ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

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: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика»*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

Москва
2024

Электронный документ

Содержание

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 – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 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 бакалавриата по направлению 02.03.01 Математика и компьютерные науки, профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений как дисциплина по выбору.

Дисциплина (модуль) изучается на 3 или 4 курсе в 5, 6 или 7 семестре на выбор. Доступна к изучению после успешного освоения дисциплин (модулей): «Machine Learning (Машинное обучение)», «Основы промышленной разработки».

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

Задачи изучения дисциплины (модуля):

— обеспечение воспроизводимости в машинном обучении через фиксацию параметров и документацию, управление жизненным циклом моделей;

— подготовка ML-моделей к масштабированию;

— внедрения и поддержки работы ML моделей в продакшене.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— как обеспечить воспроизводимость в машинном обучении;

— основные этапы жизненного цикла моделей;

— как устроен процесс управления данными для обучения, включая их лейблинг, версионирование и обеспечение качества;

— отличие между оффлайн и онлайн моделями и как планировать задачи для их обучения и внедрения;

уметь:

— подготовить ML модели к масштабированию;

— организовывать трекинг кода и результатов экспериментов;

— работать с пайплайнами обработки данных и обучения моделей;

— деплоить модели;

— автоматизировать процесс обучения и деплоя модели;

— разрабатывать тесты для проверки качества моделей и настраивать мониторинг для контроля их работы;

владеть:

— навыком внедрения и поддержки работы ML моделей в продакшене;

— навыком 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-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
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 действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и	ОПК-1.1.	Знает основные концепции и теории в области математического анализа и смежных дисциплин; методы и подходы, используемые в различных областях математики
		ОПК-1.2.	Умеет применять математические методы для решения профессиональных задач
		ОПК-1.3.	Имеет практический опыт разработки и реализации математических моделей в профессиональной деятельности

	математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности		
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области разработки, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Контактная работа		Контроль	Самостоятельная работа	
		Лекции	Семинар			
1	ML в индустрии. Базовые инструменты ML инженера	4	4		18	Домашние задания
2	Работа с данными в ML и трекинг ML экспериментов	4	4		18	Домашние задания
3	Подготовка ML моделей к деплою	4	4		18	Домашние задания
4	Тестирование и мониторинг	4	4		18	Домашние задания
5	Автоматизация обучения и деплоя ML моделей	4	4		18	Домашние задания
6	LLMOps	4	4		20	Домашние задания Стресс-тест
7	Контейнеризация ML-приложений	4	4		20	Домашние задания
	Зачет с оценкой			4		Проект
Итого:		***28***	***28***	***4***	***130***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	ML в индустрии. Базовые инструменты ML инженера	Введение. Инструменты разработки (часть 1). Инструменты разработки (часть 2)
2	Работа с данными в ML и трекинг ML экспериментов	Работа с данными в ML. Трекинг ML-экспериментов и воспроизводимость
3	Подготовка ML моделей к деплою	Деплой ML-моделей
4	Тестирование и мониторинг	Тестирование и валидация кода. Мониторинг (часть 1): стандартные инструменты. Мониторинг (часть 2): ML-специфичные инструменты
5	Автоматизация обучения и деплоя ML моделей	Использование CI/CD для ML проектов. Автоматизация пайплайна обучения моделей
6	LLMOps	LLMOps
7	Контейнеризация ML-приложений	Контейнеризация ML-приложений

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. Лакшманан, В. Машинное обучение. Паттерны проектирования : практическое пособие / В. Лакшманан, С. Робинсон, М. Мунн. - Санкт-Петербург : БХВ-Петербург, 2022. - 448 с. - ISBN 978-5-9775-6797-8. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2140204.

2. Григорьев, А. Машинное обучение. Портфолио реальных проектов : практическое руководство / А. Григорьев. - Санкт-Петербург : Питер, 2023. - 496 с. - (Серия «Библиотека программиста»). - ISBN 978-5-4461-1978-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/2123375.

3. Кацов, И. Машинное обучение для бизнеса и маркетинга : практическое руководство / И. Кацов. - Санкт-Петербург : Питер, 2019. - 512 с. - (Серия «IT для бизнеса»). - ISBN 978-5-4461-0926-5. - Текст : электронный. - URL: https://znanium.com/catalog/product/1783938.

4. Гифт, Н. Прагматичный ИИ. Машинное обучение и облачные технологии : практическое руководство / Н. Гифт. - Санкт-Петербург : Питер, 2019. - 304 с. - (Серия «Для профессионалов»). - ISBN 978-5-4461-1061-2. - Текст : электронный. - URL: https://znanium.com/catalog/product/1760806.

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.

2. 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.

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;
— механическими калькуляторами;
— 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 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, семинары, домашние задания, стресс-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 удается, 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 курсу 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.

Стресс-тест — это способ проверки системы, как она ведёт себя в экстремальных условиях, которые значительно превышают обычную нагрузку или выходят за рамки обучающих сценариев.

