ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

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

Рабочая программа дисциплины (модуля)
«Соревновательный анализ данных»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Разработка*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

Москва
2024

Электронный документ

Содержание

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Соревновательный анализ данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 Математика и компьютерные науки, профиль Разработка, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Соревновательный анализ данных» обеспечивает понимание основных принципов представления и обработки информации в различных технических и научных областях, а также формирует компетенции, необходимые для разработки современных систем обработки сигналов и звуков в прикладных задачах математики и компьютерных наук.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 Математика и компьютерные науки, профиль Разработка и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений как дисциплина по выбору.

Дисциплина (модуль) изучается на 3 или 4 курсе в 5 или 7 семестре на выбор, доступна для прохождения при условии успешного завершения дисциплин «Machine Learning (Машинное обучение)» и «Deep Learning (Глубокое обучение)».

Цель изучения дисциплины (модуля): формирование у студентов фундаментальных знаний и навыков анализа, обработки и моделирования сигналов и звуковых данных с применением математических и компьютерных методов.

Задачи изучения дисциплины (модуля):

— изучение схем валидации, базовых линий (бейзлайнов), претренированных моделей (например, BERT для NLP, ResNet для CV), принципов архитектур (CNN, Transformers), инструментов (TensorFlow, PyTorch, scikit-learn), метрик (accuracy, F1-score, BLEU) и оптимизации для больших данных;

— работа с табличными данными (pandas), feature engineering, обучение моделей (линейная регрессия, деревья решений, нейросети), стекинг, интеграция методов (ансамбли, transfer learning);

— анализ открытых источников (GitHub, Kaggle), оптимизация ресурсов (GPU на Colab), освоение новых доменов (например, медицинские данные), генерация и тестирование гипотез (А/Б-тесты, эксперименты);

— коммуникация, распределение задач, code review, управление проектами (Agile/Scrum).

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— разные схемы валидации и умеет применять нужную;
— знать как собирать бейзлайн для разных задач в ML/NLP/CV;
— какие претрейнд модели существуют и для каких задач они полезны;
— общее понимание работы сверточных сетей и трансформеров;
— инструменты, которые применяются для решения задач ML/NLP/CV;
— распространенные метрики и как их оптимизировать;
— как работать с большими датасетами и оптимизировать свой код;

уметь:

— работать с табличными данными, генерить признаки;
— обучать и тюнить ML модели разных классов;
— применять стекинг и генерить фичи, завязанные на таргет;

— искать в github, на kaggle и в интернете решения похожих задач и использовать их;
— "склеивать" разные решения/подходы вместе для достижения лучшего качества;
— хорошо структурировать свой код;
— обеспечивать воспроизводимость своих решений;
— работать в команде;
— работать с ограниченными ресурсами (вычислительными ресурсами, временем);
— использовать бесплатные вычислительные ресурсы в интернете (Colab, Kaggle, и тд);

владеть:

— разбором в незнакомом домене и сбором решения хорошего качества;
— генерацией новых гипотез и оперативной их проверкой.

2. Перечень планируемых результатов обучения

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

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области разработки, основные принципы критической оценки источников информации и их релевантности.
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики	ОПК-1.1.	Знает основные методы и подходы к решению задач прикладной и компьютерной математики, включая алгоритмы, математическое моделирование и теорию оптимизации, а также современные инструменты и технологии, используемые в этой области
		ОПК-1.2.	Умеет анализировать и формулировать математические задачи, применять

			соответствующие методы и алгоритмы для их решения, а также интерпретировать и представлять результаты в понятной и доступной форме
		ОПК-1.3.	Имеет практический опыт работы над проектами или исследованиями в области прикладной и компьютерной математики, включая участие в конкурсах, олимпиадах или научных публикациях, где были решены актуальные и значимые задачи
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Знает методы и подходы к формулированию задач, а также основные принципы математического доказательства и анализа результатов.
		ПК-1.2.	Умеет корректно ставить и формулировать математические задачи, применять строгие методы доказательства и анализировать полученные результаты.
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач
ПК-2.	Способен решать типовые задачи профессиональной деятельности в области разработки, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области разработки, включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Контактная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары (практические занятия)			
1	Машинное обучение	7	7		33	Домашние задания
2	Глубокое обучение	7	7		33	Домашние задания
3	Обработка естественного языка	7	7		32	Домашние задания
4	Компьютерное зрение	7	7		32	Домашние задания
	Зачет с оценкой			4		Проект
Итого:		***28***	***28***	***4***	***130***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Машинное обучение	Вводная лекция. EDA и лики в данных. Модели классического ML и генерация признаков. Метрики, валидация, ансамблирование. Ускорение и оптимизация вычислений
2	Глубокое обучение	Hardware: GPU, docker, code competitions. Трансформеры. Задачи ранжирования, рекомедаций. Задачи с временными рядами, нейросети для табличных данных
3	Обработка естественного языка	Задачи на NLP - введение, задача классификации. Задачи на NLP - продолжение, LLM, RAG
4	Компьютерное зрение	Задачи на CV - введение, задача классификации. Задачи на CV - object detection и сегментация. Задачи на CV - матчинг. Задачи на RL/агентов.

