта по машинному обучению. Визуализация, интерпретация, представление и анализ результатов машинного обучения. Работа с разными типами данных - преобразование графической информации, методы векторизации текста.

5.2. Учебно-тематический план

№	Наименование те-	Трудоемкость в часах	Формы
---	------------------	----------------------	-------

п/п	мы (раздела) дисциплины	Всего	Контактная работа- Аудиторная работа			Самостоятельная работа	текущего контроля успеваемости
			Общая, в т.ч.:	Лекции	Семинары, практические занятия		
1	Введение в машинное обучение	18	4	1	3	14	опрос, проверка лабораторных работ
2	Классификация.Регрессия	18	8	2	6	10	опрос, проверка лабораторных работ
3	Методы обучения с учителем. Задачи обучения без учителя	18	8	2	6	10	опрос, проверка лабораторных работ
4	Диагностика систем машинного обучения	18	6	2	4	12	опрос, проверка лабораторных работ
5	Предварительный анализ и обработка данных	18	6	2	4	12	опрос, проверка лабораторных работ
6	Практическое использование моделей машинного обучения	18	8	1	7	10	опрос, проверка лабораторных работ
	В целом по дисциплине	108	40	10	30	68	Согласно учебному плану: контрольная работа
	Итого в %	100	37	9	28	63	

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8,9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Введение в машинное обучение	Входной контроль. Изучение технологического стека анализа данных, построенного на базе языка программирования Python. Знакомство с библиотеками numpy, pandas, matplotlib. Выполнение задания по простому статистическому анализу данных инструментальными средствами, включающему продвинутое владение соответствующими инструментами. [1-3]	Решение и обсуждение задач
Классификация. Регрессия	Реализация модели классификации (логистической регрессии) своими руками наподобие линейной регрессии. Использование библиотечных функций. [1-3] Построение модели регрессии методом градиентного спуска своими руками. Знакомство с основами обучения параметров модели. Построение модели множественной регрессии своими руками. Построение модели библиотечными средствами. Сравнение результатов. Реализация модели регрессии на примере, близком к реальному (датасет boston). Интерпретация результатов моделирования. [1-3]	Решение и обсуждение задач
Методы обучения с учителем. Задачи обучения без учителя	Построение модели классификации на простых данных. Построение нескольких моделей, сравнение их эффективности. Построение модели на сложных данных с нелинейными и слабыми зависимостями. Оценка эффективности, выбор модели. Построение модели регрессии на сложных данных. Построение разных типов регрессоров. Оценка эффективности, выбор модели. Выполнение контрольной работы. [1-3] Решение задачи кластеризации на искусственных данных. Метод К-средних. Метод локтя. Обнаружение аномалий на искусственных данных. Обнаружение аномалий как задача предварительного анализа данных. Понижение размерности на искусственных данных. Интерпретация результата. Понижение размерности для визуализации данных. [1-3]	Решение и обсуждение задач

<p>Диагностика систем машинного обучения</p>	<p>Оценка эффективности модели регрессии с помощью разных метрик. Оценка эффективности классификации с помощью разных метрик. Построение диагностических кривых (PR, ROC) для бинарной классификации. Диагностика недо-и переобучения в модели классификации. Построение кривых обучения, их интерпретация. Тестовый набор данных Демонстрация необходимости кросс-валидации. Использование кросс-валидированных оценок качества модели. Демонстрация оптимизации гиперпараметров модели, использование валидационного набора для выбора модели. Поиск модели от простых к сложным. [1-3]</p>	<p>Решение и обсуждение задач</p>
<p>Предварительный анализ и обработка данных</p>	<p>Сбор и интеграция данных из разных источников. Основные численные характеристики датасета, анализ на чистоту. Нормализация признаков и заполнение отсутствующих значений на искусственном примере. Преобразование категориальных признаков в численные. Краткий алгоритм анализа данных на реальном примере. Интерпретация результатов. Отбор и инжиниринг признаков на примере, приближенном к реальному (датасет titanic). [1-3]</p>	<p>Решение и обсуждение задач</p>
<p>Практическое использование моделей машинного обучения</p>	<p>Использование простых ансамблей моделей. Оптимизация гиперпараметров моделей. Визуализация, интерпретация и представление результатов машинного обучения исходя из бизнес-задачи моделирования. Построение конвейера машинного обучения. Построение модели классификации текстов. Знакомство с разными методами векторизации текстов. Построение модели классификации изображений. Представление изображений в виде вектора. [1-3]</p>	<p>Решение и обсуждение задач</p>

6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Машинное обучение».

