

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное
образовательное учреждение высшего образования
«Казанский (Приволжский) федеральный университет»
Институт физики



УТВЕРЖДАЮ

Проректор по образовательной
деятельности КФУ

Е.А. Турилова

20__ г.

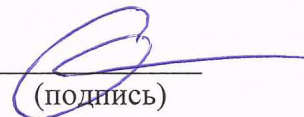


**Дополнительная профессиональная программа
профессиональной переподготовки
Информатика новых материалов**

Утверждена Учебно-методической комиссией Института физики КФУ

(протокол № 9 от «04» мая 2023 г.)

Председатель Учебно-методической комиссией Недопекин Олег Владимирович


(подпись)

Руководитель подразделения,
реализующего ДПП ПП


(подпись)

М.Р. Гафуров
(инициалы, фамилия)

«__» _____ 20__ г.

Казань, 2023 г.

I. Общие положения

1. Дополнительная профессиональная программа (программа профессиональной переподготовки) ИТ-профиля «Информатика новых материалов» (далее – Программа) разработана в соответствии с нормами Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации», с учетом требований приказа Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», с изменениями, внесенными приказом Минобрнауки России от 15 ноября 2013 г. № 1244 «О внесении изменений в Порядок организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденный приказом Министерства образования и науки Российской Федерации от 1 июля 2013 г. № 499», приказа Министерства образования и науки РФ от 23 августа 2017г. N 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ» (указать при необходимости); паспорта федерального проекта «Развитие кадрового потенциала ИТ-отрасли» национальной программы «Цифровая экономика Российской Федерации»; постановления Правительства Российской Федерации от 13 мая 2021 г. № 729 «О мерах по реализации программы стратегического лидерства «Приоритет-2030» (в редакции постановления Правительства Российской Федерации от 14 марта 2022 г. № 357 «О внесении изменений в постановление Правительства Российской Федерации от 13 мая 2021 г. № 729»); приказа Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации от 28 февраля 2022 г. № 143 «Об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» и признании утратившими силу некоторых приказов Министерства цифрового

развития, связи и массовых коммуникаций Российской Федерации об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» (далее – приказ Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации № 143); федерального государственного образовательного стандарта 09.03.02 Информационные системы и технологии (далее вместе – ФГОС ВО), а также Профессионального стандарта 06.001 «Программист», утвержденного приказом Министерства труда и социальной защиты Российской Федерации от 20.07.2022 № 424н.

2. Профессиональная переподготовка заинтересованных лиц (далее – Слушатели), осуществляемая в соответствии с Программой (далее – Подготовка), имеющей отраслевую направленность «Информационно-коммуникационные технологии», проводится в ФГАОУ ВО "Казанский (Приволжский) федеральный университет" (далее – Университет) в соответствии с учебным планом в очной форме обучения с применением электронного обучения и дистанционных образовательных технологий.

3. Разделы, включенные в учебный план Программы, используются для последующей разработки календарного учебного графика, учебно-тематического плана, рабочей программы, оценочных и методических материалов. Перечисленные документы разрабатываются Университетом самостоятельно, с учетом актуальных положений законодательства об образовании, законодательства в области информационных технологий и смежных областей знаний ФГОС ВО и профессионального стандарта 06.001 «Программист».

4. Программа регламентирует требования к профессиональной переподготовке в области разработки программного обеспечения для информационно-вычислительных систем различного назначения.

Срок освоения Программы составляет 9 месяцев, трудоемкость – 252 часа.

К освоению Программы в рамках проекта допускаются лица:

- получающие высшее образование по очной форме, лица, освоившие основную профессиональную образовательную программу (далее – ОПОП

ВО) бакалавриата – в объеме не менее первого курса (бакалавры 2-го курса),

ОПОП ВО специалитета – не менее первого и второго курсов (специалисты 3-го курса), а также магистратуры, обучающиеся по ОПОП ВО, не отнесенным к ИТ-сфере.

5. Область профессиональной деятельности 06. Связь, информационные и коммуникационные технологии (в сфере индустриального производства программного обеспечения для информационно-вычислительных систем различного назначения).

II. Цель

6. Целью подготовки слушателей по Программе является получение компетенции, необходимой для выполнения нового вида профессиональной деятельности в области информационных технологий в сфере индустриального производства программного обеспечения для информационно-вычислительных систем различного назначения; приобретение новой квалификации – «Программист» обучающимися по специальностям и направлениям подготовки, не отнесенным к ИТ-сфере.

III. Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и (или) уровней квалификации

7. Виды профессиональной деятельности, трудовая функция, указанные в профессиональном стандарте по соответствующей должности «Программист», представлены в таблице 1:

Таблица 1

Характеристика новой квалификации, связанной с видом профессиональной деятельности и трудовыми функциями в соответствии с профессиональным стандартом 06.001 «Программист»

Область профессиональной деятельности	Тип задач профессиональной деятельности	Код и наименование профессиональной компетенции	Трудовые действия	Трудовая функция	Обобщенная трудовая функция	Вид профессиональной деятельности
<p>Связь, информационные и коммуникационные технологии (в сфере индустриального производства программного обеспечения для информационно-вычислительных систем различного назначения).</p>	<p>Проектный</p>	<p>ПК -1. Способен разрабатывать программный код на основе поставленной задачи.</p>	<p>Разработка алгоритмов решения поставленных задач в соответствии с требованиями технического задания Создание программного кода в соответствии с техническим заданием (готовыми спецификациями) Структурирование исходного программного кода Комментирование и разметка программного кода в соответствии с установленными в организации требованиями Форматирование исходного программного кода в соответствии с установленными в организации требованиями Анализ и проверка исходного программного кода Отладка программного кода на уровне программных модулей Сохранение сделанных изменений программного кода в соответствии с регламентом контроля версий</p>	<p>A/01.3. Формализация и алгоритмизация поставленных задач A/02.3. Написание программного кода с использованием языков программирования, определения и манипулирования данными A/03.3. Оформление программного кода в соответствии с установленными требованиями A/04.3. Проверка и отладка программного кода</p>	<p>A. Разработка и отладка программного кода</p>	<p>Разработка программного обеспечения</p>

	<p><i>Проектный</i></p>	<p>ПК-2. <i>Способен применять информационные технологии для решения профессиональных задач в сфере естественных наук</i></p>	<p><i>Разработка процедуры проверки работоспособности программного обеспечения</i> <i>Разработка процедуры сбора диагностических данных</i> <i>Разработка процедуры измерения требуемых характеристик программного обеспечения</i> <i>Подготовка тестовых наборов данных в соответствии с выбранной методикой</i> <i>Проверка работоспособности программного обеспечения на основе разработанных тестовых наборов данных</i> <i>Сбор и анализ полученных результатов проверки работоспособности программного обеспечения</i> <i>Анализ программного кода на соответствие требованиям по читаемости и производительности</i> <i>Установление причин возникновения дефектов</i></p>	<p><i>В/01.4. Разработка процедур проверки работоспособности и измерения характеристик программного обеспечения</i> <i>В/02.4. Разработка тестовых наборов данных</i> <i>В/03.4. Проверка работоспособности программного обеспечения</i> <i>В/04.4. Рефакторинг и оптимизация программного кода</i> <i>В/04.5. Исправление дефектов, зафиксированных в базе данных</i> <i>дефектов</i></p>	<p><i>В. Проверка работоспособности и рефакторинг кода программного обеспечения</i></p>	<p><i>Проверка работоспособности программного обеспечения</i></p>
--	-------------------------	--	--	---	---	---

Таблица 2

Характеристика новой и развиваемой цифровой компетенции в ИТ-сфере, связанной с уровнем формирования и развития в результате освоения Программы «Информатика новых материалов»

Наименование сферы	Код и наименование профессиональной компетенции	Пример инструментов	0 — способность не проявляется/ проявляется в степени, недостаточной для отнесения к 1 уровню сформированности компетенции	1 — способность проявляется под внешним контролем / при внешней постановке задачи/ обучающийся пользуется готовыми, рекомендованным и продуктами	2 — способность проявляется, но обучающийся эпизодически прибегает к экспертной консультации/ самостоятельно подбирает и пользуется готовыми продуктами	3 — способность проявляется системно / обучающийся модифицирует способность под определенные задачи / создает новый продукт, обучает других
<i>Искусственный интеллект и машинное обучение</i>	ПК-3. <i>Применяет Искусственный интеллект и машинное обучение (id-37)</i>	<i>Python</i>	(+)	(+)	(-)	(-)

<i>Средства программной разработки</i>	ПК-4. <i>Применяет языки программирования для решения профессиональных задач (id-28)</i>	<i>Python</i>	(+)	(+)	(+)	(-)
--	---	---------------	-----	-----	-----	-----

IV. Характеристика новых и развиваемых цифровых компетенций, формирующихся в результате освоения программы

8. В ходе освоения Программы Слушателем приобретаются следующие профессиональные компетенции:

— ПК-1. Способен разрабатывать программный код на основе поставленной задачи;

— ПК-2. Способен применять информационные технологии для решения профессиональных задач в сфере естественных наук;

9. В ходе освоения Программы Слушателем совершенствуются следующие профессиональные компетенции:

— ПК-3. Применяет Искусственный интеллект и машинное обучение;

— ПК-4. Применяет языки программирования для решения профессиональных задач.

