

**УТВЕРЖДЕНА**

Приказом Ректора АНО ВО  
«Центральный университет»  
Е.В. Ивашкевич  
от «26» июня 2025 г. № 0626.32

**Рабочая программа дисциплины (модуля)  
«Production ML (Машинное обучение в продакшене)»  
дополнительной профессиональной программы – программы  
профессиональной переподготовки «Академия data science»**

**Траектория: Машинное обучение**

**Москва  
2025**

## Содержание

<b>1. Краткая характеристика дисциплины (модуля) .....</b>	<b>3</b>
<b>2. Тематический план .....</b>	<b>4</b>
<b>3. Содержание дисциплины (модуля) .....</b>	<b>4</b>
<b>4. Учебно-методическое обеспечение .....</b>	<b>5</b>
<b>5. Материально-техническое обеспечение .....</b>	<b>5</b>
<b>6. Методические и оценочные материалы .....</b>	<b>7</b>

## 1. Краткая характеристика дисциплины (модуля)

Изучение дисциплины (модуля) «Production ML (Машинное обучение в продакшене)» позволяет слушателям 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.

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

### **Задачи изучения дисциплины (модуля):**

- научиться обеспечивать повторяемость результатов экспериментов в машинном обучении через систематизацию процессов;
- освоить последовательность стадий создания и эксплуатации моделей машинного обучения в реальных условиях;
- разработать навыки организации хранения и контроля версий данных для эффективного обучения моделей;
- изучить различия в подходах к обучению и развертыванию моделей в оффлайн и онлайн режимах;
- приобрести умения по адаптации моделей машинного обучения для работы в масштабируемых системах.

### **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 версионирование и обеспечение качества;
- отличия между оффлайн и онлайн моделями и как планировать задачи для их обучения и внедрения;
- как подготовить ML модели к масштабированию.

#### **уметь:**

- 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 выстраивания MLOps процесса (внедрение и поддержка работы ML моделей в продакшене);
- навыком работы с MLOps инструментам.

## 2. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		<i>Очная форма</i>				
		Аудиторная работа		Контроль	Самостоя тельная работа	
Лекции	Семинары (практичес кие занятия)					
1	ML в индустрии. Базовые инструменты ML инженера	4	4		13	Домашние задания
2	Работа с данными в ML и трекинг ML экспериментов	4	5		13	Домашние задания
3	Подготовка ML моделей к деплою	5	5		13	Домашние задания
4	Тестирование и мониторинг	5	5		13	Домашние задания
5	Автоматизация обучения и деплоя ML моделей	5	5		14	Домашние задания
6	LLMOps	5	5		14	Домашние задания Стресс-тест
7	Контейнеризация ML- приложений	5	5		14	Домашние задания
	<i>Зачет с оценкой</i>			4		Проект
	<b>Итого:</b>	<b>33</b>	<b>34</b>	<b>4</b>	<b>94</b>	
	<b>Объем дисциплины (модуля) (в ак. ч.)</b>	<b>165</b>				

## 3. 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. Часть 1 Деплой ML-моделей. Часть 2
4	Тестирование и мониторинг	Тестирование и валидация кода Мониторинг. Часть 1: стандартные инструменты Предзащита проекта Мониторинг. Часть 2: ML-специфичные инструменты
5	Автоматизация обучения и деплоя ML моделей	Автоматизация пайплайна обучения моделей CI/CD для ML
6	LLMOps	LLMOps
7	Контейнеризация ML- приложений	Контейнеризация ML-приложений

#### 4. Учебно-методическое обеспечение

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

Каждый слушатель 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.

Слушателям 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](https://github.com/Dyakonov/MLDM_BOOK/blob/main/README.md).

#### 5. 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	<a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a>
2.	База данных для IT-специалистов	<a href="https://habr.com">https://habr.com</a>
3.	База данных ScienceDirect	<a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	<a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>
5.	Федеральный портал «Российское образование»	<a href="https://www.edu.ru/">https://www.edu.ru/</a>
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	<a href="http://window.edu.ru/">http://window.edu.ru/</a>
7.	Единая коллекция цифровых образовательных ресурсов	<a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a>
8.	Федеральный центр информационно - образовательных ресурсов	<a href="http://fcior.edu.ru/">http://fcior.edu.ru/</a>

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
<b>Операционные системы:</b>		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
<b>Браузеры:</b>		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
<b>Офисные приложения:</b>		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
<b>Программное обеспечение для планирования и учета времени:</b>		
Toggle app	зарубежное	свободно распространяемое
<b>Системы управления проектами:</b>		
Microsoft Imagine (Project)	зарубежное	лицензионное
<b>Системы управления базами данных:</b>		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
<b>Системы резервного копирования (backup):</b>		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
<b>Справочно-правовые системы:</b>		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
<b>Средства антивирусной защиты:</b>		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
<b>Среды разработки:</b>		
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
<b>Пакеты программных средств и библиотек:</b>		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
<b>Системы управления библиографической информацией:</b>		
Zotero	зарубежное	свободно распространяемое
<b>Сервисы и службы:</b>		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

## 6. 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 слушателя на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре слушателям 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.

*Стресс-тест* — это способ проверить, как система ведёт себя в экстремальных условиях, которые значительно превышают обычную нагрузку или выходят за рамки обучающих сценариев

*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 слушателей, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы слушатели взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи слушателя 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)**

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**.

