

**УТВЕРЖДЕНА**

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

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

**Направление подготовки:** 02.04.01 Математика и компьютерные науки

**Направленность (профиль) подготовки:** Машинное обучение

**Квалификация (степень) выпускника:** магистр

**Форма обучения:** очная (с применением ДОТ)

**Срок освоения программы:** 2 года

**Год набора:** 2025

**Москва  
2025**

## Содержание

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

## 1. Краткая характеристика дисциплины (модуля)

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

Изучение дисциплины (модуля) Machine Learning (Машинное обучение) позволяет сформировать системное понимание методов анализа данных и разработки интеллектуальных моделей, применяемых для решения прикладных и бизнес-задач, а также научиться применять современные алгоритмы и инструменты машинного обучения для повышения качества принимаемых решений и создания конкурентоспособных цифровых продуктов.

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

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

Дисциплина (модуль) изучается на 1 курсе во 2 семестре, доступна для прохождения при условии успешного завершения дисциплины (модуля) «Основы Python».

**Цель изучения дисциплины (модуля):** формирование системных теоретических знаний и практических навыков разработки, обучения и оценки моделей машинного обучения для решения прикладных задач на основе данных.

### Задачи изучения дисциплины (модуля):

- освоить ключевые понятия, типы и алгоритмы машинного обучения, а также принципы их применимости к различным типам данных и задач;
- сформировать умение работать с данными, включая их сбор, очистку, предобработку и визуализацию с использованием Python и профильных библиотек;
- научиться выбирать, настраивать и обучать модели машинного обучения, применяя методы кросс-валидации и подбора гиперпараметров;
- развить навыки оценки качества моделей, интерпретации полученных результатов и выявления проблем переобучения и недообучения;
- сформировать практические навыки использования современных библиотек машинного обучения и обоснованного выбора алгоритмов для решения прикладных задач.

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

#### **знать:**

- ключевые понятия и термины (например, обучающая и тестовая выборки, переобучение, регуляризация);
- различные типы машинного обучения (обучение с учителем, обучение без учителя, обучение с подкреплением);
- основные алгоритмы машинного обучения (линейные модели, деревья решений, ансамбли моделей, нейронные сети и др.);
- принципы работы алгоритмов и их применимости к различным типам данных;
- методы оценки производительности моделей (точность, полнота, F-мера, ROC-AUC и др.);
- принципы кросс-валидации и настроек гиперпараметров;
- популярные библиотеки для машинного обучения (например, Scikit-Learn, PyTorch, CatBoost);

- основы работы с языком программирования Python и его библиотеками для работы с данными (NumPy, Pandas, Matplotlib).

***уметь:***

- собирать, очищать и преобразовывать данные для машинного обучения;
- визуализировать данные для выявления закономерностей и аномалий;
- выбирать и применять подходящие алгоритмы машинного обучения для решения конкретных задач;
- настраивать гиперпараметры моделей и проводить их оценку;
- интерпретировать результаты работы моделей и делать выводы на основе полученных данных;
- выявлять и устранять проблемы, такие как переобучение и недообучение.

***владеть:***

- языком программирования Python и навыком работы с библиотеками для анализа данных и машинного обучения;
- навыками строить и обучать модели, выбирая корректные методы под задачу и данные;
- навыками интерпретации и критического анализа результатов и обоснования выбора алгоритмов и методов, используемых в проекте.

## 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-3.	Способен самостоятельно создавать прикладные программные средства на основе современных информационных технологий и сетевых ресурсов, в том числе отечественного производства	ОПК-3.1.	Знает основные принципы программирования, архитектуры программного обеспечения и современные языки программирования, а также особенности отечественных информационных технологий и сетевых ресурсов
		ОПК-3.2.	Умеет разрабатывать прикладные программные средства, используя современные инструменты и технологии, а также интегрировать их с сетевыми ресурсами для решения конкретных задач
		ОПК-3.3.	Имеет практический опыт разработки программных средств, используемых при построении математических моделей в естественных науках
ПК-3.	Способен решать задачи профессиональной деятельности, формулировать результат, увидеть следствия полученного результата	ПК-3.1.	Знает основные принципы и методы решения задач профессиональной деятельности, а также способы формулирования и представления результатов, включая анализ последствий и их значимость в контексте проекта
		ПК-3.2.	Умеет применять математические и компьютерные методы для решения конкретных задач, формулировать четкие и обоснованные результаты, а также анализировать их последствия для дальнейших действий и решений
		ПК-3.3.	Имеет практический опыт в решении профессиональных задач, включая участие в проектах, где были получены результаты и проанализированы их следствия, что способствовало принятию обоснованных решений
ПК-6.	Способен разрабатывать программное обеспечение для решения прикладных задач в сфере машинного обучения	ПК-6.1.	Знает основные языки программирования, методы разработки программного обеспечения, а также принципы проектирования и архитектуры программных систем, применяемых в машинном обучении