V. Планируемые результаты обучения по ДПП III

10. Результатами подготовки слушателей по Программе является получение компетенции, необходимой для выполнения нового вида профессиональной деятельности в области информационных технологий в сфере индустриального производства программного обеспечения для информационно-вычислительных систем различного назначения – обучающимися по специальностям и направлениям подготовки, не отнесенным к ИТ-сфере.

11. В результате освоения Программы слушатель должен:

Знать:

— методы и приемы формализации задач;

— языки формализации функциональных спецификаций;

— методы и приемы алгоритмизации поставленных задач;

- программные продукты для графического отображения алгоритмов;
- алгоритмы решения типовых задач, области и способы их применения;
- синтаксис выбранного языка программирования, особенности программирования на этом языке, стандартные библиотеки языка программирования;
- методологии разработки программного обеспечения;
- особенности выбранной среды программирования и системы управления базами данных;
- нормативные документы, определяющие требования к оформлению программного кода;
- установленный регламент использования системы контроля версий;
- методы и приемы отладки программного кода;
- языки, утилиты и среды программирования, и средства пакетного выполнения процедур;
- типовые метрики программного обеспечения;
- методы создания и документирования контрольных примеров и тестовых наборов данных;
- методы и средства проверки работоспособности программного обеспечения;
- языки программирования и среды разработки;
- типовые ошибки, возникающие при разработке программного обеспечения, и методы их диагностики и исправления.

Уметь:

- использовать методы и приемы алгоритмизации поставленных задач;
- использовать программные продукты для графического отображения алгоритмов;
- применять стандартные алгоритмы в соответствующих областях;

- применять выбранные языки программирования для написания программного кода;
- применять нормативные документы, определяющие требования к оформлению программного кода;
- использовать вспомогательные инструментальные программные средства для обработки исходного текста программного кода;
- интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов;
- использовать выбранную среду программирования для разработки процедур проверки работоспособности программного обеспечения на выбранном языке программирования;
- подготовка наборов данных, используемых в процессе проверки работоспособности программного обеспечения;
- интерпретировать диагностические данные (журналы, протоколы и др.);
- применять инструментальные средства коллективной работы над программным кодом;
- интерпретировать сообщения об ошибках, предупреждения, записи технологических журналов, возникающих при выполнении дефектного кода.

Иметь навыки:

- применения языка программирования Python для решения профессиональных задач.

VI. Организационно-педагогические условия реализации ДПП

12. Реализация Программы должна обеспечить получение компетенции, необходимой для выполнения нового вида профессиональной деятельности в области информационных технологий в сфере индустриального производства программного обеспечения для информационно-вычислительных систем различного назначения – обучающимися по специальностям и направлениям

подготовки, не отнесенным к ИТ-сфере.

13. Учебный процесс организуется с применением электронного обучения и дистанционных образовательных технологий, инновационных технологий и методик обучения, способных обеспечить получение слушателями знаний, умений и навыков в области Об. Связь, информационные и коммуникационные технологии (в сфере индустриального производства программного обеспечения для информационно-вычислительных систем различного назначения).

14. Реализация Программы обеспечивается научно-педагогическими кадрами Университета, допустимо привлечение к образовательному процессу высококвалифицированных специалистов ИТ-сферы и/или дополнительного профессионального образования в части, касающейся профессиональных компетенций в области создания алгоритмов и программ, пригодных для практического применения, с обязательным участием представителей профильных организаций-работодателей. Возможно привлечение региональных руководителей цифровой трансформации (отраслевых ведомственных и/или корпоративных) к проведению итоговой аттестации, привлечение работников организаций реального сектора экономики субъектов Российской Федерации. Не менее 50% общего объема аудиторных часов в рамках ДПП ПП реализуются научно-педагогическими работниками, отвечающими следующим критериям:

- наличие высшего профильного образования в ИТ-сфере и/или дополнительного профессионального образования в части, касающейся профессиональных компетенций в области создания алгоритмов и программ, пригодных для практического применения;

- наличие стажа педагогической работы в образовательных организациях высшего образования Российской Федерации и/или стажа практической работы в профильной организации ИТ-отрасли не менее 3 лет.

Не менее 20% от общего объема аудиторных часов в рамках ДПП ПП реализуются лицами, имеющими подтвержденный стаж в профессии в ИТ-

сфере или в отрасли цифровой экономики не менее двух лет, полученный не более четырех лет назад.

№ п/п	Список сотрудников ИТ-компаний	Перечень модулей (с указанием часов), которые они будут реализовывать.
1.	Скобельцын Константин Владимирович - директор ИТ-компания «Научная студия»	Модуль "Машинное обучение" (8 часов) Практика (20 часов)

VII. Учебный план ДПП

15. Объем Программы составляет 9 месяцев, трудоемкость – 252 часа.

16. Учебный план Программы определяет перечень, последовательность, общую трудоемкость разделов и формы контроля знаний.

Учебный план программы профессиональной переподготовки «Информатика новых материалов»

№ п/п	Наименование раздела (модуля)	Общая трудоемкость (252 часов)	Форма контроля
Общий блок			
1.	Основы программирования на Python	22	Зачет
2.	Основы программирования: Объектно-ориентированное программирование	28	Зачет
3.	Машинное обучение	28	Зачет
Вариативная часть			
4.1.	Хемоинформатика	28	Зачет
4.2.	Информатика новых материалов	28	Зачет
4.3.	Биоинформатика	28	Зачет
	Практика	54	Зачет
	Итоговая аттестация	36	Защита проекта
	Итого:	252	

VIII. Календарный учебный график

18. Календарный учебный график представляет собой график учебного процесса, устанавливающий последовательность и продолжительность обучения и итоговой аттестации по учебным дням.

IX. Рабочая программа учебных предметов, курсов, дисциплин (модулей), практик

19. Рабочая программа содержит перечень разделов и тем, а также рассматриваемых в них вопросов с учетом их трудоемкости.

Рабочая программа разрабатывается Университетом с учетом профессионального стандарта 06.001 «Программист».

№ п/п	Наименование и краткое содержание раздела(модуля)	Объем, часов
Общий блок		
1.	Основы программирования на Python	22
1.1.	<p>Введение в программирование на языке Python</p> <p><i>Краткое содержание:</i> Введение в Python. Базовые конструкции Python. Логические операции. Ветвления. Циклы. Базовые типы данных: числа, строки, списки, кортежи, словари, множества. Функции. Итераторы и генераторы. Jupyter Notebook. Подключение библиотек. Элементы функционального программирования (lambda, map, zip, reduce, filter). Импорт и экспорт данных в текстовом формате. Работа с внешними репозиториями: Git, GitHub. Строки. Базовые алгоритмы работы со строками. Списки. Методы работы со списками. Словари. Задачи на использование структуры данных Словарь. Функции. Создание собственных функций на Python. Работа с файлами. Чтение и запись в файлы.</p>	12
1.2.	<p>Специализированные библиотеки Python для анализа данных</p> <p><i>Краткое содержание:</i> Введение в анализ табличных данных в Python. Пакет pandas, NumPy. Объекты Series (последовательность) и DataFrame (таблица). Чтение-запись данных в различных форматах. Запросы к таблицам: выборка строк/столбцов по заданным критериям. Переформатирование данных: очистка, преобразование, слияние, изменение формы. Фильтрация отсутствующих данных. Агрегирование данных и групповые операции.</p>	6
1.3.	<p>Построение графиков и визуализация данных</p> <p><i>Краткое содержание:</i> Обзор библиотек: matplotlib, pandas, seaborn, plotly. Построение графиков, столбчатых диаграмм, гистограмм, точечных диаграмм (scatter plots), ящиков с усами. Комбинирование различных графических элементов. Построение интерактивных и трехмерных диаграмм с помощью plotly.</p>	4
2.	Основы программирования: Объектно-ориентированное программирование	28

