

Сибирский институт управления – филиал РАНХиГС
Кафедра бизнес-аналитики и статистики

ПРОГРАММА БАКАЛАВРИАТА

*«Информационные технологии в государственном и муниципальном
управлении»*

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ МОДУЛЯ

ЦИФРОВОЙ МОДУЛЬ
Б1.В.24

краткое наименование дисциплины: нет

по направлению подготовки:
38.03.04. Государственное и муниципальное управление

формы обучения: очная

Год набора – 2023

1. Цель освоения модуля:

Основанием изучения модуля является минимально необходимый объем теоретических знаний, а также приобретенные ранее знания, умения и навыки в области базового курса информационных и цифровых технологий, полученные в средних образовательных учреждениях, на первом курсе образовательной программы.

Модуль «Цифровой модуль» является общим теоретическим и методологическим основанием для подготовки к процедуре защиты и защите выпускной квалификационной работы.

2. План курса:

Б1.В.24.01 «Основы алгоритмизации и программирования»

Раздел 1. Введение в алгоритмизацию и программирование

Тема 1.1. Введение в теорию алгоритмов. Алгоритмизация. Формы представления алгоритма.

Информационный процесс. Понятие алгоритма. Свойства алгоритмов. Применение алгоритмов. Связь между алгоритмом и понятием функции. Виды алгоритмов и способы их формализации. Нумерация. Языки программирования и алгоритмизация: классификация языков программирования, особенности реализации алгоритмов в различных языках программирования

Функциональное описание алгоритмов. Псевдокод, форма Бэкуса- Наура, блок-схемы, ДРАКОН-схемы, структурограммы Насси-Шнейдермана. Краткое введение в универсальный язык моделирования алгоритмов UML.

Итерация, рекуррентность и рекурсивность. Технические и программные средства реализации информационных процессов, модели решения функциональных и алгоритмических задач, алгоритмизация и программирование.

Алгоритмические задачи. Алгоритмически неразрешимые задачи. Анализ алгоритмов: эффективность алгоритмов, функция времени работы алгоритма.

Тема 1.2. Языки программирования. Парадигмы программирования.

Классификация языка программирования: базовые свойства, признаки, особенности синтаксических конструкций, семантика.

Особенности проектирования, разработки и сопровождения программных проектов на языке высокого уровня. Виды и инструменты коллективной разработки. Сопровождение и локализация.

Документирование, тестирование и верификация программного кода. Технологии разработки программ. Гибкие технологии разработки. Экстремальное программирование. Современные технологии разработки программного обеспечения.

Тема 1.3. Типы данных, основные модели структур данных. Элементы теории множеств в программировании.

Абстрактные структуры данных: стек, очередь, очередь с приоритетом, ассоциативный массив. Отображение абстрактных структур данных на структуры хранения: массивы, списки. Базы данных. Нормирование баз данных. Нормальные формы баз данных. Проектирование баз данных – приведение к нормальным формам. Мотивы и автоматы Кортэжи, вектора, домены. Применение нечетких множеств.

Специальные типы данных: перечисления, списки, хеши, календарный тип (дата). Пользовательский ввод, обработка файлов.

Тема 1.4. Введение в программирование на языке Python. Синтаксис языка.

Возможности языка Python. Загрузка и установка Python, Anaconda3 и Jupyter Notebook. Основной принцип Python. Парадигмы программирования в Python. Основные

архитектурные черты.

Числовые данные: целые, вещественные, логические. Числовые литералы. Точность представления целых чисел и длинные целые. Числа с фиксированной точностью. Формат отображения чисел. Работа с числовыми данными и операции над ними. Смешанные операторы и определение старшинства операторов. Битовые операции. Системы счисления. Встроенные функции и модули для работы с числами. Строковые данные. Неизменяемость. Литералы строк. Экранированные последовательности, служебные символы. Юникод. Преобразование и смешивание типов.

Имя переменной. Инициализация переменных. Динамическая типизация. Операторы: присваивания, приращения. Ввод данных. Логические операции: конъюнкция, дизъюнкция, отрицание. Операции сравнения: строго больше, строго меньше, больше или равно, меньше или равно, равно, не равно.

Тема 1.5. Циклы, строки, списки, кортежи.