		ПК-6.2.	Умеет анализировать прикладные задачи, разрабатывать алгоритмы и реализовывать их в виде программного обеспечения, используя современные инструменты и технологии, а также проводить тестирование и отладку созданных решений
		ПК-6.3.	Имеет практический опыт разработки программного обеспечения в рамках реальных проектов, включая участие в командах, где были успешно реализованы решения для конкретных прикладных задач в сфере профессиональной деятельности

### 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	Алгоритмы машинного обучения		15		70	Домашние задания Тесты
2	Введение в нейросети		6		21	Домашние задания Тесты
3	Оценка качества		9		40	Домашние задания Соревнование
4	Разбор и практика решения задач с собеседований		6		21	Домашние задания Тесты
	<i>Зачет с оценкой</i>			2		
	<b>Итого:</b>		<b>36</b>	<b>2</b>	<b>152</b>	
	<b>Объем дисциплины (модуля) (в ак. ч.)</b>	<b>190</b>				
	<b>Объем дисциплины (модуля) (в зач. ед.)</b>	<b>5</b>				

### 4. Содержание дисциплины (модуля)

№ п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Алгоритмы машинного обучения	Линейные модели: методы обучения линейных моделей, борьба с переобучением Линейные модели: метод опорных векторов, ядра и двойственные задачи в SVM Деревья и ансамбли: случайный лес, беггинг Деревья и ансамбли: градиентный бустинг Обзор других методов, Работа с признаками и обучение без учителя
2	Введение в нейросети	Введение в нейросети Q&A по пройденным темам
3	Оценка качества	Оценка качества на исторических данных Оценка качества в продакшене
4	Разбор и практика решения задач с собеседований	Задачи про классические методы и метрики Задачи на базовые понятия и математическую статистику

## 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 семинары, домашние задания, соревнование, тесты, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

*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.

*Соревнование* – организованное мероприятие, в рамках которого участники соперничают друг с другом для достижения определенной цели, демонстрируя свои навыки, знания или способности в заданной области.

*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.

*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 *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:

<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	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	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,
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			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.
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

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

Активность	Вес	Количество	Описание
Домашние задания	70%	7	Набор задач по темам недели
Соревнование	20%	1	Kaggle-style соревнование с задачей на ML
Тесты	10%	3	Письменная работа, состоящая из вопросов с вариантами ответов, на которые нужно ответить за ограниченное время

**Формула расчёта итоговой оценки по дисциплине (модулю) «Machine Learning (Машинное обучение)»:**  $\langle 0,7 \times \text{среднее за домашние задания} + 0,2 \times \text{соревнование} + 0,1 \times \text{среднее за тесты} \rangle$ .

## Текущий контроль успеваемости обучающихся по дисциплине (модулю)

### Примерные домашние задания

#### Домашнее задание 1

##### Часть 1. Линейные модели и SVM

1. Обучите линейную регрессию и логистическую регрессию на выбранном датасете (например, из sklearn).
2. Сравните модели с регуляризацией L1 и L2 и проанализируйте влияние коэффициента регуляризации на переобучение.
3. Обучите SVM с линейным и нелинейным (RBF) ядром.
4. Сравните качество моделей и опишите, в каких случаях применение ядра даёт прирост.

##### Часть 2. Деревья и ансамбли

5. Обучите дерево решений и случайный лес, сравните их по качеству и устойчивости.
6. Обучите модель градиентного бустинга (например, CatBoost или GradientBoosting).
7. Проведите сравнение всех моделей по выбранной метрике и времени обучения.