2.1.	<p>Парадигмы программирования. Основы объектно-ориентированного программирования. <i>Краткое содержание:</i> Парадигма объектно-ориентированного программирования. Понятия объектов и классов. Конструктор и деструктор. Модификаторы доступа. Создание простых классов на объектно-ориентированном языке программирования. Разграничение доступа к элементам класса. Описание полей и методов класса.</p>	4
2.2.	<p>Разработка классов, создание экземпляров класса. Указатели на классы. <i>Краткое содержание:</i> Разработка собственных классов. Инкапсуляция. Абстракция. Конструктор копий. Переопределение операторов. Реализация методов класса. Переопределение методов. Перегрузка методов. Переопределение операторов. Дружественные функции. Разработка программы, содержащей класс для реализации прикладной математической задачи.</p>	6
2.3.	<p>Организация иерархии классов. Интерфейсы. Множественное наследование. <i>Краткое содержание:</i> Наследование. Виртуальные методы. Полиморфизм. Динамическое связывание. Абстрактный класс. Создание классов-наследников. Выполнение проекта с организацией иерархии классов.</p>	4
2.4.	<p>Создание и использование шаблонов классов. <i>Краткое содержание:</i> Обобщенное программирование. Шаблоны классов и особенности их применения. Создание программ, содержащих обобщенные классы. Применение обобщенных классов для организации хранения объектов произвольного типа.</p>	6
2.5.	<p>Использование стандартных библиотек классов при создании программ на языке высокого уровня. <i>Краткое содержание:</i> Применение классов стандартной библиотеки для создания проектов, содержащих графический интерфейс пользователя. Создание проектов с графическим интерфейсом пользователя для решения прикладных математических задач и их визуализации.</p>	4
2.6.	<p>Основы параллельного программирования. <i>Краткое содержание:</i> Классы стандартной библиотеки, предназначенные для формирования дополнительных потоков и распределения вычислительной нагрузки между ними. Создание приложений для организации вычислений с использованием параллельных алгоритмов.</p>	4
3.	Машинное обучение	28
3.1.	<p>Основы теории статистического обучения <i>Краткое содержание:</i> Задачи и способы машинного обучения. Понятие о сложности модели. Параметры и гиперпараметры модели. Смещенность и причины его возникновения. Понятие о регуляризации моделей. Валидация моделей. Метрики качества классификационных и регрессионных моделей.</p>	4

3.2.	<p>Линейные методы машинного обучения</p> <p><i>Краткое содержание:</i></p> <p>Множественная линейная регрессия. Гребневая регрессия. Логистическая регрессия. Знакомство с библиотекой sklearn. Построение модели на основе гребневой регрессии, подбор коэффициента регуляризации в кросс-валидации, влияние регуляризации на качество моделей (построить график зависимости качества моделей от коэффициента регуляризации).</p>	6
3.3.	<p>Линейные методы ортогонализации пространства</p> <p><i>Краткое содержание:</i></p> <p>Метод главных компонент (PCA). Метод частичных наименьших квадратов (PLS). Анализ данных с помощью методов PCA и PLS. Подбор оптимального числа латентных переменных. Графики счетов и нагрузок в анализе данных.</p>	4
3.4.	<p>Нелинейные методы машинного обучения.</p> <p><i>Краткое содержание:</i></p> <p>Деревья принятия решений. Метод случайного леса и градиентного бустинга. Использование метода случайного леса в классификационном моделировании, оценка влияния числа деревьев и иных гиперпараметров. Использование метрик качества классификации (TPR, TNR, ACC, PPV, F-мера, AUC ROC).</p>	4
3.5.	<p>Искусственные нейронные сети</p> <p><i>Краткое содержание:</i></p> <p>Общие представления. Слои нейронной сети. Понятие о графе вычислений. Обратное распространение ошибки, обучение нейронной сети. Основные активационные функции. Нейронные сети как конструктор. Обзор основных типов архитектур нейронных сетей (многослойный персептрон, рекуррентные, конволюционные, состязательные, обучение с подкреплением) и их применение. Обзор основных достижений. Знакомство с библиотекой PyTorch, написание простой нейронной сети для регрессии/классификации. Написание многослойного персептрона, подбор функции активации и иные трюки в обучении сети</p>	4
3.6.	<p>Байесов подход к машинному обучению.</p> <p><i>Краткое содержание:</i></p> <p>Наивный байесов классификатор. Гауссовы процессы (кригинг) как метод регрессии. Kernel trick. Использование гауссовых процессов в моделировании, влияние параметров ядра на качество результатов.</p>	4
3.7.	<p>Обучение без учителя. Кластеризация</p> <p><i>Краткое содержание:</i></p> <p>Иерархические и неиерархические методы. Методы визуализации. Иерархическая кластеризация методами single и complete-link. Визуализация методом t-SNE</p>	2
Вариативная часть		
4.1	Хемоинформатика	28
4.1.1	<p>Представления химических объектов.</p> <p><i>Краткое содержание:</i></p> <p>Текстовые, линейные, матричные, графовые. Представление молекул. Типичные представления молекул в химии (структурная формула, химическая формула, тривиальное имя). Особенности представления в хемоинформатике, требования к представлениям.</p>	6

	<p>Виды представлений. Линейные представления (имена, WLN, SMILES, SLN, InChI). Представление молекулярных графов. Битовые строки (структурные ключи, отпечатки пальцев, хэшированные отпечатки пальцев). Матричное представление, виды матриц. Табличное представление. Трехмерные представления.</p>	
4.1.2.	<p>Маркуши, 3D представления. Реакции. <i>Краткое содержание:</i> Типичные форматы файлов (MDL, Sybyl, PDB). Конвертация между представлениями. Конвертация структура-имя и имя-структура. Конвертация структуры в линейные представления. 2D-3D конвертация. Представления реакций: основные виды. Представления реакции как набора реагентов и продуктов. Представления реакции как характеристик реакционного центра. Представления реакции как разности продуктов и реагентов.</p>	4
4.1.3.	<p>Базы данных. Алгоритмы. 3D поиски. <i>Краткое содержание:</i> Структура баз данных. Виды поиска в химических базах данных. Поиск по структуре, подструктуре, суперструктуре и по молекулярному сходству в базах данных различных типов. Основные алгоритмы поиска. Использование скринов. Рекурсивный подход. Ульмановский подход. Поиск в 3D базах данных. Жесткий и гибкий поиск. Основные химические базы данных в хемоинформатике</p>	4
4.1.4.	<p>Молекулярное разнообразие и молекулярное сходство. Дескрипторы. <i>Краткое содержание:</i> Дизайн библиотек данных. Использование для виртуального скрининга и для высокопроизводительного скрининга. Теоретическая комбинаторная химия. Разбросанные и сфокусированные библиотеки. Генерация структур. RECAP. Fragmenter. Кластеризация молекул. Иерархические подходы. Неиерархические подходы. Отбор молекул без кластеризации. Определение и использование дескрипторов. Роль дескрипторов в хемоинформатике. Многообразие дескрипторов. Классификация дескрипторов по функциональности. Физико-химические дескрипторы. Топологические индексы. Трехмерные дескрипторы. Фрагментные дескрипторы. Фармакофорные дескрипторы.</p>	4
4.1.5.	<p>QSAR: классический QSAR, лучшие практики, подготовка данных. <i>Краткое содержание:</i> История моделирования "структура-свойство": SAR/QSAR/QSPR. Построение и валидация моделей. Предобработка данных. Удаление смесей, неорганических и металлоорганических соединений. Конвертация структур, удаление солей и выбор состояния ионизации. Нормализация специфических хемотипов, резонансных форм и таутометов. Выявление дубликатов. Заключительная ручная проверка. Общие принципы построения моделей. Метод наименьших квадратов. Понятие об оверфиттинге и принцип оптимальной сложности моделей. Принципы отбора дескрипторов. Проблемы, связанные с отбором дескрипторов. Общие принципы валидации моделей.</p>	10
4.2.	Информатика новых материалов	28