Фонд оценочных средств включает контрольные материалы для проведения текущего контроля и промежуточной аттестации.

Вопросы для оценки знаний и умений, характеризующих формирование компетенций

Шифр компетенции	Вопросы	Правильный ответ
ПKN-4	1. Какие основные задачи решаются в рамках машинного обучения?	Задачи классификации, регрессии и кластеризации
	2. Что такое набор данных (датасет)?	Это коллекция данных
	3. Для чего используется линейная регрессия?	Для предсказания непрерывной целевой переменной
	4. Что называют обучением с подкреплением?	Обучение, основанное на собственном опыте
	5. В чем заключается основная идея генетических алгоритмов?	Использование эволюционных принципов для поиска оптимальных решений
ПК-2	6. Какие методы имеет предварительный анализ данных?	Сплошное наблюдение, выборочное наблюдение, сравнение данных
	7. Какие три типа пропусков данных существуют?	MCAR, MAR, MNAR
	8. Что такое NearMiss?	Метод недостаточной выборки
	9. Как называются параметры, значения которых задается до начала обучения модели и не изменяется в процессе обучения?	Гиперпараметры модели
	10. Какие модели машинного обучения являются сетевыми?	Рекуррентная, сверточная нейронная сети, наивный Байесовский классификатор
ПК-4	11. Напишите формулу такого показателя эффективности диагностики модели машинного обучения	Правильность = количество правильных прогнозов / общее количество прогнозов
	12. Какая функция активации обычно используется в выходном слое нейронной сети для задачи регрессии?	Линейная

	13. Какую роль играет переобучение в задаче машинного обучения и разработке прикладных средств анализа текста?	Снижение точности модели
	14. Что такое FastText?	Расширение Word2Vec
	15. Какие существуют методы визуализации данных?	Гистограммы, круговые и столбчатые диаграммы
ПК-5	16. В каком типе задач машинного обучения используется метод опорных векторов?	Классификация
	17. Какую роль играет регуляризация в задаче машинного обучения?	Помогает улучшить обобщение модели
	18. Какие есть принципы нормирования данных?	Масштабирование, стандартизация, нормализация
	19. Какие методы используются для преобразования категориальных признаков в числовые?	One-Hot Encoding, Label Encoding
	20. Какие есть виды ансамблевых методов?	Стекинг, бэггинг, бустинг

Практико-ориентированные задания

Шифр компетенции	Практико-ориентированные задания	Правильный ответ
ПKN-4	1. В базе данных имеется 10 000 записей. Сколько информации (в битах) необходимо для кодирования одной записи, если на каждый атрибут выделяется по 3 бита?	29 бит информации.
	2. В задаче регрессии имеется 500 точек данных. Сколько параметров требуется настроить в линейной регрессии, если каждая точка представлена двумя координатами (x, y)?	Необходимо настроить 1000 параметров.
	3. Имеется набор данных для классификации, состоящий из 50 000 примеров. Сколько раз необходимо выполнить итерацию обучения алгоритма k-ближайших соседей (k-NN), если в каждой итерации используется 10% набора данных для проверки?	Необходимо выполнить 10 итераций.
ПК-2	4. Ваша задача разработать систему рекомендаций для электронной коммерции. На основе какой фильтрации будет оптимально это сделать?	Коллаборативной фильтрации

