

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Учебный центр по ДПО ИМИТ



УТВЕРЖДАЮ

Проректор по ВРиДО

Е.Н. Гончарова

20 21 г.

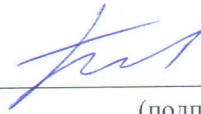
ПРОГРАММА ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

«Основы Data Science и визуализация данных на Python»

Программа рассмотрена и одобрена на заседании методической комиссии Института математики и информационных технологий

протокол № 5 от «15» 02 2024 г.,

Председатель методической
комиссии ИМИТ



/И.В. Пономарев/

(подпись)

1. Общая характеристика программы

Наука о данных (Data Science) – это практическая дисциплина, которая занимается изучением методов извлечения знаний из данных, основывается на методах и теориях из многих областей знаний, объединяет методы по обработке данных, статистические методы, методы интеллектуального анализа данных и приложения искусственного интеллекта для работы с данными.

Программа «Основы Data Science и визуализация данных на Python» направлена на получение слушателями теоретических знаний и новых компетенций, необходимых для решения широкого круга прикладных задач анализа данных с помощью языка программирования Python.

Python – это универсальный, простой, гибкий и очень популярный язык программирования, для которого создано большое количество модулей и библиотек, содержащих уже реализованные методы и алгоритмы для анализа данных, визуализации и машинного обучения. Python позволяет организовать процесс анализа данных без применения глубоких математических знаний, дает возможность легко визуализировать данные и строить модели машинного обучения.

В результате освоения данного курса слушатели получают теоретические знания и практические навыки, которые позволят им самостоятельно обрабатывать и преобразовывать «сырые» данные в полезные знания, что станет хорошим фундаментом для карьеры в области Data Science.

В основу курса «Основы Data Science и визуализация данных на Python» заложены принципы модульности, последовательности освоения материала и практической направленности. Содержание модулей направлено на:

- знакомство с базовыми конструкциями языка Python;
- изучение методов анализа, визуализации и обработки данных, алгоритмов машинного обучения;
- формирование умений и навыков применения модулей и пакетов Python для решения прикладных задач.

1.1. Цель реализации программы

Целью реализации программы повышения квалификации является получение слушателями новых компетенций в области анализа данных и машинного обучения, освоение языка программирования Python, модулей и библиотек, содержащих инструменты для анализа и обработки данных.

1.2. Планируемые результаты обучения

В результате освоения данного курса слушатели должны:

– **знать:** базовые конструкции языка Python, основные методы обработки, визуализации и анализа данных, алгоритмы машинного обучения;

– **уметь:** применять базовые конструкции языка Python, основные методы обработки, визуализации и анализа данных, алгоритмы машинного обучения для решения поставленных задач;

– **владеть:** основными методами обработки, визуализации и анализа данных, алгоритмами машинного обучения, навыками применения модулей и пакетов Python для решения прикладных задач.

1.3. Категория слушателей.

К освоению курса ДПО допускаются лица, имеющие и (или) получающие СПО и (или) ВО и все заинтересованные лица. Требования к уровню подготовки поступающего на обучение, необходимые для освоения программы, определяются с целью оценки возможности освоения программы ДПО (высшее образование или среднее специальное).

Курс ориентирован на представителей нетехнических специальностей или начинающих IT-специалистов. Особых требований к уровню подготовки поступающего на обучение – нет.