4.2.1.	<p>Компьютерные технологии в материаловедении</p> <p><i>Краткое содержание:</i></p> <p>Потребности современной экономики в новых многофункциональных материалах для строительства, электроники, пищевой промышленности, химии и медицины. Классификация материалов. Физические характеристики материалов.</p> <p>Теоретическое предсказания структуры новых материалов, обладающих заданными свойствами; основы компьютерного дизайна материалов; теория функционала плотности; молекулярная динамика; использование веб-технологий и искусственного интеллекта для прогнозирования новых материалов с заданными физическими свойствами</p>	10
4.2.2	<p>Компьютерный дизайн материалов с использованием пакета программ MedeA</p> <p><i>Краткое содержание:</i></p> <p>Расчеты структуры и свойств материалов из первых принципов. Знакомство с оболочкой MedeA; построение ячейки; выбор потенциала плотности в зависимости от изучаемого материала; модуль VASP – основные возможности и характеристики; визуализация структуры и физических свойств; программирование физических свойств и изменения структуры материалов под действием внешних условий (давление, внедрение примесей); расчет физических свойств материалов из первых принципов; расчет фазовых диаграмм и характеристик фазовых переходов.</p>	10
4.2.3.	<p>Администрирование базы данных по физическим свойствам материалов.</p> <p><i>Краткое содержание:</i></p> <p>Роль баз данных в современном технологическом процессе синтеза структур. Методы поиска материалов с заданными свойствами. Работа с существующими базами данных по свойствам материалов. Обработка и систематизация больших данных.</p> <p>Реляционные базы данных. Регулярные выражения</p> <p>Использование консольных программ Linux (grep, awk) для обработки текстовых файлов</p> <p>Обработка массивов данных в программных пакетах MATLAB и на языке Python с применением библиотек SciPy, NumPy. Экспорт данных из базы данных физических свойств для применения в задаче кластерного анализа (k-means). Основы работы с интерфейсом командной строки в Linux. Решение задач на поиск строк в тексте по шаблону (на примере какого-либо языка программирования). Преобразование текстовых файлов одного формата в другой при помощи grep, awk. Поиск и экспорт строк в текстовом файле по шаблону (gawk).</p> <p>Рассмотрение примера и выполнение индивидуальных заданий по интерполяции, аппроксимации данных и сглаживанию данных при помощи оконных фильтров. Рассмотрение примера и выполнение индивидуальных заданий по переводу величин из одних единиц измерения в другие. Форматированный вывод на примере таблицы значений физического параметра материала.</p> <p>Рассмотрение примера и выполнение индивидуальных заданий по выводу данных в текстовый файл с выравниванием, заданным количеством символов после запятой и т. п. На основе данных физических свойств в формате CSV решить задачу кластерного</p>	8

	анализа методом k-средних: подготовка датасета, чтение датасета программой, обработка данных методом k-means, построение графиков. Выполнение удаленного доступа по SSH. Построение графиков с использованием по материалам при помощи пакета gnpurplot. При помощи библиотек компьютерного зрения (OpenCV) получить координаты частиц со снимка микроскопа.	
4.3.	Биоинформатика	28
4.3.1.	<p>Генетический аппарат клетки. Основы геномных и транскриптомных технологий</p> <p><i>Краткое содержание:</i></p> <p>Клетка прокариот, животных, растений и биологические макромолекулы, строение и функции - принципы формирования вторичной, третичной структур белка, РНК, взаимодействие биомолекул.</p> <p>Биологические макромолекулы, строение и функции - углеводы, липиды, белки и нуклеиновые кислоты</p> <p>Понятия генома, транскриптома, протеома. Копирование и реализация генетической информации</p> <p>Геном бактерий, геном эукариот, геном человека. Различия в организации и функционировании генома у про- и эукариот</p> <p>Высокопроизводительное секвенирование. Сборка генома. Метагеномный анализ.</p>	8
4.3.2.	<p>Работа с нуклеотидными последовательностями ДНК/РНК</p> <p><i>Краткое содержание:</i></p> <p>Обзор и классификация молекулярных баз данных</p> <p>Базы данных нуклеотидных последовательностей GenBank NCBI / ENA EMBL / DNA DDBJ. Форматы представления нуклеотидных последовательностей</p> <p>Молекулярная эволюция нуклеотидных последовательностей и принципы выравнивания последовательностей. Парное и множественное выравнивание нуклеотидных последовательностей ДНК, Clustal Omega</p> <p>Поиск гомологичных нуклеотидных последовательностей, blastn.</p> <p>Молекулярно-филогенетические деревья. Реконструкция и анализ молекулярно-филогенетических деревьев, MEGA</p>	10
4.3.3.	<p>Работа с аминокислотными последовательностями и пространственными структурами белков</p> <p><i>Краткое содержание:</i></p> <p>Поиск и выравнивание полипептидных последовательностей, белковый blast , программы множественного выравнивания</p> <p>Компьютерные методы предсказания вторичной структуры белка</p> <p>Пространственная структура белка, Protein Data Bank</p> <p>Методы предсказания пространственной структуры белка по его аминокислотной последовательности</p> <p>Компьютерное моделирование структуры и динамики белков, методы молекулярной динамики и молекулярного докинга.</p>	10
5.	<p>Практика</p> <p><i>Краткое содержание:</i></p> <p>Практика проводится на базе организаций различных организационно-правовых форм и форм собственности или их основных структурных подразделений, осуществляющих деятельность, соответствующую виду (видам) деятельности, к которому (которым) готовится обучающийся. Организации, с которыми заключен договор о прохождении обучающимися КФУ практики на их базе.</p>	54
6.	<p>Итоговая аттестация</p> <p><i>Краткое содержание:</i></p> <p>В ходе итоговой аттестации обучающиеся обеспечивают</p>	36

	<p>презентацию (защиту) разработанного цифрового решения (проекта), а также перечня решаемых им проблем и эффектов, ожидаемых от его реализации (внедрения) в отрасль. Проектное решение должно отвечать критериям актуальности, законченности, а также возможности интеграции его компонентов в иные системы и сервисы. Выдача диплома о профессиональной переподготовке о прохождении ДПП ПП осуществляется по итогам успешного прохождения итоговой аттестации.</p>	
--	--	--

20. Учебно-тематический план Программы определяет тематическое содержание, последовательность разделов и (или) тем и их трудоемкость.

№ п/п	Наименование раздела(модуля)	Количество часов		
		аудиторных		самостоятельной работы (выполнение тестовых заданий)
		Лекции	Семинары	
Общий блок				
1.	Основы программирования на Python	8	10	4
2.	Основы программирования: Объектно-ориентированное программирование	8	10	10
3.	Машинное обучение	8	10	10
Вариативная часть				
4.1.	Хемоинформатика	10	8	10
4.2.	Информатика новых материалов	10	8	10
4.3.	Биоинформатика	10	8	10
5.	Практика	54		
	Итоговая аттестация	36		

X. Формы аттестации

21. Слушатели, успешно выполнившие все элементы учебного плана, допускаются к итоговой аттестации.

Итоговая аттестация по Программе проводится в форме защиты проекта.

22. Лицам, успешно освоившим Программу (в области создания алгоритмов и программ, пригодных для практического применения, или навыков использования и освоения цифровых технологий, необходимых для

выполнения нового вида профессиональной деятельности) и прошедшим итоговую аттестацию в рамках проекта «Цифровые кафедры», выдается документ о квалификации: диплом о профессиональной переподготовке.

При освоении ДПП ПП параллельно с получением высшего образования диплом о профессиональной переподготовке выдается не ранее получения соответствующего документа об образовании и о квалификации (за исключением лиц, имеющих среднее профессиональное или высшее образование).

23. Лицам, не прошедшим итоговую аттестацию или получившим на итоговой аттестации неудовлетворительные результаты, а также лицам, освоившим часть Программы и (или) отчисленным из Университета, выдается справка об обучении или о периоде обучения по образцу, самостоятельно устанавливаемому Университетом.

XI. Оценочные материалы

24. Контроль знаний, полученных слушателями при освоении разделов (модулей) Программы, осуществляется в следующих формах:

- текущий контроль успеваемости – обеспечивает оценивание хода освоения разделов Программы, проводится в форме устных опросов/тестовых заданий/ситуационных заданий;

- промежуточная аттестация – завершает изучение отдельного модуля Программы, проводится в форме зачета;

- итоговая аттестация – завершает изучение всей программы.