	5. Перед вами стоит задача построить модель прогнозированных временных рядов. Какие модельные формы вы будете для этого использовать?	ARIMA или LSTM
	6. Дан набор данных, состоящий из последовательностей действий пользователей на веб-сайте. Требуется классифицировать эти действия, то есть определить, какие из них являются целевыми, а какие – нет. Что для этого нужно сделать?	Определить цели, собрать данные, создать модель классификации и внедрить её на веб-сайте.
ПК-4	7. Доктор хочет построить модель, прогнозирующую возможность развития осложнений у пациента после операции - все ли будет благополучно в течение нескольких следующих месяцев или нет. К какому типу относится данная задача?	Классификация
	8. Астроном хочет построить модель, которая сможет разбить известные науке звезды на группы по их характеристикам, чтобы лучше изучить их особенности. К какому типу относится данная задача?	Кластеризации
	9. Заполните пропуск: «Для выпуклой функции указывает сторону наискорейшего убывания»	Градиент
ПК-5	10. Что может являться критерием останова в градиентном спуске?	Достижение определенной точности
	11. Рассмотрим два объекта: у первого отступ линейного классификатора равен -10, у второго -1000. Как соотносятся два этих объекта?	Второй объект находится дальше от разделяющей поверхности, чем первый.
	12. Рассмотрим пользователя социальной сети как объект в задаче машинного обучения. Что может являться задачей классификации?	Предсказание пола пользователя
	17. Рассмотрим признак “Образовательная программа” при анализе данных по студентам университета. Этот признак может принимать три значения: “Экономика”, “Математика”, “Философия”. Воспользуемся one-hot кодированием и заменим этот признак на три бинарных, которые будут соответствовать категориям в том порядке, в котором они перечислены выше. Как будет закодирован признак со значением “Философия”?	(0,0,1)
	18. Чему будет равен корень из среднеквадратичной ошибки для набора из 3 наблюдений, где отклонение предсказания линейной регрессии от реальных значений равны: -1, 2, 2?	2

Тесты

Шифр компетенции	Тестовые задания	Правильный ответ
ПКН-4	1. Какой из следующих методов машинного обучения используется для решения задачи классификации? а. Метод k-средних б. Наивный Байесовский метод с. Линейная регрессия d. Метод опорных векторов	d
	2. Какая из следующих моделей машинного обучения является регрессионной моделью? а. Наивный Байесовский метод б. Метод k-ближайших соседей с. Деревья решений d. Линейная регрессия	d
	3. Какой из следующих алгоритмов машинного обучения основан на использовании деревьев решений? а. Метод опорных векторов б. Случайный лес с. Наивный Байесовский метод d. k-ближайших соседей	b
ПК-2	4. Что является основным недостатком наивного Байесовского метода? а. Сложность в масштабировании б. Низкая точность с. Высокая вычислительная сложность d. Зависимость от распределения данных	d
	5. Что используется в методе k-ближайших соседей для определения расстояния между точками? а. Евклидово расстояние б. Манхэттенское расстояние с. Расстояние Чебышёва d. Расстояние Хемминга	a
	6. Какой из перечисленных методов относится к группе методов глубокого обучения? а. Метод k-means б. Метод опорных векторов с. Сверточная нейронная сеть d. Дерево решений	c
ПК-4	7. Что такое предобучение в машинном обучении? а. Процесс обучения модели на большом наборе данных б. Процесс улучшения качества модели после ее обучения с. Процесс получения начальных весов модели для дальнейшего обучения d. Процесс отбора признаков для обучения модели	c
	8. Что такое метод градиентного спуска? а. Это метод оптимизации функции потерь в машинном обучении б. Это метод выбора признаков в модели с. Это метод кластеризации данных d. Это метод оценки качества модели	a

	<p>9. Что такое ранняя остановка в обучении нейронных сетей?</p> <p>а. Остановка обучения после определенного количества эпох</p> <p>б. Остановка обучения, если качество модели на валидационной выборке не улучшается в течение определенного количества эпох</p> <p>с. Остановка обучения, если качество модели на обучающей выборке достигло максимума</p> <p>д. Остановка обучения, если модель достигла нулевого значения функции потерь</p>	b
ПК-5	<p>10. В чем заключается принцип работы автоэнкодера?</p> <p>а. В сжатии данных</p> <p>б. В восстановлении данных</p> <p>с. В классификации данных</p> <p>д. В предсказании временных рядов</p>	a
	<p>11. Что из следующего является особенностью сверточной нейронной сети?</p> <p>а. Наличие слоев с большим количеством нейронов</p> <p>б. Наличие слоёв, состоящих из сверточных фильтров</p> <p>с. Наличие только одного полносвязного слоя</p> <p>д. Отсутствие слоёв с нелинейными функциями активации</p>	b
	<p>12. Каким образом DropConnect улучшает обобщающую способность модели?</p> <p>а. Путем случайного отключения нейронов</p> <p>б. Путем случайного изменения весов между нейронами</p> <p>с. Путем уменьшения количества слоёв</p> <p>д. Путем увеличения количества скрытых слоев</p>	a

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

7.1. Основная литература

1. Колдаев, В. Д. Структуры и алгоритмы обработки данных : учебное пособие / В. Д. Колдаев. - Москва : РИОР : ИНФРА-М, 2021. - 296 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1230215>. - Текст: электронный.