1.4. *Трудоемкость обучения – 72 часа.*

1.5. *Форма обучения – дистанционная.*

2. Содержание программы

1.6. Учебный план программы повышения квалификации «Основы Data Science и визуализация данных на Python»

№ п/п	Наименование разделов, (дисциплин, модулей)	Всего час.	Дистанционные занятия, час.			СРС, час.	Форма контроля
			лекции	практич., семинары	Лаборатор.		
1.	Введение в науку о данных	10	2	-	2	6	Отчет по лабораторной работе
2.	Элементы Python	12	2	-	4	6	Отчет по лабораторной работе
3.	Библиотека Pandas для анализа данных	12	2	-	4	6	Отчет по лабораторной работе
4.	Визуализация с помощью библиотеки Matplotlib	12	2	-	4	6	Отчет по лабораторной работе
5.	Статистика. Подготовка и исследование данных	10	2	-	2	6	Отчет по лабораторной работе
6.	Машинное	14	2	-	6	6	Отчет по лабораторной

	обучение в Scikit-learn						работе
7.	Итоговая аттестация	2	-	2	-	-	Защита проекта в форме онлайн-собеседования
	Итого	72	12	2	22	36	

1.7. Учебно-тематический план

№ п/п	Наименование разделов, тем	Всего час.	Дистанционные занятия, час.			СРС, час.
			лекции	практич., семинары	Лаборатор.	
1.	Наука о данных: основные понятия и определения. Язык программирования Python для анализа данных. Среды разработки.	10	2	-	2	6
2.	Элементы Python: переменные и операторы, ввод и вывод, инструкции, структуры данных; условные операторы и циклы; встроенные и пользовательские функции, lambda-выражения, пакеты и модули функций	12	2	-	4	6
3.	Библиотека Pandas для анализа данных. Объекты Series и DataFrame. Основные функции для работы с DataFrame	12	2	-	4	6
4.	Визуализация с помощью библиотеки Matplotlib. Виды графиков и функции для их построения; настройка графиков	12	2	-	4	6
5.	Элементы статистики. Методы подготовки и исследования данных: очистка, фильтрация, поиск	10	2	-	2	6

	закономерностей.					
6.	Машинное обучение. Задачи кластеризации, классификации и регрессии. Библиотека Scikit-learn. Построение и оптимизация моделей.	14	2	-	6	6
7.	Итоговая аттестация	2	-	2	-	-
	Итого	72	10	2	24	36

1.8. Календарный учебный график.

Дата	Время (часов)	Тема	Преподаватель
15.03.2021	6	Введение в науку о данных	Кротова Ольга Сергеевна Козлов Денис Юрьевич Хворова Любовь Анатольевна
16.03.2021	6	Введение в науку о данных. Элементы Python	Кротова Ольга Сергеевна
17.03.2021	6	Элементы Python	Кротова Ольга Сергеевна
18.03.2021	6	Элементы Python. Библиотека Pandas для анализа данных	Кротова Ольга Сергеевна
19.03.2021	6	Библиотека Pandas для анализа данных.	Кротова Ольга Сергеевна
20.03.2021	6	Библиотека Pandas для анализа данных. Визуализация с помощью библиотеки Matplotlib	Кротова Ольга Сергеевна
21.03.2021	6	Визуализация с помощью библиотеки Matplotlib.	Кротова Ольга Сергеевна
22.03.2021	6	Визуализация с помощью библиотеки Matplotlib. Элементы статистики. Методы подготовки и исследования данных	Кротова Ольга Сергеевна Хворова Любовь Анатольевна

23.03.2021	6	Элементы статистики. Методы подготовки и исследования данных.	Хворова Любовь Анатольевна
24.03.2021	6	Элементы статистики. Методы подготовки и исследования данных. Машинное обучение в Scikit-learn	Хворова Любовь Анатольевна Козлов Денис Юрьевич
25.03.2021	6	Машинное обучение в Scikit-learn	Козлов Денис Юрьевич
26.03.2021	4	Машинное обучение в Scikit-learn	Козлов Денис Юрьевич
27.03.2021	2	Итоговая аттестация	Кротова Ольга Сергеевна Козлов Денис Юрьевич Хворова Любовь Анатольевна

1.9. Рабочая программа дисциплины «Основы Data Science и визуализация данных на Python»

Раздел 1. Введение в науку о данных (10 часов)

Тема 1.3. Наука о данных: основные понятия и определения (4 часов)