25. В ходе освоения Программы каждый слушатель выполняет следующие отчетные работы:

№ п/п	Наименование раздела (модуля)	Задание	Критерии оценки
Общий блок			
1.	Основы программирования на Python	Тестовые задания	50% выполнения тестовых заданий

2.	Основы программирования: Объектно-ориентированное программирование	Устные вопросы	Корректный ответ на вопрос с использованием соответствующей терминологии
3.	Машинное обучение	Задание на проверку усвоенных знаний	50% выполнения задания
Вариативная часть			
4.1.	Хемоинформатика	Тестовые задания	50% выполнения тестовых заданий
4.2.	Информатика новых материалов	Задание на проверку усвоенных знаний	50% выполнения тестовых заданий
4.3.	Биоинформатика	Тестовые задания	50% выполнения тестовых заданий
5	Практика		Отчет по практике
6	Промежуточная аттестация	Вопросы к зачету/ Задание на проверку усвоенных знаний	Корректный ответ на вопрос с использованием соответствующей терминологии/50% выполнения задания
7	Итоговая аттестация	Выполнение и защита проекта	Защита проекта

26. Текущий контроль. Перечень примерных устных/тестовых/типовых заданий

Модуль Основы программирования на Python

1. Что выведет данный код?

```
import math
print(math.sqrt(36))
```

- a) 3
- b) 6
- c) -6
- d) Ошибка

2. Что выведет данный код?

```
import math
print(math.floor(1.8))
```

- a) 1
- b) 2
- c) -1.8
- d) Ошибка

3. Что выведет данный код?

```
import math  
print(math.log2(8))
```

- a) 4.0
- b) 2.0
- c) 3.0
- d) Ошибка

4. Что выведет данный код?

```
a = False - True  
print(a)
```

- a) -1
- b) 0
- c) 1

5. Что выведет данный код?

```
b = bool(False + 1)  
print(b)
```

- a) False
- b) True
- c) 2

6. Что выведет данный код?

```
c = True + 0.1  
print(c)
```

- a) False
- b) 1

c) 1.1

7. Что выведет данный код?

```
d = True ** False
```

```
print(d)
```

a) 0

b) 1

c) False

8. Что выведет данный код?

```
e = 1 == True
```

```
print(e)
```

a) True

b) 1

c) False

Модуль Основы программирования: Объектно-ориентированное программирование

1. Что называют парадигмой программирования?

2. Какую парадигму реализует язык Python?

3. Язык Python считается низкоуровневым или высокоуровневым?

4. Что такое ООП?

5. Что подразумевает абстракция с точки зрения ООП?

6. Что такое инкапсуляция?

7. Что такое наследование?

8. Что такое полиморфизм?

9. Какие существуют виды полиморфизма?

10. В чем отличие компилируемых и интерпретируемых языков программирования?

11. Что такое статическая и динамическая типизация?

12. Что такое слабая и сильная типизация?

13. Опишите процесс преобразования исходного кода в исполняемый файл.

14. В чем отличие ссылки от указателя?

15. Что такое указатель на функцию и как он может быть использован?

16. Какие способы группировки данных в Python вам известны?

17. Для чего предназначены структуры?

Модуль Машинное обучение

1. Моделирование "структура-реакционная способность" с использованием множественной линейной регрессии в программа ISIDA-QSPR. Используя в качестве исходных данных файлы ANIV TIBO.SDF и TEST_ANIV-TIBO.SDF, содержащие экспериментальные значения анти-ВИЧ активности ($\log(1 / IC_{50})$) 57 и 12 производных тетрагидроимидазобензодиазепинона (TIBO) соответственно, постройте индивидуальную MLR модель. В качестве дескрипторов используйте SMF-дескрипторы (фрагментные дескрипторы), включающие последовательности атомов и связей длиной от 2 до 12. В качестве уравнения регрессии используйте следующее: $Y = \text{SUM}(A_i * X_i)$. С помощью полученной модели предскажите контрольную выборку. Проанализируйте результаты. Оцените R² и RMSE.

2. Изучение распределения молекул в химическом пространстве и распределения экспериментальных сигналов с использованием анализа главных компонент в программе Weka. Используя в качестве исходных данных файлы ANIV TIBO.SDF и TEST_ANIV-TIBO.SDF, содержащие экспериментальные значения анти-ВИЧ активности ($\log(1 / IC_{50})$) 57 и 12 производных тетрагидроимидазобензодиазепинона (TIBO) соответственно, визуализируйте распределение молекул в химическом пространстве с помощью метода главных компонент (в 2D и 3D пространстве). Для каких целей, кроме визуализации, применяется метод главных компонент?

3. Использование деревьев алгоритма случайного леса для решения регрессионных задач (на примере программы Weka). Используя в качестве исходных данных файлы alkan-bp-connect.arff (содержит дескрипторы и точки кипения для 74 алканов), alkan-mp-connect.arff (содержит дескрипторы и точки плавления для 74 алканов), selwood.arff (содержит дескрипторы и значения активностей для 33 антифиляриатозных аналогов антимицина (из базы данных selwood)), shapiro.arff (содержит дескрипторы и значения

активностей для 124 фенольных ингибиторов бактерий полости рта (из базы данных sharipo)), постройте модели "структура-свойство" (первые 2 файла) и модели "структура-активность" (оставшиеся 2 файла). В качестве метода машинного обучения используйте "случайный лес" (число деревьев = 100). Используйте пятикратную кросс-валидацию. Проанализируйте результаты.

4. Использование алгоритма k-ближайших соседей (kNN) для решения классификационных задач (на примере программы Weka). Используя в качестве исходных данных файлы A2AC_train.arff, A2AC_test.arff, A2AC_external.arff, A2AC_train.sdf, A2AC_test.sdf, A2AC_external.sdf, постройте модель "структура-активность". Набор данных содержит лиганды аденозинового рецептора человека (A2A). Соединения были классифицированы как активные, если pK_i было больше 5. Обучающая выборка содержит 385 активных и 2089 неактивных молекул, контрольная выборка - 371 активную и 2103 неактивных молекул. Класс активности (0 для неактивных, 1 для активных) записан в поле SDF Activity. Файлы с расширением arff содержат фрагментные дескрипторы. В качестве метода машинного обучения используйте метод k ближайших соседей (параметры по умолчанию). Вычислите основные количественные показатели качества классификационных моделей - сбалансированную точность, чувствительность, специфичность. Постройте ROC-кривую.

Проанализируйте результаты.

5. Использование метода опорных векторов для решения классификационных задач (на примере программы Weka). Используя в качестве исходных данных файлы A2AC_train.arff, A2AC_test.arff, A2AC_external.arff, A2AC_train.sdf, A2AC_test.sdf, A2AC_external.sdf, постройте модель "структура-активность". Набор данных содержит лиганды аденозинового рецептора человека (A2A). Соединения были классифицированы как активные, если pK_i было больше 5. Обучающая выборка содержит 385 активных и 2089 неактивных молекул, контрольная выборка - 371 активную и 2103 неактивных молекул. Класс активности (0 для неактивных, 1 для активных) записан в поле SDF Activity. Файлы с расширением arff содержат фрагментные дескрипторы. В качестве метода

машинного обучения используйте метод опорных векторов (со следующими параметрами - $C = 1$, $\text{kernel} = \text{rbf}$). Вычислите основные количественные показатели качества классификационных моделей - сбалансированную точность, чувствительность, специфичность. Постройте ROC-кривую. Проанализируйте результаты.

6. Использование метода "наивного Байеса" для решения классификационных задач (на примере программы Weka). Используя в качестве исходных данных файлы A2AC_train.arff, A2AC_test.arff, A2AC_external.arff, A2AC_train.sdf, A2AC_test.sdf, A2AC_external.sdf, постройте модель "структура-активность". Набор данных содержит лиганды аденозинового рецептора человека (A2A). Соединения были классифицированы как активные, если pK_i было больше 5. Обучающая выборка содержит 385 активных и 2089 неактивных молекул, контрольная выборка - 371 активную и 2103 неактивных молекул. Класс активности (0 для неактивных, 1 для активных) записан в поле SDF Activity. Файлы с расширением arff содержат фрагментные дескрипторы. В качестве метода машинного обучения используйте метод "наивного Байеса" (параметры по умолчанию). Вычислите основные количественные показатели качества классификационных моделей - сбалансированную точность, чувствительность, специфичность. Постройте ROC-кривую. Проанализируйте результаты.

7. Сравните ранее полученные результаты классификации выборки A2AC с помощью различных методов машинного обучения. Какой метод оказался лучше? Предположите, почему?

8. С помощью Logistic Regression была построена модель и предсказаны вероятности принадлежности к "классу 1" для 4 молекул. По имеющимся данным постройте ROC-кривую.

Предсказанная вероятность Истинный класс

Mol 1 0,1 1

Mol 2 0,2 0

Mol 3 0,5 0

Mol 4 0,9 1

9. В конце построения регрессионной модели, Weka вывела следующее сообщение:

SMOreg

Support vectors:

-0.01*k[12]

-0.01*k[42]

+0.01*k[51]

+1.15

Объясните, как оценить значение для нового объекта, используя модель, построенную с помощью Weka.

