

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол №1

**Рабочая программа дисциплины (модуля)
«Production ML (Машинное обучение в продакшене)»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Машинное обучение

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2024

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	4
3. Тематический план.....	6
4. Содержание дисциплины (модуля).....	6
5. Учебно-методическое обеспечение	8
6. Материально-техническое обеспечение	8
7. Методические и оценочные материалы	10

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Production ML (Машинное обучение в продакшене)» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 02.04.01 Математика и компьютерные науки, профиль Машинное обучение, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Production ML (Машинное обучение в продакшене)» позволяет студентам получить практические навыки, необходимые для успешной интеграции машинного обучения в бизнес-процессы, что критически важно для достижения конкурентных преимуществ. Кроме того, оно способствует пониманию вызовов и решений, связанных с масштабированием и поддержкой ML-моделей в условиях реального времени.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль Машинное обучение и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений.

Дисциплина (модуль) изучается на 2 курсе в 3 семестре, является дисциплиной по выбору.

Цель изучения дисциплины (модуля): освоение студентами методов и практик внедрения, развертывания и поддержки моделей машинного обучения в реальных производственных системах.

Задачи изучения дисциплины (модуля):

- формирование знаний, как обеспечить воспроизводимость в машинном обучении;
- формирование знаний основных этапов жизненного цикла моделей;
- формирование знаний, как устроен процесс управления данными для обучения, включая их лейблинг, версионирование и обеспечение качества;
- формирование знаний, в чем отличие между оффлайн и онлайн моделями и как планировать задачи для их обучения и внедрения;
- формирование знаний, как подготовить ML модели к масштабированию;
- формирование умения организовывать трекинг кода и результатов экспериментов;
- формирование умения работать с пайплайнами обработки данных и обучения моделей;
- формирование умения деплоить модели;
- формирование умения автоматизировать процесс обучения и деплоя модели;
- формирование умения разрабатывать тесты для проверки качества моделей и настраивать мониторинг для контроля их работы;
- формирование навыка внедрения и поддержки работы ML моделей в продакшене;
- формирование навыка работы с MLOps инструментами.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-6.	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1.	Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания
		УК-6.2.	Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки
		УК-6.3.	Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области
		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие

			в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен определять общие формы и закономерности области машинного обучения	ПК-1.1.	Знает основные теоретические концепции и принципы, относящиеся к области машинного обучения, а также ключевые закономерности и модели, которые помогают в анализе и интерпретации данных
		ПК-1.2.	Умеет проводить систематический анализ области разработки, выявлять и формулировать общие закономерности и тенденции, а также применять методы исследования для получения новых знаний и понимания
		ПК-1.3.	Имеет практический опыт работы в области машинного обучения, включая участие в научных проектах, исследованиях или практических заданиях, где были выявлены и описаны общие формы и закономерности
ПК-3.	Способен решать задачи профессиональной деятельности, формулировать результат, увидеть следствия полученного результата	ПК-3.1.	Знает основные принципы и методы решения задач профессиональной деятельности, а также способы формулирования и представления результатов, включая анализ последствий и их значимость в контексте проекта
		ПК-3.2.	Умеет применять математические и компьютерные методы для решения конкретных задач, формулировать четкие и обоснованные результаты, а также анализировать их последствия для дальнейших действий и решений
		ПК-3.3.	Имеет практический опыт в решении профессиональных задач, включая участие в проектах, где были получены результаты и проанализированы их следствия, что способствовало принятию обоснованных решений