Тема 1.1. Python для анализа данных (2 часа)

Тема 1.2. Среды разработки. Среды разработки для анализа данных: дистрибутив Anaconda, Jupyter Notebook, Google Colaboratory (4 часа)

Раздел 2. Элементы Python (12 часов)

Тема 2.1. Переменные и операторы. Ввод и вывод данных. Инструкции и структура программы. Доступ к документации. (2 часа)

Тема 2.2. Структуры данных (2 часа)

Тема 2.3. Условные операторы и циклы (4 часов)

Тема 2.4. Встроенные и пользовательские функции, lambda-выражения, пакеты и модули функций (4 часа)

Раздел 3. Библиотека Pandas для анализа данных (12 часов)

Тема 3.1. Библиотека Pandas. Структуры данных Series и DataFrame (4 часа)

Тема 3.2. Основные функции для работы с DataFrame (4 часа)

Тема 3.3. Операции над данными. Комбинирование данных из разных источников. Обработка пропущенных значений (4 часа)

Раздел 4. Визуализация с помощью библиотеки Matplotlib (12 часов)

Тема 4.1. Библиотеки Python для визуализации данных: Matplotlib, Seaborn, Plotly (4 часа)

Тема 4.2. Виды графиков и функции для их построения с помощью Matplotlib (4 часа)

Тема 4.3. Настройка графиков (4 часа)

Раздел 5. Элементы статистики. Методы подготовки и исследования данных (10 часов)

Тема 5.1. Элементы статистического анализа данных (4 часа)

Тема 5.2. Подготовка данных: методы очистки, фильтрации и устранения ошибок в данных (2 часа)

Тема 5.3. Исследование данных: поиск закономерностей в данных на основе визуальных и описательных методов (4 часа)

Раздел 6. Машинное обучение в Scikit-learn (14 часов)

Тема 6.1. Введение в машинное обучение. Обучение с учителем и без учителя (2 часа)

Тема 6.2. Задачи кластеризации, классификации и регрессии (4 часа)

Тема 6.3. Библиотека Scikit-learn для машинного обучения (4 часа)

Тема 6.4. Построение и оптимизация моделей машинного обучения (4 часа)

Итоговая аттестация (2 часа)

Перечень лабораторных работ и практических (семинарских) занятий

№ темы	Наименование практических (семинарских) занятий лабораторных работ	Дистанционных занятий, час.
1.	Элементы языка программирования Python	4
2.	Обработка данных в Pandas	4
3.	Визуализация данных с помощью Matplotlib	4
4.	Статистический анализ. Подготовка и исследование данных	4
5.	Кластерный анализ	2
6.	Задачи классификации и регрессии	4
7.	Построение и оптимизация моделей в Scikit-learn	2

Виды самостоятельной работы слушателей (СРС)

№ п/п	Вид СРС	Трудоемкость, час.
1.	Самостоятельная работа реконструктивно-вариативного типа (Индивидуальное задание)	20
2.	Эвристическая самостоятельная работа (Работа над индивидуальным	16

2. Условия реализации программы (организационно-педагогические)

2.1. Материально-технические условия

Компьютер с комплектацией:

- процессор не ниже Core 2 DUO 2.4Hz и ОЗУ: не ниже 4GB;
- ОС: WINDOWS 7 или выше;
- Браузер: Google Chrome, Yandex.Browser;
- Дистрибутив Anaconda
- доступ к образовательному пространству, расположенному по адресу: <https://public.edu.asu.ru/>, где расположены материалы для самостоятельного изучения и выполнения.

2.2. Учебно-методическое и информационное обеспечение программы

Литература

Список основной литературы:

1. Доусон М. Програмируем на Python. – СПб: Питер, 2014. – 416 с.
2. Грас Дж. DataScience. Наука о данных с нуля. – СПб: БХВ-Петербург, 2017. – 336 с.
3. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных. – СПб: Питер, 2017. – 336 с.
4. Рашка С. Python и машинное обучение. – М.: ДМК Пресс, 2017. – 418 с.