7.2. Дополнительная литература

1. Нагаева, И. А. Основы алгоритмизации и программирования: практикум : учебное пособие / И. А. Нагаева, И. А. Кузнецов. - Москва : Берлин : Директ- Медиа, 2021. - 169 с. - ЭБС Университетская библиотека ONLINE. - URL: <https://biblioclub.ru/index.php?page=book&id=598404>. - Текст : электронный.

2. Златопольский, Д. М. Программирование: типовые задачи, алгоритмы, методы: учебное пособие / Д. М. Златопольский. — 4-е изд. (эл.). — Москва : Лаборатория знаний, 2020. — 226 с.: ил. — ЭБС Университетская библиотека ONLINE. — URL: <https://biblioclub.ru/index.php?page=book&id=222873>. - Текст: электрон-

ный.

3. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2024. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://ezpro.fa.ru:2058/bcode/511121>

4. Лимановская, О. В. Основы машинного обучения : учебное пособие / О. В. Лимановская, Т. И. Алферьева. — 2-е изд. — Москва : ФЛИНТА, 2022. — 88 с. — ISBN 978-5-9765-5006-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/231677>. — Режим доступа: для авториз. Пользователей.

5. Гладилин, П. Е. Технологии машинного обучения : учебно-методическое пособие / П. Е. Гладилин, К. О. Боченина. — Санкт-Петербург : НИУ ИТМО, 2020. — 75 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/190885>. — Режим доступа: для авториз. пользователей.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОН-ЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система Znanium
5. <http://www.znaniy.com>
6. Электронно-библиотечная система издательства «ЮРАЙТ»
7. <https://urait.ru/>
8. Электронно-библиотечная система издательства Проспект
9. <http://ebs.prospekt.org/books>
10. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>
11. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
12. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>
13. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
14. Национальная электронная библиотека Б.Ир://нэб.рф/
15. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>
16. Pyru 1.0.9 [Электронный ресурс]: сайт. - Режим доступа: <https://pypi.python.org/pypi/pyru>
17. Python Data Analysis Library [Электронный ресурс]: сайт. - Режим доступа: <http://pandas.pydata.org/>
18. Python Documentation [Электронный ресурс]: сайт. - Режим доступа:

<http://python.org/doc/>

19. Python Standard Library [Электронный ресурс]: сайт. - Режим доступа: <https://docs.python.org/2/library/>

20. Scikit-learn Machine Learning in Python [Электронный ресурс]: сайт. - Режим доступа: <http://scikit-learn.org>

21. The Python Tutorial // <https://docs.python.org/3/tutorial/index.html>

22. NumPy User Guide // <http://docs.scipy.org/doc/numpy/user/index.html>

23. Pandas User Guide <http://pandas.pydata.org/pandas-docs/stable/>

9. Методические указания для обучающихся по освоению дисциплины

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного списка. Кроме этого, необходимо активно работать с Интернет- источниками и пособиями других авторов, помогающими усвоить материал отдельных разделов программы.

Необходимо конспектировать лекции, помечая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на консультации.

При подготовке к семинарским занятиям необходимо изучить вопросы, вынесенные на самостоятельное изучение, так как семинарские занятия предполагают их обсуждение и дискуссию по теме; кроме того, задания для самостоятельной работы необходимы для того, чтобы успешно выполнить самостоятельные задания на семинарах.

Индивидуальные задания для работы на компьютере, файлы с выполненными заданиями необходимо хранить в личной сетевой папке в компьютерной сети вуза.

Для выполнения кейса задачи, рекомендуется использовать язык программирования Python и специализированные библиотеки для анализа данных и машинного обучения.

Студент должен прислать рабочий notebook с решением кейса задачи, комментариями кода и аналитическими выводами до конца экзамена на электронную почту преподавателя.

Контрольные работы проводятся в форме письменного экспресс-опроса, открытых и закрытых тестовых заданий, решения ситуационных и практических задач по темам учебной дисциплины

Критерии оценки практической экзаменационной работы:

1. Структурированность отчета. В работе должна прослеживаться четкая структура - подготовительный этап, анализ данных, построение простых моделей, сравнение и анализ моделей, выводы, построение моделей с учетом выводов, итоговый результат.

