

Министерство науки и высшего образования РФ Ульяновский государственный университет Ф-Рабочая программа по дисциплине	Форма	
---	-------	--

УТВЕРЖДЕНО
 решением Ученого совета ИЭиБ
 от « 23 » июня 2022 г., протокол № 09 / 252
 Председатель Е.М.Белый
 « 23 » июня 2022 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

М.1.2.6 Машинное обучение

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

38.04.01 Экономика

Квалификация выпускника

Магистр

(бакалавр/магистр/специалист)

Программа магистратуры

Искусственный интеллект в финансово-экономических
системах

Курс 1
 Семестр 2

Распределение учебного времени

Трудоемкость по учебному плану	108 / 3	часов/зачетных единиц
Лекции	18	часов
Лабораторные работы	-	часов
Практические занятия	18	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	36	часов
Контактная работа по экзамену	-	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	72	часов
Самостоятельная работа по подготовке к экзамену	-	часов
Экзамен	-	семестр
Зачет	2	семестр
БРК, ДЗ	-	семестр

(год)

Раздел 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является достижение планируемых результатов обучения, соответствующих установленным в ОПОП индикаторам достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения
1. ПК-5 Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях	ПК-5.1 Ставит задачи по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области	знания: Классы методов и алгоритмов машинного обучения умения: Ставить задачи и адаптировать методы и алгоритмы машинного обучения навыки: Умение ставить задачи по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области
2. ПК-6 Способен руководить проектами по созданию систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения со стороны заказчика	ПК-6.1 Руководит разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика	знания: Возможности современных инструментальных средств и систем программирования для решения задач машинного обучения умения: Проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения навыки: Способность руководить разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика
	ПК-6.2 Осуществляет руководство созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения	знания: Функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения Принципы построения систем искусственного интеллекта, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта, методы интеллектуального планирования экспериментов умения: Применять современные инструментальные средства и системы программирования для разработки новых моделей и методов машинного обучения навыки: Способность осуществлять руководство коллективной проектной деятельностью для создания, поддержки и использования систем искусственного интеллекта

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к части, формируемой участниками образовательных отношений ОПОП.

Дисциплина является обязательной

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих практиках: Производственная практика. Практика по профилю профессиональной деятельности (ПК-5), Производственная практика. Практика по профилю профессиональной деятельности (ПК-6), Преддипломная практика (ПК-5), Преддипломная практика (ПК-6); государственной итоговой аттестации в форме: Подготовка к процедуре защиты и защита выпускной квалификационной работы (ПК-6), Подготовка к процедуре защиты и защита выпускной квалификационной работы (ПК-5)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Для формирования заявленных компетенций используются методологические технологии, реализующие деятельностный, личностно-ориентированный, практико-ориентированный подходы.

Основными стратегическими технологиями являются: лекционные занятия, практические занятия, процедуры самообучения

На достижение конкретных целей обучения направлены применяемые тактические технологии: case-study, задания, классическая лекция, проблемная лекция

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2 семестр

Виды и тематика занятий	Количество часов	Формируемые компетенции
ВВЕДЕНИЕ в МАШИННОЕ ОБУЧЕНИЕ	54	ПК-5, ПК-6
Лекция. Основные понятия и пирамида задач машинного обучения. Машинное обучение и ИИ. Архитектура систем ИИ.	2	
Лекция. Постановки задач регрессионного анализа и классификации. Постановки задач анализа структуры данных и кластеризации. Примеры задач из области финансов и экономики. Подходы к их решению.	2	
Лекция. Сбор, фильтрация, подготовка и предварительный анализ данных. Разметка данных для последующего применения методов обучения с учителем.	2	
Лекция. Онтология предметной области. Графы знаний. Применение графов знаний для решения задач. Извлечение знаний из данных.	3	
Практическое занятие. Разбор примеров задач, возникающие при построении финансово-экономических систем	2	
Практическое занятие. Знакомство со средой решения задач. Jupyter notebook. Google Colab. Использование готовых решений.	2	
Практическое занятие. Изучаем базовые конструкции языка Python. Работа с коллекциями данных.	2	
Практическое занятие. Чтение данных из различных источников, в том числе из репозитория данных. Предварительный анализ данных. Библиотеки pandas и matplotlib.	3	
Задания для самостоятельной работы, в том числе выполнение Тесты контроля усвоения теории Практические задания Кейс	36	
Решение задач машинного обучения	54	ПК-5, ПК-6

Лекция. Решение задач регрессионного анализа	3
Лекция. Решение задач классификации	3
Лекция. Снижение размерности. Анализ структуры данных	3
Практическое занятие. Решение задач регрессионного анализа с использованием библиотеки <code>skikit-learn</code>	3
Практическое занятие. Решение задач классификации с использованием библиотеки <code>skikit-learn</code>	3
Практическое занятие. Решение задач снижения размерности и кластеризации с использованием библиотеки <code>skikit-learn</code>	3
Задания для самостоятельной работы, в том числе выполнение Тесты контроля усвоения теории Практические задания Кейс	36
Иная контактная работа:	0

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Изучение дисциплины рекомендуется начать с ознакомления с рабочей программой, ее структурой и содержанием разделов. Учебный материал структурирован, изучение дисциплины осуществляется в тематической последовательности.