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
Лекции	Семинары (практические занятия)					
1	ML в индустрии. Задачи и инструменты ML инженера	2	2		12	Домашние задания
2	Работа с данными в ML	2	2		12	Домашние задания
3	Постановка и трекинг ML экспериментов	2	2		12	Домашние задания
4	Подготовка ML моделей к деплою	2	2		12	Домашние задания
5	Автоматизация обучения и деплоя ML моделей	2	2		12	Домашние задания
6	Тестирование и мониторинг	2	2		10	Домашние задания Тест
7	Презентация проектов и итоги курса	3	3		10	Домашние задания
	<i>Зачет с оценкой</i>			4		Проект
	Итого:	15	15	4	80	
	Объем дисциплины (модуля) (в ак. ч.)	114				
	Объем дисциплины (модуля) (в зач. ед.)	3				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	ML в индустрии. Задачи и инструменты ML инженера	Основные области применения машинного обучения в бизнесе. Роли и обязанности ML инженера. Обзор инструментов и технологий для разработки ML решений
2	Работа с данными в ML	Сбор и очистка данных. Анализ и визуализация данных. Подготовка данных для обучения моделей
3	Постановка и трекинг ML экспериментов	Формулирование гипотез и задач для экспериментов. Использование систем трекинга для управления экспериментами. Оценка результатов и метрик моделей
4	Подготовка ML моделей к деплою	Оптимизация моделей для продакшена. Форматы сохранения и развертывания моделей. Создание API для взаимодействия с моделями
5	Автоматизация обучения и деплоя ML моделей	Использование CI/CD для ML проектов. Инструменты для автоматизации процессов обучения. Управление версиями моделей и данными
6	Тестирование и мониторинг	Методы тестирования ML моделей. Мониторинг производительности и качества моделей в продакшене. Обработка сбоев и управление рисками

7	Презентация проектов и итоги курса	Подготовка и структура финальной презентации. Обсуждение результатов и выводов проекта. Оценка и обратная связь по проектам участников курса
---	------------------------------------	--

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Стружкин, Н. П. Базы данных: проектирование. Практикум : учебник для вузов / Н. П. Стружкин, В. В. Годин. — Москва : Издательство Юрайт, 2025. — 291 с. — (Высшее образование). — ISBN 978-5-534-00739-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561215>.

Дополнительная литература:

1. Дьяконов, А.Г. Машинное обучение и анализ данных / А.Г. Дьяконов. — URL: https://github.com/Dyakonov/MLDM_BOOK/blob/main/README.md.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое

Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Production ML (Машинное обучение в продакшене)» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, практические занятия, домашние задания, тест, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (практическом занятии) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Тест – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины (модуля).

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины (модуля), понимать смысл научных категорий и уметь их использовать в профессиональной лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания студентами важных методологических категорий.

Проект – исследовательская работа по дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Production ML (Машинное обучение в продакшене)»

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета с оценкой*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			задачами.
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	Зачтено	
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Production ML (Машинное обучение в продакшене)» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	50%	Набор задач по темам недели
Проекты	30%	Исследовательская работа по дисциплине (модулю) и презентация результатов
Тест	20%	Набор заданий по теме на проверку знаний

Формула расчёта итоговой оценки по дисциплине (модулю) «Production ML (Машинное обучение в продакшене)»: $\langle 0,5 \times \text{среднее за домашние задания} + 0,3 \times \text{среднее за проекты} + 0,2 \times \text{тест} \rangle$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание: ML в индустрии. Задачи и инструменты ML инженера

1. Опишите три основные области применения машинного обучения в современной индустрии и приведите по одному примеру для каждой.
2. Расскажите о ключевых ролях и обязанностях ML инженера в проекте по разработке ML-системы.
3. Составьте список из пяти популярных инструментов или библиотек, используемых ML инженерами, и кратко опишите назначение каждого.
4. Найдите и проанализируйте статью или кейс о внедрении ML в бизнес-процесс: какие задачи решались, и какую роль играл ML инженер?
5. Опишите, какие навыки и знания необходимы ML инженеру для эффективной работы с большими данными и развертыванием моделей.

Домашнее задание: Работа с данными в ML

1. Соберите небольшой датасет (например, с открытых источников) и опишите процесс его очистки: какие шаги вы предприняли и почему.
2. Проведите базовый анализ данных: рассчитайте основные статистики (среднее, медиану, стандартное отклонение) для выбранных числовых признаков.
3. Постройте визуализации данных (гистограммы, scatter plot, boxplot) и сделайте выводы о распределении и взаимосвязях признаков.
4. Опишите методы обработки пропущенных значений и выбросов, которые вы применили к своему датасету.
5. Подготовьте данные для обучения модели: выполните нормализацию или стандартизацию признаков и объясните выбор метода.