10. Определите, что представляет собой F-мера, значение которой рассчитывается в программе Weka для классификационной модели.

Модуль Хемоинформатика

1. Какие из указанных SMILES соответствует молекуле аспирина?

а) CC(=O)Oc1ccccc1C(O)=O

б) c1c(C(O)=O)cc(OC(O)C)ccc1

в) OC(=O)c(cccc1)c1OC(=O)C

г) c1(C(=O)O)ccccc1OC(=O)C

2. Какие InChI для приведенной молекулы гуанина соответствуют молекуле и являются стандартными?

а) InChI=1/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-

3/h1H,(H4,6,7,8,9,10,11)/f/h8,10H,6H2

б) InChI=1S/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)

в) InChI=1S/C6H6N5O2/c6-5-9-3-2(4(11)10-5)7-1-8-3/h1H,(H4,6,7,8,9,10,11)

г) InChI=1/C5H5N5O/c6-5-9-3-2(4(11)10-5)7-1-8-

3/h1H,(H4,6,7,8,9,10,11)/f/h7,9H,6H2

3. Какой SMARTS запроса будет определять выделенную подструктуру (NcccN) в приведённой молекуле (5-butyl-4-methylbenzene-1,3-diamine)?

Атомы водорода не принимать во внимание.

а) N~*~*~*~N

б) NcccN

в) [#7]ccc[#7]

г) [NH2]aaa[NH2]

4. Какой из приведенных SMILES удовлетворяет приведенной структуре Маркуша?

а) OCCc1c(C)cccc1

б) OCCCCCCCCc1cc(C(=O)O)ccc1

в) OCCc1ccc(C(C)=O)cc1

г) OCCc1ccc(COC=O)cc1

5. Что такое трехмерный граф:

а) граф, в котором указаны все атомы, но для каждой вершины заданы координаты в про-пространстве

б) полный граф, в котором меткой ребра являются расстояния между данными атомами

в) молекулярный граф, в котором указана стереохимия с помощью клиновидных связей

г) молекулярный граф, в котором вместо порядка связей указана длина связей

6. Каким преимуществом НЕ обладает дизайн библиотек с помощью маркушей:

а) структуры в полученной библиотеке обычно синтетически доступны

б) большое количество соединений может быть сгенерировано очень эффективным путем

в) для создания библиотеки не требуется указания большого количества фрагментов и реагентов, из которых будут сформированы желаемые структуры

г) исходные данные легко подготовить вручную и очень простым путем

7. Наиболее широко используемыми квантовохимическими дескрипторами являются:

а) энергии молекулярных орбиталей

б) энтальпии образования водородной связи

в) значения молекулярной рефракции

г) автокорреляции электростатического потенциала

8. Для построения Regression Error Characteristic, или REC-кривой, по оси абсцисс откладывают..., а по оси ординат?

а) абсолютную ошибку (модуль отклонения предсказанного значения от экспериментально-го); процент объектов, предсказанный с ошибкой не более, чем абсолютная ошибка

б) R^2 ; RMSE

в) чувствительность; 1-селективность

г) R^2 ; Q^2

9. Какое из приведенных выражений содержит формулу для вычисления индекса схожести Тверского? а-число активных бит в одной молекуле, б - число включенных бит в другой молекуле, с - число бит, которые являются активными в обеих молекулах.

а) $c/(a+b-c)$

б) $2c/(a+b)$

в) $(a+b-2c)/(a+b-c)$

г) $c/(c+\alpha(a-c)+\beta(b-c))$

10. В какой из приведенных баз данных можно найти информацию, характеризующую прочность связывания данного химического соединения с различными белками?

а) CAS

б) PubChem

в) ChEMBL

г) ZINC

Модуль Информатика новых материалов

1. Ознакомиться с интерфейсом базы данных Materials Project

(<https://materialsproject.org>). Выполнить выгрузку и систематизацию данных для различных типов неорганических материалов (металлов, оксидов и др.) из этой базы данных.

2. Выполнить поиск информации по свойствам различных материалов в базе данных NOMAD (<https://nomad-lab.eu/>).

Модуль Биоинформатика

1. Гены домашнего хозяйства – это (выберите один или несколько правильных ответов)

- а) гены, характерные только для этого организма
- б) гены, активные в данный момент времени
- в) гены, активные независимо от внешних условий
- г) гены, которые активны только в определенных условиях
- д) структурные гены, обеспечивающие основные жизненные функции клетки
- е) гены, экспрессия которых наблюдается постоянно во всех клетках организма
- ж) гены, обеспечивающие осуществление специализированных функций некоторыми типами клеток

2. Гены роскоши – это (выберите один или несколько правильных ответов)

- а) гены, характерные только для этого организма
- б) гены, активные в данный момент времени
- в) гены, активные независимо от внешних условий
- г) гены, которые активны только в определенных условиях
- д) структурные гены, обеспечивающие основные жизненные функции клетки
- е) гены, экспрессия которых наблюдается постоянно во всех клетках организма
- ж) гены, обеспечивающие осуществление клеткой специализированных функций

3. Вторичная структура белка (выберите один или несколько правильных ответов)

- а) образуется за счет ковалентных связей между аминокислотами
- б) образуется за счет водородных связей между аминокислотами
- в) образуется за счет дисульфидных между цистеинами
- г) зависит от первичной последовательности белка (последовательности аминокислот)
- д) формирует домены и мотивы
- е) является трехмерной
- ж) является линейной

з) описывается графиком Рамачандрана

4. Третичная структура белка (выберите один или несколько правильных ответов)

а) образуется за счет ковалентных связей между аминокислотами

б) образуется за счет водородных связей между аминокислотами

в) образуется за счет дисульфидных между цистеинами

г) зависит от первичной последовательности белка (последовательности аминокислот)

д) формирует домены и мотивы

е) является трехмерной

ж) является линейной

з) описывается графиком Рамачандрана

5. Технологии секвенирования нового поколения (NGS) связаны с использованием методов (выберите один или несколько верных ответов):

а) обрыва цепи.

б) дробовика (Shotgun).

в) случайного фрагментирования ДНК.

г) создания библиотеки случайных последовательностей ДНК.

д) сборки ридов.

е) сборки контиг.

ж) сборки скаффолдов.

6. Какие платформы секвенирования позволяют проводить секвенирование геномов (выберите один или несколько правильных ответов)

а) секвенирование по Сэнгеру

б) Illumina

в) Ion Torrent

г) PacBio

д) Oxford nanopore

7. Выход секвенирования это (выберите один правильный ответ)

а) отражение качества (точности прочтения)

б) объем последовательностей, прочтенных с высоким качеством

в) объем всех прочтенных последовательностей

г) полученный после секвенирования результат (собранный геном, метагеном и т.д.)

8. Какой вид репарации восстанавливает ошибки в ДНК, возникающие НЕ в процессе репликации (выберите один или несколько правильных ответов)

а) эксцизионная

б) Mismatch

в) фотореактивация

г) рекомбинационная

д) SOS-репарация

9. Точечные мутации образуются за счет (выберите один или несколько правильных ответов):

а) ошибок в ходе транскрипции

б) ошибок в ходе репликации

в) ошибок в ходе репарации

г) ошибок в ходе рекомбинации

д) действия химических мутагенов

е) встраивания вирусов

ж) действия ультрафиолета

10. Методами метагеномики можно получить (выберите один верный ответ):

а) количественный состав микробиоты высших организмов.

б) частотное распределение видов в состав микробиоты высших организмов.

в) качественный состав микробиоты высших организмов.

г) геномы микроорганизмов в чистых культурах и их функции.

д) геномы некультивируемых микроорганизмов и их функции.

27. Промежуточная аттестация. Перечень примерных устных/типовых заданий

Модуль Основы программирования на Python

Вопросы к зачету:

1. Язык Python и особенности его стиля программирования.

Интерактивный режим Python. IPython. Jupyter Notebook.

2. Синтаксис и управляющие конструкции языка Python. Переменные, значения и их типы. Типы данных в Python.
3. Встроенные операции и функции. Основные алгоритмические конструкции.
4. Условный оператор. Множественное ветвление.
5. Циклы и счетчики.
6. Определение функций. Параметры и аргументы. Вызовы функций. Оператор возврата.
7. Конструкции *args, **kwargs.
8. Списки, кортежи и словари.
9. Операторы общие для всех типов последовательностей.
10. Специальные операторы и функции для работы со списками. Срезы.
11. Работа со словарями. Методы словарей.
12. Случайные числа. random, randrange, choice.
13. Функции обработки строк. join, replace, split.
14. Стандартная библиотека и pip. Модули и пакеты в Python. Основные стандартные модули.
15. Импортирование модулей. Создание собственных модулей и их импортирование. Специализированные модули и приложения.
16. Файлы и исключения. Работа с внешними источниками данных.
17. Исключения, обработка исключений, вызов исключений (try-except-finally).
18. Утверждения (assert). Открытие, чтение, запись. (open, инструкция with).
19. Работа с текстовыми файлами, xml и csv - файлами.
20. Функциональное программирование. Лямбда-функции.
21. Использование функций map, filter, reduce, zip.