Курс состоит из двух модулей. Каждый модуль включает в себя несколько тем. Освоение каждой темы состоит из следующих работ:

- работа с лекциями, дополнительным теоретическим материалом;
- закрепление нового материала с использованием теста темы (5-7 тестовых заданий);
- выполнение практических заданий.

В конце каждого модуля обучающийся выполняет итоговое задание раздела - кейсовое задание с использованием всего изученного материала.

Практические задания и кейсы выполняются с использованием бесплатного ПО - интерпретатора Python-3 и библиотеки машинного обучения `skikit-learn` и вспомогательных библиотек работы с массивами данных и графиками `numpy`, `pandas`, `matplotlib`, `scipy`. Используется среда вычислений Jupyter Notebook или PyCharm.

Занятия лекционного типа дают систематизированные знания по дисциплине, концентрируют внимание на наиболее сложных и важных вопросах. Следует начать освоение дисциплины с изучения основных понятий и постановок задач машинного обучения. Знание постановки задачи и проблем подбора методов решения позволит в дальнейшем легко ориентироваться при решении практических задач. Подготовка к **занятиям практики** включает ознакомление с планом практического занятия; работу с конспектом лекций, работу с электронными ресурсами курса: прохождение теста темы, выполнение практических заданий текущей темы, работу с учебной и учебно-методической литературой курса.

При выполнении кейсов рекомендуется дополнительно познакомиться с рекомендуемыми научными статьями, рекомендованными в методических материалах темы.

Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплин представлены в рабочей программе. Условия аттестации приведены в технологической карте, входящей в состав рабочей программы дисциплины.

Раздел 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющихся в библиотеке, или электронный адрес издания (ресурса) в сети Интернет
УЧЕБНЫЕ, УЧЕБНО-МЕТОДИЧЕСКИЕ И НАУЧНЫЕ ИЗДАНИЯ		
1.	Есипов, Борис Алексеевич. Методы исследования операций [Текст] : учеб. пособие / Б. А. Есипов. Санкт-Петербург [и др.]: Лань, 2010. - 253 с. ISBN 978-5-8114-0917-4. Экземпляры: всего 6.	6
2.	Есипов, Б. А. Методы исследования операций [Электронный ресурс] : учебное пособие / Б. А. Есипов. 2-е изд., испр. и доп. Санкт-Петербург: Лань, 2013. - 304 с. ISBN 978-5-8114-0917-4.	https://e.lanbook.com/books/element.php?pl1_id=68467
3.	Рутковская, Данута. Нейронные сети, генетические алгоритмы и нечеткие системы [Текст] / Д. Рутковская, М. Пилиньский, Л. Рутковский ; пер. с пол. И. Д. Рудинского. М.: Горячая линия - Телеком, 2007. - 383 с. ISBN 5-93517-103-1. Экземпляры: всего 4.	4
4.	Хайкин, Саймон. Нейронные сети [Текст] : Полный курс / С. Хайкин ; [пер. с англ. Н. Н. Куссуль, А. Ю. Шелестова ; под ред. Н. Н. Куссуль]. 2-е изд. Москва [и др.]: Вильямс, 2006. - 1103 с. ISBN 5-8459-0890-6.	8
ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ		
1.	Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.	https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html
2.	Машинное обучение и нейросетевой анализ данных в Python	https://mooped.net/course/view.php?id=393
3.	Программирование в Python и методы вычислений	https://mooped.net/course/view.php?id=351
4.	Boutaba, R., Salahuddin, M.A., Limam, N. et al. A comprehensive survey on machine learning for networking: evolution, applications and research opportunities. J Internet Serv Appl 9, 16 (2018). https://doi.org/10.1186/s13174-018-0087-2	https://jisajournal.springeropen.com/articles/10.1186/s13174-018-0087-2
5.	API design for machine learning software: experiences from the scikit-learn project, Buitinck et al., 2013.	https://arxiv.org/abs/1309.0238

6.2. Материально-техническая база и программное обеспечение

№№ п/п	Аудитории для проведения учебных занятий, самостоятельной работы и проведения государственной итоговой аттестации	Перечень основного оборудования	Программное обеспечение
1.	517 (I)	Персональный компьютер 1 (1), Персональный компьютер 2 (20), Комплект учебной мебели (1)	Microsoft Windows Enterprise, Microsoft Office Standard, Агент Dr.Web, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач

Раздел 7. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ/ ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценивания индикаторов достижения компетенций направлены на:

- усвоение теоретического материала (объем знаний, глубина усвоения), предусмотренного рабочей программой;
- умение излагать материал (четкость, грамотность изложения материала, точность и полнота воспроизведения учебного материала);
- умение применять теоретические знания при решении практических заданий.

