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: *Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика»*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2024*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2024

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 8
4. **Содержание дисциплины (модуля)** 8
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Соревновательный анализ данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «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 «Программа двух дипломов НИУ ВШЭ и ЦУ «Прикладная математика и информатика» и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

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

— изучить основные типы задач анализа данных и особенности их постановки в условиях соревнований, включая классификацию, регрессию, ранжирование и прогнозирование временных рядов;

— освоить этапы решения соревновательной задачи: от анализа данных и feature engineering до обучения моделей, валидации и подготовки финального решения;

— научиться применять современные методы машинного обучения, метрики качества и стратегии валидации для построения эффективных и воспроизводимых решений;

— развить навыки разработки и сравнения моделей с использованием библиотек python, таких как numpy, pandas, scikit-learn, а также инструментов визуализации и анализа результатов;

— освоить практику работы с соревновательными платформами, организации экспериментов и ансамблирования моделей для достижения высокого качества предсказаний.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— основные принципы организации соревнований по анализу данных и машинному обучению;

— этапы решения соревновательной задачи: постановка задачи, анализ данных, построение признаков, обучение моделей, оценка качества и подготовка решения;

— основные типы задач анализа данных (классификация, регрессия, ранжирование, кластеризация, прогнозирование временных рядов и др.);

— основные метрики качества моделей и особенности их применения в соревновательных задачах;

— современные методы и модели машинного обучения, используемые при решении соревновательных задач;

— принципы валидации, предотвращения переобучения и обеспечения воспроизводимости результатов.

уметь:

— анализировать постановку соревновательной задачи и формулировать стратегию её решения;

— выполнять разведочный анализ данных и выявлять закономерности в данных;

— формулировать, проверять и интерпретировать гипотезы о данных и моделях;

— разрабатывать признаки и подготавливать данные для обучения моделей;

— обучать, настраивать и сравнивать модели машинного обучения;

— использовать локальную валидацию для оценки качества решений;

— разрабатывать базовые и улучшенные решения для соревновательных задач;

— анализировать результаты экспериментов и выбирать наиболее эффективные подходы.

владеть:

— инструментами анализа и обработки данных на языке Python;

— библиотеками NumPy, Pandas, Scikit-Learn для решения задач анализа данных;

— инструментами визуализации данных и анализа результатов экспериментов;

— методами построения признаков (feature engineering);

— методами кросс-валидации, подбора гиперпараметров и ансамблирования моделей;

— практическими навыками работы с соревновательными платформами по анализу данных;

— подходами к организации экспериментов, воспроизводимости вычислений и оформлению итоговых решений.

2. Перечень планируемых результатов обучения

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

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области искусственного интеллекта, основные принципы критической оценки источников информации и их релевантности.
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен 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 искусственного интеллекта, опираясь на информационную и библиографическую культуру, используя информационно-коммуникационные технологии и учитывая основные требования информационной безопасности	ПК-2.1.	Знает основы информационной и библиографической культуры, а также принципы информационной безопасности и применения информационно-коммуникационных технологий в профессиональной деятельности
		ПК-2.2.	Умеет эффективно использовать информационно-коммуникационные технологии для решения стандартных задач профессиональной деятельности, учитывая требования информационной безопасности
		ПК-2.3.	Имеет опыт работы с информационными ресурсами и технологиями в области искусственного интеллекта,

			включая соблюдение норм информационной безопасности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Лекции	Практические занятия			
1	Введение в соревнования	6	6		24	Соревнование
2	Табличные соревнования	8	8		34	Соревнование Доклад
3	Обработка естественного языка	8	8		34	Соревнование
4	Компьютерное зрение	8	8		34	Соревнование
	Зачёт с оценкой			4		Проект
	Итого:	***30***	***30***	***4***	***126***	
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	Введение в соревнования	Введение в соревнования, вводная лекция, открытое занятие
2	Табличные соревнования	Решение соревнования, особенности и трюки табличных соревнований
3	Обработка естественного языка	Решение соревнования, особенности и трюки текстовых соревнований
4	Компьютерное зрение	Решение соревнования, особенности и трюки визуальных соревнований

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 : учебник для вузов / под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2026. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/583032.

2. Платонов, А. В. Машинное обучение : учебник для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2026. — 78 с. — (Высшее образование). — ISBN 978-5-534-22200-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/600893.

Дополнительная литература:

1. Криволапов, С. Я. Введение в анализ данных. Поиск структуры данных с применением языка Python : учебное пособие / С.Я. Криволапов. — Москва : ИНФРА-М, 2026. — 177 с. — (Высшее образование). — DOI 10.12737/2082643. - ISBN 978-5-16-019001-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2262008.

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, семинары, домашние задания, контесты, 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.

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.

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

Доклад – это структурированное устное или письменное представление информации по определенной теме, предназначенное для информирования или обсуждения.

Методические рекомендации по подготовке к докладу включают тщательное изучение темы и сбор необходимых материалов, а также создание четкой структуры и практику представления с акцентом на основные моменты и визуальные материалы.

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) «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 ***зачёта 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, если иное не указано в силлабусе дисциплины:

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. 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,
6	Хорошо	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		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.
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) «Соревновательный анализ данных» оценивается следующим образом:

Активность	Вес	Количество	Описание
Соревнование	70%	7	Организованное мероприятие, в рамках которого участники соперничают друг с другом для достижения определенной цели
Доклад	15%	1	Структурированное устное или письменное представление информации по определенной теме
Зачёт с оценкой	15%	1	Проект - Исследовательская работа по дисциплине (модулю) и презентация результатов

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

Формула расчёта итоговой оценки по дисциплине (модулю) «Соревновательный анализ данных»: «0,7 × среднее за соревнования + 0,15 × доклад + 0,15 × зачёт с оценкой».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерное описание задания и критерии оценивания к соревнованию
Соревнование 1.

