

**УТВЕРЖДЕНА**

Решением Ученого совета  
АНО ВО «Центральный университет»  
«24» июня 2025 г.  
Протокол № 2

**Рабочая программа дисциплины (модуля)  
«Machine Learning (Машинное обучение)»**

Направление подготовки: 38.04.05 Бизнес-информатика

Направленность (профиль) подготовки: Продуктовый менеджмент

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2025

**Москва  
2025**

## Содержание

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

## 1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Machine Learning (Машинное обучение)» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 38.04.05 Бизнес-информатика, профиль Продуктовый менеджмент, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 990 от 12.08.2020 года.

Изучение дисциплины (модуля) Machine Learning (Машинное обучение) обеспечивает формирование системного понимания методов и инструментов анализа данных, необходимых для разработки, оценки и внедрения ML-решений в прикладных и бизнес-контекстах. Освоение дисциплины позволяет научиться переводить бизнес-задачи в технические модели, обоснованно выбирать и интерпретировать алгоритмы машинного обучения и эффективно работать в рамках полного жизненного цикла ML-проекта.

### Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 38.04.05 Бизнес-информатика, профиль Продуктовый менеджмент и входит в обязательную часть Блока 1.

Дисциплина (модуль) изучается на 1 курсе во 2 семестре, доступна для прохождения при условии успешного завершения дисциплин (модулей) «Python для анализа данных», «Основы статистики».

**Цель изучения дисциплины (модуля):** формирование теоретических знаний и практических навыков применения методов машинного обучения для решения прикладных и бизнес-задач на основе данных на всех этапах жизненного цикла ML-решения.

### Задачи изучения дисциплины (модуля):

- освоить базовые понятия, типы и принципы машинного обучения, а также области применимости основных алгоритмов и моделей;
- научиться работать с данными для задач машинного обучения, включая сбор, очистку, предобработку, анализ и визуализацию данных с использованием Python и профильных библиотек;
- сформировать умение выбирать, обучать и настраивать модели машинного обучения, применяя методы подбора гиперпараметров и кросс-валидации;
- освоить методы оценки качества и интерпретации результатов моделей, а также критический анализ полученных метрик и выводов;
- развить навыки перевода бизнес-задач в технические решения машинного обучения и эффективного взаимодействия с командой в рамках ML lifecycle.

### В результате освоения дисциплины (модуля) обучающийся должен:

#### *знать:*

- основы теории машинного обучения: ключевые понятия и термины (например, обучающая и тестовая выборки, параметры модели, feature engineering), различные типы машинного обучения (обучение с учителем, обучение без учителя) и их особенностей;
- алгоритмы и модели: основные алгоритмы машинного обучения (линейные модели, деревья решений, ансамбли моделей, нейронные сети и др.); базовые принципы работы алгоритмов и их применимости в различных задачах машинного обучения;
- методы оценки моделей: методы оценки производительности моделей (accuracy, precision, recall, MAE, MSE и др.); принципы кросс-валидации и настроек гиперпараметров;

- инструменты и библиотеки: популярные библиотеки для машинного обучения (например, Scikit-Learn); основы работы с языком программирования Python и его библиотеками для работы с данными (NumPy, Pandas, Matplotlib);
- особенности бизнес-процесса разработки ML решения - ML lifecycle.

***уметь:***

- работать с данными: собирать, очищать и преобразовывать данные для машинного обучения, визуализировать данные для выявления закономерностей и аномалий;
- строить модели: выбирать и применять подходящие алгоритмы машинного обучения для решения конкретных задач, настраивать гиперпараметры моделей;
- анализировать результаты: интерпретировать результаты работы моделей и делать выводы на основе полученных данных;
- переводить бизнес-задачу в техническое представление, общаться с командой разработки, используя техническую терминологию и контекст.

***владеть:***

- навыками программирования: базовое владение языком программирования Python и умение работать с библиотеками для анализа данных и машинного обучения;
- навыками критического мышления: критически анализировать результаты и обосновывать выбор алгоритмов и методов, используемых в проекте;
- навыками постоянного обучения: навык самообразования (следить за трендами в сфере ИИ - как с бизнес, так и с технической точки зрения);
- навыками продуктового видения: навык перевода бизнес необходимости в технические требования и обратно.