Шкала оценивания представлена ниже.

Уровень сформированности элементов компетенции	Критерии оценивания	Шкала оценивания
Пороговый уровень	Обучающийся имеет знания основного материала, проявляет умение логично его излагать, но может допускать неточности в изложении материала, недостаточно правильные формулировки, испытывает затруднения в выполнении практических заданий.	удовлетворительно
Продвинутый уровень	Обучающийся твердо знает программный материал, излагает его грамотно и по существу, не допускает существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения	хорошо
Высокий уровень	Обучающийся глубоко и прочно усвоил программный материал, грамотно и логически стройно его излагает, дает исчерпывающие ответы на поставленные вопросы. В ответе тесно увязывается теория с практикой, при этом обучающийся не затрудняется с ответом при видоизменении задания, свободно справляется с задачами, вопросами и другими видами применения знаний, показывает знакомство с монографической литературой, периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	отлично

7.1. Текущий контроль успеваемости

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины (модуля) и производится с применением технологии рейтингового контроля в соответствии с технологической картой дисциплины. Порядок составления технологической карты и алгоритм проведения процедуры оценивания видов деятельности обучающихся, направленных на освоение знаний, умений, навыков и/или опыта деятельности, по накопительной системе в баллах устанавливается положением о системе РИТМ в ФГБОУ ВО «ПГТУ»

7.2. Промежуточная аттестация обучающихся

Промежуточная аттестация обучающихся направлена на оценивание результатов обучения по дисциплине (модулю) и проводится с использованием фондов оценочных средств.

Примеры типовых контрольных заданий из базы фонда оценочных средств по образовательной программе.

Раздел 1. Введение в машинное обучение

Пороговый уровень.

Выберите наиболее точное определение машинного обучения (из урока 1.1)

Выберите один ответ:

Машинное обучение - это когда компьютерная программа решает некоторую задачу, обучаясь на опыте в соответствии с некоторым критерием качества работы.

Машинное обучение - это когда компьютерная программа решает некоторую задачу, обучаясь на накапливаемых данных

Машинное обучение - это когда компьютерная программа решает некоторую задачу, обучаясь на данных, оптимизируя свое решение в соответствии с некоторым критерием качества работы.

Машинное обучение - это когда компьютерная программа решает некоторую задачу, обучаясь на опыте, если её работа по решению этой задачи улучшается с опытом в соответствии с некоторым критерием качества работы.

Продвинутый уровень

Определите - какие задачи из перечисленных относятся к типу задач обучения с учителем

Выберите один или несколько ответов:

Вы создали онлайн-курс и запустили его в этом семестре. Однако результаты обучения студентов на курсе слишком различаются. Хотелось бы понять - в чем причины таких разных результатов? Кто такие отстающие? И как с ними построить работу?

Последнее время операторы вашей организации не справляются с потоком электронных писем с вопросами от клиентов. Нужно попробовать построить систему автоматизированных ответов на типовые запросы граждан. Вы получили задание выявить типы запросов граждан

Последнее время слишком много спамных писем к вам начало приходить (корпоративный

фильтр не справляется), вы устали их читать и складывать в папку "Спам". Решили, что вы сами сделаете программу, которая будет отфильтровывать спам, все, что не смог отфильтровать корпоративный фильтр .

Вы апробировали свой онлайн-курс более чем на 1000 студентах, проанализировали закономерности обучения и задумались над тем, чтобы создать программу, которая сможет дать обратную связь студенту, чтобы он видел - на какую оценку он выходит при завершении данного курса, если так будет продолжать работать.

Высокий уровень

Сформулируйте задачу из вашей предметной области как задачу машинного обучения с учителем. Определите источник данных, критерий успешности решения задачи.

Раздел 2. Решение задач машинного обучения

Базовый уровень.

Укажите все задачи, которые имеет смысл и возможно трактовать как задачи классификации:

Выберите один или несколько ответов:

Запомнить дни рождения и другие данные сотрудников университета

Распознать рукописный текст

Провести группировку учащихся по уровням знаний на основании результатов тестирования, если заранее возможные уровни знаний неизвестны

Реализовать логический вывод с использованием базы знаний

Фильтровать спам, если достаточно большое количество электронных писем уже вручную расклассифицировано на спам и не-спам

Отметьте галочками только те постановки задач, которые можно рассматривать как задачи регрессионного анализа

Выберите один или несколько ответов:

Химический процесс в реакторе протекает во времени. Через каждую минуту измеряется концентрация химического реагента в реакторе. Измерения выполняются с определенной точностью. Требуется построить зависимость концентрации от времени, которая бы старалась исправить ошибки измерения (устранить ошибки).

