
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол № 2

**Рабочая программа дисциплины (модуля)
«Машинное обучение в структурной биоинформатике»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: III в биотехе

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2026

**Москва
2026**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	5
3. Тематический план.....	9
4. Содержание дисциплины (модуля).....	9
5. Учебно-методическое обеспечение	11
6. Материально-техническое обеспечение	11
7. Методические и оценочные материалы	13

1. Краткая характеристика дисциплины (модуля)

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

Изучение дисциплины (модуля) «Машинное обучение в структурной биоинформатике» позволяет сформировать у студентов современные цифровые компетенции, необходимые для анализа сложных структурных биологических данных, включая трёхмерные модели белков, ДНК и РНК; обеспечивает междисциплинарную подготовку, объединяющую биологию, биоинформатику и информатику, и способствует эффективному применению методов машинного обучения в задачах предсказания функций молекул, дизайна лекарств и анализа белковых взаимодействий.

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 ИИ в биотехе и входит в обязательную часть Блока 1 как дисциплина по выбору.

Дисциплина (модуль) изучается на 1 курсе в 1 семестре.

Цель изучения дисциплины (модуля): формирование студентов системного понимания методов машинного обучения и развитие практических навыков их применения для анализа пространственных структур биомолекул, прогнозирования свойств белков и нуклеиновых кислот, а также выявления закономерностей в структурных биологических данных.

Задачи изучения дисциплины (модуля):

- освоить этапы жизненного цикла проектов по машинному обучению в контексте структурной биоинформатики;
- изучить принципы работы ключевых алгоритмов и методов оценки их качества применительно к биомолекулярным данным;
- научиться проводить разведочный анализ структурных данных, генерировать признаки на основе 3д-моделей и предотвращать утечки информации;
- овладеть навыками построения и валидации моделей для задач классификации, регрессии, кластеризации и прогнозирования взаимодействий биомолекул;
- развить компетенции в интерпретации моделей и применении машинного обучения для решения задач вычислительной биологии и структурной биоинформатики.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- основы методологии науки о данных, машинного обучения, включая этапы жизненного цикла ML-проектов;
- принципы работы ключевых алгоритмов: линейные модели, kNN, деревья решений, ансамбли (Random Forest, бустинг), кластеризация и снижение размерности;
- метрики качества, функции потерь и методы валидации моделей для задач регрессии, классификации и аплифт-моделирования;

- концепции смещения и разброса, переобучения, регуляризации, а также подходы к интерпретации моделей (SHAP, LIME);
- типы признаков и методы работы с ними, включая генерацию, отбор и предотвращение утечек данных.

уметь:

- проводить разведочный анализ данных (EDA), формулировать гипотезы и готовить данные для моделирования;
- обучать, настраивать и оценивать модели машинного обучения с использованием кросс-валидации, пайплайнов и метрик;
- реализовывать задачи кластеризации, детекции аномалий и снижения размерности, интерпретируя результаты;
- строить и валидировать аплифт-модели, а также генерировать признаки для сложных данных;
- интерпретировать модели и визуализировать влияние признаков с помощью SHAP, LIME и других инструментов;
- применять базовые методы машинного обучения для решения задач структурной и вычислительной биологии;
- использовать методы машинного обучения и искусственного интеллекта для анализа биомолекулярных структур и биоинформатических данных;
- разрабатывать и адаптировать модели машинного обучения с учетом особенностей структурных биологических данных, включая белки, нуклеиновые кислоты и их взаимодействия.

владеть:

- инструментами Python для анализа и моделирования: pandas, numpy, scikit-learn;
- продвинутыми библиотеками ML: XGBoost, CatBoost, lightgbm, а также SHAP, LIME для интерпретации;
- навыками построения воспроизводимых ML-пайплайнов и работы с разнотипными данными;
- методами предобработки, валидации и оптимизации моделей в реальных задачах;
- практиками применения и интерпретации моделей в бизнес-контексте, включая причинно-следственный анализ;
- навыками построения моделей для решения простых задач структурной биоинформатики, включая анализ и интерпретацию структурных биологических данных.