Список дополнительной литературы:

1. Хейдт М. Изучаем pandas. – М.: ДМК Пресс, 2018. – 438 с.
2. Вандер П. Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер, 2018. – 576 с.

Интернет-ресурсы:

1. Google Colaboratory – URL: <https://colab.research.google.com/notebooks/intro.ipynb>
2. Документация Python – URL: <https://www.python.org/doc/>

3. Оценка качества освоения программы (форма аттестации, оценочные и методические материалы)

Мониторинг степени освоения учебного материала слушателями осуществляется в форме отчетов по лабораторным работам и итогового зачета (защита проекта в формате онлайн-собеседования).

Практические задания (лабораторные работы) к каждому занятию предметно-методического раздела являются обязательными и подлежат размещению на образовательном портале. Результаты проверенных заданий обучающихся также размещаются на образовательном портале с возможными комментариями по поводу выявленных ошибок.

Оценка качества освоения программы осуществляется в виде итогового зачета (защита проекта в форме онлайн-собеседования).

4. Кадровые условия (составители программы)

Составитель рабочей программы – Козлов Денис Юрьевич, к.ф.-м.н., доцент, зав.каф. информатики Алтайского государственного университета, Кротова Ольга Сергеевна, преподаватель кафедры теоретической кибернетики и прикладной математики Алтайского государственного университета, Хворова Любовь Анатольевна, к.т.н., доцент кафедры теоретической кибернетики и прикладной математики Алтайского государственного университета.

Директор ИДПО

 Т.Г. Строителева

Директор Учебного центра по ДПО ИМИТ

 О.А. Жданова

УТВЕРЖДАЮ

Директор ИДПО

 Т.Г. Строителева

«___» _____ 20__ г.

РАСПИСАНИЕ ЗАНЯТИЙ

Программа: «Основы Data Science и визуализация данных на Python»

(кол-во час. - 72, форма обучения - дистанционная)

Сроки обучения: 15-27 марта 2021 года

Дата	Время (часов)	Тема	Преподаватель
15.03.2021	6	Введение в науку о данных	Кротова Ольга Сергеевна Козлов Денис Юрьевич Хворова Любовь Анатольевна
16.03.2021	6	Введение в науку о данных. Элементы Python	Кротова Ольга Сергеевна
17.03.2021	6	Элементы Python	Кротова Ольга Сергеевна
18.03.2021	6	Элементы Python. Библиотека Pandas для анализа данных	Кротова Ольга Сергеевна
19.03.2021	6	Библиотека Pandas для анализа данных.	Кротова Ольга Сергеевна
20.03.2021	6	Библиотека Pandas для анализа данных. Визуализация с помощью библиотеки Matplotlib	Кротова Ольга Сергеевна
21.03.2021	6	Визуализация с помощью библиотеки Matplotlib.	Кротова Ольга Сергеевна
22.03.2021	6	Визуализация с помощью библиотеки Matplotlib. Элементы статистики. Методы подготовки и исследования данных	Кротова Ольга Сергеевна Хворова Любовь Анатольевна

23.03.2021	6	Элементы статистики. Методы подготовки и исследования данных.	Хворова Любовь Анатольевна
24.03.2021	6	Элементы статистики. Методы подготовки и исследования данных. Машинное обучение в Scikit-learn	Хворова Любовь Анатольевна Козлов Денис Юрьевич
25.03.2021	6	Машинное обучение в Scikit-learn	Козлов Денис Юрьевич
26.03.2021	4	Машинное обучение в Scikit-learn	Козлов Денис Юрьевич
27.03.2021	2	Итоговая аттестация	Кротова Ольга Сергеевна Козлов Денис Юрьевич Хворова Любовь Анатольевна

Директор Учебного центра по ДПО ИМИТ
(подпись)


И.О.Ф. О.А. Жданова