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. Анализ данных и процессов: Учебное пособие / Барсегян А.А., Куприянов М.С., Холод И.И. - СПб:БХВ-Петербург, 2009. - 512 с. - ISBN 978-5-9775-0368-6. - Текст : электронный. - URL: https://znanium.com/catalog/product/350638.

2. Карау, Х. Изучаем Spark. Молниеносный анализ данных : практическое руководство / Х. Карау, Э. Конвински, П. Венделл, З. Матей ; пер. с англ. - 2-е изд. - Москва : ДМК Пресс, 2023. - 305 с. - ISBN 978-5-89818-320-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2102607.

3. Малов, Д. А. Глубокое обучение и анализ данных. Практическое руководство : практическое руководство / Д. А. Малов. - Санкт-Петербург : БХВ-Петербург, 2023. - 272 с. - ISBN 978-5-9775-1172-8. - Текст : электронный. - URL: https://znanium.com/catalog/product/2123365.

4. Apache Kafka. Потоковая обработка и анализ данных : практическое руководство / Г. Шапира, Т. Палино, Р. Сиварам, К. Петти. - 2-е изд. - Санкт-Петербург : Питер, 2023. - 512 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-2288-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2123357.

Дополнительная 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.

2. Высоконагруженные приложения. Программирование, масштабирование, поддержка. — СПб.: Питер, 2018. — 640 с.: ил. — (Серия «Бестселлеры O'Reilly»). ISBN 978-5-4461-0512-0.

6. Материально-техническое обеспечение

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

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

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

— столами и стульями;
— компьютерной техникой;
— механическими калькуляторами;
— специализированным оборудованием, включая демонстрационное оборудование.

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

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

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

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

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное

Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

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

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

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

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

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

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы

избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Проект – исследовательская работа по курсу и презентация результатов.

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

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) «Соревновательный анализ данных»

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

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

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

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

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

Дисциплина (модуль) «Соревновательный анализ данных» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	75%	Набор задач по темам недели
Проекты	25%	Исследовательская работа по курсу и презентация результатов

Формула расчёта итоговой оценки по дисциплине (модулю) «Соревновательный анализ данных»: «0,75 × среднее за домашние задания + 0,25 × среднее за проект».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание «Машинное обучение»

1. Реализовать линейную регрессию для предсказания непрерывной переменной на заданном наборе данных.
2. Построить модель дерева решений и проанализировать её работу на классификационной задаче.
3. Сравнить качество моделей ансамбля (Random Forest и Gradient Boosting) на одном и том же датасете.
4. Провести кросс-валидацию для выбранной модели и рассчитать основные метрики качества (accuracy, precision, recall, F1-score).
5. Реализовать стекинг двух различных моделей и оценить прирост качества по сравнению с базовыми моделями.
6. Проанализировать влияние несбалансированности классов на качество модели и применить методы борьбы с переобучением (например, регуляризацию или балансировку классов).

Домашнее задание «Компьютерное зрение»

1. Выполнить фильтрацию изображения с помощью различных фильтров (гауссовский, медианный) и сравнить результаты.
2. Реализовать алгоритм выделения признаков (например, градиенты или гистограммы ориентированных градиентов) на изображениях.
3. Построить и обучить простую сверточную нейронную сеть (CNN) для задачи классификации изображений.
4. Выполнить детекцию объектов на изображении с использованием предобученной модели (например, YOLO или SSD).
5. Реализовать сегментацию изображения с помощью метода пороговой обработки или более сложных алгоритмов (например, U-Net).
6. Использовать популярные библиотеки (OpenCV, PyTorch/TensorFlow) для обработки и визуализации изображений и результатов моделей.

Домашнее задание «Обработка естественного языка»

1. Выполнить токенизацию и очистку текста на заданном корпусе данных.
2. Реализовать различные методы векторизации текста: мешок слов (Bag of Words), TF-IDF и эмбеддинги (Word2Vec или GloVe).
3. Обучить рекуррентную нейронную сеть (RNN) или LSTM для задачи классификации текстов.
4. Использовать трансформерную модель (например, BERT) для задачи Named Entity Recognition (NER).
5. Выполнить машинный перевод с использованием предобученной модели или простого seq2seq подхода.
6. Оценить качество моделей с помощью соответствующих метрик (например, accuracy, F1-score, BLEU) и проанализировать результаты.

Примерное описание к проекту

Проект по теме: Машинное обучение — Построение и оптимизация модели для классификации с использованием ансамблей и методов борьбы с несбалансированными данными

Цели проекта:

- Освоить основные алгоритмы машинного обучения (линейная регрессия, деревья решений, ансамбли).
- Научиться применять методы оценки и валидации моделей, включая кросс-валидацию и расчет метрик качества.
- Изучить техники повышения качества моделей: стекинг, бустинг, подбор гиперпараметров.
- Разобраться с проблемой несбалансированных данных и методами борьбы с переобучением.