##### Часть 3. Работа с признаками и unsupervised

8. Выполните базовую генерацию или отбор признаков.
9. Примените один метод обучения без учителя (например, K-means или PCA) и интерпретируйте результат.

#### Домашнее задание 2

1. Реализуйте простую полносвязную нейронную сеть (на PyTorch или Keras) для задачи классификации.
2. Проанализируйте влияние:
  - количества слоев,
  - функции активации,
  - learning rate.
3. Зафиксируйте признаки переобучения и примените один способ борьбы с ним (dropout или регуляризация).
4. Подготовьте список из 5 вопросов по темам дисциплины, которые вызывают наибольшие затруднения, и предложите попытку ответа на каждый.

#### Домашнее задание 3

##### Часть 1. Исторические данные

1. Разделите данные на обучающую и тестовую выборки.
2. Примените кросс-валидацию.
3. Рассчитайте не менее 3 метрик (например, accuracy, F1, ROC-AUC).
4. Проанализируйте различия между метриками.

##### Часть 2. Продакшен-контекст

5. Опишите возможные источники деградации модели в продакшене.
6. Предложите план мониторинга модели (метрики, частота проверки).
7. Опишите, что такое data drift или concept drift, и как их выявлять.

#### Домашнее задание 4

##### Часть 1. Классические методы и метрики

1. Решите 3 задачи на выбор модели под конкретный кейс (классификация/регрессия/кластеризация).
2. Обоснуйте выбор алгоритма и метрики.

##### Часть 2. Базовые понятия и матстатистика

3. Решите 3 теоретические задачи на:
  - переобучение,

- доверительные интервалы,
- проверку статистических гипотез.

4. Объясните решение в логике устного ответа на собеседовании (кратко и структурированно).

### Примерные задания для тестов

1. Что нужно усреднить в задаче бинарной классификации, чтобы получить сбалансированную точность ?
  - (a) точность и полноту
  - (b) чувствительность и специфичность
  - (c) полноту класса 1 и класса 0
  - (d) точность класса 1 и класса 0
2. Требуется по описанию проекта предсказать его доходность (в рублях) через год после начала реализации. Предполагается, что имеется база подобных проектов с подробной статистикой их дальнейшей судьбы.
  - (a) это задача классификации
  - (b) это задача регрессии
  - (c) это задача обучения с размеченными данными
  - (d) это задача обучения без меток
3. Какой метод кластеризации актуален для иерархических структур?
  - (a) K-Means
  - (b) DBSCAN
  - (c) Агломеративная кластеризация
  - (d) PCA
4. Что представляет собой проблема "bias-variance trade-off"?
  - (a) Компромисс между простотой и функциональной выразимостью модели
  - (b) Выбор между различными алгоритмами машинного обучения
  - (c) Определение оптимального порога для классификации данных
5. Требуется по описанию участников торгов выявить потенциальных мошенников. При этом есть истории предыдущих торгов, случаев доказанного мошенничества пока нет.
  - (a) это задача классификации
  - (b) это задача регрессии
  - (c) это задача обучения с размеченными данными
  - (d) это задача обучения без меток
6. Что происходит при увеличении  $k$  в методе kNN (отметьте все подходящие варианты)?
  - (a) Улучшается качество
  - (b) Алгоритм превращается в константный (для заданной обучающей выборки)
  - (c) Увеличивается число гиперпараметров
  - (d) Алгоритм "упрощается" (ответы становятся более стабильными при небольших изменениях обучающей выборки)
7. Какая из перечисленных тем не относится к методам кластеризации в машинном обучении?
  - (a) K-Means
  - (b) Support Vector Machines
  - (c) DBSCAN
  - (d) Hierarchical Clustering