## 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)
ОПК-1.	Способен разрабатывать стратегию развития информационных технологий инфраструктуры предприятия и управлять ее реализацией	ОПК-1.1.	Знает принципы стратегического планирования в области информационных технологий, архитектуру ИТ-инфраструктуры и современные технологические тренды
		ОПК-1.2.	Умеет разрабатывать и обосновывать стратегию развития ИТ-инфраструктуры предприятия с учетом бизнес-целей, рисков и ресурсов
		ОПК-1.3.	Имеет практический опыт управления реализацией ИТ-стратегии, включая координацию проектов, взаимодействие с заинтересованными сторонами и оценку эффективности внедрённых решений
ОПК-3.	Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта	ОПК-3.1.	Знает современные методы анализа данных, прогнозирования и поддержки принятия решений, включая технологии машинного обучения и искусственного интеллекта
		ОПК-3.2.	Умеет применять программные инструменты и интеллектуальные системы для сбора, обработки и интерпретации данных в целях стратегического планирования и управления продуктом
		ОПК-3.3.	Имеет практический опыт применения аналитических инструментов и технологий искусственного интеллекта для принятия обоснованных стратегических решений и прогнозирования развития бизнес-продуктов
ОПК-5.	Способен проводить исследования, организовывать самостоятельную и коллективную научно-исследовательскую, проектную и учебно-профессиональную деятельность для поиска, выработки и применения	ОПК-5.1.	Знает методы научных исследований и проектной деятельности, применяемые для анализа и разработки инновационных решений в сфере ИКТ
		ОПК-5.2.	Умеет планировать и организовывать исследовательскую и проектную работу как индивидуально, так и в

	новых решений в области информационно-коммуникационных технологий		составе команды, с учётом целей и задач цифровой трансформации
		ОПК-5.3.	Имеет практический опыт участия в научно-исследовательских и прикладных проектах, направленных на создание и внедрение новых ИКТ-продуктов и сервисов, включая презентацию результатов
ПК-2.	Способен разрабатывать и внедрять цифровые продукты с учетом принципов устойчивого развития и цифровой инклюзивности	ПК-2.1.	Знает принципы устойчивого развития и цифровой инклюзивности, а также их применение в процессе создания цифровых продуктов
		ПК-2.2.	Умеет проектировать цифровые решения, учитывающие экологические, социальные и этические аспекты, включая доступность для различных категорий пользователей
		ПК-2.3.	Имеет практический опыт внедрения цифровых продуктов с соблюдением стандартов инклюзивного дизайна и оценки их воздействия на устойчивое развитие

### 3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контр оль	Самосто ятельна я работа	
Лекции	Семинары (Практичес кие занятия)					
1	Введение	4	4		16	Домашние задания
2	Регрессия	4	4		16	Домашние задания
3	Классификация	4	4		16	Домашние задания
4	Кластеризация	4	4		16	Проект
5	Нейронные сети	2	2		9	Домашние задания, Проект
6	Изображения	2	2		9	Проект
7	Звук	2	2		9	Домашние задания, Проект
8	Текст	2	2		9	Проект
9	Генерация	2	2		8	Домашние задания, Проект
10	Рекомендательные системы	2	2		8	Проект
11	Заключение	2	2		8	Домашние задания
	<i>Экзамен</i>				<i>6</i>	
	<i>Итого:</i>	<i>30</i>	<i>30</i>		<i>124</i>	
	<i>Объем дисциплины (модуля) (в ак. ч.)</i>	<i>190</i>				
	<i>Объем дисциплины (модуля) (в зач. ед.)</i>	<i>5</i>				

### 4. Содержание дисциплины (модуля)

№ п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Введение	Введение. Machine Learning lifecycle Введение. Основы оптимизации Machine Learning алгоритмов
2	Регрессия	Знакомство с задачей Регрессии Построение ML пайплайна для прототипирования решения задачи Регрессии
3	Классификация	Знакомство с задачей Классификации Построение ML пайплайна для прототипирования решения задачи Классификации
4	Кластеризация	Знакомство с задачей Кластеризации Построение ML пайплайна для прототипирования решения задачи Кластеризации
5	Нейронные сети	Введение. Нейронные сети
6	Изображения	Работа с изображениями
7	Звук	Работа со звуком
8	Текст	Работа с текстом
9	Генерация	Задача генерации
10	Рекомендательные системы	Построение рекомендательных систем
11	Заключение	Заключение. ИИ разработка в реальной жизни

## 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. Бринк, Х. Машинное обучение : практическое руководство / Х. Бринк, Д. Ричардс, М. Феверолф. - Санкт-Петербург : Питер, 2018. - 336 с. - (Серия «Библиотека программиста»). - ISBN 978-5-496-02989-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1766396>.