Домашнее задание: Постановка и трекинг ML экспериментов и подготовка моделей к деплою

1. Сформулируйте гипотезу для ML эксперимента на основе выбранного датасета и опишите критерии успеха модели.
2. Опишите, как можно использовать систему трекинга экспериментов (например, MLflow или Weights & Biases) для управления экспериментами.
3. Проведите эксперимент по обучению модели с разными параметрами и зафиксируйте результаты (метрики качества). Сделайте выводы о влиянии параметров на качество.
4. Опишите основные шаги по оптимизации модели для продакшена (например, уменьшение размера, ускорение инференса).
5. Создайте простой REST API (например, с использованием Flask или FastAPI) для развертывания обученной модели и опишите процесс её вызова.

Примерные задания для теста

Тест по методам тестирования ML моделей

- 1. Какой из следующих методов используется для оценки производительности ML модели?**
 - A) A/B тестирование
 - B) Кросс-валидация
 - C) Гиперпараметрическая оптимизация
 - D) Все вышеперечисленные
- 2. Что такое "drift" в контексте мониторинга ML моделей?**
 - A) Изменение структуры данных
 - B) Изменение распределения входных данных
 - C) Падение производительности модели
 - D) Все вышеперечисленные
- 3. Какой из следующих метрик не является метрикой качества для классификационных моделей?**
 - A) Точность (Accuracy)
 - B) Полнота (Recall)
 - C) Среднеквадратичная ошибка (MSE)
 - D) F1-меры
- 4. Какой подход используется для мониторинга производительности модели в реальном времени?**
 - A) Логирование
 - B) Статистический контроль
 - C) A/B тестирование
 - D) Все вышеперечисленные
- 5. Что такое "обработка сбоев" в ML?**
 - A) Устранение ошибок в коде
 - B) Реакция на сбои в работе модели
 - C) Оптимизация модели
 - D) Создание резервной копии данных
- 6. Какой инструмент часто используется для мониторинга ML моделей в продакшене?**
 - A) TensorBoard
 - B) Jupyter Notebook
 - C) Microsoft Excel
 - D) PowerPoint
- 7. Какой метод позволяет оценить влияние изменений в модели на ее производительность?**
 - A) A/B тестирование

- B) Кросс-валидация
- C) Bootstrap
- D) Монте-Карло

8. Какой из следующих вариантов является примером риска, связанного с ML моделями?

- A) Падение производительности
- B) Неправильная интерпретация результатов
- C) Изменение входных данных
- D) Все вышеперечисленные

9. Какой метод может использоваться для автоматического обнаружения "дрифта" в данных?

- A) Упрощение модели
- B) Статистический тест
- C) Увеличение объема данных
- D) Снижение сложности модели

10. Что такое "реакция на сбой" в контексте ML?

- A) Предотвращение ошибок в коде
- B) Автоматическая корректировка модели при сбоях
- C) Восстановление модели после сбоя
- D) Все вышеперечисленные

Примерное описание к проекту

Задание для проекта: "Полный цикл разработки ML модели"

Описание задания

В рамках данного проекта вам предстоит пройти полный цикл разработки ML модели, начиная с работы с данными и заканчивая деплоем и автоматизацией процессов. Проект включает в себя несколько этапов, каждый из которых будет оцениваться по определенным критериям.

Этапы подготовки проекта

1. Сбор и очистка данных

- **Сбор данных:** Найдите и соберите набор данных, подходящий для вашей задачи (например, из открытых источников, таких как Kaggle, UCI Machine Learning Repository и т.д.).
- **Очистка данных:** Обработайте пропущенные значения, выбросы и дублирующиеся записи. Опишите методы, которые вы использовали.
- **Документация:** Подготовьте отчет о процессе сбора и очистки данных.

2. Анализ и визуализация данных

- **Анализ данных:** Проведите разведочный анализ данных (EDA). Рассчитайте основные статистики и выявите взаимосвязи между признаками.
- **Визуализация:** Постройте графики и диаграммы для визуализации данных и их распределений (гистограммы, boxplot, scatter plot и т.д.).
- **Документация:** Создайте отчет с визуализациями и выводами.

