

РАБОЧАЯ ПРОГРАММА
модуля(курса)
«Программирование на Python»

1. АННОТАЦИЯ

Дисциплина «Программирование на Python» включает учебный материал, способствующий получению практического опыта, навыков и умений программирования на языке программирования Python. Изучаются базовые алгоритмы и простейшие структуры данных, а также практика применения базовых возможностей и библиотек языка для решения прикладных задач.

Цель дисциплины – повысить профессиональную компетенцию специалиста в области разработки программных приложений на языке Python, в следующих областях: работа с таблицами и структурами данных, их чтение, анализ и последующая обработка в задачах статистики и машинного обучения. Дисциплина рассматривается, как один из основных курсов программы профессиональной переподготовки «Искусственный интеллект и глубокое обучение».

2. СОДЕРЖАНИЕ

Учебная программа курса

| № п/п | Наименование модуля, разделов и тем | Содержание обучения (по темам в дидактических единицах), наименование и тематика лабораторных работ, практических занятий (семинаров), самостоятельной работы с указанием кол-ва часов, используемых образовательных технологий и рекомендуемой литературы |
|----------|--|---|
| 1. | 2. | 3. |
| 1 | Тема 1. Введение в Python и Jupyter Notebook. Python без программирования. | 1.1. Основная цель - привить навыки работы в среде Jupyter Notebook и Google Colab. Язык программирования Python. Среда Jupyter Notebook. Установка дистрибутива Anaconda (или другой эквивалентной системы). Работа в Google Colab (или другой эквивалентной среде). 1.2. Python без программирования. Основная цель - дать представление и привить навыки по выполнению арифметических операций и простейших операций по обработке текста в Python (Jupyter Notebook). Целые числа, вещественные числа, строки. Функции ввода и вывода в Python. (2 часа) |
| 2 | Тема 2. Контейнеры в языке Python. Основные конструкции языка Python. | 2.1. Контейнеры в языке Python. Основная цель - дать представление о списках (list), кортежах (tuple) и словарях (dict) и привить навыки работы с ними. Использование контейнеров для обработки текстовых данных. 2.2. Основные конструкции языка Python. Основная цель - дать представление о синтаксических конструкциях языка питон для организации ветвлений и циклов и привить навыки работы с ними. Булевский (логический) тип данных. Логические выражения. Ветвления (конструкции if, if else, if elif else). Цикл for. Цикл while. Команды break и continue. (2 часа) |
| 3 | Тема 3. Функции, модули и библиотеки в языке Python. | 3.1. Функции, модули и библиотеки в языке Python. Основная цель - познакомить слушателей с концепцией |

| | | |
|---|--|---|
| | | функции, модуля и библиотеки, формальных параметров, областью видимости переменных и сформировать навыки по работе с функциями, модулями и библиотеками и созданию своих функций и библиотек. (2 часа) |
| 4 | Тема 4. Работа с текстовыми данными. Работа с табличными данными. | 4.1 Работа с текстовыми данными. Основная цель - привить навыки работы с простейшими функциями работы с текстовыми строками и текстовыми файлами. 4.2. Работа с табличными данными. Основная цель - дать представление об особенностях работы с табличными данными и привить навыки работы с функциями библиотеки Pandas. Библиотека Pandas и работа с табличными данными. Обзор библиотеки Pandas. Возможности библиотеки для решения задач обработки данных. (2 часа) |
| 5 | Практические занятия (семинары) | Практические задания с использованием необходимых библиотек языка Python в приложении Jupyter Notebook). (24 часа) |
| 6 | Самостоятельная работа | Выполнение домашних заданий по теме занятия (14 часов) |
| 7 | Зачет | Лабораторная работа (2 часа) |

3. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

(формы аттестации, оценочные и методические материалы)

Промежуточная аттестация представляет собой собеседование, которое проводится по результатам практических занятий (семинаров). Для сдачи зачета слушатель должен сделать все практические задания:

1. Загрузите из файла heart.csv (<https://www.kaggle.com/ronitf/heart-diseaseuci?select=heart.csv>) данные о сердечных заболеваниях.