Модуль Основы программирования: Объектно-ориентированное программирование

Вопросы к зачету:

1. Генераторы, декораторы, рекурсия.
2. Модификация функций с помощью декораторов.
3. Итерируемые объекты. Использование генераторов (yield).
4. ООП в Python. Классы, объекты и экземпляры классов. Наследование.
5. Магические методы. Переопределение операторов. Методы классов.
6. Инкапсуляция. Условно частные и строго частные методы.
7. Регулярные выражения. Использование регулярных выражений.
8. Наука о данных и Python. Библиотеки: NumPy, pandas, matplotlib, SciPy.
9. Основы NumPy: массивы и векторные вычисления.
10. Инструменты визуализации данных для Python.
11. Введение в API библиотеки matplotlib.
12. Библиотека pandas. Введение в структуры данных pandas.
13. Объекты Dataframe и Series.
14. Визуализация данных в pandas. Seaborn.
15. Агрегирование данных и групповые операции.
16. Сбор и подготовка данных в Python: извлечение данных с web-страниц (web-scraping).
17. Работа с динамическими сайтами с помощью Selenium.
18. Массовый скрепинг с помощью scrapy.
19. Работа со структурированными данными: JSON и XML.

Модуль Машинное обучение

Вопросы к зачету:

1. Основные принципы машинного обучения. Задачи машинного обучения.
2. Основные принципы машинного обучения. Принципы построения моделей в классической статистике.
3. Функция правдоподобия
4. Метод наименьших квадратов (МНК).
5. Оверфиттинг, обучающие и контрольные выборки, внутренний и внешний перекрестный и скользящий контроле.
6. Оценка прогнозирующей способности моделей. Критерии оценки качества регрессионных моделей. Критерии оценки качества

классификационных моделей.

7. Предобработка данных.
8. Множественная линейная регрессия (MLR). Основы метода.
9. Понятие о регуляризации. Основы метода гребневой регрессии.
10. Метод главных компонент (PCA). Построение PCA.
11. Метод частичных наименьших квадратов (PLS). Геометрический смысл. Визуализация результатов.
12. Искусственные нейронные сети. Основные понятия.
13. Самоорганизующиеся карты Кохонена.
14. Деревья принятия решений. Выращивание, переобучение и обрезка дерева.
15. Метод случайного леса.
16. Метрические методы прогнозирования. Метод k ближайших соседей (kNN) и его обобщения.
17. Метод опорных векторов (SVM).
18. Решение регрессионных задач (SVR). Метод одноклассовой классификации (1-SVM, SVDD).
19. "Наивный" Байесов классификатор.
20. Генеративные топографические карты (GTM).
21. Обучение без учителя. Кластеризация.

Модуль Хемоинформатика

22. Построить QSAR-модель для предсказания указанного свойства предоставленного и подготовленного для моделирования набора соединений. Какие именно дескрипторы были выбраны для моделирования? Почему? Как подбирались гиперпараметры? Какие гиперпараметры оптимизировались? Почему? Оценить описательную и предсказательную способность этой модели. Предсказать свойства для дополнительного тестового набора данных. Объяснить полученные результаты.
23. Подготовить предоставленный набор соединений для последующего QSAR-моделирования их свойств. Сравнить размеры исходной и подготовленной выборки. Объяснить, почему некоторые объекты и их свойства были изменены / удалены. На сколько выборок нужно разделить подготовленный набор данных? Почему? Сгенерировать для

подготовленных выборок дескрипторы указанного типа. Объяснить, как именно происходила генерация, на каких выборках и почему именно так. Требуется ли уменьшение количества дескрипторов? Почему? Как именно можно это осуществить?

24. Из химической базы данных (PubChem, ChEMBL или др.) извлечь набор соединений, подструктурой / суперструктурой которых является указанное соединение (или для которых индекс схожести Танимото с указанным соединением составит представленное значение) и значение указанного свойства для которых больше / меньше / находится в указанных границах. Стандартизировать соединения полученной выборки и преобразовать их в уникальные SMILES / InChI. Какие еще этапы подготовки набора данных для последующего QSAR моделирования требуется провести? Почему?

Модуль Информатика новых материалов

1. Построить QSAR-модели для анализа структуры материалов на основе баз данных Materials Project, NOMAD и др.
2. Использовать QSAR-модели для поиска количественных соотношений структура-свойство на основе баз данных Materials Project, NOMAD и др.

Модуль Бионформатика

1. Проведите сборку *de novo* генома бактерии. Определите филогенетическое положение организма. Прочтения в формате FASTQ прилагаются.
2. Проведите сборку *de novo* генома вируса. Определите филогенетическое положение организма. Прочтения в формате FASTQ прилагаются.
3. Проведите сборку генома бактерии на референсном геноме. Прочтения в формате FASTQ прилагаются.
4. Проведите сборку генома бактерии на референсном геноме. Прочтения в формате FASTQ прилагаются.
5. Постройте третичные структуры белков и определите вероятный сайт их связывания. Аминокислотные последовательности белков прилагаются.

27. Модуль Практика

Студент заполняет отчет о проделанной работе в заданной форме. Отчет

должен включать: 1) титульный лист; 2) содержание; 3) введение; 4) основную часть; 5) заключение. Во введении указывается наименование организации, где студент проходил практику, подразделение, выполняемая работа, руководитель практики от организации. В основной части отчета по проектной практике необходимо осветить следующие вопросы: - краткая характеристика предприятия, на котором студент проходил практику; - описание используемых информационных подсистем и информационных технологий; - описание изученных в ходе практики информационных подсистем и информационных технологий; - тексты и описание изученных или разработанных в ходе проектной практики программных модулей. В заключении подводятся итоги проектной практики, фиксируются выполненные и невыполненные разделы задания на проектную практику. В приложениях следует помещать вспомогательные материалы, к которым относятся: промежуточные расчеты, таблицы вспомогательных цифровых данных, иллюстрации вспомогательного характера, учредительные документы, инструкции, заполненные формы отчетности, договоры и другие документы, анкеты обследования потребителей и т.д.

28. Итоговая аттестация. Перечень примерных тем проектов

1. Подсчёт количества частиц на фото микроскопии СЭМ, ТЭМ, контрастный микроскопии с помощью ИИ.
2. Разработка метода поиска эффективного потенциала межмолекулярного взаимодействия для воды на основе метода главных компонент.
3. Создание модуля поиска для сайта базы данных материалов.
4. Разработка методики дискриминантного анализа для молекулярных систем.
5. Предсказание физико-химических свойств материалов с помощью методов машинного обучения.
6. Программирование механизмов химических реакций на языке Питон
7. Отечественные операционные системы в применении к научным исследованиям.
8. Разработать и валидировать модель для предсказания ингибирования

- белка-переносчика желудочного-кишечного тракта MDR1.
9. Разработать и валидировать модель для предсказания ингибирования белка-переносчика желудочного-кишечного тракта MRP1.
 10. Разработать и валидировать модель растворимости химических соединений.
 11. Разработать и валидировать модель липофильности химических соединений.

XII. Материально-техническое и учебно-методическое обеспечение Программы

Электронная информационно-образовательная среда КФУ (ЭИОС) представляет собой совокупность электронных информационных ресурсов, электронных образовательных ресурсов, информационных технологий, телекоммуникационных технологий, соответствующих технологических средств, обеспечивающих освоение обучающимися образовательных программ или их частей, а также взаимодействие между всеми участниками образовательного процесса независимо от места их нахождения;

ЭИОС обеспечивает:

— доступ к учебным планам, рабочим программам дисциплин (модулей), практик, к изданиям электронных библиотечных систем и электронным образовательным ресурсам;

— фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;

— проведение всех видов занятий, процедур оценки результатов обучения, реализация которых предусмотрена с применением электронного обучения, дистанционных образовательных технологий; <https://edu.kpfu.ru/>

— формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок на эти работы со стороны любых участников образовательного процесса;

— взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети

«Интернет».

Система «Антиплагиат.ВУЗ» и другие ресурсы позволяющие обеспечивать освоение обучающимися образовательных программ в полном объеме независимо от места нахождения обучающихся.