Условная инструкция `if`. Множественное ветвление. Блоки, отступы. Проверка истинности. Трехместное выражение `if/else`. Циклы `while`. Общий формат цикла. Инструкции `break`, `continue`, `pass` и `else`. Циклы `for`. Общий формат. Присваивание кортежа в цикле `for`. Вложенные циклы `for`. Счетные циклы: `while` и `range`. Изменение списков: `range`. Параллельный обход: `zip` и `map`. Генерирование индексов и элементов: `enumerate`.

Операции над последовательностями. Функции и методы строк. Способы представления строк. Поиск по шаблону. Доступ по индексам и извлечение подстроки. Преобразование строк. Изменение строк. Форматирование строк. Строковые методы. Модули для работы со строками.

Списки. Литералы списков. Операции над последовательностями. Индексы, срезы, матрицы. Изменение списка. Методы, специфичные для типа. Проверка выхода за границы. Вложенные списки. Дополнительные операции над списками. Генераторы списков. Синтаксис расширенного генератора списков.

Тема 1.6. Множества, словари. Функции.

Основы функций. Создание функций. Инструкция `def`. Полиморфизм в языке Python. Локальные переменные. Области видимости и аргументы. Правила видимости. Основы видимости имен в языке Python. Инструкция `global`. Способы доступа к глобальным переменным. Вложенные функции. Фабричные функции. `lambda`-выражения. Передача аргументов. Специальные режимы сопоставления аргументов.

Словари. Литералы словарей. Операции отображения. Операции над словарями. Изменение словарей. Способы создания словарей. Сортировка по ключам. Отсутствующие ключи. Использование словарей для имитации гибких списков. Использование словарей для разреженных данных. Кортежи. Особенности синтаксиса кортежей. Преобразования и неизменяемость. Множества.

Раздел 2. Основы работы с библиотеками языка Python.

Тема 2.1. Чтение и запись данных, форматы файлов.

Хранение массивов на диске в двоичном формате. Сохранение и загрузка текстовых файлов. Чтение текстовых файлов порциями. Вывод данных в текстовом формате. Ручная обработка данных в формате с разделителями. Данные в формате JSON. XML и HTML: разбор веб-страниц. Взаимодействие с базами данных.

Тема 2.2. Форматирование данных.

Комбинирование и слияние наборов данных. Слияние объектов `DataFrame` как в базах данных. Слияние по индексу. Конкатенация вдоль оси. Комбинирование перекрывающихся данных. Изменение формы и поворот. Устранение дубликатов.

Преобразование данных с помощью функции или отображения. Замена значений. Переименование индексов осей. Дискретизация и раскладывание. Обнаружение и фильтрация выбросов. Перестановки и случайная выборка. Вычисление индикаторных переменных. Регулярные выражения. Векторные строковые функции в pandas.

Тема 2.3. Введение в библиотеки языка Python. Библиотека NumPy.

Обзор популярных библиотек. Особенности типы данных в NumPy. Работа с векторами и матрицами. Операции между массивами и скалярами. Индексирование и вырезание. Булево индексирование. Прихотливое индексирование. Транспонирование массивов и перестановка осей. Запись логических условий в виде операций с массивами. Математические и статистические операции. Методы булевых массивов. Сортировка. Устранение дубликатов и другие теоретико-множественные операции. Генерация случайных чисел. Моделирование случайных блужданий.

Тема 2.4. Библиотека Pandas.

Объекты Series, DataFrame. Индексные объекты. Базовая функциональность библиотеки. Вычисление описательных статистик. Корреляция и ковариация. Уникальные значения, счетчики значений и членство. Обработка и фильтрация отсутствующих данных. Иерархическое индексирование. Сводная статистика по уровню. Доступ по целочисленному индексу. Структура данных Panel. Преобразование словарей в табличный формат Pandas, загрузка данных из внешних источников. Особенности фильтрации и обращения к данным.

Тема 2.5. Аналитические возможности библиотек Python.

Фильтрация. Сортировка и ранжирование. Группировки и ряды распределения. Сводные таблицы. Аналитические отчеты.

Тема 2.6. Построение графиков и визуализация данных в Python. Библиотека matplotlib.

Введение в API библиотеки matplotlib. Рисунки и подграфики. Цвета, маркеры и стили линий. Риски, метки и надписи. Аннотации и рисование в подграфике. Сохранение графиков в файле. График плотности распределения. Функции построения графиков в pandas. Линейные графики. Столбчатые диаграммы. Гистограммы и графики плотности. Диаграммы рассеяния. Изображение нескольких графиков в одной области.