Имеются данные (Рост, вес) для 1000 студентов 1-го курса ПГТУ. Требуется найти зависимость

Роста первокурсника от его Веса.

Даны значения коэффициентов a, b, c полинома $f(x) = a + b \cdot x + c \cdot x^4$ и значение y . Требуется найти минимум функции $\Phi(x) = (y - f(x))^2$.

Продвинутый уровень.

Построить многочлен вида
 $y(x) = a + b \cdot x^2$,
используя МНК по таблице исходных данных:

X	-1	0	2
Y	-2	3,9	-6

Высокий уровень:

Имеется обучающая выборка:

Xtrain	-3	-2	0	1	3	5
Ytrain	7.437	1.563	-2.437	-0.563	7.437	23.437

И тестовая выборка:

Xtest	-1	2	4
Ytest	-0.563	1.563	13.563

Строится линейная модель

вида $y_m(x) = a_0 + a_1 * x + a_2 * x^2 + a_3 * x^3 + a_4 * \cos(2 * \pi * x)$ с помощью МНК.

Если попытаться построить линейную модель "в лоб", используя нормальную систему уравнений $F^T F * a = F^T Y$ на обучающей выборке, то ничего не получится. Система будет вырожденной.

Необходимо, применяя параметр регуляризации α , перейти к решению регуляризованной системы: $(F^T F + \alpha * Id) * a = F^T Y$. Тогда решение будет существовать.

Определить оптимальное значение параметра α , если в качестве списка исследуемых значений параметра α использовался список: [1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9].

В качестве критерия оптимальности использовался критерий минимума максимальной абсолютной ошибки на ВСЕЙ выборке (включая и обучающую и тестовую): $\max |y_k - y_{mk}|$.

В ответе указать оптимальное значение параметра α

Высокий уровень

В период пандемии важную роль стало играть онлайн-обучение. Для организации учебного процесса в большом количестве стали использоваться онлайн-курсы. Одной из задач построения интеллектуального тьютора онлайн-обучения, является задача прогноза конечного результата обучения на онлайн-курсе: сможет ли студент успешно завершить курс, в зависимости от его результатов работы в первой части курса.

Один из способов - решать задачу классификации и попытаться определить к какому классу следует отнести данного слушателя, к классу успешно завершивших или к противоположному классу? Отнести его к классу тех, кто завершил курс на отлично или на удовлетворительно или не завершил.

При решении этого кейса, вам потребуется обучить модель `KNeighboursClassifier` из пакета `sklearn`. В качестве исходных данных вам будут предоставлены результаты выполнения начальных заданий онлайн-курса слушателями (объектно-признаковая матрица) и их результат окончания курса (столбец "label").

Перечень вопросов для проведения промежуточной аттестации

1. Введение в машинное обучение.

Подход к управлению, основанный на данных или что такое машинное обучение.

Пирамида задач построения систем ИИ. Основные типы задач машинного обучения

Архитектура систем ИИ. Архитектура систем ИИ на основе машинного обучения. Управление большими данными.

Нейронные сети.

Анализ больших данных, используемых в финансово-экономической деятельности предприятия.

Нечеткое представление информации по проекту и эвристическое представление решения задачи финансово-

экономической сферы.

Задачи обучения в проекте по разработке системы поддержки принятия решения в финансово-экономической сфере.

Реализация технологии машинного обучения.

Обучение с учителем. Обучение без учителя. Обучение с подкреплением. Примеры разных типов задач.

Подготовка и очистка данных. Первичный анализ данных. Виды графического анализа данных.

Задачи классификации и анализа.

2. Решение задач машинного обучения.

Постановка задачи регрессионного анализа. Критерий минимизации MSE. Геометрическая интерпретация метода наименьших квадратов. Применение МНК. Критерий точности MAE. Критерий множественной регрессии R².

Основные API в библиотеке scikit-learn для решения задач регрессионного анализа.

Постановка задачи классификации. Функции потерь в задачах классификации. Метрики качества. Матрица (не)соответствия. Accuracy, Precision, Recall. F1-метрика. ROC AUC - метрика.

Методы классификации. Метод k-ближайших соседей (kNN). Достоинства и недостатки метода.

Деревья решений. Логистическая регрессия. Байесовские классификаторы.

Основные API в библиотеке scikit-learn для решения задач регрессионного анализа.