3. Постановка и трекинг ML экспериментов

- **Формулирование гипотез:** Определите гипотезы и задачи, которые вы хотите проверить с помощью модели.
- **Трекинг экспериментов:** Используйте систему трекинга (например, MLflow или Weights & Biases) для управления экспериментами и хранения результатов.
- **Документация:** Подготовьте отчет о ваших гипотезах и результатах экспериментов.

4. Подготовка ML моделей к деплою

- **Оптимизация моделей:** Обучите несколько моделей и оптимизируйте их для продакшена (например, с использованием методов регуляризации или уменьшения размерности).
- **Сохранение и развертывание:** Сохраните модели в подходящих форматах (например, pickle, ONNX) и подготовьте API для их развертывания (например, с использованием Flask или FastAPI).
- **Документация:** Создайте отчет о процессе подготовки моделей.

5. Автоматизация обучения и деплоя ML моделей

- **CI/CD:** Настройте CI/CD для автоматизации процессов обучения и деплоя моделей (например, с использованием GitHub Actions или Jenkins).
- **Управление версиями:** Организуйте систему управления версиями моделей и данных (например, DVC).
- **Документация:** Подготовьте финальный отчет о процессе автоматизации.

Защита проекта

- **Формат защиты:** Презентация проекта (15-20 минут) с демонстрацией работы модели и API.
- **Структура презентации:**
 - Введение и цели проекта.
 - Процесс сбора и очистки данных.
 - Результаты анализа и визуализации данных.
 - Гипотезы и результаты экспериментов.
 - Подготовка и оптимизация моделей.
 - Демонстрация API и процесса автоматизации.
 - Заключение и выводы.

Критерии оценивания

1. **Качество данных (20%):** Полнота и корректность собранных данных, качество очистки.
2. **Анализ и визуализация (20%):** Глубина анализа, качество визуализаций и выводов.
3. **Постановка экспериментов (20%):** Четкость формулировки гипотез, использование системы трекинга.
4. **Подготовка моделей (20%):** Оптимизация моделей, качество API, правильность форматов сохранения.

5. **Автоматизация процессов (20%):** Эффективность CI/CD, управление версиями, документация.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Какой из следующих инструментов чаще всего используется для разработки ML моделей? A) Excel B) TensorFlow C) Microsoft Word D) Notepad	В	ОПК-1
2.	Какова основная роль ML инженера? A) Разработка веб-сайтов B) Оптимизация баз данных C) Создание и внедрение ML моделей D) Администрирование серверов	С	ОПК-1
3.	Какое из следующих направлений не является областью применения машинного обучения в бизнесе? A) Обработка естественного языка B) Анализ финансовых данных C) Уборка помещений D) Рекомендательные системы	С	ПК-1
4.	Какой метод используется для очистки данных от выбросов? A) Нормализация B) Удаление дубликатов C) Стандартизация D) Логарифмирование	В	ПК-3
5.	Какой из следующих инструментов используется для визуализации данных? A) NumPy B) Pandas C) Matplotlib D) Scikit-learn	С	ПК-3
6.	Назовите одну из основных областей применения машинного обучения в бизнесе.	Рекомендательные системы	УК-6
7.	Как называется роль специалиста, который отвечает за внедрение ML моделей в продакшен?	ML инженер	УК-6
8.	Какой процесс включает в себя удаление ненужных или ошибочных данных?	Очистка данных	ОПК-1
9.	Какой тип графика часто используется для визуализации распределений данных?	Гистограмма	ОПК-1

10.	Какой метод используется для преобразования категориальных данных в числовые?	Кодирование	ОПК-1
11.	Как называется система, используемая для управления экспериментами в ML?	Система трекинга	ПК-1
12.	Какой метрикой часто оценивается качество классификационной модели?	Точность/accuracy	ПК-1
13.	Какой формат часто используется для сохранения обученных моделей?	Pickle	ПК-3
14.	Какой фреймворк часто используется для создания API для ML моделей?	Flask	ПК-3
15.	Как называется процесс автоматизации развертывания и обучения моделей?	CI/CD	ПК-3