Основными элементами ЭИОС КФУ являются:

а) электронные информационные ресурсы:

— официальный сайт КФУ (<https://kpfu.ru/>);

— личные кабинеты участников образовательного процесса, обеспечивающие доступ к компонентам ЭИОС КФУ;

— корпоративная электронная почта;

— сайт Научной библиотеки им. Н.И. Лобачевского;

— информационно-аналитическая система управления образовательным процессом;

— система автоматического поиска текстовых заимствований;

— другие базы данных и файловые системы, используемые в образовательном процессе;

б) электронные образовательные ресурсы:

— система управления обучением Moodle;

— сайт дистанционного обучения (<https://edu.kpfu.ru/>), содержащий более 3500 цифровых образовательных ресурсов;

— площадка для создания и тестирования курсов (<https://do.kpfu.ru/>);

в) электронные библиотечные системы:

— внутренняя электронная библиотечная система КФУ, обеспечивающая доступ к информационным ресурсам, включающая печатные и электронные документы на русском и иностранных языках;

— внешние электронные библиотечные системы и электронные библиотеки, доступ к которым осуществляется на договорной основе;

г) средства вычислительной техники:

— серверное оборудование КФУ;

— компьютеры, эксплуатируемые в КФУ;

— ноутбуки, планшеты, смартфоны и другие портативные, мобильные персональные компьютеры;

- средства организационной и множительной техники;
- мультимедийное оборудование.

Система электронного (дистанционного) обучения (далее – СДО) – электронная информационно-образовательная среда в виде системно-организованной совокупности информационно-коммуникационных средств и технологий, процессов программно-аппаратного и организационно-методического обеспечения, деятельности научно-педагогического, педагогического, учебно-вспомогательного и инженерного персонала (работников), ориентированная на реализацию системы сопровождения учебного процесса с целью удовлетворения образовательных потребностей обучающихся независимо от места их нахождения

Доступ в СДО обеспечивается непрерывно (в круглосуточном режиме с коэффициентом доступности всех компонентов среды не ниже 99,5 %) и из любой точки подключения к сети Интернет с заданными характеристиками канала связи.

Доступ ко всем сервисам СДО является персонализированным (под единой учетной записью).

Освоение ДПП ПП предполагает использование следующего программного обеспечения и информационно-справочных систем:

- Программное обеспечение для доступа к платформе Microsoft Teams
- Браузер: Яндекс Браузер, Mozilla Firefox
- Пакет офисного программного обеспечения: МойОфис, Microsoft Office 365
- ПО для работы с документами в формате PDF: Окуляр ГОСТ, Adobe Acrobat Reader DC
- ЭБС Издательства "Лань"

Учебно-методическая литература для ДПП ПП также имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны,

высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

ХIII. Список литературы

1. Введение в хемоинформатику : учебное пособие / Т. И. Маджидов, И. И. Баскин, И. С. Антипин, А. А. Варнек .- Казань : [Казанский университет], 2013 . [Ч. 1]: Компьютерное представление химических структур. - 2013. - 173 с.

2. Введение в хемоинформатику : учебное пособие / Т. И. Маджидов, И. И. Баскин, И. С. Антипин, А. А. Варнек .- Казань : [Казанский университет], 2013 . Ч. 2: Химические базы данных / Т. И. Маджидов, И. И. Баскин, А. А. Варнек ; Казан. федер. ун-т .- Казань : [Издательство Казанского университета], 2015. - 185 с.

3. Введение в хемоинформатику : учебное пособие / Т. И. Маджидов, И. И. Баскин, И. С. Антипин, А. А. Варнек. - Казань : [Казанский университет], 2013 . Ч. 3: Моделирование 'структура-свойство' / И. И. Баскин, Т. И. Маджидов, А. А. Варнек ; Казан. федер. ун-т .- Казань : [Издательство Казанского университета], 2015. - 302 с.

4. Омельченко В.П., Медицинская информатика : учебник / В.П. Омельченко, А.А. Демидова. - Москва: ГЭОТАР-Медиа, 2018. - 528 с. - ISBN 978-5-9704-4320-0 - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <http://www.studentlibrary.ru/book/ISBN9785970443200.html>

5. Кохонен, Т. Самоорганизующиеся карты : учебное пособие / Т. Кохонен ; под редакцией Ю. В. Тюменцева ; перевод В. Н. Агеева. - 3-е изд. -

Москва : Лаборатория знаний, 2017. - 660 с. - ISBN 978-5-00101-461-4. - Текст: электронный // Лань: электронно-библиотечная система. - URL: <https://e.lanbook.com/book/94143> (дата обращения: 11.03.2020). - Режим доступа: для авториз. пользователей.

6. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. - 2-е изд., стер. - Санкт-Петербург : Лань, 2019. - 212 с. - ISBN 978-5-8114-4493-9. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/120063> (дата обращения: 11.03.2020). - Режим доступа: для авториз. пользователей.

7. Маккинли У., Python и анализ данных [Электронный ресурс] / Уэс Маккинли - М. : ДМК Пресс, 2015. - 482 с. - ISBN 978-5-97060-315-4 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785970603154.html>

8. Гуриков С. Р. Основы алгоритмизации и программирования на Python : учеб. пособие / С.Р. Гуриков. М. : ФОРУМ : ИНФРА-М, 2018. 343 с. (Высшее образование: Бакалавриат). - Режим доступа: <http://znanium.com/catalog/product/924699>

9. Саммерфилд М., Python на практике [Электронный ресурс] / Марк Саммерфилд - М. : ДМК Пресс, 2014. - 338 с. - ISBN 978-5-97060-095-5 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785970600955.html>

10. Спирин, А. С. Молекулярная биология. Рибосомы и биосинтез белка : учебное пособие / А. С. Спирин. - Москва : Лаборатория знаний, 2019. - 594 с. - ISBN 978-5-00101-623-6. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/110208> (дата обращения: 13.04.2022). - Режим доступа: для авториз. пользователей.

11. Концепции современного естествознания: учебник для студентов вузов / под ред. В.Н. Лавриненко, В.П. Ратникова. - 4-е изд., перераб. и доп. - М. : ЮНИТИ-ДАНА, 2017. -319 с. - ISBN 978-5-238-01225-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1028500> (дата обращения: 13.04.2022). - Режим доступа: по подписке.

12. Применение молекулярных методов исследования в генетике: Учебное пособие/Нефедова Л. Н. - М.: НИЦ ИНФРА-М, 2016. - 104 с. ISBN

978-5-16-009872-2 - Режим доступа: <http://znanium.com/catalog/product/460545>

13. Основы генетики : учебник / В.В. Иванищев. ? М. : РИОР : ИНФРА-М, 2017. ? 207 с. - Режим доступа: <http://znanium.com/catalog/product/557529>

14. Волкова В.М., Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python [Электронный ресурс]: учебное пособие / Волкова В.М. - Новосибирск : Изд-во НГТУ, 2017. - 74 с. – ISBN 978-5-7782-3183-2 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785778231832.html>

15. Прохоренок Н. А. и др. Python 3. Самое необходимое: Пособие / Прохоренок Н.А., Дронов В.А. - СПб:БХВ-Петербург, 2016. - 464 с. ISBN 978-5-9775-3631-8 - Режим доступа: <http://znanium.com/catalog/product/944129>

16. Сузи Р. А. Python: Пособие / Сузи Р.А. - СПб:БХВ-Петербург, 2015. - 759 с. ISBN 978-5-9775-1417-0 - Режим доступа: <http://znanium.com/catalog/product/939857>

17. Ребриков Д.В., NGS: высокопроизводительное секвенирование [Электронный ресурс] / Д. В. Ребриков [и др.] ; под общей редакцией Д. В. Ребрикова. - 2-е изд. (эл.). - М. : БИНОМ, 2015. - 235 с. - ISBN 978-5-9963-3024-9 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785996330249.html>

18. Уилсон К., Принципы и методы биохимии и молекулярной биологии [Электронный ресурс] / редакторы К. Уилсон и Дж. Уолкер. - М. : БИНОМ, 2013. - 848 с. (Методы в биологии) - ISBN 978-5-9963-2126-1 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785996321261.html>

19. Основы биохимии: Учебное пособие / Т.Л. Ауэрман, Т.Г. Генералова, Г.М.

20. Сусялянок. - М.: НИЦ ИНФРА-М, 2014. - 400 с. ISBN 978-5-16-005295-3 - Режим доступа: <http://znanium.com/catalog/product/460475>