Для подготовки к стресс-тесту начните с определения сценариев тестирования, включая крайние случаи (например, шумные данные, перегрузки или adversarial атаки), и соберите соответствующие тестовые наборы данных. Разработайте метрики оценки, такие как точность, время отклика и устойчивость к ошибкам, и настройте инфраструктуру для симуляции условий стресса, используя инструменты вроде TensorFlow Model Analysis или PyTorch. Проведите предварительные тесты на меньших масштабах, документируйте результаты и итеративно улучшайте модель, чтобы минимизировать риски перед полноценным развертыванием. Наконец, организуйте команду для анализа результатов и планирования улучшений, обеспечивая соответствие этическим и регуляторным стандартам.

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. 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 задачами.
9	Отлично	Зачтено	
8	Отлично	Зачтено	
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, но испытывает трудности при самостоятельных ответах и использует
4	Удовлетворительно	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			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.
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	60%	Набор задач по темам недели
Проект	30%	Исследовательская работа по дисциплине (модулю) и презентация результатов
Стресс-тест	10%	Набор заданий по теме на проверку знаний

Формула расчёта итоговой оценки по дисциплине (модулю) «Production ML (Машинное обучение в продакшене)»: «0,6× среднее за домашние задания + 0,3 × среднее за проекты + 0,20 × за стресс-тест».

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

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 стресс-теста

Задание для стресс-теста по теме: "Подготовка ML моделей к деплою. Деплой ML-моделей"

Цель задания: разработать и провести стресс-тест для модели машинного обучения (например, простой классификатор, такой как логистическая регрессия или нейронная сеть для задачи классификации изображений), чтобы оценить ее готовность к масштабируемому деплою в продакшене. Стресс-тест должен симулировать экстремальные условия, выявить уязвимости и подготовить модель к устойчивой работе.

Шаги выполнения:

1. **Выберите модель и подготовьте данные:** Возьмите готовую модель (например, из scikit-learn или TensorFlow) и тестовый датасет (например, MNIST или CIFAR-10). Подготовьте пайплайн предобработки данных, включая нормализацию и аугментацию, и обучите модель на baseline-наборе данных.

2. **Определите сценарии стресс-теста:** Создайте набор стресс-сценариев, таких как:

   — Перегрузка: Обработка большого объема запросов (например, 10x увеличение нагрузки) с измерением latency и throughput.
   — Аномальные данные: Тестирование на зашумленных, искаженных или adversarial примерах (например, добавьте шум к изображениям или используйте FGSM-атаки).
   — Сбои инфраструктуры: Симуляция потери соединения, ограничений памяти/CPU или внезапного увеличения размера входных данных.

3. **Настройте инструменты для тестирования:** Используйте фреймворки вроде Locust для нагрузочного тестирования, TensorFlow Model Analysis (TFMA) или PyTorch для анализа метрик (точность, recall, latency). Интегрируйте модель в контейнер (Docker) для изолированного тестирования.

4. **Проведите тест и проанализируйте результаты:** Запустите стресс-тест, соберите метрики (например, время отклика, ошибки, ресурсное потребление) и сравните с baseline. Выявите слабые места (например, переобучение на аномалиях) и предложите улучшения, такие как оптимизация модели или добавление fallback-механизмов.

5. **Документируйте и предложите план деплоя:** Опишите результаты теста, риски и рекомендации по масштабированию (например, использование Kubernetes для оркестрации). Подготовьте модель к деплою с учетом стресс-теста, например, через A/B-тестирование или canary-релизы.

Критерии оценки: Задание считается выполненным, если стресс-тест охватывает не менее 3 сценариев, метрики измерены объективно, а выводы включают практические рекомендации по улучшению модели для продакшена. Время на выполнение: 2 пары. Рекомендуемые инструменты: Python, Docker, Kubernetes (опционально).

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	B	УК-1
2.	Какой метод используется для очистки данных от выбросов? A) Нормализация B) Удаление дубликатов C) Стандартизация D) Логарифмирование	B	ПК-1
3.	Какой из следующих инструментов используется для визуализации данных? A) NumPy B) Pandas C) Matplotlib D) Scikit-learn	C	УК-2
4.	Назовите одну из основных областей применения машинного обучения в бизнесе.	Рекомендательные системы	УК-1
5.	Как называется роль специалиста, который отвечает за внедрение ML моделей в продакшен?	ML инженер	УК-2
6.	Какой процесс включает в себя удаление ненужных или ошибочных данных?	Очистка данных	ОПК-1
7.	Какой тип графика часто используется для визуализации распределений данных?	Гистограмма	ОПК-1
8.	Какой метрикой часто оценивается качество классификационной модели?	Точность/accuracy	ПК-1
9.	Какой формат часто используется для сохранения обученных моделей?	Pickle	ПК-2
10.	Какой фреймворк часто используется для создания API для ML моделей?	Flask	ПК-2
11.	Какое уравнение описывает обновление Q-функции в Q-learning?	Q-update	УК-1
12.	Какой алгоритм использует политику для прямого обучения в RL?	Policy gradient	УК-2
13.	Какое значение gamma в RL обозначает горизонт планирования?	discount factor	ОПК-1
14.	Какой метод используется для оценки политики в RL?	Monte Carlo	ПК-1
15.	Какой инструмент применяется для симуляции сред в RL?	OpenAI Gym	ПК-2