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 критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1.	Знает основные принципы системного подхода и методы критического анализа, а также теоретические основы стратегического планирования и принятия решений
		УК-1.2.	Умеет применять методы системного анализа для выявления ключевых проблемных ситуаций, формулировать гипотезы и разрабатывать альтернативные стратегии действий на основе полученных данных
		УК-1.3.	Имеет практический опыт в проведении анализа реальных проблемных ситуаций в рамках проектов, способен вырабатывать и обосновывать стратегии действий, учитывая различные аспекты и последствия
УК-2.	Способен управлять проектом на всех этапах его жизненного цикла	УК-2.1.	Знает основные методологии управления проектами, ключевые этапы жизненного цикла проекта и инструменты для планирования и контроля
		УК-2.2.	Умеет разрабатывать проектную документацию, устанавливать цели и задачи проекта, а также эффективно распределять ресурсы и управлять рисками на всех этапах его реализации
		УК-2.3.	Имеет практический опыт в управлении реальными проектами, включая планирование, исполнение и завершение, а также в оценке результатов и проведении анализа успешности проекта
ОПК-1.	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и	ОПК-1.1.	Знает основные методы и подходы к решению задач в области биологии, медицины и биотехнологий, включая

	компьютерной математики		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 в области биологии, медицины, биотехнологий, 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 анализ биологических данных, моделирование процессов и обработку экспериментальных результатов.
ОПК-2.	Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы	ОПК-2.1.	Знает основные математические модели и методы, используемые в области биологии, медицины, биотехнологий, включая статистическое моделирование, дифференциальные уравнения и численные методы, а также современные подходы к исследованию и анализу данных
		ОПК-2.2.	Умеет разрабатывать и адаптировать математические модели для решения конкретных проблем в области биологии, медицины, биотехнологий, проводить их анализ и верификацию, а также интерпретировать полученные результаты в контексте научных исследований
		ОПК-2.3.	Имеет практический опыт создания и исследования математических моделей в рамках научных проектов или исследований, включая участие в публикациях, конференциях или коллаборациях, где были разработаны и апробированы новые концепции и методы

ПК-2.	Способен математически корректно ставить и решать естественнонаучные и прикладные задачи в области биологии, медицины или биотехнологий	ПК-2.1.	Знает основные методы математического моделирования и теоретические основы современной биологии и биоинформатики, необходимые для корректной постановки задач
		ПК-2.2.	Умеет формулировать математические модели биологических процессов, выбирать адекватные методы решения и интерпретировать результаты в профессиональном контексте
		ПК-2.3.	Имеет практический опыт постановки и решения прикладных задач с использованием методов ИИ в области биологии, медицины или биотехнологий
ПК-3.	Способен решать задачи профессиональной деятельности и применять методы искусственного интеллекта в прикладных и фундаментальных исследованиях в области биологии, медицины или биотехнологий	ПК-3.1.	Знает принципы и методы решения профессиональных задач, включая применение профильно-специализированных знаний в области ИИ для фундаментальных и прикладных исследований в области биологии, медицины или биотехнологий
		ПК-3.2.	Умеет применять математические и компьютерные методы, формулировать результаты и анализировать их последствия для научных и практических решений
		ПК-3.3.	Имеет опыт участия в проектах, где методы ИИ применялись для решения исследовательских и прикладных задач в области биологии, медицины или биотехнологий
ПК-5.	Способен передавать результаты прикладных задач в виде профессиональных рекомендаций в области искусственного интеллекта, биоинформатики и биотехнологии	ПК-5.1.	Знает методы формулирования рекомендаций и терминологию области искусственного интеллекта и биоинформатики
		ПК-5.2.	Умеет интерпретировать результаты анализа и моделирования и формулировать рекомендации для практического применения в области биологии, медицины или биотехнологий

		ПК-5.3.	Имеет опыт разработки рекомендаций по результатам проектов в области искусственного интеллекта и биотехнологии
ПК-6.	Способен разрабатывать программное обеспечение для решения прикладных задач в области искусственного интеллекта, биоинформатики и биотехнологии	ПК-6.1.	Знает основные языки программирования, методы разработки программного обеспечения, а также принципы проектирования и архитектуры программных систем, применяемых в области искусственного интеллекта, биоинформатики и биотехнологии
		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 области искусственного интеллекта, биоинформатики и биотехнологии

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостояте льная работа	
		Лекции	Семинары (практичес кие занятия)			
1	Основные термины машинного обучения и смежных областей. Оценка качества моделей	16	8	1	22	Домашнее задание Коллоквиум Проект
2	Линейные модели	10	5	2	20	Домашнее задание Коллоквиум Проект
3	Деревья решений и модели, основанные на них	17	8	1	23	Домашнее задание Коллоквиум Проект
4	Обучение без учителя и работа с признаками	17	9	2	23	Домашнее задание Коллоквиум Проект
	Экзамен			6		
Итого:		60	30	12	88	
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	Основные термины машинного обучения и смежных областей. Оценка качества моделей	Жизненный цикл машинного обучения. Метрики и функционалы потерь в задаче регрессии Метрики и функционалы потерь в задаче классификации. Контроль качества и выбор модели Метрические алгоритмы
2	Линейные модели	Линейная регрессия Логистическая регрессия, SVM
3	Деревья решений и модели, основанные на них	Решающие деревья Сложность алгоритмов. Ансамбли алгоритмов Случайные леса Градиентный бустинг Аплифт-моделирование
4	Обучение без учителя и работа с признаками	Кластеризация и методы понижения размерности Обнаружение аномалий Генерация и отбор признаков Интерпретация моделей и диагностика сдвига данных