Выполните следующие задания:

- 1) Сколько образцов (объектов) содержит данный датасет?
- 2) Сколько атрибутов (признаков) содержит данный датасет? Подробно опишите значение каждого признака.
- 3) Опишите тип каждого признака (числовой / дискретный / непрерывный / категориальный / номинальный / бинарный / ординальный)?
- 4) Вычислите, сколько мужчин/женщин в датасете?
- 5) Вычислите описательные статистики для количественных признаков (среднее значение, медиана, мода, размах, дисперсия, среднеквадратичное отклонение, 1й/2й/3й квартили, межквартильный размах).
- 6) Постройте гистограммы для признаков age, trestbps, chol, thalach, oldpeak. Расположите гистограммы на одном графике в одну линию. Подпишите оси каждой гистограммы.
- 7) Постройте диаграммы размаха для признаков age, trestbps, chol, thalach, oldpeak.
- 8) Постройте на одном графике две кривые PDF (probability density function) для признака chol. Одна PDF для мужчин, другая - для женщин. Подпишите оси, добавьте легенду.
- 9) Для признаков, которые не были указаны в п.6-7, постройте полигоны частот (подграфики расположите в нескольких строках и столбцах). Подпишите оси.
- 10) Сгруппируйте данные по полу и вычислите для каждой группы среднее значение признака chol, применив функцию агрегации. Изобразите результаты в виде столбчатой диаграммы, где столбцы должны соответствовать полу, а высота столбцов - соответствующим средним значениям признака chol. Добавьте к каждому столбцу планку погрешности, отражающую среднеквадратичное отклонение.

11) Постройте следующие диаграммы рассеяния:

- trestbps от age
- chol от age
- thalach от age
- oldpeak от age

Изобразите точки на диаграммах разными цветами в зависимости от пола.

Подпишите оси и добавьте легенду. Попробуйте визуально определить, коррелируют ли рассматриваемые переменные с возрастом. Проверьте свои предположения, вычислив коэффициенты корреляции Спирмена. Сделайте выводы.

12) Проверьте признаки age, trestbps, chol, thalach, oldpeak на нормальность с помощью критерия Шапиро-Уилка (6 часов)

2. Загрузите датасет ISOLET (<https://archive.ics.uci.edu/ml/datasets/ISOLET>). Выполните следующие задания:

- 1) Опишите рассматриваемый датасет.
- 2) Выполните нормализацию признаков.
- 3) Разбейте данные на обучающую и тестовую выборку.
- 4) Вызовите метод k ближайших соседей. Постройте графики зависимости ошибки этого метода на обучающей и тестовой выборках от k . Сделайте вывод.
- 5) Выполните процедуру перекрестного контроля (5-fold, 10-fold, LOO) с обучающей выборкой. Постройте графики зависимости CV-ошибки от числа используемых соседей в методе k ближайших соседей. Выберите наилучшую модель и проверьте ее качество на тестовой выборке.
- 6) Примените к рассматриваемым данным
 - Линейный дискриминантный анализ
 - Квадратичный дискриминантный анализ
 - Логистическую регрессию

Для каждого метода вычислите ошибки на обучающей и тестовой выборках.

7) На рассматриваемых данных обучите следующие классификаторы:

- Random Forest
- Extremely Random Trees

Постройте графики зависимости ошибки на обучающей и тестовой выборке от количества используемых деревьев.

8) Натренируйте на рассматриваемых данных нейронную сеть с одним скрытым слоем ("vanilla") с 200 нейронами в нем (`hidden_layer_sizes = (200,)`). В качестве функции активации используйте положительную срезку (`activation = 'relu'`). Постройте графики зависимости ошибки на обучающей и тестовой выборке от параметра α (`weight_decay`). (6 часов)

Методы контроля и оценки результатов освоения модуля

| № п/п | Наименование процедуры | Основные показатели оценки | Формы и методы контроля и оценки |
|-------|---|--|----------------------------------|
| 1 | Промежуточная аттестация. Программирование на Python | Владеет базовыми алгоритмами и простейшими структурами данных, а также практикой применения базовых возможностей и библиотек языка для решения прикладных задач. | Зачет/ Лабораторная работа |

Критерии оценки

| № п/п | Наименование процедуры | Основные показатели оценки | | Формы и методы контроля и оценки |
|-------|---|----------------------------|---|----------------------------------|
| | Промежуточная аттестация. Программирование на Python | Зачтено | При выполнении задания выполнены все этапы задачи. | Зачет/ Лабораторная работа |
| | | | При выполнении задания выполнены все этапы алгоритма, но отдельные части и аргументация не уточнены или частично не были представлены. | |
| | | | Не выполнены все этапы алгоритма, допущены логические ошибки и полученный результат не обоснован. | |
| | | Незачтено | Слушатель не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки и не умеет применять базовые термины и знания при решении типовых практических задач. | |

4. УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ МОДУЛЯ

4.1 Учебно-методическое и информационное обеспечение программы:

Для эффективного освоения компетенций, формируемых учебной дисциплиной важно использование в учебном процессе активных и интерактивных форм проведения занятий.