2. 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>.

### ***Дополнительная литература:***

1. Коэльо, Л. Построение систем машинного обучения на языке Python : практическое руководство / Л. Коэльо, В. Ричарт ; пер. с англ. А. А. Слинкина. - 3-е изд. - Москва : ДМК Пресс, 2023. - 304 с. - ISBN 978-5-89818-331-8. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2102618>.

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>.

## 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	Катастрофы, стихийные бедствия, аварии, эпидемии. Солнечная и геомагнитная активность. /ежедневный обзор	<a href="http://www.disasters.chat.ru">http://www.disasters.chat.ru</a>
2	Каталог по безопасности жизнедеятельности	<a href="http://www.eun.chat.ru">http://www.eun.chat.ru</a>
3	Научная электронная библиотека elibrary.ru библиотека	<a href="https://elibrary.ru/defaultx.asp">https://elibrary.ru/defaultx.asp</a>
4	База данных для IT-специалистов	<a href="https://habr.com">https://habr.com</a>
5	База данных ScienceDirect	<a href="https://www.sciencedirect.com">https://www.sciencedirect.com</a>
6	Официальный сайт Министерства науки и высшего образования Российской Федерации	<a href="https://minobrnauki.gov.ru/">https://minobrnauki.gov.ru/</a>
7	Федеральный портал «Российское образование»	<a href="https://www.edu.ru/">https://www.edu.ru/</a>
8	Информационная система "Единое окно доступа к образовательным ресурсам"	<a href="http://window.edu.ru/">http://window.edu.ru/</a>
9	Единая коллекция цифровых образовательных ресурсов	<a href="http://school-collection.edu.ru/">http://school-collection.edu.ru/</a>
10	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>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

## 7. Методические и оценочные материалы

### Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Machine Learning (Машинное обучение)» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, проекты, 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.

*Семинар* — это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

*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.

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

Для успешной подготовки к проекту рекомендуется: четко определить цели и задачи проекта; составить план работы, разбив проект на этапы с указанием сроков выполнения каждого из них; использовать разнообразные источники информации и инструменты для

исследования темы; регулярно проверять прогресс и вносить коррективы в план, если это необходимо.

*Бонусные баллы* — это оценки, которые студенты могут получить за выполнение дополнительных заданий.

Формат бонусных баллов позволяет студентам улучшить общую оценку по дисциплине (модулю) и стимулирует углубленное изучение материала.

*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) «Machine Learning (Машинное обучение)».**

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

**Промежуточная аттестация** по дисциплине (модулю) осуществляется в форме **экзамена**, 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	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	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 задачами.
9	Отлично	
8	Отлично	
7	Хорошо	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
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		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.
5	Удовлетворительно	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	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Machine Learning (Машинное обучение)» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	30%	7	Набор задач по темам недели
Проекты	50%	7	Исследовательская работа по дисциплине (модулю) и презентация результатов
Экзамен	20%	1	Работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

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

**Формула расчёта итоговой оценки по дисциплине (модулю) «Machine Learning (Машинное обучение)»:**  $\langle 0,3 \times \text{среднее за домашние задания} + 0,5 \times \text{среднее за проекты} + 0,2 \times \text{экзамен} \rangle$ .

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

**Примерные домашние задания**

**Домашнее задание 1.**

1. Выбрать датасет для задачи регрессии (например, предсказание стоимости недвижимости, спроса или времени доставки).
2. Провести разведочный анализ данных (EDA): анализ распределений, выбросов, корреляций признаков.
3. Сформулировать задачу машинного обучения: определить целевую переменную, признаки и тип постановки (базовая модель).

4. Построить ML-пайплайн (предобработка пропусков, кодирование категориальных признаков, масштабирование, обучение модели).
5. Обучить не менее двух моделей (например, Linear Regression и Random Forest/Gradient Boosting) и сравнить их по метрикам (MAE, RMSE,  $R^2$ ).
6. Проанализировать важность признаков и сделать выводы о качестве решения.

### Домашнее задание 2.

1. Выбрать датасет для задачи классификации (например, отток клиентов, кредитный скоринг, определение спама).
2. Провести EDA: анализ баланса классов, исследование признаков и взаимосвязей.
3. Сформулировать задачу классификации (бинарная или многоклассовая), определить целевую переменную и признаки.
4. Реализовать ML-пайплайн: предобработка данных, разделение на train/test, обучение моделей.
5. Обучить не менее двух моделей (например, Logistic Regression и Random Forest/Gradient Boosting).
6. Оценить качество моделей с использованием метрик (Accuracy, Precision, Recall, F1-score, ROC-AUC) и построить confusion matrix.
7. Сделать выводы о применимости моделей и влиянии дисбаланса классов.