Б1.В.24.02 «Современные методы обработки информации и визуализации данных»

Раздел 1. Основы обработки и анализа данных

Тема 1.1. Введение в анализ данных

Предмет и задачи дисциплины. Состав современного программного обеспечения анализа данных. Основные направления и проблемы анализа массовых данных. Понятие Data Mining. Data Mining как часть рынка интеллектуальных технологий. Набор данных и их атрибутов. Измерения. Шкалы измерения признаков. Задача классификации. Задача регрессии. Задача прогнозирования. Задача кластеризации. Методы визуализации.

Процесс Data Mining. Средства извлечения данных: методы и возможности. Начальные этапы: анализ предметной области; постановка задачи, подготовка данных. Построение и использование модели. Стандарты Data Mining. Стандарт CRISP-DM. Стандарт PMML. Стандарты SQL/MM и OLE DB for Data Mining. Другие стандарты, описывающие методологию Data Mining. Категории статистики: совокупность, статистические показатели, признаки. Подготовка исходных массивов данных.

Тема 1.2. Формирование массивов данных

Технологии формирования наборов данных и информационных массивов. Сэмплинг. Парсинг данных. Типы наборов данных. Форматы хранения данных. Метаданные. Открытые данные. Базы данных.

Статистическое наблюдение, социологический опрос. Цифровые техники проведения наблюдения и сбора данных.

Тема 1.3. Описательный анализ данных

Сущность, значение, виды и способы расчета статистических величин (абсолютных, относительных и средних).

Вариация признака в совокупности, значение ее статистического изучения. Ряд распределения как метод отражения вариации признака. Атрибутивные и вариационные ряды. Дискретные и интервальные вариационные ряды. Основные структурные характеристики вариационного ряда: мода, медиана, квартили, децили, перцентили. Основные показатели вариации: размах вариации, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации. Правило сложения дисперсий. Эмпирический коэффициент детерминации. Эмпирическое корреляционное отношение.

Эмпирическое распределение. Теоретические распределения. Показатели эксцесса и асимметрии. Статистические критерии оценки согласованности эмпирических и теоретических распределений.

Тема 1.4. Подготовка и обработка данных. Качество данных

Принципы формирования массивов данных. Этапы подготовки данных. Проблемы и ошибки. Очистка данных. Обработка пропущенных значений (импутация данных). Обработка дубликатов. Группировка данных. Квантование. Виды группировок. Определение числа групп и величины интервалов группировки. Исследование совокупности на наличие выбросов и аномалий. Трансформация данных. Приведение данных к сопоставимому виду. Задача редукции данных. Основные технологии и методы редукции данных.

Тема 1.5. Визуализация данных

Методы визуализации данных в системах обработки статистических данных. Виды и классификация статистических таблиц. Формулировка заголовка таблицы: принципы и обязательные составляющие. Элементы таблицы и этапы ее построения. Электронные таблицы. Сводные таблицы. OLAP- технологии.

История развития графических методов, современные тенденции в использовании графического метода визуализации. Элементы графика и принципы его построения. Отдельные виды графиков и диаграмм: назначение и особенности построения.

Современные приемы визуализации данных. Дашборды. Инфографика в различных предметных областях.

Тема 1.6. Проверка статистических гипотез

Виды статистических гипотез. Ошибки первого и второго рода. Т-тест Стьюдента. Критические значения. Уровень значимости. Статистические гипотезы о равенстве средних.

Тема 1.7. Исследование взаимосвязей.

Понятие о взаимосвязи явлений и процессов. Задачи статистического изучения взаимосвязей. Виды и формы связей. Технологии исследования взаимосвязей переменных разных типов шкал.

Основы корреляционно-регрессионного анализа. Парная и множественная корреляция. Уравнение регрессии. Оценка параметров уравнения регрессии, интерпретация параметров. Оценка статистической значимости параметров парной линейной регрессии.

Непараметрические методы измерения взаимосвязей. Построение таблиц сопряженности. Определение силы связи номинальных переменных. Оценка

значимости связи в таблицах сопряженности. Оценка связи ранговых переменных.

