
УТВЕРЖДЕНА

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

**Рабочая программа дисциплины (модуля)
«Хемоинформатика и машинное обучение»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: III в биотехе

Специализация: Структурная биоинформатика

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.....	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) «Хемоинформатика и машинное обучение» 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) «Хемоинформатика и машинное обучение» позволяет эффективно использовать вычислительные технологии для ускорения поиска перспективных соединений, прогнозирования их свойств и биологической активности.

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 специализации «Структурная биоинформатика».

Дисциплина (модуль) изучается на 2 курсе в 3 семестре.

Цель изучения дисциплины (модуля): формирование у обучающихся системных знаний и практических навыков в области хемоинформатики и машинного обучения, необходимых для анализа химических данных, построения и валидации моделей «структура – активность/свойство» (QSAR/QSPR), а также разработки вычислительных подходов в задачах поиска новых соединений и предсказания их свойств с использованием современных программных инструментов и методов искусственного интеллекта.

Задачи изучения дисциплины (модуля):

- освоить основные принципы представления химических структур в цифровом виде и работу с форматами SMILES, InChI, молекулярными графами и 3д-моделями;
- научиться вычислять и интерпретировать различные типы молекулярных дескрипторов для описания химических соединений;
- изучить алгоритмы машинного обучения, применяемые в хемоинформатике, и освоить их реализацию на языке Python с использованием специализированных библиотек;
- овладеть методами предобработки, валидации и оценки качества моделей, включая кросс-валидацию и метрики для задач классификации и регрессии;
- развить навыки работы с открытыми химическими базами данных, визуализации структур и результатов моделирования, а также подготовки научно обоснованных отчетов по вычислительным экспериментам.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- основные понятия и определения хемоинформатики: молекулярные дескрипторы, химические графы, SMILES, InChI, базы химических соединений;
- принципы представления химических структур в цифровом виде (линейные нотации, графовые модели, 2д/3д-структуры);
- классификацию и типы молекулярных дескрипторов (топологические, электронные, геометрические, фармакофорные и др.);
- основы машинного обучения: задачи регрессии, классификации, кластеризации,

понижения размерности;

- основные алгоритмы машинного обучения, применяемые в хемоинформатике: линейная/логистическая регрессия, метод опорных векторов (SVM), случайный лес, градиентный бустинг, нейронные сети;
- принципы построения моделей QSAR/QSPR (количественные соотношения «структура – активность/свойство»);
- методы предобработки данных: нормализация, отбор признаков, борьба с пропущенными значениями и выбросами;
- принципы валидации моделей: кросс-валидация, hold-out, внешние тестовые выборки, метрики качества (R^2 , RMSE, accuracy, precision, recall, ROC-AUC);
- основы работы с химическими базами данных (ChEMBL, PubChem, ZINC и др.);
- этические и нормативные аспекты применения ИИ в химии и фармацевтике.

уметь:

- преобразовывать химические структуры в цифровые представления с использованием библиотек (RDKit, Open Babel и др.);
- вычислять и интерпретировать молекулярные дескрипторы;
- формировать обучающие и тестовые выборки для хемоинформатических задач;
- применять алгоритмы машинного обучения для предсказания свойств и биологической активности соединений;
- оценивать качество построенных моделей и интерпретировать результаты;
- осуществлять визуализацию химических структур и результатов моделирования;
- работать с открытыми химическими базами данных для извлечения и анализа данных;
- проводить валидацию QSAR-моделей в соответствии с принципами OECD;
- использовать языки программирования (Python) и специализированные библиотеки (scikit-learn, DeepChem, Mordred) для решения задач хемоинформатики;
- анализировать и критически оценивать результаты моделирования с точки зрения применимости в реальных задачах (например, разработка лекарств).