Изучение учебной дисциплины предполагает наличие аудиторной и самостоятельной видов работ слушателей. В ходе практических занятий рассматриваются практические задачи из практики с целью наиболее полного овладения умениями и навыками.

Лекции по учебной дисциплине призваны формировать знания, предусмотренные учебной программой, и включают теоретическую базу обработки данных, на базе которой строятся прикладные аспекты.

Наряду с проработкой основной литературы (глав базового учебника) предусмотрено самостоятельное чтение дополнительной литературы (статей и других научных публикаций).

Практические занятия в малых группах и самостоятельная внеаудиторная работа направлены на выработку навыков анализа данных.

Для достижения поставленных целей преподавания дисциплины реализуются следующие средства, способы и организационные мероприятия:

- изучение теоретического материала дисциплины на лекции с использованием компьютерных технологий;
- самостоятельное изучение теоретического материала дисциплины с использованием Internet-ресурсов, информационных баз, электронных библиотек, методических разработок, специальной и научной литературы;
- закрепление теоретического материала при проведении практических занятий с использованием учебного и научного оборудования, выполнения проблемно-ориентированных, поисковых, творческих заданий.

Самостоятельная работа слушателей включает:

1. Изучение учебной литературы по курсу.
2. Решение практических ситуаций и задач

3. Изучение источников управленческой информации
4. Работу с ресурсами Интернет
5. Решение практических ситуаций в виде творческих заданий
6. Изучение статистической информации
7. Подготовку к дифференцированному зачету.

4.2. Содержание комплекта учебно-методических материалов.

а) основная литература:

1. МакГрат М. (2021) Программирование на Python для начинающих Эксмо
2. Géron Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2nd Edition. O'Reilly Media. 2019. Рус. пер. 1-го издания: О. Жерон Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow. ООО Диалектика. 2018
3. A.C. Müller, S. Guido Introduction to Machine Learning with Python: A Guide for Data Scientists. Рус. пер.: А. Мюллер, С. Гвидо Введение в машинное обучение с помощью Python. Вильямс, 2017.

б) дополнительная литература:

4. Bhasin H. (2019). Python Basics : A Self-Teaching Introduction. Dulles, Virginia: Mercury Learning & Information. <http://search.ebscohost.com/login.aspx?direct=true&site=eds-live&db=edsebk&AN=1991381>
5. G. James, D. Witten, T. Hastie, R. Tibshirani An Introduction to Statistical Learning with Applications in R. Springer, 2013. <https://faculty.marshall.usc.edu/gareth-james/ISL/>
Рус. пер.: Г. Джеймс, Д. Уиттон, Т. Хасте, Р. Тибишрани Введение в статистическое обучение с примерами на языке R. ДМК Пресс, 2016.
6. I. Goodfellow, Y. Bengio, A. Courville Deep Learning The MIT Press 2016. Рус. пер.: И. Бенджио, Я. Гудфеллоу, А. Курвилль Глубокое обучение. ДМК Пресс, 2017
7. Материалы курса лекций «Основы программирования»: НОУ ИНТУИТ:
8. <http://www.intuit.ru/studies/courses/2193/67/info>, режим доступа – свободный.

в) программное обеспечение и Интернет-ресурсы

9. Anaconda: the Most World's Most Popular Data Science Platform <https://www.anaconda.com/>
10. scikit-learn: Machine Learning in Python <https://scikit-learn.org/>
11. TensorFlow: Комплексная платформа машинного обучения с открытым исходным кодом <https://www.tensorflow.org/>
12. pyTorch: An open source machine learning framework that accelerates the path from research prototyping to production deployment. <https://pytorch.org/> Лекции и практические занятия проводятся с использованием возможностей мультимедийного класса. Использование в учебном процессе активных и интерактивных форм проведения занятий (компьютерных симуляций, деловых и ролевых игр, разбор конкретных ситуаций).

4.3. Материально-технические условия реализации программы:

Материально-техническая база

| № п.п. | Наименование модуля (тем, разделов) | Материально-технические условия для реализации программ (наличие лабораторий, производственных участков и т.п. по профилю программы профессиональной переподготовки) |
|--------|--|--|
| 1. | Тема 1. Введение в Python и Jupyter Notebook. Python без программирования. | Реализация дисциплины предполагает наличие у слушателя компьютера с возможностью работы в Python |
| 2. | Тема 2. Контейнеры в языке | |

| | | |
|---|---|--|
| | Python. Основные конструкции языка Python. | |
| 3 | Тема 3. Функции, модули и библиотеки в языке Python. | |
| 4 | Тема 4. Работа с текстовыми данными. Работа с табличными данными. | |