8. Каким образом может оцениваться качество модели линейной классификации?
- (a) С помощью точности (accuracy)
  - (b) С помощью матрицы ошибок (confusion matrix)
  - (c) С помощью F1-меры (F1-score)
9. Почему наивная форма стекинга не применяется на практике (в отличие от других форм)?
- (a) из-за переобучения
  - (b) из-за сложности
  - (c) из-за того, что является частной формой бустинга
10. Как связано понятие BVD с терминами переобучения и недообучения?
- (a) Высокое смещение, низкий разброс – недообучение
  - (b) Низкое смещение, высокий разброс – переобучение
  - (c) Смещение и разброс не связаны с обучением
  - (d) Высокий разброс, низкое смещение – недообучение
11. Выберите верные утверждения для гиперпараметров:
- (a) Гиперпараметров больше, чем параметров
  - (b) Гиперпараметры настраиваются градиентным спуском
  - (c) Гиперпараметры настраиваются в результате обучения (метод fit)
  - (d) Значения гиперпараметров можно выбрать с помощью перекрестной проверки
12. Что представляет собой метод t-SNE в контексте уменьшения размерности?
- (a) Генерация новых признаков на основе комбинаций исходных признаков
  - (b) Преобразование пространства признаков в пространство меньшей размерности с сохранением относительных расстояний
  - (c) Классификация объектов на основе различия в их признаках
  - (d) Оптимизация функции потерь для снижения ошибки модели
13. Для чего может использоваться L1-регуляризация?
- (a) Для уменьшения переобучения модели
  - (b) Для нормировки коэффициентов
  - (c) Для отбора признаков
  - (d) Для подбора гиперпараметров
14. В косых деревьях (oblique decision trees)... :
- (a) Используются предикаты с линейными комбинациями признаков
  - (b) Используются не все признаки
  - (c) Используются ядра
  - (d) Используется подрезка
15. В каких схемах контроля каждый объект гарантировано ровно 1 раз попадает в тестовую выборку?
- (a) K-fold CV
  - (b) Стратифицированный K-fold CV
  - (c) Бутстреп
  - (d) LOO

16. Минимальное значение среднего внутрикластерного расстояния достигается. . .
- (a) На одном кластере
  - (b) На числе кластеров, равном числу объектов
  - (c) В точке перегиба
17. Какие два основных типа линейной регрессии существуют?
- (a) Простая линейная регрессия и множественная линейная регрессия
  - (b) Логистическая регрессия и полиномиальная регрессия
  - (c) Сигмоидная регрессия и эластичная регрессия
18. Что в стекинге называется метаалгоритмом? (выберите все правильные ответы)
- (a) базовый алгоритм
  - (b) алгоритм, который учится на ответах базового
  - (c) алгоритм, ответы которого будут итоговыми
19. Почему на практике минимизируют эмпирический риск?
- (a) теоретический риск невозможно минимизировать в общем случае
  - (b) это делает модель проще
  - (c) это позволяет уменьшить эффект переобучения / запоминания
20. На каких весах считается регуляризация в линейной регрессии и почему?
- (a) На всех, так как это позволяет избежать переобучения
  - (b) На всех, так как это позволяет нормировать признаки
  - (c) На всех кроме  $w_0$ , так как  $w_0$  отвечает за «масштаб»
  - (d) На всех кроме случайно выбранного, так как это позволяет избежать переобучения
21. Какие из функций ошибок / функционалов качества эквивалентны MSE (минимум MSE соответствует оптимуму функции)?
- (a) RMSE
  - (b) Хьюбера
  - (c) logcosh
  - (d)  $R^2$
22. Какая из компонент в BVD отвечает за чувствительность модели к конкретному набору данных?
- (a) Смещение
  - (b) Разброс
  - (c) Ошибка
  - (d) Шум
23. Какие ансамбли являются параллельными? (выберите все правильные ответы)
- (a) бустинг
  - (b) бэгинг
  - (c) случайный лес

