

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
«Тюменский государственный университет»

СОГЛАСОВАНО

Руководитель дополнительной
образовательной программы

А.Н. Шелешов

(подпись) (инициалы, фамилия)

М.П.

«11» 07 2022 г.

УТВЕРЖДАЮ

Директор Центра дополнительного
образования ТюмГУ

И.В. Ковалева

(подпись) (инициалы, фамилия)

М.П.

«14» * 07 2022 г.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА

ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ

«Анализ табличных данных»

1. ОБЩАЯ ХАРАКТЕРИСТИКА ДОПОЛНИТЕЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1.1. Цель реализации дополнительной образовательной программы

Цели курса - ознакомить слушателей с основными методами обработки данных средствами языка Python, развить компетенцию “Применяет языки программирования для решения профессиональных задач”, необходимую для профессиональной деятельности специалиста по Data Science.

Задачи курса:

- знакомство с библиотекой Pandas языка программирования Python, предназначенной для обработки таблиц с данными;
- изучение алгоритмов обработки данных;
- повышение профессиональной эффективности;
- изучение базовых моделей искусственного интеллекта, применяемых для анализа данных;

1.2. Требования к обучающемуся

Уровень образования: неоконченное высшее. Не ниже второго курса бакалавриата и третьего курса специалитета.

1.3. Трудоемкость обучения

Общая трудоемкость программы – 250 академических часов.

2. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

- Освоение программы.
- Знакомство с алгоритмами анализа данных.
- Приобретение компетенции “Применяет языки программирования для решения профессиональных задач”, уровень сформированности компетенции: базовый (Применяет языки программирования (в т.ч. скрипты) для решения профессиональных задач под контролем более опытных специалистов)

Присвоение квалификации **06.042 Специалист по большим данным** (утверждена приказом Министерства труда и социальной защиты РФ от 6 июля 2020 №405н)

3. СОДЕРЖАНИЕ ДОПОЛНИТЕЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

3.1. Учебный план с применением дистанционных образовательных технологий и электронного обучения

№ п/п	Наименование разделов и/или дисциплин	Всего часов	В том числе	
			Аудиторная работа	Самост. работа
1	Знакомство с языком Python и средой разработки	20	3	17
2	Простейшие действия над таблицами с данными.	20	3	17

3	Подсчет статистических характеристик.	20	3	17
4	Визуализация данных.	20	3	17
5	Работа с пропусками в данных, поиск выбросов	20	4	16
6	Статистические тесты на сравнение средних значений и значимость коэффициента корреляции.	20	4	16
7	Алгоритмы кластеризации.	20	4	16
8	Линейная регрессия.	20	4	16
9	Алгоритмы классификации.	20	4	16
10	Разбор кейсов из индустрии	20	8	12
	Производственная практика	50		
	Итого	250	40	160

3.2. Содержание дополнительной образовательной программы

№ п/п	Дисциплина (раздел, тема)	Содержание
1	Знакомство с языком Python и средой GoogleColab	Знакомство с языком Python и средой GoogleColab, обзорное занятие.
2	Простейшие действия над таблицами с данными.	Создание таблиц, загрузка из файла, подсчет числа строк/столбцов, селекции.
3	Подсчет статистических характеристик.	Вычисление средних значений (медиан, отклонений) по столбцам. Вычисление коэффициента корреляции.
4	Визуализация данных.	Знакомство с библиотекой matplotlib, построение гистограмм, графиков по данным.
5	Работа с пропусками в данных, поиск выбросов	Алгоритмы работы с пропусками данных, алгоритмы детекции аномалий (выбросов) в данных.
6	Статистические тесты на сравнение средних значений и значимость коэффициента корреляции.	Понятие о статистическом тесте. Проверка двух средних значений на равенство, точный тест Фишера, значимость коэффициента корреляции.
7	Алгоритмы кластеризации.	Понятие об алгоритмах кластеризации, алгоритм k-means.