Цель соревнования: применить навыки анализа табличных данных, построения признаков и обучения моделей машинного обучения для решения задачи бинарной классификации.

Задание:

1. на основе предоставленного датасета с информацией о клиентах (демография, поведение, тарифы) построить модель, предсказывающую факт оттока (churn);

2. выполнить разведочный анализ, обработать пропуски и категориальные признаки, разбить данные, обучить модель, провести локальную валидацию и отправить предсказания в заданном формате.

Критерии оценивания:

— корректность понимания постановки задачи и выбор соответствующей метрики качества;

— качество предобработки данных: обработка пропусков, кодирование категориальных признаков, нормализация;

— обоснованность выбора модели и адекватность её настройки в условиях ограниченного времени;

— наличие и правильность локальной валидации (например, train-val split или кросс-валидация);

— успешность отправки решения в требуемом формате (csv с колонками id и pred);

— интерпретация результатов: анализ ошибок, сравнение с базовым решением, объяснение ключевых шагов;

— соблюдение принципов воспроизводимости: фиксация seed, читаемость кода, структура ноутбука.

Соревнование 2.

Цель соревнования: применить методы обработки естественного языка для решения задачи сентимент-анализа на примере кратких текстовых отзывов.

Задание:

1. по текстам отзывов определить, является ли отзыв положительным или отрицательным;

2. выполнить базовую очистку текста (удаление знаков препинания, стоп-слов), токенизацию, векторизацию (tf-idf или bag-of-words), обучить модель, провести валидацию и отправить предсказания.

Критерии оценивания:

— корректность обработки текстовых данных и выбор подходящего метода векторизации;

— обоснованность выбора модели машинного обучения и её адаптация к задаче;

— наличие и качество локальной валидации для оценки устойчивости модели;

— успешность формирования и отправки предсказаний в требуемом формате (csv: id, pred);

— интерпретация результатов: анализ наиболее важных признаков, выявление типичных ошибок модели;

— соблюдение ограничений (например, запрет на использование предобученных эмбеддингов);

— читаемость кода и структура ноутбука: наличие комментариев, логичная последовательность шагов.

Соревнование 3.

Цель соревнования: применить базовые методы компьютерного зрения для классификации изображений предметов одежды по категориям.

Задание:

1. по изображениям размером 28x28 в оттенках серого определить категорию одежды (футболка, брюки, платье и др.);

2. загрузить и нормализовать данные, обучить модель (логистическая регрессия, случайный лес или простая нейросеть), провести валидацию и отправить предсказания в систему.

Критерии оценивания:

— корректность загрузки и предобработки изображений (нормализация, формат);

— обоснованность выбора архитектуры модели и её адаптация к задаче;

— наличие и качество локальной валидации (разделение данных, оценка переобучения);

— успешность отправки решения в требуемом формате (csv: id, pred);

— интерпретация результатов: анализ ошибок, визуализация предсказаний, сравнение с базовым уровнем;

— соблюдение принципов воспроизводимости: фиксация случайных зерен, читаемость кода;

— использование или отказ от аугментации изображений (опционально, с пояснением).

Примерные темы и критерии оценивания для доклада

Темы докладов

1. Основные платформы для соревнований по анализу данных Kaggle Competitions ai Zindi и их особенности
2. Типичная структура соревнования постановка задачи данные метрика лидерборд сроки
3. Стратегии участия в соревнованиях от baseline до medal как выстраивать путь улучшения
4. Разведочный анализ данных EDA в табличных соревнованиях выявление аномалий распределений и скрытых паттернов
5. Методы feature engineering для табличных данных создание агрегатов взаимодействий временных признаков
6. Борьба с утечкой данных data leakage в соревнованиях типичные ошибки и способы их избежания
7. Кросс-валидация в соревновательных задачах стратегии разбиения stratified time-series group
8. Ансамблирование моделей стекинг блендинг и их применение для улучшения качества
9. Использование LightGBM XGBoost и CatBoost в табличных соревнованиях сравнение настройка особенности
10. Предобработка текстовых данных токенизация нормализация удаление шума в NLP-соревнованиях
11. Использование предобученных эмбеддингов и трансформеров BERT RoBERTa в текстовых соревнованиях
12. Аугментация текстовых данных техники для увеличения обучающей выборки в NLP
13. Предобработка изображений нормализация аугментация центрирование в CV-соревнованиях
14. Использование предобученных моделей ResNet EfficientNet и fine-tuning в соревнованиях по компьютерному зрению
15. Современные тренды в соревновательном анализе данных AutoML нейросетевые архитектуры MLOps в соревнованиях

Критерии оценивания доклада

— соответствие содержания доклада заявленной теме и глубина раскрытия ключевых аспектов;

— логичность структуры выступления: наличие введения, основной части, выводов и примеров;

— использование актуальных примеров из реальных соревнований (kaggle, competitions.ai и др.) для иллюстрации идей;

— качество визуального сопровождения: читаемость слайдов, уместность графиков, таблиц, схем;
— грамотность изложения: отсутствие фактических ошибок, корректное использование терминологии;
— умение отвечать на вопросы и вести дискуссию по теме доклада;
— соблюдение регламента выступления (обычно 7–10 минут).

Примерное описание задания и критерии оценивания к проекту

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/	ПК-1

		сегментация/ матчинг	
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