24. Что такое разброс (variance) модели?
- (a) Ошибка, связанная с неправильным представлением о целевой функции или недостаточной сложностью модели
  - (b) Ошибка, связанная с шумом в данных
  - (c) Ошибка, связанная с переобучением модели
25. Что пытается оптимизировать SNE (Stochastic Neighbor Embedding)?
- (a) Снижение размерности с сохранением глобальной структуры
  - (b) Вероятностное приближение расстояний между точками в высокоразмерном и низкоразмерном пространствах
  - (c) Увеличение плотности кластеров
  - (d) Максимизация линейной зависимости между переменными
26. Как работает алгоритм градиентного бустинга?
- (a) Каждое следующее дерево итерационно настраивает на градиент ошибки
  - (b) Каждое следующее дерево итерационно настраивается на антиградиент ошибки
  - (c) Сумма деревьев совместно настраивается на градиент ошибки
  - (d) Сумма деревьев совместно настраивается на антиградиент ошибки
27. Какой метод кластеризации дает возможность определять количество кластеров в данных автоматически?
- (a) K-Means
  - (b) Hierarchical Clustering
  - (c) DBSCAN
  - (d) Agglomerative Clustering
28. Выберите все верные утверждения про ROC-AUC
- (a) Отражает качество ранжирования объектов
  - (b) Не зависит от масштаба значений, которые выдает алгоритм
  - (c) Если инвертировать порядок объектов в модели с  $\text{ROC-AUC} = 0$ , то получим  $\text{ROC-AUC} = 1$
  - (d) Можно считать, как долю верно отранжированных пар
29. Какой метод кластеризации позволяет размечать выбросы?
- (a) Mini-Batch K-Means
  - (b) K-Means
  - (c) DBSCAN
  - (d) Иерархическая кластеризация
30. Какие из перечисленных алгоритмов ленивые?
- (a) 1NN
  - (b) 3NN
  - (c) ближайший центроид
  - (d) линейная регрессия

### Примерное задание и критерии оценивания к соревнованию

#### Задание:

На основе предоставленных данных необходимо:

1. Определить целевую переменную и тип задачи.
2. Разделить данные на обучающую и валидационную выборки.
3. Обучить и сравнить три модели:
  - линейную модель (линейная или логистическая регрессия);
  - SVM;
  - k ближайших соседей (kNN).

4. Выбрать лучшую модель по валидационной метрике.
5. Сделать финальное предсказание для тестовой выборки.

Разрешается:

- использовать масштабирование признаков;
- подбирать гиперпараметры в разумных пределах;
- использовать кросс-валидацию (по возможности).

**Ограничения:**

- Используются только изученные модели: линейные, SVM, kNN.
- Нельзя использовать ансамбли и бустинг.
- Тестовая выборка используется только один раз для финальной проверки.

**Критерии оценки:**

Оценивание проводится по следующим критериям:

- Итоговое значение основной метрики на тестовой выборке.
- Корректность выбора метрики в соответствии с типом задачи.
- Наличие сравнения трёх обязательных моделей.
- Корректность разбиения данных и процедуры валидации.
- Обоснование выбора финальной модели.
- Использование корректной терминологии машинного обучения.

**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-3
2.	Как называется средняя ошибка модели на обучающей выборке? <b>Ответ запиши в виде словосочетания на русском языке.</b>	Эмпирический риск/ эмпирический риск	ПК-6
3.	Верно ли то, что L1-регуляризация (LASSO) склонен к отбору признаков? <b>Ответ запиши в виде Да/Нет.</b>	да / Да / верно / Верно	ПК-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	ПК-3
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-6
6.	Среднеквадратичная _____ - это квадратный корень из MSE. Какое слово пропущено? <b>Ответ запиши в виде одного слова на русском языке.</b>	ошибка/Ошибка	ПК-6
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	ПК-6
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	ПК-6

9.	<p>Что такое "метка" в обучении с учителем?</p> <p>А. Значение, которое модель предсказывает</p> <p>Б. Признак объекта</p> <p>В. Функция ошибки</p> <p>Г. Гиперпараметр модели</p>	А	ОПК-3
10.	<p>Как называется вариант kNN, где ближайшие соседи влияют сильнее?</p> <p>А. kNN</p> <p>Б. Взвешенный kNN</p> <p>В. Soft-Margin SVM</p> <p>Г. Бэггинг</p>	Б	ОПК-3
11.	<p>Что решает проблему вырожденности матрицы в линейной регрессии?</p> <p>А. Увеличение обучающей выборки</p> <p>Б. Регуляризация (например, Ridge)</p> <p>В. Уменьшение числа признаков</p> <p>Г. Использование kNN</p> <p>Д. Стекинг</p> <p>Е. PCA</p>	Б	ОПК-3
12.	<p>Как в случайном лесе вычисляется важность признака?</p> <p>А. Через корреляцию с целевой переменной</p> <p>Б. Через уменьшение ошибки после расщепления по признаку</p> <p>В. Через веса в линейной модели</p> <p>Г. Через расстояние между объектами</p>	Б	ОПК-3