владеть:

- навыками программирования на Python в контексте хемоинформатических задач;
- работой с хемоинформатическими библиотеками (RDKit, Mordred, CDK);
- техниками построения и отладки пайплайнов обработки химических данных;
- методами визуализации данных (matplotlib, seaborn, Plotly) и химических структур (NGLview, PyMOL, RDKit в Jupyter);
- навыками работы в средах разработки (Jupyter Notebook, Google Colab, PyCharm);
- подготовкой отчётов и презентаций по результатам вычислительных экспериментов;
- работой с облачными платформами и инструментами автоматизации экспериментов (MLflow, Weights & Biases — опционально);
- навыками командной работы при решении междисциплинарных задач (химия + ИИ).

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
ПК-1.	Способен определять общие формы и закономерности области искусственного	ПК-1.1.	Знает основные теоретические концепции и принципы искусственного интеллекта в биотехнологии, а также методы

	интеллекта в биологии, медицине или биотехнологии, обрабатывать и интерпретировать данные биологических исследований		обработки, анализа и визуализации данных в биологии, медицине или биотехнологии
		ПК-1.2.	Умеет проводить систематический анализ области профессиональной деятельности, выявлять закономерности и тенденции, моделировать основные биологические процессы и интерпретировать полученные результаты с использованием методов биоинформатики и искусственного интеллекта
		ПК-1.3.	Имеет практический опыт работы в области искусственного интеллекта в биотехнологии, включая участие в научных проектах, исследованиях или практических заданиях
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

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	Основы хемоинформатики и подготовка данных	9	9		14	Домашнее задание Тест Мини-проект
2	Методы анализа и построения моделей	8	8		12	Домашнее задание Тест Мини-проект
3	Продвинутые методы и моделирование	8	8		11	Домашнее задание Тест Мини-проект
4	Практическое и этическое применение	5	5		11	Домашнее задание Тест Мини-проект
	Экзамен			6		
Итого:		30	30	6	48	
Объем дисциплины (модуля) (в ак. ч.)		114				
Объем дисциплины (модуля) (в зач. ед.)		3				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Основы хемоинформатики и подготовка данных	Введение в хемоинформатику Цифровое представление химических структур Химические базы данных и источники данных Молекулярные дескрипторы: теория и классификация Работа с библиотеками хемоинформатики (RDKit) Основы машинного обучения для химии Предобработка химических данных
2	Методы анализа и построения моделей	Методы отбора и генерации признаков Построение и валидация QSAR/QSPR-моделей Алгоритмы машинного обучения в хемоинформатике
3	Продвинутые методы и моделирование	Нейронные сети и глубокое обучение в химии Генеративные модели и дизайн новых молекул Кластеризация и анализ химических пространств Прогнозирование ADMET-свойств
4	Практическое и этическое применение	Этические, регуляторные и практические аспекты применения ИИ в химии

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. Биотехнология. Практический курс : учебник и практикум для вузов / под редакцией А. А. Красноштановой. — Москва : Издательство Юрайт, 2026. — 174 с. — (Высшее образование). — ISBN 978-5-534-20448-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/589750>.

2. Биотехнология : учебник и практикум для вузов / под редакцией Н. В. Загоскиной, Л. В. Назаренко. — 4-е изд., испр. и доп. — Москва : Издательство Юрайт, 2025. — 384 с. — (Высшее образование). — ISBN 978-5-534-16026-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567470>.

Дополнительная литература:

1. Иванищев, В. В. Молекулярная биология : учебник / В. В. Иванищев. — 2-е изд. — Москва : РИОР : ИНФРА-М, 2026. — (Высшее образование). — 233 с. — DOI: <https://doi.org/10.29039/01857-6>. - ISBN 978-5-369-01857-6. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2199640>.

2. Стефанов, В. Е. Биоинформатика : учебник для вузов / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. — Москва : Издательство Юрайт, 2026. — 249 с. — (Высшее образование). — ISBN 978-5-534-20194-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/583663>.

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:
Электронный документ

- 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) «Хемоинформатика и машинное обучение» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, тесты, мини-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

Электронный документ

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.