### Домашнее задание 3.

1. Выбрать датасет для задачи кластеризации (например, сегментация клиентов или пользователей).
2. Провести предварительный анализ данных и подготовку признаков (масштабирование, обработка пропусков).
3. Обосновать выбор признаков для кластеризации.
4. Реализовать ML-пайплайн для кластеризации с использованием как минимум двух алгоритмов (например, K-Means и DBSCAN или Agglomerative Clustering).
5. Определить оптимальное число кластеров (метод локтя, silhouette score и др.).
6. Визуализировать полученные кластеры (при возможности снижения размерности — PCA/UMAP).
7. Интерпретировать полученные сегменты и описать их возможное прикладное значение.

## Примерные описания заданий для проектов

### Проект 1.

#### Цель проекта:

Разработать MVP-решение для задачи кластеризации данных, продемонстрировав понимание принципов кластерного анализа и умение выстраивать полный ML-пайплайн от подготовки данных до интерпретации результатов в прикладном и бизнес-контексте.

#### Задачи проекта:

- Ознакомьтесь с постановкой задачи кластеризации и определите цель сегментации для выбранного датасета или бизнес-кейса.
- Подберите и проанализируйте набор данных, выполните очистку, предобработку и при необходимости feature engineering.
- Выберите подходящие алгоритмы кластеризации (например, K-means, DBSCAN, иерархическую кластеризацию) и обоснуйте выбор.
- Постройте ML-пайплайн для прототипирования решения, включая этапы подготовки данных, обучения модели и получения кластеров.
- Подберите и настройте ключевые гиперпараметры моделей, используя эвристики и/или метрики качества кластеризации.

- Оцените качество полученных кластеров с применением внутренних метрик (silhouette score, inertia и др.).
- Визуализируйте результаты кластеризации и проинтерпретируйте полученные группы.
- Сформулируйте выводы и рекомендации, показав, как результаты кластеризации могут быть использованы в продуктовых или бизнес-решениях.

#### **Результат проекта (MVP):**

Рабочий прототип ML-решения для задачи кластеризации, реализованный в виде воспроизводимого пайплайна (код + описание), демонстрирующий корректную постановку задачи, применение алгоритмов кластеризации, оценку и интерпретацию результатов.

### **Проект 2.**

#### **Цель проекта:**

Разработать Feature Card для MVP-решения на основе нейронной сети, продемонстрировав понимание принципов работы нейронных сетей, областей их применения и ограничений в контексте реальной бизнес-задачи.

#### **Задачи проекта:**

- Опишите бизнес-задачу, для решения которой целесообразно применение нейронных сетей.
- Сформулируйте ML-задачу и определите тип нейронной сети, подходящий для её решения (полносвязная, сверточная, рекуррентная и др.).
- Определите входные и выходные данные, источники данных и требования к их качеству.
- Опишите архитектуру модели на концептуальном уровне и обоснуйте ключевые решения.
- Зафиксируйте метрики качества модели и критерии успеха MVP.
- Опишите ограничения, риски и предположения, связанные с использованием нейронной сети.
- Сформулируйте ожидаемую бизнес-ценность и сценарии использования результата модели.

#### **Результат проекта:**

Подготовленная Feature Card MVP-решения на основе нейронной сети, содержащая описание бизнес-задачи и ML-постановки, концептуальную архитектуру модели, требования к данным, метрики качества и критерии успеха, а также анализ ограничений, рисков и ожидаемой бизнес-ценности, позволяющая использовать документ как основу для последующей разработки и внедрения решения.

### **Проект 3.**

#### **Цель проекта:**

Сформировать Feature Card для MVP рекомендательной системы, отразив ключевые элементы построения, валидации и использования рекомендаций в продукте.

#### **Задачи проекта:**

- Опишите продуктовый контекст и проблему, которую должна решать рекомендательная система.
- Определите тип рекомендательной системы (content-based, collaborative filtering, hybrid) и обоснуйте выбор.
- Сформулируйте ML-задачу и определите входные данные (пользователи, объекты, взаимодействия).
- Опишите логику формирования рекомендаций и предполагаемый ML-пайплайн.
- Определите метрики качества рекомендаций и продуктовые метрики эффективности.
- Зафиксируйте ограничения данных, cold-start проблемы и потенциальные риски.