2. Наличие выводов. Работа должна содержать текстовые замечания, поясняющие каждый шаг работы студента: что делается, зачем и какую информацию это нам дает. Оценивается полнота и адекватность выводов.

3. Визуализация. Работа должна демонстрировать навыки студента визуализировать информацию. Особенно на этапах описательного анализа и анализа обучаемости модели. Оценивается разнообразие, наглядность и информативность визуализации.

4. Использование метрик эффективности. Оценивается разнообразие и адекватность задаче примененных метрик эффективности (включая время обучения) а также полнота сравнения и правильность выводов из сравнения моделей по разным метрикам.

5. Валидность результатов. Студент должен продемонстрировать умение оценивать достоверность измерения метрик моделей и повышать ее с использованием перекрестной проверки (кроссвалидации). Использование k-fold cross validation является предпочтительным методом измерения эффективности модели. Если происходит выбор модели, то ее итоговая эффективность должна измеряться на чистом наборе данных.

Список датасетов, использующихся на экзамене:

1. <https://www.kaggle.com/uciml/mushroom-classification>
2. <https://www.kaggle.com/lodetomasi1995/income-classification>
3. <https://www.kaggle.com/uciml/glass>
4. <https://www.kaggle.com/uciml/german-credit>
5. <https://www.kaggle.com/zaurbegiev/my-dataset>
6. <https://www.kaggle.com/kaushikuresh147/customer-segmentation>
7. <https://www.kaggle.com/deepul109/star-dataset>
8. <https://www.kaggle.com/vinesmsuic/star-categorization-giants-and-dwarfs>
9. <https://www.kaggle.com/shebrahimi/financial-distress>
10. https://www.kaggle.com/teej_mahal20/airline-passenger-satisfaction
11. <https://www.kaggle.com/amir75/caesarean-section-classification>
12. <https://www.kaggle.com/sachinsharma1123/performance-prediction>
13. <https://www.kaggle.com/ninzaami/loan-predication>
14. <https://www.kaggle.com/caparrini/beatsdataset>
15. <https://www.kaggle.com/zhiruo19/covid19-symptoms-classification>
16. <https://www.kaggle.com/kunalvsingh93/banking-modelclassification>
17. <https://www.kaggle.com/mansoordaku/ckdisease>

18. <https://www.kaggle.com/mnassrib/telecom-churn-datasets>
19. <https://www.kaggle.com/akshayksingh/kidney-disease-dataset>
20. <https://www.kaggle.com/henriqueyamahata/bank-marketing>
21. <https://www.kaggle.com/maajdl/yeh-concret-data>
22. <https://www.kaggle.com/hellbuoy/car-price-prediction>
23. <https://www.kaggle.com/rhuebner/human-resources-data-set>
24. <https://www.kaggle.com/loveall/appliances-energy-prediction>
25. <https://www.kaggle.com/elikplim/forest-flres-data-set>
26. <https://www.kaggle.com/nandvard/microsoft-data-science-capstone>
27. <https://www.kaggle.com/shebrahimi/financial-distress>
28. <https://www.kaggle.com/aungpyaeap/beauty>
29. <https://www.kaggle.com/vbmokin/ammonium-prediction-in-river-water>
30. <https://www.kaggle.com/veer06b/marrket-mix-dataset>

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

1. ОС Astra Linux
 2. Libre Office
 3. Антивирус Kaspersky
 4. Дистрибутив языка Python 3.6 (или более поздней версии) Anaconda
 5. Облачная среда разработки Google Colab или аналогичная
 6. Среда разработки Jupyter Notebook
 7. Современные профессиональные базы данных и информационные справочные системы
- не используются
- Сертифицированные программные и аппаратные средства защиты информации
- не используются

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Материально-техническая база Краснодарского филиала Финансового университета соответствует действующим противопожарным правилам и нормам, обеспечивает проведение всех видов дисциплинарной и междисциплинарной

плинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Образовательный процесс обеспечивается специальными помещениями, которые представляют собой аудитории для проведения занятий лекционного типа, занятий семинарского типа, выполнения курсовых работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы студентов и помещения для хранения и профилактического обслуживания учебного оборудования. Специальные помещения укомплектованы специализированной мебелью и техническими средствами обучения, обеспечивающими представление учебной информации большой аудитории, демонстрационным оборудованием.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, обеспечивающей доступ к сети Интернет и электронной информационно-образовательной среде Краснодарского филиала Финансового университета.