Тест – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины.

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины, понимать смысл научных категорий и уметь их использовать в профессиональной лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания студентами важных методологических категорий.

Мини-проект — это краткосрочная практическая работа в малых группах на семинарах, направленная на применение теоретических знаний для решения конкретных задач.

В ходе мини-проекта студенты проводят анализ проблемы, разрабатывают техническое решение и представляют результаты в виде отчета, модели или презентации. Такой формат способствует развитию навыков самостоятельной работы, критического мышления и командного взаимодействия.

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) «Хемоинформатика и машинное обучение»

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
Электронный документ

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, 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, 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.

Дисциплина (модуль) «Хемоинформатика и машинное обучение» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	20%	10	Набор задач по темам недели
Тесты	10%	10	Форма проверки знаний студентов через ответы на вопросы с выбором из нескольких вариантов ответов
Мини-проект	30%	1	Краткосрочная практическая работа в малых группах на занятиях, направленная на решение конкретных задач
Экзамен	40%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по дисциплине (модулю)

Формула расчёта итоговой оценки по дисциплине (модулю) «Хемоинформатика и машинное обучение»: $0,2 \times \text{Среднее за домашние задания} + 0,1 \times \text{Среднее за тесты} + 0,3 \times \text{Мини-проект} + 0,4 \times \text{Экзамен}$.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1. Работа с химическими структурами и дескрипторами

1. Выберите 10 органических соединений (например, из PubChem или ChEMBL).
2. Преобразуйте их структуры в форматы SMILES и InChI с помощью RDKit.
3. Постройте 2д- и 3д-представления молекул.
4. Вычислите 5 топологических и 3 физико-химических дескриптора (например, LogP, молекулярная масса, TPSA, количество Н-доноров/акцепторов).
5. Оформите результаты в виде таблицы и краткого отчёта с интерпретацией дескрипторов.

Домашнее задание 2. Построение и валидация QSAR-модели

1. Скачайте набор данных (например, активность ингибиторов киназы из ChEMBL).
2. Очистите данные: удалите дубликаты, пропущенные значения, стандартизируйте структуры.
3. Вычислите молекулярные дескрипторы с помощью Mordred или RDKit.
4. Постройте модель регрессии (например, случайный лес или SVM) для предсказания IC_{50} .
5. Оцените качество модели: R^2 , RMSE, кросс-валидация.
6. Проанализируйте важность признаков и оформите отчёт.

Домашнее задание 3. Прогнозирование ADMET-свойств

1. Используя готовый датасет (например, Tox21 или ADME dataset из DeepChem), выберите одно свойство (например, растворимость, проницаемость через BBB, токсичность).
2. Подготовьте данные: нормализация, кодирование, разделение на обучающую и тестовую выборки.
3. Обучите модель классификации (например, логистическая регрессия, градиентный бустинг).

4. Оцените качество: accuracy, precision, recall, ROC-AUC.
5. Визуализируйте ROC-кривую и важность признаков.
6. Сделайте выводы о применимости модели.

Примерные задания для теста

Вопрос 1.

Какой формат позволяет однозначно кодировать химическую структуру в виде строки?

1. PNG
2. SMILES
3. CSV
4. JSON

Правильный ответ: 2

Вопрос 2.

Какой из перечисленных дескрипторов относится к геометрическим?

1. Количество ароматических колец
2. Энергия наивысшей занятой молекулярной орбитали (HOMO)
3. Момент инерции молекулы
4. LogP

Правильный ответ: 3

Вопрос 3.

Какая метрика используется для оценки качества модели регрессии в QSAR?

1. Accuracy
2. ROC-AUC
3. R^2
4. Precision

Правильный ответ: 3

Вопрос 4.

Какой алгоритм машинного обучения часто используется для классификации активности соединений?

1. Метод главных компонент (PCA)
2. К-средних (k-means)
3. Случайный лес (Random Forest)
4. Линейная регрессия

Правильный ответ: 3

Вопрос 5.