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

<b>Десятибалльная оценка</b>	<b>Пятибалльная оценка</b>	<b>Оценка за зачет</b>	<b>Общая характеристика результата обучения по дисциплине (модулю)</b>
10	Отлично	Зачтено	Слушатель 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. Слушатель 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	Слушатель 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
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			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. Слушатель хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
5	Удовлетворительно	Зачтено	Слушатель 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. Слушатель способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Слушатель 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
Стресс-тест	10%	Способ проверки, как система ведёт себя в экстремальных условиях
Зачет с оценкой	30%	Проект – исследовательская работа по дисциплине (модулю) и презентация результатов

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

**Формула расчёта итоговой оценки по дисциплине (модулю) «Production ML (Машинное обучение в продакшене)»:** « $0,6 \times$  среднее за домашние задания +  $0,1 \times$  тест +  $0,3 \times$  зачет с оценкой».

**Текущий контроль успеваемости обучающихся по дисциплине (модулю)**

**Примерные домашние задания**

## Домашнее задание 1.

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 2.

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 3.

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.

## Примерное описание задания к стресс-тесту

### Описание:

Стресс-тест моделирует реальные условия работы ML-инженера в индустрии, где системы подвергаются экстремальным нагрузкам, значительно превышающим обычные обучающие сценарии. Вы должны интегрировать знания из всех тем курса для создания, развертывания и мониторинга полнофункционального ML-пайплайна. Задание

выполняется индивидуально или в небольшой команде (2–3 человека) в ограниченное время. "Стресс" заключается в следующих экстремальных условиях:

- **Высокая нагрузка данных:** Работа с датасетом в 10–100 млн записей (например, симуляция big data), требующим оптимизации обработки.
- **Ограниченные ресурсы:** Ограниченный доступ к вычислительным мощностям (например, только CPU без GPU, или лимит на память/диск).
- **Неожиданные сбои:** Симуляция реальных проблем, таких как внезапные изменения в данных (drift), сетевые сбои, или атаки (например, DDoS-подобная нагрузка на API).
- **Временное давление:** Короткие дедлайны с промежуточными чекпоинтами, где система должна выдерживать пиковые нагрузки (например, 1000+ одновременных запросов).
- **Выход за рамки обучения:** Задание требует творческого применения знаний, без пошаговых инструкций — вы должны адаптироваться к непредвиденным ситуациям, как в реальном продакшене.

#### **Цели стресс-теста:**

- Проверить устойчивость вашего ML-пайплайна к экстремальным условиям.
- Оценить интеграцию навыков: от работы с данными и трекинга до контейнеризации, автоматизации и мониторинга.
- Развить навыки быстрого реагирования на сбои, оптимизации и масштабирования.

#### **Требуемые инструменты и технологии:**

- Базовые инструменты ML-инженера: Python, Jupyter, Git, Docker и т.д.
- Работа с данными и трекинг: Pandas, DVC, MLflow для обработки больших объемов и воспроизводимости.
- Подготовка к деплою: FastAPI/Flask для API, ONNX/TensorFlow Serving для моделей.
- Тестирование и мониторинг: Pytest для unit-тестов, Prometheus/Grafana для стандартного мониторинга, Evidently или custom метрики для ML-специфичного (drift, accuracy).
- Автоматизация: GitHub Actions/Jenkins для CI/CD пайплайнов обучения и деплоя.
- LLMops: Если модель включает LLM (например, для генерации текста), интегрировать инструменты вроде Hugging Face Transformers с MLOps.
- Контейнеризация: Docker/Kubernetes для упаковки и оркестрации приложений.

#### **Шаги задания (выполняются последовательно с симуляцией стресса):**

##### **1. Подготовка данных и модели:**

- Загрузите и обработайте экстремально большой датасет (например, синтетический набор с 50 млн записей для задачи классификации или регрессии). Оптимизируйте ETL-процесс (Pandas/Dask) для обработки в условиях ограниченной памяти.
- Обучите модель (например, XGBoost или нейросеть) с трекингом экспериментов в MLflow. Симулируйте "стресс": внезапно измените данные (добавьте шум или drift), и адаптируйте модель без полной переобучения.

##### **2. Деплой модели:**

- Подготовьте модель к продакшену: сериализуйте, оптимизируйте (квантизация) и упакуйте в контейнер (Docker). Разверните API (FastAPI) в Kubernetes-подобной среде (или локально с симуляцией кластера).
- Симулируйте стресс: Ограничьте ресурсы (например, 1 CPU core, 2GB RAM) и протестируйте на пиковой нагрузке (1000 запросов/сек с помощью Locust или JMeter).

##### **3. Тестирование и мониторинг:**

- Напишите тесты (Pytest) для кода и модели (unit, integration). Настройте мониторинг: стандартные метрики (latency, throughput) в Grafana/Prometheus, плюс ML-специфичные (data drift, model performance) с Evidently.

- Симулируйте сбои: Введите искусственные ошибки (например, сетевой таймаут или bad input), и обеспечьте, чтобы система автоматически восстанавливалась (алерты, retries).

#### 4. Автоматизация и LLMOps:

- Автоматизируйте пайплайн: Создайте CI/CD (GitHub Actions) для обучения, тестирования и деплоя. Если модель включает LLM (например, для постобработки), интегрируйте LLMOps (Hugging Face + custom pipelines).

- Симулируйте стресс: Запустите пайплайн под высокой нагрузкой, с внезапными изменениями кода (merge conflicts) или данными, требующими быстрой итерации.

#### 5. Финализация и стресс-оценка:

- Демонстрируйте систему в действии: Покажите, как она выдерживает экстремальные условия (например, 10-кратный рост нагрузки). Подготовьте отчет с метриками (uptime, ассурасу под стрессом) и анализом сбоев.

- Время на исправления: Если система "ломается" (например, OOM error), у вас есть 30–60 минут на фикс.

### 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.