		Подведение итогов курса. Частые вопросы на собеседованиях по машинному обучению
--	--	---

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. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. — 2-е изд. — Москва : ДМК Пресс, 2023. — 401 с. — ISBN 978-5-89818-300-4. — URL: <https://znanium.ru/catalog/product/2085038>.

2. Мэрфи, К. П. Вероятностное машинное обучение: введение : практическое руководство / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. — Москва : ДМК Пресс, 2023. — 990 с. — ISBN 978-5-93700-119-1. — URL: <https://znanium.ru/catalog/product/2109489>.

Дополнительная литература:

1. Гудфеллоу, И. Глубокое обучение / И. Гудфеллоу, Й. Бенджио, А. Курвилль ; пер. с англ. — Москва : ДМК Пресс, 2018. — 652 с. — ISBN 978-5-97060-618-6. — URL: <https://znanium.ru/catalog/product/2109488>.

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
Электронный документ

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) «Машинное обучение в структурной биоинформатике» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, коллоквиум, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

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

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

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.

Коллоквиум – устные ответы на вопросы, список которых известен студенту заранее.

В процессе подготовки к коллоквиуму необходимо проанализировать учебные материалы, ознакомившись с лекциями, учебниками и дополнительными источниками, акцентируя внимание на ключевых темах. Рекомендуется создать структурированные конспекты, выделяя основные идеи, термины и формулы.

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.

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) «Машинное обучение в структурной биоинформатике»

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 **экзамена**, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, 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 (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя
9	Отлично	
8	Отлично	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
		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.
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать 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	
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 задаст дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Машинное обучение в структурной биоинформатике» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	30%	10	Устные ответы на вопросы, список которых известен студенту заранее
Коллоквиум	30%	3	Набор задач по темам недели

Активность	Вес	Количество	Описание
Проект	20%	1	Исследовательская работа по дисциплине (модулю) и презентация результатов
Экзамен	20%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Машинное обучение в структурной биоинформатике»: $0,3 \times \text{Среднее за домашние задания} + 0,3 \times \text{Коллоквиум} + 0,2 \times \text{Проект} + 0,2 \times \text{Экзамен}$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1. Прогнозирование стабильности белковых мутаций

1. Используйте набор данных с известными мутациями и их влиянием на стабильность белка (например, из ProThermDB).
2. Сформируйте признаки: тип аминокислоты, её свойства (заряд, гидрофобность), положение в структуре (ядро/поверхность), консервативность.
3. Постройте модель регрессии для предсказания изменения энергии стабильности ($\Delta\Delta G$).
4. Оцените качество модели (MSE, R^2) с помощью перекрёстной проверки.
5. Оформите отчёт с анализом наиболее важных признаков.

Домашнее задание 2. Классификация вторичной структуры белка по аминокислотной последовательности

1. Возьмите данные из базы PDB с аннотированной вторичной структурой (α -спирали, β -слои, нерегулярные участки).
2. Преобразуйте последовательность в окна по 7–15 аминокислот, закодируйте их (one-hot, BLOSUM).
3. Обучите модель (логистическая регрессия, случайный лес) для предсказания типа структуры.
4. Оцените качество (ассигасу, F1) и сравните с baseline.
5. Проанализируйте, какие признаки наиболее информативны.

Домашнее задание 3. Кластеризация белковых доменов по структурному сходству

1. Используйте данные о 3д-структурах белковых доменов (например, из CATH или SCOP).
2. Извлеките признаки: RMSD, матрицы расстояний между α -углеродами, векторы из PCA.
3. Примените методы кластеризации (k-средних, иерархическая) и оцените качество (силуэт).
4. Сопоставьте кластеры с известными суперсемействами.
5. Сделайте вывод о соответствии структурного и эволюционного сходства.