Раздел 2. Основы интеллектуального анализа данных

Тема 2.1. Задача регрессии

Регрессия, метрики качества регрессии. Линейная регрессия. Библиотека sklearn. Пример: линейная регрессия на одном признаке, её геометрия, идея обучения. Идея обучения для многомерной регрессии. Градиентный спуск. Проблемы градиентного спуска. Метрики качества. Интерпретируемость линейной модели. Групповая линейная регрессия.

Тема 2.2. Задача классификации данных

Идея классификации. Логистическая регрессия. Меры расстояния. Алгоритмы кластеризации (иерархические алгоритмы, алгоритмы квадратичной ошибки, выделение связанных компонент). Бинарные классификации. Множественные классификации. Технологии проведения дискриминантного анализа. Пробит- и логит-регрессии. Нейросетевые техники. Оценка качества проведения классификации, метрики качества. ROC-кривые. Технологии получения деревьев классификации и регрессии.

Тема 2.3. Задача кластеризации данных

Алгоритмы и программы кластерного анализа признаков приемами оптимизации, подготовка исходных данных, процедура решения, анализ результатов. Агломеративные и дивизивные процедуры. Расстояния в признаковом пространстве. Алгоритмы и программы аналитической многомерной группировки, подготовка исходных данных, варианты задания групп, интерпретация результатов.

Тема 2.4. Основы работы с временными рядами и прогнозирования

Основы работы с временными рядами. Индексирование, выборка, подмножества. Временные ряды с неуникальными индексами. Диапазоны дат, частоты и сдвиг. Генерация диапазонов дат. Частоты и смещения дат. Сдвиг данных с опережением и с запаздыванием. Графики временных рядов. Скользящие функции: экспоненциально взвешенные функции, бинарные скользящие функции.

Тема 2.5. Обработка текстовых данных

Задачи анализа текстовых данных. Обработка естественного языка. Мешок слов. Стемминг и лемматизация. Классификация текстов. Семантический анализ текстов. Семантическое ядро. Показатели анализа текстов: тошнота, доля уникальных и значимых слов.

Основная литература

Дисциплина «Основы алгоритмизации и программирования»

1. Костюкова, Н. И. Графы и их применение : учебное пособие / Н. И. Костюкова. — 3-е изд. — Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 147 с. — ISBN 978-5-4497-0367-5. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/89435.html>

2. Костюкова, Н.И. Комбинаторные алгоритмы для программистов : учебное пособие / Н.И. Костюкова. — 2-е изд. — Москва : ИНТУИТ, 2016. — 216 с. — ISBN 978-5-9556-0069-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/100621>

3. Федоров, Д. Ю. Программирование на языке высокого уровня python : учебное пособие для прикладного бакалавриата / Д. Ю. Федоров. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2019. — 161 с. — (Бакалавр. Прикладной курс). — ISBN 978-5-534-10971-9. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://www.biblio-online.ru/bcode/437489> (дата обращения: 01.11.2019).

4. Гниденко, И. Г. Технологии и методы программирования : учебное пособие

для прикладного бакалавриата / И. Г. Гниденко, Ф. Ф. Павлов, Д. Ю. Федоров. — Москва : Издательство Юрайт, 2019. — 235 с. — (Бакалавр. Прикладной курс). — ISBN 978-5-534-02816-4. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://www.biblio-online.ru/bcode/433611> (дата обращения: 01.11.2019).

5. Основы алгоритмизации и программирования на Python : учеб. пособие / С.Р. Гуриков. — М. : ФОРУМ : ИНФРА-М, 2018. — 343 с. - Текст : электронный. - URL: <http://znanium.com/catalog/product/924699>

Дисциплина «Современные методы обработки информации и визуализации данных»

6. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — URL: <https://urait.ru/bcode/450166>

7. Елисеева И. И., Боченина М. В., Капралова Е. Б., Курышева С. В., Нерадовская Ю. В., Флуд Н. А., Науменко Н. В., Щирина А. Н., Парик И. Ю. Статистика. Практикум [Электронный ресурс]: Учебное пособие Для академического бакалавриата. - Москва: Юрайт, 2019. - 514 с – Режим доступа: <https://urait.ru/bcode/425262>

8. Мхитарян В. С., Агапова Т. Н., Суринов А. Е., Луппов А. Б., Миронкина Ю. Н. Статистика. В 2 ч. Часть 1 [Электронный ресурс]: Учебник и практикум для вузов. - Москва: Юрайт, 2021. - 249 с – Режим доступа: <https://urait.ru/bcode/475170>