- Опишите, как MVP-решение будет интегрировано в продукт и использоваться пользователями.

**Результат проекта:**

Сформированная Feature Card MVP рекомендательной системы, в которой зафиксированы продуктовый контекст, тип и логика рекомендаций, источники и структура данных, ключевые метрики качества и эффективности, ограничения и риски, а также сценарии интеграции решения в продукт, обеспечивающая единое понимание задачи между бизнесом и технической командой.

**Проект 4.**

**Цель проекта:**

Разработать Feature Card для MVP-решения по обработке изображений, продемонстрировав понимание специфики данных, моделей и бизнес-ценности computer vision-подходов.

**Задачи проекта:**

- Опишите бизнес-или продуктовую задачу, решаемую с помощью анализа изображений.
- Сформулируйте ML-задачу (классификация, детекция, сегментация и др.) и ожидаемый результат.
- Определите требования к изображениям, источники данных и объём датасета.
- Опишите подход к предобработке изображений и выбор модели (CNN, transfer learning и др.).
- Зафиксируйте метрики качества и критерии успешности MVP.
- Опишите ограничения, риски и этические аспекты работы с изображениями.
- Сформулируйте ценность решения для бизнеса и возможные направления масштабирования.

**Результат проекта:**

Готовая Feature Card MVP-решения по работе с изображениями, включающая формализованную ML-задачу, требования к данным и качеству изображений, выбранный подход и тип модели, метрики оценки, ограничения и этические аспекты, а также описание бизнес-ценности и потенциала масштабирования решения в продукте.

**Задания для промежуточной аттестации по дисциплине (модулю)**

№ п/п	Задание	Ответ	Компетенция
1.	Как называется множество всех возможных объектов в задаче машинного обучения? <b>Ответ запиши в виде словосочетания на русском языке.</b>	Пространство объектов/ пространство объектов	ПК-2
2.	Как называется средняя ошибка модели на обучающей выборке? <b>Ответ запиши в виде словосочетания на русском языке.</b>	Эмпирический риск/ эмпирический риск	ПК-2
3.	Верно ли то, что L1-регуляризация (LASSO) склонен к отбору признаков? <b>Ответ запиши в виде Да/Нет.</b>	да / Да / верно / Верно	ОПК-3
4.	Как называется метод оценки качества модели через генерацию подвыборок с возвращением? <b>Ответ запиши в виде одного слова на английском или русском языках.</b>	Бутстреп/бутстреп/ bootstrap/Bootstrap	ОПК-1
5.	Средняя ..... ошибка - это функция ошибки, вычисляемая как средний модуль отклонений <b>Ответ запиши в виде одного слова на русском языке.</b>	абсолютная/Абсолютная	ПК-2

6.	Среднеквадратичная _____ - это квадратный корень из MSE. Какое слово пропущено? <b>Ответ запиши в виде одного слова на русском языке.</b>	ошибка/Ошибка	ПК-2
7.	Какая компания создала наиболее популярную библиотеку градиентного бустинга Categorical Boosting (CatBoost)? <b>Ответ запиши в виде одного слова на русском языке.</b>	Яндекс/яндекс	ОПК-1
8.	Верно ли то, что метод главных компонент - это метод, который проецирует данные на второстепенные компоненты? <b>Ответ запиши в виде Да/Нет.</b>	нет / Нет / не верно / Не верно	ОПК-1
9.	Что такое "метка" в обучении с учителем? А. Значение, которое модель предсказывает Б. Признак объекта В. Функция ошибки Г. Гиперпараметр модели	А	ОПК-3
10.	Как называется вариант kNN, где ближайшие соседи влияют сильнее? А. kNN Б. Взвешенный kNN В. Soft-Margin SVM Г. Бэггинг	Б	ОПК-3
11.	Что решает проблему вырожденности матрицы в линейной регрессии?  А. Увеличение обучающей выборки Б. Регуляризация (например, Ridge) В. Уменьшение числа признаков Г. Использование kNN Д. Стекинг Е. PCA	Б	ОПК-5
12.	Как в случайном лесе вычисляется важность признака?  А. Через корреляцию с целевой переменной Б. Через уменьшение ошибки после расщепления по признаку В. Через веса в линейной модели Г. Через расстояние между объектами	Б	ОПК-5