8	Линейная регрессия.	Построение модели линейной регрессии в языке Python. Ее применение в анализе данных.
9	Алгоритмы классификации.	Обзор алгоритмов классификации (kNN, решающие деревья, случайны лес, бустинг итд.)
10	Разбор кейсов из индустрии	Мастер-классы от индустриальных партнеров по разбору кейсов и задач из ИТ-отрасли.

3.3. Календарный учебный график

Период обучения (день / неделя / месяц)	Дисциплина (раздел, тема)
19.09-02.10	Знакомство с языком Python и средой GoogleColab
03.10-16.10	Простейшие действия над таблицами с данными.
17.10-30.10	Подсчет статистических характеристик.
31.10-13.11	Визуализация данных.
14.11-27.11	Работа с пропусками в данных, поиск выбросов
28.11-11.12	Статистические тесты на сравнение средних значений и значимость коэффициента корреляции.
12.12-25.12	Алгоритмы кластеризации.
13.02-26.02	Линейная регрессия.
27.02-12.03	Алгоритмы классификации.
13.03-26.03	Разбор кейсов
27.03-28.05	Производственная практика
29.05-04.06	Итоговая аттестация

4. УСЛОВИЯ РЕАЛИЗАЦИИ ДОПОЛНИТЕЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

4.1. Квалификация преподавателей, участвующих в реализации дополнительной образовательной программы

№ п/п	Преподаватель	Ученая степень, ученое звание, должность
1.	Шевляков Артём Николаевич	Доктор физико-математических наук, зам. директора Института математики и компьютерных наук Тюменского государственного университета.

4.2. Материально-технические условия реализации программы

№ п/п	Наименование помещения	Вид занятий	Наименование оборудования, программного обеспечения
1.	Аудитория	Лекции, практические занятия, консультации	Компьютер, мультимедийный проектор, экран, доска, флипчарт

4.3. Учебно-методическое обеспечение программы

4.3.1. Основная литература

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. – Москва : Издательство Юрайт, 2022. – 490 с. – (Высшее образование). – ISBN 978-5-534-00616-2. – Текст : электронный - URL: <https://urait.ru/bcode/489100> (дата обращения: 14.06.2022).
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. – Москва : Издательство Юрайт, 2022. – 174 с. – (Высшее образование). – ISBN 978-5-9916-5009-0. – Текст: электронный URL: <https://urait.ru/bcode/469306> (дата обращения: 14.06.2022).

4.3.2. Дополнительная литература

1. Кремер, Н. Ш. Математическая статистика : учебник и практикум для вузов / Н. Ш. Кремер. – Москва: Издательство Юрайт, 2022. – 259 с. – (Высшее образование). – ISBN 978-5-534-01654-3. – Текст : электронный // URL: <https://urait.ru/bcode/489976> (дата обращения: 14.06.2022).
2. Малугин, В. А. Математическая статистика: учебное пособие для вузов / В. А. Малугин. – Москва: Издательство Юрайт, 2022. – 218 с. – (Высшее образование). – ISBN 978-5-534-06965-5. – Текст: электронный // URL: <https://urait.ru/bcode/493394> (дата обращения: 14.06.2022).
3. Трофимов, А. Г. Математическая статистика: учебное пособие для вузов / А. Г. Трофимов. – 2-е изд. – Москва: Издательство Юрайт, 2022. – 257 с. – (Высшее образование). – ISBN 978-5-534-08874-8. – Текст: электронный // URL: <https://urait.ru/bcode/494524> (дата обращения: 14.06.2022).

5. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ДОПОЛНИТЕЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.

Предусмотрена итоговая аттестация. Форма итоговой аттестации: демонстрационный экзамен.

6. РАЗРАБОТЧИКИ ДОПОЛНИТЕЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Шевляков Артём Николаевич, доктор физико-математических наук, зам. директора Института математики и компьютерных наук Тюменского государственного университета.