Примерные вопросы к коллоквиуму

1. Какие признаки можно извлечь из 3д-структуры белка для задач машинного обучения?
2. Как оценить качество предсказания изменения стабильности белка при мутации?
3. Почему важно учитывать структурное окружение при генерации признаков для аминокислот?
4. Как избежать утечки данных при использовании структурных баз (PDB, AlphaFold)?
5. Как методы понижения размерности (PCA, t-SNE) могут помочь в анализе ансамблей конформаций белков?
6. Как можно использовать кластеризацию для анализа динамики белков (например, по данным молекулярной динамики)?
7. Какие ограничения имеет применение линейных моделей к структурным данным?
8. Почему градиентное бустинговое может быть эффективно для предсказания функций белков?
9. Как интерпретировать важность признаков в модели, предсказывающей связывание белка с лигандом?
10. Как аплифт-моделирование может быть применено в контексте дизайна мутаций с заданным эффектом?

Критерии оценивания участия в коллоквиуме:

- Полнота и точность ответа на поставленный вопрос.
- Умение использовать терминологию машинного обучения и биоинформатики корректно.
- Способность логично излагать мысли и строить аргументированный ответ.
- Наличие примеров из биологических или медицинских задач.
- Готовность к диалогу, ответам на уточняющие вопросы и критике собственных рассуждений.

Примерное описание задания и критерии оценивания к проекту

«Предсказание влияния мутаций на связывание белка с ДНК с использованием методов машинного обучения»

Описание задания: студент разрабатывает модель для предсказания, усилит или ослабит ли мутация в ДНК-связывающем домене белка его аффинность к целевой последовательности. Проект включает сбор данных, генерацию признаков на основе структуры, построение и интерпретацию модели.

Этапы выполнения:

- Сбор данных из баз (PDB, HTRI, UniProt) о белках, связывающихся с ДНК, и их мутациях.
- Извлечение структурных и физико-химических признаков: расстояние до ДНК, заряд, консервативность, гидрофобность.
- Построение бинарной классификации: «усиление» vs «ослабление» связывания.
- Обучение и сравнение моделей (логистическая регрессия, случайный лес, XGBoost).

- Интерпретация с помощью SHAP, выделение ключевых признаков.
- Подготовка отчёта (8–10 страниц) и презентации (10 слайдов).

Критерии оценивания:

- Актуальность и обоснованность постановки задачи.
- Качество и корректность извлечения признаков из структурных данных.
- Обоснованность выбора моделей и метрик, отсутствие утечек данных.
- Глубина интерпретации и биологическая значимость результатов.
- Чёткость изложения, логичность структуры отчёта и презентации.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Какой тип данных используется для анализа 3D-структуры белка?	Координаты атомов	УК-1
2.	Как называется база данных, содержащая трёхмерные структуры биомолекул?	PDB (Банк данных белков)	УК-2
3.	Какой метод позволяет снизить размерность структурных данных?	Метод главных компонент	ОПК-1
4.	Какой признак может указывать на участок белка, участвующий в связывании с ДНК?	Положительный заряд поверхности	ОПК-2
5.	Какой функционал потерь используется для бинарной классификации?	Логарифмический потерь	ПК-2
6.	Какой алгоритм строит композицию деревьев на ошибках предыдущих?	Градиентное бустингование	ПК-3
7.	Какой метод позволяет оценить вклад признаков в предсказание модели?	SHAP	ПК-5
8.	Какой алгоритм кластеризации требует задания числа кластеров?	К-средних	ПК-6
9.	Как называется процесс автоматического выявления ошибок в структуре белка?	Обнаружение аномалий	ОПК-1
10.	Какой тип регуляризации удаляет незначимые признаки?	L1-регуляризация	ПК-2
11.	Какой показатель измеряет долю правильно предсказанных положительных случаев?	Точность (precision)	УК-1
12.	Как называется этап, на котором проверяется, не зависит ли модель от утечек данных?	Диагностика сдвига данных	УК-2
13.	Какой метод позволяет предсказать вторичную структуру по последовательности?	Модель на основе окон	ОПК-2
14.	Какой инструмент визуализирует влияние признаков на предсказание?	LIME	ПК-3
15.	Какой алгоритм подходит для задачи классификации типов белковых доменов?	Случайный лес	ПК-5

16.	Как называется модель, объединяющая множество деревьев по подвыборкам?	Случайный лес	ПК-6
17.	Какой метод позволяет сгенерировать признаки на основе соседних аминокислот?	Скользящее окно	ОПК-1
18.	Какой показатель используется для оценки регрессии в задаче $\Delta\Delta G$?	R^2	ПК-2
19.	Какой подход позволяет оценить эффект мутации на функцию белка?	Аплифт-моделирование	ПК-3
20.	Какой метод позволяет визуализировать кластеры структур белков?	t-SNE	ОПК-2