Задачи проекта:

1. Выбрать и подготовить датасет с несбалансированными классами для задачи классификации.
2. Реализовать и обучить базовые модели: линейную регрессию (логистическую регрессию), дерево решений.
3. Построить ансамблевые модели: Random Forest и Gradient Boosting.
4. Провести кросс-валидацию и оценить модели по метрикам accuracy, precision, recall, F1-score.
5. Реализовать стекинг на основе нескольких моделей для улучшения качества предсказаний.
6. Применить методы борьбы с несбалансированностью (например, oversampling, undersampling, взвешивание классов) и регуляризацию для борьбы с переобучением.
7. Подобрать гиперпараметры с помощью Grid Search или Random Search.
8. Сравнить результаты, сделать выводы о влиянии методов на качество модели.

Этапы выполнения:

1. **Подготовка данных:** загрузка, анализ, очистка, разметка классов.
2. **Реализация базовых моделей:** обучение и первичная оценка.
3. **Построение ансамблей:** обучение Random Forest и Gradient Boosting.
4. **Валидация и оценка:** кросс-валидация, расчет метрик.
5. **Оптимизация:** стекинг, подбор гиперпараметров, методы борьбы с несбалансированностью.
6. **Анализ результатов:** сравнительный анализ моделей, визуализация метрик.
7. **Подготовка отчёта и презентации:** описание методов, результатов и выводов.

Критерии оценивания:

- Корректность и полнота подготовки данных (10%).
- Правильная реализация и обучение базовых моделей (15%).
- Качество построенных ансамблевых моделей и стекинга (20%).
- Применение и обоснование методов борьбы с несбалансированностью и переобучением (15%).
- Качество и полнота оценки моделей с использованием метрик и кросс-валидации (15%).

- Глубина анализа результатов и обоснованность выводов (15%).
- Качество оформления отчёта и презентации (10%).

Критерии защиты проекта:

- Четкое и структурированное изложение цели и задач проекта.
- Демонстрация этапов работы и полученных результатов.
- Обоснование выбора моделей, методов оптимизации и борьбы с несбалансированностью.
- Ответы на вопросы по теории используемых алгоритмов и практическим аспектам.
- Способность критически оценить качество моделей и предложить дальнейшие улучшения.
- Качество презентации, умение ясно донести основные идеи и результаты.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1	Назовите первый этап в процессе разведочного анализа данных (EDA).	загрузка данных/ импорт данных/ сбор данных	УК-1
2	Приведите пример типа лика в данных, который можно обнаружить при анализе.	data leakage/ утечка данных/leak	УК-1
3	Назовите одну из основных моделей классического машинного обучения, изучаемых в курсе.	линейная регрессия/random forest/SVM/логистическая регрессия	УК-1
4	Укажите одну ключевую метрику для оценки качества классификации в машинном обучении.	Accuracy/precision/recall/F1-score	УК-2
5	Приведите название одной из схем валидации моделей, используемых для оценки качества.	k-fold/holdout/ cross-validation	УК-2
6	Назовите один из методов ансамблирования моделей, изучаемых в курсе.	Bagging/boosting/ stacking	УК-2
7	Укажите один из способов ускорения вычислений в машинном обучении.	Параллелизация/ оптимизация алгоритмов/ использование GPU	ОПК-1
8	Приведите пример архитектуры нейросети, применяемой для обработки временных рядов.	RNN/ LSTM/GRU	ОПК-1
9	Назовите одну из задач ранжирования, рассматриваемых в курсе глубокого обучения.	Поиск/рекомендация/ retrieval	ОПК-1
10	Приведите название одного из основных компонентов трансформеров.	Attention/encoder/ decoder	ПК-1
11	Назовите первую задачу обработки естественного языка (NLP), изучаемую в курсе.	Классификация/генерация/перевод	ПК-1
12	Приведите пример одной из задач компьютерного зрения (CV), изучаемых в курсе.	классификация изображений/object detection/ сегментация/ матчинг	ПК-1
13	Назовите один из подходов к работе с большими языковыми моделями (LLM) в NLP.	fine-tuning/prompt engineering/ RAG	ПК-2

14	Приведите пример одного из методов матчинга в компьютерном зрении.	feature matching/ template matching/ descriptor-based	ПК-2
15	Укажите одно из требований информационной безопасности при работе с данными в AI.	Конфиденциальность / целостность/ доступность	ПК-2
16	Назовите один из инструментов для оптимизации вычислений в больших данных.	NumPy/Pandas/ Dask	УК-1
17	Приведите один из принципов критической оценки источников информации в AI.	Релевантность/ авторитетность/ актуальность	УК-1
18	Назовите один из способов синтеза информации из различных источников.	мета-анализ/обзор литературы/интеграция данных	УК-2
19	Укажите одно из ограничений при выборе способов решения задач в AI.	вычислительные ресурсы/время/ правовые нормы	ОПК-1
20	Приведите один из математических методов оптимизации, применяемых в ML.	градиентный спуск/ генетические алгоритмы/линейное программирование	ПК-1