Какая база данных содержит информацию о биологической активности тысяч соединений?

1. UniProt
2. ChEMBL
3. KEGG
4. PDB

Правильный ответ: 2

Примерное описание задания и критерии оценивания к мини-проекту

Выполните вычислительный проект по построению и анализу QSAR-модели для
Электронный документ

предсказания биологической активности или физико-химического свойства соединений.

Этапы проекта:

- Выбор задачи: активность против белка, растворимость, токсичность и др.
- Сбор данных из открытых источников (ChEMBL, PubChem, DeepChem).
- Предобработка: стандартизация структур, расчёт дескрипторов, очистка данных.
- Построение модели (регрессия/классификация) с использованием scikit-learn или DeepChem.
- Валидация: кросс-валидация, hold-out, оценка метрик.
- Интерпретация результатов: анализ важности признаков, визуализация.
- Оформление: Jupyter-ноутбук или отчёт (5–7 стр.) с кодом, графиками и выводами.
- Формат: Jupyter-ноутбук с комментариями + презентация (5–7 слайдов) или письменный отчёт.

Критерии оценивания мини-проекта:

- Качество сбора и подготовки данных — корректная стандартизация, очистка, расчёт дескрипторов.
- Обоснованность выбора модели и алгоритма — адекватный выбор метода в зависимости от задачи.
- Качество валидации и оценки модели — использование правильных метрик, кросс-валидации, внешней выборки.
- Глубина анализа и интерпретации — анализ важности признаков, визуализация, выводы о применимости модели.
- Оформление и воспроизводимость — читаемость кода, наличие комментариев, логичность структуры, возможность повторения эксперимента.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Что означает аббревиатура QSAR?	Количественное соотношение «структура – активность»	УК-1
2.	Какой формат позволяет однозначно восстановить химическую структуру из строки?	InChI	УК-2
3.	Какой тип дескриптора учитывает распределение электронной плотности?	Электронный	ПК-1
4.	Что такое TPSA?	Полярная поверхность молекулы, связанная с проницаемостью	ПК-2
5.	Какой алгоритм используется для построения дерева решений?	CART	ПК-3
6.	Что означает метрика ROC-AUC?	Площадь под ROC-кривой, мера	ПК-5

		качества бинарной классификации	
7.	Какой метод понижает размерность данных, сохраняя дисперсию?	PCA (метод главных компонент)	УК-1
8.	Какой библиотека Python используется для хемоинформатики?	RDKit	УК-2
9.	Что такое overfitting?	Переобучение модели на обучающих данных	ПК-1
10.	Какой принцип OECD важен для валидации QSAR-моделей?	Наличие внешней тестовой выборки	ПК-2
11.	Какой дескриптор отражает липофильность молекулы?	LogP	ПК-3
12.	Что означает hold-out валидация?	Разделение данных на обучающую и тестовую выборки один раз	ПК-5
13.	Какой тип нейронной сети подходит для обработки графов молекул?	GNN (графовые нейронные сети)	УК-1
14.	Какой метод используется для генерации новых молекул?	Генеративные модели (VAE, GAN, трансформеры)	УК-2
15.	Что такое молекулярный граф?	Модель молекулы в виде вершин (атомы) и рёбер (связи)	ПК-1
16.	Какой база данных содержит миллионы биологически активных соединений?	ChEMBL	ПК-2
17.	Что означает precision в классификации?	Доля правильно предсказанных положительных объектов среди всех предсказанных	ПК-3
18.	Какой алгоритм часто используется для отбора признаков?	LASSO (регуляризация L1)	ПК-5
19.	Что такое фармакофор?	Набор структурных фрагментов, ответственных за биологическую активность	УК-1

20.	Почему важно учитывать этические аспекты при использовании ИИ в химии?	Чтобы избежать создания токсичных или контролируемых веществ	УК-2
-----	--	--	------